Resilience of Energy Infrastructure and Services: Modeling, Data Analytics, and Metrics

This paper discusses data analytics-based power grid resilience modeling and the relevant metrics.

By CHUANYI JI, *Senior Member IEEE*, YUN WEI, AND H. VINCENT POOR, *Fellow IEEE*

ABSTRACT | Large-scale power failures induced by severe weather have become frequent and damaging in recent years, causing millions of people to be without electricity service for days. Although the power industry has been battling weather-induced failures for years, it is largely unknown how resilient the energy infrastructure and services really are to severe weather disruptions. What fundamental issues govern the resilience? Can advanced approaches such as modeling and data analytics help industry to go beyond empirical methods? This paper discusses the research to date and open issues related to these questions. The focus is on identifying fundamental challenges and advanced approaches for quantifying resilience. In particular, the first aspect of this problem is how to model large-scale failures, recoveries, and impacts, involving the infrastructure, service providers, customers, and weather. The second aspect is how to identify generic vulnerability in the infrastructure and services through large-scale data analytics. The third aspect is to understand what resilience metrics are needed and how to develop them.

KEYWORDS | Data analytics; failure; nonstationary spatiotemporal models; power distribution infrastructure; recovery; resilience metrics; services to customers

C. Ji is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: jichuanyi@gatech.edu).
Y. Wei is with Cooperfund, Oak Brook, IL 60523 (e-mail: ywei@cooperfund.com).
H. V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

I. INTRODUCTION

Severe weather events such as storms, flooding, and extreme temperatures have been occurring across the United States and the world in recent years, increasingly threatening places where large populations and economic activities are heavily concentrated [1]–[4]. Among the most affected are the energy infrastructure and services to customers, where weather-induced failures have affected millions of people for days [2], [5], [6]. In response to these disruptions, nationwide efforts focused on resilience have been initiated [2], [4], [6]–[8]. Here resilience refers to the ability to reduce failures under external disruptions and to recover rapidly once failures occur [2], [4], [8].

However, as pointed out by the taskforce report [8], the current understanding of resilience is limited for the energy infrastructure under severe weather. In particular, it is largely unknown how resilient our infrastructure really is to severe weather [2]. Moreover, the problem is not only about fixing the physical infrastructure. Services (i.e., electricity supplies to customers) are pertinent that involve users, service providers (i.e., distribution system operators (DSOs)) and policy makers [9].

This paper discusses research to date and challenges on resilience. The focus is on how to quantify resilience of the energy infrastructure and services to customers. Here the infrastructure refers to power distribution grids that deliver electricity directly to users. Power distribution grids are particularly vulnerable to severe weather disruptions, where 90% of failures have occurred [4]. Furthermore, the current power distribution infrastructure is not yet fully equipped with the state-of-the-art technologies for efficient monitoring and protection

[10]–[12]. Services of power distribution grids that involve a large number of customers in disjoint service regions managed by disparate distribution system operators and regulators are particularly challenging. A severe weather event can have a wide geographical span. For example, Superstorm Sandy affected eight million customers in 21 DSO service regions in the United States. In the face of these issues, quantifying resilience remains a challenging problem, involving both infrastructure and services [2], [9], [13]–[15].

Notably, resilience centers on complex and interacting networks involving weather, the power distribution infrastructure, and a community of customers, service providers, and policy makers [2], [4]. The failure aspect of resilience relates to the interactions between the physical infrastructure and weather. The recovery aspect relates mainly to services. Services depend on complex factors, not only the infrastructure but also DSOs, customers, and policies. Both aspects require advanced modeling to incorporate a large number of dependent variables, and data analysis to gain knowledge about what determines resilience. Thus resilience involves a multitude of complex factors from weather to the physical infrastructure, customers, service providers, and policy makers. These issues call for not only new thinking and actions from industry but also research that can address fundamental problems underlying the challenges.

We identify three technical challenges within this context. The first challenge is how to model complex interactions among weather, failures at the infrastructure, and recoveries by service providers governed by policies. Mathematical models are needed to describe these important factors, from a large number of local and dependent failures and recoveries, to customer responses and weather [9], [14], [16], [17]. Meanwhile, it is pertinent to incorporate multiple spatiotemporal scales in these models that span a distribution system locally to service areas regionally [9], [14], [16]. The second challenge is the development of data analytics that can learn how resilient the infrastructure and services really are from measurements. The power industry has been collecting data on failures and restoration. Such data can potentially be turned into knowledge to guide resilience enhancement. A challenge is the availability of detailed data at a large scale across service territories. Detailed data are owned by DSOs. Large-scale data studies thus call for active participation of DSOs and policy makers. Here, we describe the granularity, scale, paucity, and inaccuracy of existing data, from which we learn what new information should be collected. Furthermore, data analytics, although at an early stage for resilience, suggests what knowledge can be learned from the available measurements. Finally, the third challenge is how to measure resilience in a way that incorporates the infrastructure, services, customers, and weather. Such resilience metrics are needed for the community to be able to quantify threats and system-wide performance.

These three aspects are interrelated as illustrated in Fig. 1. Modeling lays a foundation to guide data analytics
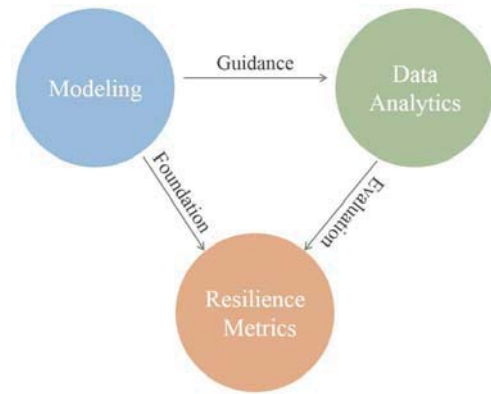
Fig. 1. *Illustration of three interrelated challenges.*

(i.e., on what simple quantities to evaluate from complex failure and recovery processes, and what data to use to gain knowledge on resilience). Meanwhile, modeling provides a basis from which resilience metrics can be derived as system-wide performance. Data analytics provide knowledge and insights on resilience of operational distribution grids and services. Data also underlay measures of resilience metrics and model parameters. Such knowledge should in turn improve modeling and data analytics.

The rest of this paper is organized as follows. The fundamental aspects of the problem and challenges are described in Section II. Modeling is discussed in Section III. Data analytics are reviewed in Section IV. Resilience metrics are described in Section V. Challenges and open issues are discussed throughout the paper and are summarized in Section VI.

II. PROBLEM DESCRIPTION AND FUNDAMENTAL CHALLENGES

The quantification of resilience depends on characterization of the performance of power systems. Resilience can be understood as the ability of the power systems to avoid or reduce failures and to recover quickly after failure occurrences. These two aspects are interrelated through the concepts of resilience across multiple spatiotemporal scales as stated below.

A. Infrastructure

The first aspect of resilience is that of reducing failures in the energy infrastructure. Here, as noted above, the infrastructure refers to power distribution systems, the last stage of the grid [18]. Severe weather events (e.g., high winds and flooding) damage power components such as down wires caused by falling debris, damaged transformers, or nonfunctional distribution substations. Component failures in a power distribution system are local (i.e., do not cascade for radial topology) but can involve large numbers of customers and span a wide geographical area [19]. Protective devices,

activated by failures or fault currents, are also considered as infrastructural failures since they interrupt electricity supplies to customers [18]. Examples of activated protective devices include open switches and blown fuses [20]. Outages are further caused by failures within a distribution system, where devices downstream lose power but are not damaged [9], [16].

B. Services

The second aspect of resilience relates to services (i.e., maintaining electricity supplies to customers) [21], [22]. Recovery from failures thus signifies the service aspect of resilience, involving complex factors of infrastructures, service providers, and policies [23]. In particular, DSOs are responsible for restoring electricity supplies to customers when disruptions occur. Thus, services are provided in a decentralized fashion, where individual DSOs are responsible for their own managed territories. Services are also governed by policies in the form of guidelines from state and federal governments [24]. Policy makers also participate actively in recovery processes (i.e., help guide restoration crews as shown in Superstorm Sandy and Hurricane Matthew). Hence, customers, DSOs, and policy makers form a community relevant to resilient energy services.

C. Multiple Spatiotemporal Scales

Resilience involves interactions among power distribution grids, services to customers, the community, and weather, as illustrated in Fig. 2. Such interactions occur dynamically across multiple spatiotemporal scales. For example, high winds cause fallen debris that induce failures to overhead power distribution lines in minutes [14]. Outages caused by failures occur in seconds or subseconds within a distribution infrastructure [16], [25]. Recovery occurs in seconds for restoring outages and in days for difficult manual repairs [9], [14]. Spatial scales vary from components in a local distribution system to townships, single service regions, and multiple service territories [26].
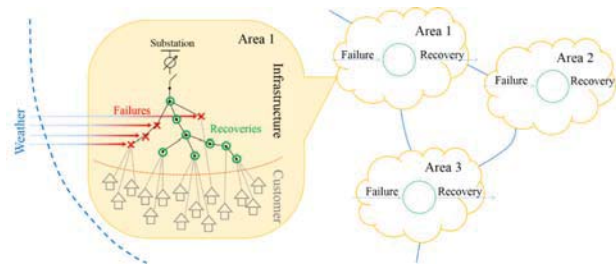
Fig. 2. *Illustration of interactions among weather, infrastructure, and community.*

D. Challenges and Open Issues

The following challenges emerge.

1) Modeling: How can we quantify resilience, i.e., starting with modeling, for large-scale weather events? As resilience should be a property of dynamic and dependent networks, models are necessary for integrating local but dependent failures, recoveries, the community, and weather at a large scale. Such models are challenging to obtain across multiple spatiotemporal scales for the infrastructure and services jointly (see Section III for detailed discussions).

2) Data analysis: How resilient are our power distribution grids and services in the first place? What should be enhanced for resilience? Data are needed from operational power distribution grids and services to aid in understanding what fundamentally governs resilience. A challenge is to obtain both detailed and large-scale data from the grid and about weather. Data are owned by individual DSOs governed by policies. Hence, data analytics call for collaboration from DSOs and policy makers (see Section IV).

3) Resilience metrics: How can we measure resilience at the network level involving both the infrastructure and services? Reliability metrics have been used as standards [27]. However, these metrics are designed for daily operations rather than severe weather events. Resilience metrics are required to incorporate dynamic characteristics at multiple spatiotemporal scales (see Section V).

III. MODELING

An objective of modeling is to characterize relationships among a large number of dependent variables. Such variables include weather, failures at the distribution grid level, recoveries, and impacts on customers and the community overall. To date, there does not exist such a model that incorporates all these pertinent factors. Different aspects have been studied in prior work, from static models to nonstationary spatiotemporal random processes.

A. Static Models in Machine Learning

A large body of prior work addresses the modeling of how severe weather induces initiating failures [20], [28]–[30]. These models are static, focusing on finding a mapping between weather variables and failures. Such models pioneered the work in this area, starting with one node (e.g., a power distribution line or a component), and one to multiple weather variables [20], [29], [31]. For example, the failure rate, which is the average number of new failures occurring per mile per hour of overhead power lines, is modeled as a quadratic function of wind intensity in [18]. Fragility, which is the conditional probability of a
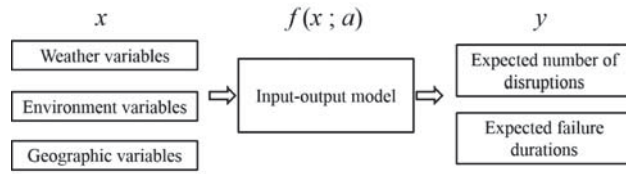
Fig. 3. *Machine learning view of static models.*

component failure given weather variables, is modeled as a function of wind intensity or gust, precipitation, and surge elevation, respectively, in [32] and [33].

These models, although diverse, can be unified through machine learning as illustrated in Fig. 3. Consider an n-dimensional vector of exogenous variables $x \in R^n$ at a given location for $1 \leq n$. Modeling can be viewed as finding a static mapping, $f(x; a): x \rightarrow y$, between exogenous variables x and targets $y \in R^m$ for $1 \leq m$. Here y describes failures (i.e., as the number of failures or failure durations, or the probability of failures/durations), and $a \in R^l$ is a vector of unknown parameters with $1 \leq l$. The data set $D = \{x^{(k)}, y^{(k)}\}$ is obtained on pairs of exogenous and failure variables, where $(x^{(k)}, y^{(k)})$ denotes the kth sample of the input x and desired output y of the learning machine. The goal of learning is to obtain either f or the parameter a for a chosen f using data D so that f approximates an underlying mapping from x to y. This is clearly a context of supervised learning [34], [35]. The models are static, where neither f nor the parameters a nor the inputs and outputs (x, y) vary with time.

The input variables x mainly represent weather phenomena, including wind intensity, speed, and gust, as well as precipitation. Several example models for f have been studied.

1) Poisson generalized linear model (GLM) [36]: The number of power failures in a grid cell is modeled as a Poisson random variable with mean μ, where $\ln(\mu)$ is assumed to be a linear function of the weather variables x.

2) Negative binomial generalized linear model (NBGLM) [20]: An error term ϵ is introduced into the GLM to model the dispersion (i.e., the inconsistency between the mean and the variance).

3) Generalized additive model (GAM) [31]: The linear function is replaced by a nonlinear mapping f, with examples including cubic splines and nonparametric models.

4) Spatially dependent Poisson linear models [37]: Spatial correlation is included in the GLM as a multivariate normal distribution across different grid cells.

The above models have been used widely in subsequent works [38]–[41]. Resulting models identify fallen trees as major causes of power failures [20], [42], and transformers as the components most affected by severe storms [36], [38].

Other learning methods have been applied, including classification and regression trees, Bayesian

additive regression trees, and multivariate adaptive regression splines [29]. Bayesian additive regression trees have been found to be the most accurate in predicting durations of failures given weather variables [29]. Principal component analysis (PCA) has been found to be effective in learning from correlated weather variables [38].

The static models assume that failures occur independently of time and locations [20], [37]. This assumption on temporal independence is reasonable if evolution of failures is not considered. The assumption of spatial independence can be invalid since sufficiently close locations may experience similar weather impacts [37]. In addition, certain geolocations exhibit a higher likelihood of weather-induced failures than the others [2]. Due to these assumptions, certain static models are obtained by aggregating over time and service regions [20]. However, the aggregation may lose spatiotemporal information needed for failure and recovery studies (see Section IV for further discussions).

B. Spatiotemporal Random Processes

When sufficiently fine spatial and temporal scales are taken into consideration, failures and recoveries need to be modeled as spatiotemporal random processes [9], [14], [43]–[45]. Such models characterize dynamic interactions of infrastructural failures, services, and customers, which are not quantified by static models [9], [14].

1) Dynamic Models for Cascading Failures: The prior works motivate such modeling albeit the problem they consider is on cascading failures that occur in power transmission rather than distribution grids [46], [47]. For example, dynamic models have been developed for cascading failures through branching processes [46], [48], Markov decision processes [49], hybrid system models for random and sporadic failures [50], and other probabilistic temporal or spatial models (see [51]–[54] and references therein).

These models are based on stationary probability distributions while weather-induced failure–recovery processes are nonstationary [9], [14]. Furthermore, severe-weather-induced disruptions span a wide geographical area. Therefore, spatiotemporal processes are needed for weather-induced failures and recoveries.

2) Nonstationary Failure–Recovery–Impact Processes: Some recent work has developed a spatiotemporal nonstationary model for dependent failure–recovery processes [14]–[17]. This model is motivated by nonstationary queues [55]. Such models have been applied to failure–recovery processes from severe weather events [14], [17]. However, the queuing model is inapplicable when a finer spatiotemporal scale is considered for impacts on customers of each failure and recovery, and restoration is conducted with priorities for critical customers [22].

A formulation is then developed from bottom-up, starting with failures in the power distribution infrastructure, incorporating service recovery through failure durations,

and impacts on customers [9]. Such models integrate a large number of interdependent variables at the finest spatiotemporal scale.

To consider this model, assume failures are already detected. $I_i^{(d)}(t)$ $(d = f, o)$ is an indicator function, representing a failure (f) or an outage (o) for $I_i^{(d)}(t) = 1$; otherwise, $I_i^{(d)}(t) = 0$. i includes the type, geolocation, and system location of device i. Service is characterized by how rapidly power is restored to customers [2], and is thus represented by the downtime duration $D_i(v)$ for failure or outage i that occurs at time v. An indicator function $I[D_i(v) > t - v]$ represents the recovery event, where failure or outage i is not yet recovered at t, $0 < v < t$. Finally, the impact on customers evaluated at time t is modeled via a function $G_i(v, t)$ for disruption i, which occurs at time v for $v < t$. As a simple example, $G_i(v, t)$ is the customer downtime resulting from failure or outage i.

Failures, outages, recoveries, and costs are dependent for a given weather event, evolving in time and locations. Incorporating randomness from weather disruptions, the spatiotemporal nonstationary random processes model a collection of dependent infrastructural failures, recoveries, and costs as coupled processes:

1) failure (and outage): $\{I_q^{(d)}(v), q \in S(v), v > 0\}$;
2) recovery: $\{I[D_k(v) > t - v], k \in \overline{S(t)}, 0 < v < t\}$;
3) cost: $\{G_j(v, t), j \in S(v), 0 < v < t\}$.

Here, $S(v)$ and $\overline{S}(t)$ consist of nodes in normal operation at time v and disruptions at time t, respectively. While such a model starts from the finest spatial scale of individual components and customers, aggregation can be done at the scales of an area, a township, one service region, and multiple DSO territories, as illustrated in Fig. 2.

Quantifying completely the spatiotemporal nonstationary random processes is prohibitive since that requires joint probability distributions at all time epochs. The first moments can be used in an initial effort, including the time-varying failure rates and marginal conditional probability of downtime duration given failure occurrence time [9]. These simple quantities guide data analytics in Section IV.

C. Open Issues and Challenges

The first open issue that emerges is how to characterize two generic properties of resilience (Section II): 1) dependent "networks" from the infrastructure to customers and service providers impacted by weather; and 2) nonstationarity across multiple spatiotemporal scales. Different models have characterized different aspects of the problem. There are no models yet that characterize all the generic properties.

The second open issue is the complexity of modeling in characterizing the uncertainty. Models can become prohibitively complex when involving spatiotemporal uncertainty in the physical infrastructure and services [56], [57]. For example, questions arise whether power flows should

be incorporated for studying resilience [52], [58]–[60]; or whether failures can be assumed as detected already so that failure and recovery processes are on (failed, outaged or normal) states of components [9], [14]. The aforementioned models do not involve power flows, which makes the formulation analytically tractable. Such models also allow data analytics to be conducted using available measurements (see Section IV). However, starting from power flows enables failure detection and grid reconfiguration [11], [58]. A challenge is to identify appropriate granularity in modeling-dependent variables based on objectives from weather, the infrastructure, services, and impacts on customers.

The third open issue is on the community. While the roles of DSOs and policy makers are embedded in services (and infrastructure enhancement), approaches are lacking to model such influences explicitly. Although not specific for the resilience problem, approximate dynamic programming is one of the approaches that choose optimal control policies towards a self-healing grid [61]. An open question is how to quantify policies that impact services in a real-world setting.

Finally, a further open issue is how to combine strengths of machine learning approaches and spatiotemporal random processes to incorporate diverse factors at a large scale. For example, using data from phasor measurement units (PMUs) (see Section IV), failures can be identified from power flows in conjunction with state estimation and optimal sensor (PMU) placement [11]. Such an approach can be extended to power distribution grids, where learning algorithms can potentially learn statistical characteristics from PMU measurements and probability distributions of failure hypotheses from historical data. Another related application, which increases the visualization of power distribution grids, is topology inference through learning from data collected only at the edge nodes [62]. In a broader context, machine learning is expected to play a major role in how to quantify uncertainty and data analysis [63], which are the key issues of modeling and the resilience problem overall. In this context, the availability of data at a sufficiently fine spatiotemporal scale is crucial for enabling learning approaches [63].

IV. DATA ANALYTICS

Data analytics learn knowledge from measurements on failures, recoveries, and weather variables. Knowledge learned helps to answer such questions as how resilient the infrastructure and services really are; and what governs resilience or the lack thereof. Here modeling provides a pertinent role of guiding data analytics (i.e., on what measurements to use and what quantities to estimate).

A. Data

Data determines significantly what knowledge can be learned about resilience.

1) Data on Failure and Recovery: The power industry has been collecting data on failures, restoration, and impacts to customers [18]. The objectives of this collection have been outage management, customer communication, and reporting [24]. Data analysis is yet to become a focus.

As a typical example of granular data, an item on a failure includes "occurrence- and restoration-time, geolocation, system-location, the type of activated protective device, the number of customers affected" [9], [14], [37]. Here activated protective devices signify the actual failed components, and themselves interrupt electricity service to customers [36], [37].

One minute is the finest temporal resolution of the available data for failure occurrence and restoration time. Resulting from customer reports on service interruptions, such a time scale is consistent with that of weather-induced failures from dynamic evolution of severe storms [9], [14], [20], [36]. Failures can further cause outages in a distribution system, where certain components lose power but are not damaged [16]. Outages occur in seconds or less [16], [25]. Therefore, the temporal scale of failure data should ideally be less than a second. Such a fine temporal scale is not yet achieved in current data collection. Advanced technologies are required to attain such granularity beyond customer reports. The current resolution on restoration time is also a minute, where failure durations vary from minutes to hours and days.

Geolocations of failures and outages are provided in exact coordinates of (latitude, longitude) [9]. The location information is usually available on activated protective devices rather than actual failed components [9], [16], [37]. Data on locations of failed power components are more informative for studying relationships between failures and weather variables [36], [39]. Although not usually available for research, the information on failed devices is known in principle after restoration. Detecting power failures and identifying their locations in real time have been of research interest, especially with deployment of smart meters, (micro) PMUs for distribution systems, field sensors, and control units [11], [64]–[67]. Overall, accurate geolocations together with downtime of both failed components and activated protective devices are desirable for studying spatiotemporal variability of severe weather impacts (Fig. 4).

Accuracy of the data is another pertinent issue. Existing collection methods can fail to generate high-resolution data, especially in severely impacted service regions [68]. For example, a large number of customer calls within a short time duration hinders failure isolation [68]. Repair crews are typically busy fire-fighting to restore services to customers; data collection on recovery time is thus not a priority [68]. Therefore, automated approaches are pertinent for accurate data collection.
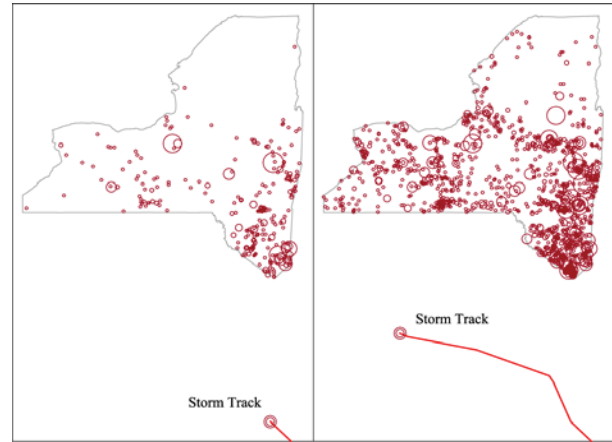
Fig. 4. *Geolocations and occurrences of failures, and the storm track during Superstorm Sandy.*

Impacts on customers provide another important source of information on resilience. Available data on impact are currently measured as the number of customers affected by each failure [9], [31], [36]. Total customer downtime can then be obtained, reflecting impacts from both failures and recoveries [9].

B. Data on Weather Variables

Data on weather variables offer pertinent information on external causes of failures and delays in recovery. Commonly used data on severe storms have been collected on wind intensity and gust, precipitation, moisture, and temperature [28], [31], [36]. Such data are usually provided by additional sources outside DSO service regions. While an extensive survey of weather data is beyond the scope of this work, well-known example data in wind and precipitation are from the National Oceanic and Atmospheric Administration (NOAA) [9], [36], [38], [69]. The spatiotemporal granularity of the data varies. For example, the wind speed is measured in minutes and at the centroid of each zip code [42]. The resolution for gust wind speed is estimated at three-second intervals and each 3.66-km × 2.44-km grid cell [38]. Weather data with a coarse spatiotemporal resolution can be insufficient for terrains with dynamically varying weather conditions. Recent data collection and forecasting techniques improve spatial resolution by incorporating community weather stations [69], [70].

Several DSOs have installed densely distributed weather stations in their service regions, where existing regional weather service is insufficient for dynamic local conditions. This allows data to be collected on both weather and power failures at comparable spatiotemporal scales [28]. For instance, National Grid has deployed weather stations, each of which covers five square miles in a service area [28]. Central Hudson Electric and Gas has weather stations

needed for the varying terrain conditions in its service region [71]. San Diego Gas and Electric has sensors and mini weather stations for predicting wildfires and its resulting power failures [72]. Furthermore, a commercial product (Deep Thunder) by IBM offers localized weather prediction at a spatial scale of city blocks [70].

Storm surge and flooding result in damages to power components [68]. However, data on surge and flooding are available from fewer sources [73]. Synthetic data have been generated from simulation on storm surge and flooding as well as high winds when detailed data are difficult to obtain [74], [75]. A challenge is for simulated data to sufficiently accurately relate to failures in the infrastructure.

Other exogenous variables include land cover, where residential, forest, commercial, industrial, and transportation land use seem to be related to impacts of service interruptions on customers [20], [36], [37]. For example, tree density is an exogenous variable used in prior work that results in downed wires and other initiating failures from high winds [20], [36], [37].

C. Analytics

As modeling and available measurements lay foundations for data analytics, a pertinent question is what knowledge can be learned from data.

1) Outage Management: DSOs have long been collecting data on failures and recoveries [18]. A major use of data is for outage management [18], [68]. For example, customer reports of service interruptions are combined with outage management systems to localize failures. Such information is then used to guide repair crews. Outage maps are generated for customers showing the evolution of failures and restoration [76]. Aggregated information on failure and recovery is used for reporting impacts and performance of service restoration [77]. Data analytics have not been a significant part of standard outage management in practice.

2) Failure Prediction in One Service Region: Failure prediction has been studied by the prior works [28], [29], [35], [30]. One of the first works, although mainly focused on reliability rather than resilience, applies machine learning to predict equipment failures in the New York City power grid [30]. An objective is to enable proactive maintenance for reducing severe impacts of power failures resulting in events such as explosions or fires. Data on failures and assets have been collected from manholes in multiple years, and reactive point processes are used to learn model parameters from the data [78]. Although the power grid in New York City is complex, the data analytics showed promise for failure prediction [30], [78].

Several other prior works pioneered failure prediction using weather data and regression models (see Section III). The premise is that if the likelihood of failures can be obtained given weather variables, failures can be predicted through weather forecasts [28], [29], [36]. Data on failures and weather from one service region are used for parameter estimation and model validation [20], [28], [29], [31]–[33], [38], [40], [42].

A challenge is that detailed data are often unavailable on failures due to security issues [39]. As such, the early works have had to use aggregated failure data [20], [31]–[33], [38]. Temporal aggregation yields the number of failures, ranging from one day to an entire period of a hurricane [20], [32], [79]. Spatial aggregation of failure locations ranges from a small grid cell of 0.42 km^2 to an area specified by a geocode or zip code [20], [32], [40], [80].

On other occasions failure and weather data have different granularity [20]. Failure data are then aggregated to match the coarser geographic resolution of measurements on weather and other exogenous variables [20]. Overall, aggregated information over time and locations cannot specify exactly when and where individual failures occur and recover. Thus, data in weather and failures, when either are aggregated, can affect the accuracy of a learned model and consequently prediction.

With densely installed weather stations in a service territory, several recent works have been able to use detailed data on both weather and power failures, resulting in a few failure prediction systems for DSOs [28], [70], [72].

D. Regression Study at the National Scale

As a severe weather event often spans multiple service regions, a question is how to extend data analysis from one service territory to a regional or national scale. Granular data on power failure and recovery are owned privately by individual DSOs. A recent work explores a novel option of publicly available data [81]. Such data result from DSOs' annual reports on the IEEE standard reliability indices: system average interruption frequency index (SAIFI) and system average interruption duration index (SAIDI) [27]. SAIFI and SAIDI are, respectively, the average number of power failures and downtime durations per customer per year. Thus such data are aggregated with a spatial resolution of a service region and temporal scale of a year [77].

The data have been collected at the national scale across the United States over the past 13 years [81]. Data on exogenous variables have also been obtained on weather, DSO expenditures on reliability, and the density of power lines. The data from all sources are used to learn parameters of regression models [81]. The failure and duration indices are found to correlate with weather variables, especially when major weather events occurred.

While this approach explores new large-scale data sources, stationarity of the variables may be required so that regression using aggregated data can be equivalent to using detailed measurements. Intuitively, the approach is expected to perform well for daily operations when the stationarity is natural for failures and restorations. When

including a severe weather event, detailed data are needed at sufficiently fine spatiotemporal scales.

E. Infrastructure and Service Resilience in Multiple Regions

Owned by individual DSOs, detailed and large-scale data require collaboration from multiple service providers. As a first step, four DSOs and policy makers have collaborated in a recent work, providing detailed data on failure and recovery in Upstate New York [9]. The data span four service regions over 50 000 square miles that serve two million customers. The granularity of the data is the finest that the current collection can offer [9]. The data consist of failures and recoveries that occurred during Superstorm Sandy and daily operations in 2012. Weather data were not used for the study.

Such detailed data at the large scale enable complementary questions to be studied: whether data analysis can help identify generic vulnerabilities (i.e., nonresilience) in the infrastructure and services. If so, how can data collected by DSOs benefit long-term prediction of resilience through enhancement? A recent work studying these questions is summarized below [9].

Guided by the nonstationary spatiotemporal model (Section III-B), the analysis of detailed data focuses on a few simple model parameters such as failure rates and conditional probabilities of downtime durations. The data analysis reveals infrastructural vulnerability illustrated in Fig. 5(a), where local failures, although they do not cascade

within power distribution systems, have nonlocal impacts (i.e., they affect a large number of customers) [9]. A scaling law further characterizes systematically how local failures impact customers: A point of scaling obeys the 20–80 rule [82], where the top 20% of failures affect 80% of the customers. Importantly, such infrastructural vulnerability was not caused by Superstorm Sandy but exists regularly in daily operations. The hierarchical structure of power distribution systems relates to the infrastructural vulnerability, where the majority of the top failures were seen to occur at the higher level of the hierarchy and thus affected a large number of customers. Superstorm Sandy exacerbated such infrastructural vulnerability by increasing the likelihood of such failures. The large scale spanning the four service regions confirms that the vulnerability is common across four DSO service territories.

In contrast, recovery patterns on customer services are different from those failures [Fig. 5(b)]. A large number (i.e., 89% of all) small failures that affected the bottom 34% of customers aggregate to 56% of total customer interruption time. This illustrates challenges for services and requires further study [9].

F. Challenges and Possible Directions

The first challenge is how to obtain sufficiently detailed and accurate data. Research to date suggests the necessity of collecting data with sufficient spatiotemporal granularity. A temporal resolution should be comparable to that of failure and outage occurrences, i.e., at seconds for outages, and
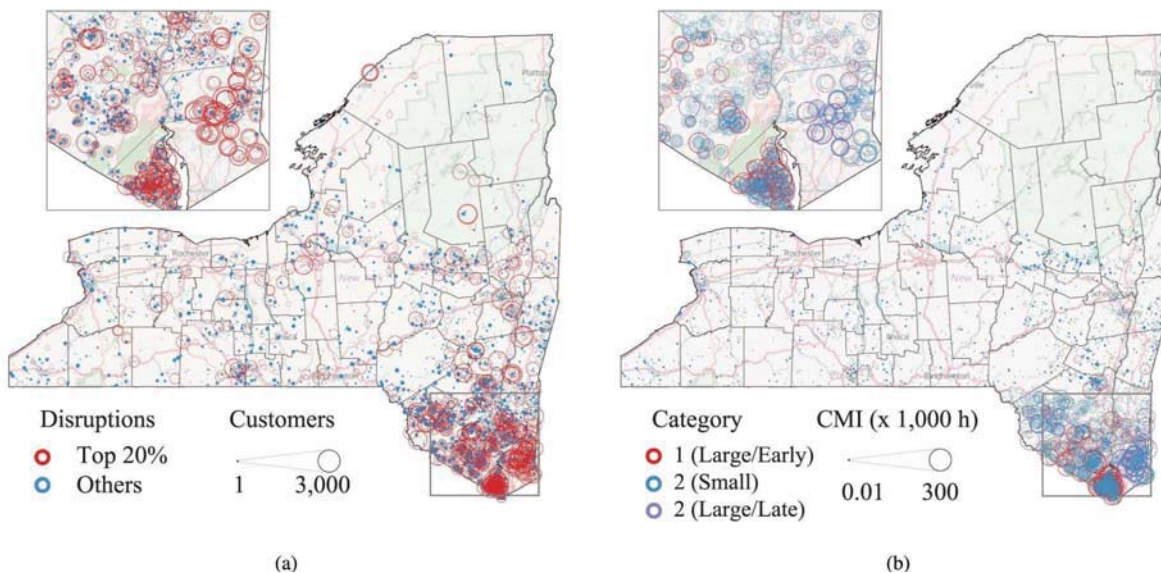
Fig. 5. *Geographical distribution of failures and customer downtime in Upstate New York during Superstorm Sandy. (a) Top 20% and the remaining failures. (b) Customer interruption hours (CMI), where the colors represent the downtime for large (top percentage) failures that recovered rapidly versus the remaining disruptions. Each marker represents a failure or an outage. Map reproduced using OpenStreetMap and ArcGIS software. The figures are adopted from [9].*

minutes for weather-induced failures. Desired spatial resolution corresponds to exact failure and restoration locations. Additionally, large-scale (and detailed) data are needed to understand whether knowledge learned is widely applicable. The best available data so far attain accurate spatial resolution, a minute as the time scale, and across several multiple service regions. Aggregated data have been used at the national level. Data acquisition for power recoveries is also insufficient. This is because repair crews naturally place a higher priority on restoring services than on data collection.

Advanced collection systems can help circumvent data paucity and inaccuracy. A large-scale deployment of low-cost sensors pioneers this approach for monitoring voltage sags and communicating event information across service regions in the United States [83]. A next-generation collection system can potentially gather high-resolution data, using widely deployed micro PMUs in power distribution systems, advanced metering infrastructure, and intelligent devices [12], [15], [65], [66], [85]–[87]. Advanced detection algorithms can then accurately detect and locate failures in real time to replace customer reporting [11].

The second challenge is for data on weather and other exogenous variables to have a sufficiently fine spatiotemporal scale. When weather data have a coarser spatial temporal resolution compared to that from the grid [20], the pertinent and detailed information on infrastructural failures may be wasted [14], [20]. An open issue is how to obtain data with sufficient resolution from diverse sources (i.e., the infrastructure, services, and weather).

The third challenge is how to enable collaboration on data analysis. Large-scale and detailed data from multiple DSOs are severely lacking for research due to security and privacy concerns [2]. An important issue is to actively involve DSOs and policy makers to collaborate on data analytics. As shown in the prior study [9], resilience is for everyone; thus, it is possible for DSOs, policy makers, and academia to collaborate on data analytics. Procedures that support security and privacy will help enable and expand such collaboration.

Finally, data analytics, although showing promise of learning knowledge on generic vulnerabilities and predicting failures, are still at an early stage. The potential of data analytics for resilience is yet to be fully explored when measurements become more and more available.

V. RESILIENCE METRICS

Modeling and data analytics lay a foundation for resilience metrics. Such metrics are expected to characterize system-wide performance by including all factors from weather to the infrastructure, services, and community. Thus, resilience metrics need to be derived from models and measured from data. While such metrics are yet to be fully developed, the research to date shows the necessity for new metrics. Open issues include what variables are pertinent and how to derive the performance metrics.

A. Invalidity of Standard Metrics

Are new metrics needed for resilience in the first place? Two IEEE reliability standards for daily operations (SAIFI and SAIDI) have been extended to severe weather events: storm average interruption frequency index (STAIFI) and storm average interruption duration index (STAIDI) [88], [89]. Given a severe weather event, STAIFI and STAIDI are defined as [89]

$$\text{STAIFI} = \frac{\text{total number of customers interrupted}}{\text{total number of customers served}} \quad (1)$$

$$\text{STAIDI} = \frac{\text{total customer storm interruption minutes}}{\text{total number of customers served}}. \quad (2)$$

Now consider a data set from a service region during Superstorm Sandy in Upstate New York. Each data sample corresponds to a failure signified by an activated protective device with the occurrence time, duration, and number of customers affected. There are 1334 failures from October 28, 2012, to October 31, 2012 (see [9] for details). The STAIFI and STAIDI values and their standard deviations are obtained using (1) and (2) and shown in Fig. 6. The standard deviations are so large for both indices that they allow a negative quantity when the error bars are taken into consideration. Such large deviations suggest that STAIFI and STAIDI exhibit too much uncertainty to be valid for characterizing resilience. Therefore, extending, by brute force, the reliability standards to resilience metrics is not viable.

Having the large standard deviation is not a coincidence but results naturally from nonstationary failure–recovery processes. Nonstationary random processes can of course exhibit time-varying mean functions [9], [14], [53]. This is clearly shown by the time-varying version of STAIFI and STAIDI from (1) and (2) that are computed using samples at 1-h intervals in Fig. 6. In contrast, the STAIFI and STAIDI are static sample averages by definition, and thus insufficient for representing non-stationary failure and recovery processes.

B. Other Metrics and On-Going Studies

Estimated time of restoration (ETR) is another metric used by industry [68]. ETR informs customers of the expected time needed for restoring services after failures. While appealing to users, ETR is difficult to estimate accurately because of the uncertainty and dynamics from nonstationary failure and recovery processes.

Fragility and its variations relate failures to weather variables [40], [80], [89], [90]. Such a relationship is necessary to view resilience through potential threats, and thus is promising to characterize a performance metric. A challenge is how to include dynamics and system-wide performance in such a resilience measure.

Dynamic metrics such as quality and its variations (i.e., robustness and rapidity) characterize over time parts of a system or the number of customers in normal operations
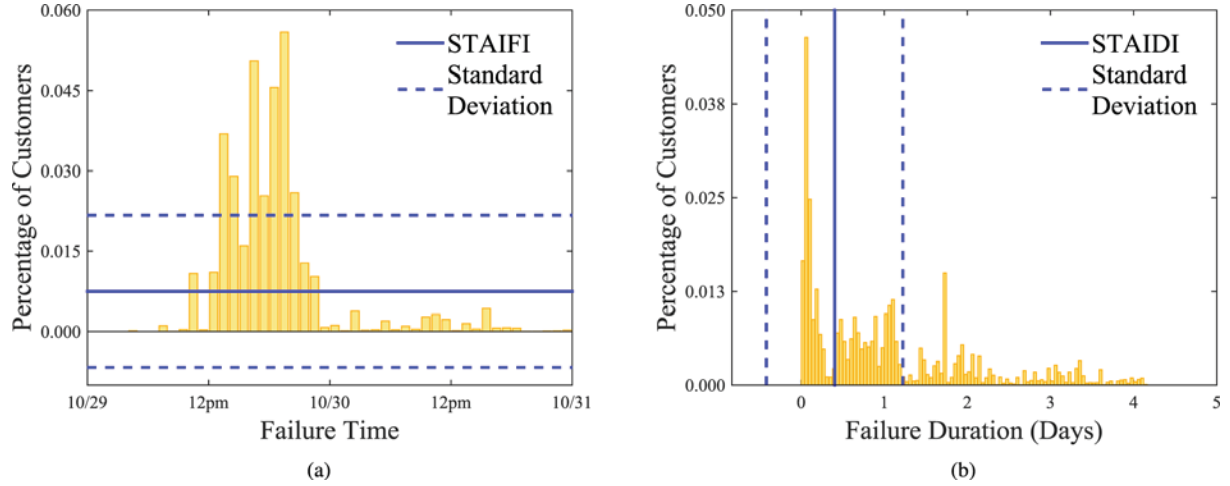
Fig. 6. *STAIFI and STAIDI from a service region in Upstate New York during Hurricane Sandy. The histograms show the percentage of affected customers versus (a) time and (b) interruption duration. The error bars correspond to the standard deviations.*

[91]–[93]. These metrics include dynamic evolution of resilience but are based on pre-assumed functions of time.

Questions arising include, what factors should a resilience metric include; and how to derive such a metric. In principle, a resilience measure should include pertinent factors from weather to failures, recoveries, and impacts on customers [94]. As resilience quantifies system-wide performance, a fundamental approach is to derive such a metric from bottom-up based on modeling, including weather variables as potential causes, then failures, recoveries, and impacts as consequences [16], [17].

Grounded by the spatiotemporal model, resilience metrics in recent work are derived from the bottom-up to incorporate some of the factors: nonstationary failure–recovery processes and impacts to customers [16], [17]. However, weather and other exogenous variables are not included. A metric $R(t)$ is defined as

$$R(t) = 1 - \frac{1}{C_0} E\{C(t; d)\} \qquad (3)$$

where $E\{C(t; d)\}$ is the expected cost/impact at time t [16], [17]; $d > 0$ is a threshold on tolerable delays for recovery; C_0 is a normalization factor; and $R(t) = 1$ indicates the best resilience and $R(t) = 0$ is nonresilience, so that $1 - R(t)$ is the percentage cost or impact, evolving with occurrences of failures and recoveries.

The impact/cost has been derived using failure–recovery processes developed through nonstationary queuing models [16], [17]. Data from Hurricane Ike have been used to obtain the value of the metric for an operational power distribution grid [16], [17]. The failure–recovery–cost processes in Section III [9] can potentially be used to evaluate the impact/cost on customers. Thus, resilience metrics depend critically on modeling and data.

C. Challenges and Discussions

Despite the progress made to date, there are yet to be performance measures that incorporate all three intrinsic characteristics at the system level: spatiotemporal nonstationary failure–recovery; weather variables; and service providers, customers, and the community overall. Static metrics such as STAIFI and STAIDI characterize average behaviors of failures and recoveries but not the spatiotemporal evolutions during a severe weather event. The dynamic metrics recently developed include failure–recovery processes and impacts on customers, but not weather. The following research questions relating to resilience metrics arise.

1) What resilience metrics can encompass cohesively the three pertinent characteristics: exogenous weather variables, spatiotemporal nonstationary failures in the infrastructure, and recoveries of services for customers?
2) What approaches can lead to such resilience metrics at the system level, combining weather with failure–recovery–impact processes?
3) What (additional) data are needed to evaluate resilience of the infrastructure and services?

Answers to those questions are expected to result from both development of system-wide metrics and modeling that incorporates the variables from bottom-up. Extensive data analytics are also needed to obtain values of newly developed metrics and to compare them with the standards.

VI. CONCLUSION

Quantifying resilience of the energy infrastructure and services under severe weather is pertinent but understudied, as shown by the prior works. An immediate need is to understand how resilient the energy infrastructure and services really are. Such understanding will enable fundamental enhancement of

resilience beyond responding to severe weather. In this context, unique characteristics emerge involving weather, failure, and recovery processes in the infrastructure and services, as well as impacts on communities. These characteristics are interrelated and impact resilience together. Therefore, modeling, data analytics, and resilience metrics need to be studied cohesively and at a large scale.

Models, when developed for separate aspects of the problem, are found incapable of characterizing these unique properties jointly. Formulating the problem from bottom-up through spatiotemporal random processes has the potential to characterize the interactions of failure–recovery–impact processes. While the modeling framework extends to customers as parts of communities, roles of service providers and policy makers have been insufficiently studied. Relationships between weather variables and failures have been studied in a separated context. Models are yet to be developed to incorporate all pertinent factors.

Data analytics, although at an early stage, have started to show promise in learning knowledge about resilience. Data collected by DSOs, when sufficiently detailed, have shown potential in identifying generic vulnerabilities of the infrastructure and services. Large-scale and detailed data are particularly needed from multiple service territories. This provides an opportunity for collaboration among DSOs, policy makers, and researchers. After all, resilience is for the benefit of the entire community. As big data are prospering in many fields of engineering, the resilience problem presents a new application area.

Modeling lays a foundation for deriving resilience metrics. Widely used IEEE standards are mostly developed for reliability in daily operations. Those metrics, when directly extended to severe weather disruptions, are found to exhibit too much uncertainty to be reliable. Resilience metrics developed from bottom-up can incorporate nonstationary spatiotemporal failure–recovery–impact processes across multiple spatiotemporal scales. However, such metrics do not yet include weather variables. The metrics with weather variables have been developed in a separate context. An open issue is how to derive resilience metrics, combining weather with the infrastructure and services.

Full of open issues and challenges, the problem of resilience provides a fertile ground for technical study. Modeling, data analytics, and metrics are still very much open for development. Models and metrics are yet to include cohesively a wide range of exogenous variables. More details from the infrastructure such as power flows may be needed in the study of resilience.

In a broader context, while the methods discussed here focus on the power distribution infrastructure and services under severe weather, the approaches for quantifying resilience can be generalized to other dependent networks in a natural environment. ∎

REFERENCES

[1] P. Hoffman and W. Bryan, "Comparing the impacts of the 2005 and 2008 hurricanes on U.S. energy infrastructure," Office Electr. Del. Energy Rel. U.S. Dept. Energy, OE/ISER Rep., Feb. 2009.

[2] M. R. Bloomberg, "A stronger, more resilient New York," New York, NY, USA, PlaNYC Rep., Jun. 2013.

[3] J. A. Sathaye *et al.*, "Estimating impacts of warming temperatures on California's electricity system," *Global Environ. Change*, vol. 23, no. 2, pp. 499–511, 2013.

[4] "Economic benefits of increasing electric grid resilience to weather outages," President's Council of Econ. Adv., U.S. Dept. Energy's Office Electr. Del. Energy Rel., Executive Office President, Washington, DC, USA, Tech. Rep., Aug. 2013.

[5] "U.S. energy sector vulnerabilities to climate change and extreme weather," U.S. Dept. Energy, Washington, DC, USA, Tech. Rep. DOE-PI-0013, Jul. 2013.

[6] H. Rudnick, "Impact of natural disasters on electricity supply," *IEEE Power Energy Mag.*, vol. 9, no. 2, pp. 22–26, Mar./Apr. 2011.

[7] *Committee on Increasing National Resilience to Hazards and Disasters, Committee on Science, Engineering, and Public Policy, Disaster Resilience: A National Imperative.* National Academies Press, Washington, DC, USA, 2012.

[8] "Weathering the storm: Report of the grid resiliency task force," Office of Governor Martin O'Malley, Tech. Rep., Sep. 2012.

[9] C. Ji *et al.*, "Resilience of power grid: Large-scale data analysis across multiple service regions," *Nature Energy*, vol. 1, no. 5, p. 16052, 2016.

[10] A. G. Phadke and J. S. Thorp, *Synchronized Phasor Measurements and Their Applications.* New York, NY, USA: Springer-Verlag, 2008.

[11] Y. Zhao, J. Chen, A. Goldsmith, and H. V. Poor, "Identification of outages in power systems with uncertain states and optimal sensor locations," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 6, pp. 1140–1153, Dec. 2014.

[12] B. P. Wiseman, Y. Chen, L. Xie, and P. R. Kumar, "PMU-based reduced-order modeling of power system dynamics via selective modal analysis," in *Proc. IEEE/PES Transmiss. Distrib. Conf. Expo.*, May 2016, pp. 1–5.

[13] *Workshop on Resilient Smart Grid Customers*, Brookhaven National Laboratory, Long Island, NY, USA, Apr. 2014. [Online]. Available: http://bnl.gov/rcsg2014/

[14] Y. Wei, C. Ji, F. Galvan, S. Couvillon, G. Orellana, and J. Momoh, "Learning geotemporal nonstationary failure and recovery of power distribution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 1, pp. 229–240, Jan. 2014.

[15] Z. Zhong *et al.*, "Power system frequency monitoring network (FNET) implementation," *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 1914–1921, Nov. 2005.

[16] Y. Wei, C. Ji, F. Galvan, S. Couvillon, and G. Orellana, "Dynamic modeling and resilience for power distribution," in *Proc. IEEE SmartGridComm*, Vancouver, BC, Canada, Oct. 2013, pp. 85–90.

[17] Y. Wei, C. Ji, F. Galvan, S. Couvillon, G. Orellana, and J. Momoh, "Non-stationary random process for large-scale failure-recovery and resilience of power distribution," *Appl. Math.*, vol. 7, no. 3, pp. 233–249, 2016.

[18] R. E. Brown, *Electric Power Distribution Reliability*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2008.

[19] P. Hoffman *et al.*, "Hardening and resiliency: U.S. energy industry response to recent hurricane seasons," Dept. Office Electr. Del. Energy Rel., U.S. Dept. Energy, OE/ISER Final Rep., Aug. 2010.

[20] H. Liu, R. A. Davidson, D. V. Rosowsky, and J. R. Stedinger, "Negative binomial regression of electric power outages in hurricanes," *J. Infrastruct. Syst.*, vol. 11, no. 4, pp. 258–267, 2005.

[21] M. Panteli and P. Mancarella, "The grid: Stronger, bigger, smarter?: Presenting a conceptual framework of power system resilience," *IEEE Power Energy Mag.*, vol. 13, no. 3, pp. 58–66, May/Jun. 2015.

[22] E. Yamangil, R. Bent, and S. Backhaus, "Resilient upgrade of electrical distribution grids," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 1233–1240.

[23] Y. Wang, C. Chen, J. Wang, and R. Baldick, "Research on resilience of power systems under natural disasters—A review," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1604–1613, Mar. 2016.

[24] " Before and after the storm: A compilation of recent studies, programs, and policies related to storm hardening and resiliency," Edison Elect. Inst., Washington, DC, USA, Tech. Rep., Jan. 2013.

[25] M. Amin and J. Stringer, "The electric power grid: Today and tomorrow," *MRS Bull.*, vol. 33, no. 4, pp. 399–407, Apr. 2008.

[26] "2011 Electric reliability performance report," Dept. Public Ser. State New York, Office of Electric, Gas, and Water, Tech. Rep., Jun. 2012.

[27] *IEEE Draft Guide for Electric Power Distribution Reliability Indices,* IEEE Standard P1366/D6, Nov. 2011, pp. 1–40.

[28] M. Angalakudati *et al.*, "Improving emergency storm planning using machine learning," in *Proc. IEEE PES T&D Conf. Expo.*, Apr. 2014, pp. 1–6.

[29] R. Nateghi, S. D. Guikema, and S. M. Quiring, "Comparison and validation of statistical methods for predicting power outage durations in the event of hurricanes," *Risk Anal.*, vol. 31, no. 12, pp. 1897–1906, Dec. 2011.

[30] C. Rudin *et al.*, "Machine learning for the New York City power grid," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 2, pp. 328–345, Feb. 2012.

[31] S. D. Guikema, S.-R. Han, and S. Quiring, "Estimating power outages during hurricanes using semi-parametric statistical methods," in *Proc. Struct. Congr.*, 2008, pp. 1–9.

[32] D. A. Reed, K. C. Kapur, and R. D. Christie, "Methodology for assessing the resilience of networked infrastructure," *IEEE Syst. J.*, vol. 3, no. 2, pp. 174–180, Jun. 2009.

[33] D. A. Reed, M. Powell, and J. Westerman, "Energy supply system performance for hurricane katrina," *J. Energy Eng.*, vol. 136, no. 4, pp. 95–102, 2010.

[34] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Hoboken, NJ, USA: Wiley, Nov. 2000.

[35] Y. Abu-Mostafa, M. Magdon-Ismail, and H. T. Lin, *Learning from Data.* AML Book, 2012.

[36] R. A. Davidson, H. Liu, I. K. Sarpong, P. Sparks, and D. V. Rosowsky, "Electric power distribution system performance in carolina hurricanes," *Natural Hazards Rev.*, vol. 4, no. 1, pp. 36–45, 2003.

[37] H. Liu, R. A. Davidson, and T. V. Apanasovich, "Spatial generalized linear mixed models of electric power outages due to hurricanes and ice storms," *Rel. Eng. Syst. Safety*, vol. 93, no. 6, pp. 897–912, 2008.

[38] S.-R. Han, S. D. Guikema, S. M. Quiring, K.-H. Lee, D. Rosowsky, and R. A. Davidson, "Estimating the spatial distribution of power outages during hurricanes in the Gulf coast region," *Rel. Eng. Syst. Safety*, vol. 94, no. 2, pp. 199–210, 2009.

[39] Y. J. Park, "Reliability assessment of the Florida electric power network system against hurricanes," Ph.D. dissertation, Univ. Florida, Gainesville, FL, USA, 2012.

[40] J. Winkler, L. Dueñas-Osorio, R. Stein, and D. Subramanian, "Performance assessment of topologically diverse power systems subjected to hurricane events," *Rel. Eng. Syst. Safety*, vol. 95, no. 4, pp. 323–336, Apr. 2010.

[41] B. J. Cerruti and S. G. Decker, "A statistical forecast model of weather-related damage to a major electric utility," *J. Appl. Meteorol. Climatol.*, vol. 51, no. 2, pp. 191–204, Feb. 2012.

[42] H. Liu, R. A. Davidson, and T. V. Apanasovich, "Statistical forecasting of electric power restoration times in hurricanes and ice storms," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 2270–2279, Nov. 2007.

[43] R. G. Gallager, *Stochastic Processes: Theory for Applications.* Cambridge, U.K.: Cambridge Univ. Press, 2014.

[44] B. Hajek, *Random Processes for Engineers*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, Mar. 2015.

[45] T. Kailath and H. V. Poor, "Detection of stochastic processes," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2230–2231, Oct. 1998.

[46] I. Dobson, B. A. Carreras, V. E. Lynch, and D. E. Newman, "Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 17, no. 2, p. 026103, 2007.

[47] D. Bienstock, *Electrical Transmission System Cascades and Vulnerability: An Operations Research Viewpoint.* Philadelphia, PA, USA: SIAM, 2015.

[48] I. Dobson, B. A. Carreras, and D. E. Newman, "A branching process approximation to cascading load-dependent system failure," in *Proc. 37th Hawaii Int. Conf. Syst. Sci.*, Jan. 2004, p. 10.

[49] Z. Wang, A. Scaglione, and R. J. Thomas, "A Markov-transition model for cascading failures in power grids," in *Proc. 45th Hawaii Int. Conf. Syst. Sci. (HICSS)*, 2012, pp. 2115–2124.

[50] I. A. Hiskens, "Power system modeling for inverse problems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 3, pp. 539–551, Mar. 2004.

[51] A. Bernstein, D. Bienstock, D. Hay, M. Uzunoglu, and G. Zussman, "Power grid vulnerability to geographically correlated failures—Analysis and control implications," in *Proc. IEEE INFOCOM*, May 2014, pp. 2634–2642.

[52] P. Hines, J. Apt, and S. Talukdar, "Large blackouts in North America: Historical trends and policy implications," *Energy Policy*, vol. 37, no. 12, pp. 5249–5259, Dec. 2009.

[53] M. Ilic, L. Xie, and Q. Liu, "Engineering IT-enabled sustainable electricity services: The tale of two low-cost green azores islands," in *Power Electronics and Power Systems*, vol. 30. New York, NY, USA: Springer-Verlag, Jun. 2013.

[54] H. Xiao and E. Yeh, "Cascading link failure in the power grid: a percolation-based analysis," *Proceeding of IEEE International Conference on Communications (ICC) Workshop*, 2011, pp. 1–6.

[55] D. Bertsimas and G. Mourtzinou, "Transient laws of non-stationary queueing systems and their applications," *Queueing Syst. Theory Appl.*, vol. 25, no. 1, pp. 115–155, Jan. 1997.

[56] J. Guckenheimer and J. M. Ottino, "Foundations for complex systems research in the physical sciences and engineering," Nat. Sci. Found., Colombo, Sri Lanka, NSF Workshop, Tech. Rep., Sep. 2008.

[57] C. D. Brummitt, P. D. H. Hines, I. Dobson, C. Moore, and R. M. D. Souza, "Transdisciplinary electric power grid science," *Proc. Nat. Acad. Sci.*, vol. 110, no. 30, p. 12159, 2013.

[58] L. Gan and S. H. Low, "An online gradient algorithm for optimal power flow on radial networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 625–638, Mar. 2016.

[59] J. Lavaei, D. Tse, and B. Zhang, "Geometry of power flows and optimization in distribution networks," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 572–583, Mar. 2014.

[60] H. Zhu and G. B. Giannakis, "Sparse overcomplete representations for efficient identification of power line outages," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 2215–2224, Nov. 2012.

[61] R. N. Anderson, A. Boulanger, W. B. Powell, and W. Scott, "Adaptive stochastic control for the smart grid," *Proc. IEEE*, vol. 99, no. 6, pp. 1098–1115, Jun. 2011.

[62] D. Deka, S. Backhaus, and M. Chertkov, "Tractable structure learning in radial physical flow networks," in *Proc. IEEE 55th Conf. Decision Control (CDC)*, Vancouver, BC, Canada, Dec. 2016, pp. 6631–6638.

[63] *Committee on Analytical Research Foundations for the Next-Generation Electric Grid; Board on Mathematical Sciences and Their Applications; Division on Engineering and Physical Sciences; National Academies of Sciences, Engineering, and Medicine, Analytic Research Foundations for the Next-Generation Electric Grid.* National Academies Press, Washington, DC, USA, 2016.

[64] T. Overbye, P. Sauer, C. DeMarco, B. Lesieutre, and M. Venkatasubramanian, "Using PMU Data to Increase Situational Awareness," Dept. Power Syst. Eng. Res. Center, Final Project Rep. PSERC Publication, Tech. Rep. 10–16, Sep. 2010.

[65] A. von Meier, D. Culler, A. McEachern, and R. Arghandeh, "Micro-synchrophasors for distribution systems," in *Proc. IEEE PES 5th Innov. Smart Grid Technol. (ISGT) Conf.*, Feb. 2014, pp. 1–5.

[66] H. Tram, "Technical and operation considerations in using smart metering for outage management," in *Proc. IEEE/PES Transmiss. Distrib. Conf. Expo.*, Apr. 2008, pp. 1–3.

[67] G. Madingou, M. Zarghami, and M. Vaziri, "Fault detection and isolation in a DC microgrid using a central processing unit," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Feb. 2015, pp. 1–5.

[68] "Post Sandy enhancement plan," Consolidated Edison Co. New York and Orange and Rockland Utilities, Tech. Rep., Jun. 2013.

[69] National Center for Environmental Information. (Sep. 8, 2016). *Climate Monitoring.* [Online]. Available: http://www.ncdc.noaa.gov/climate-monitoring/

[70] E. Berger. (2016). *Meet Deep Thunder: IBM's Next Step in the Automation of Forecasting.* [Online]. Available: http://arstechnica.com/science/2016/06/meet-deep-thunder-ibmsnextstep-in-the-automation-of-forecasting/

[71] J. Foerster, "Hyper-local weather data provides accuracy to Central Hudson Gas & Electric," *Power Eng.*, vol. 120, no. 4, Apr. 2016.

[72] San Diego Gas & Electric. (2016). *SDG&E Weather Awareness System.* [Online]. Available: http://www.sdgeweather.com/

[73] Federal Emergency Management Agency (FEMA). (2016). *FEMA Flood Map Service Center.* [Online]. Available: https://msc.fema.gov/portal

[74] N. Lin, K. Emanuel, M. Oppenheimer, and E. Vanmarcke, "Physically based assessment of hurricane surge threat under climate change," *Nature Clim. Change*, vol. 2, no. 6, pp. 267–462, 2012.

[75] F. Zhang, Y. Weng, J. A. Sippel, Z. Meng, and C. H. Bishop, "Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter," *Monthly Weather Rev.*, vol. 137, pp. 2105–2125, Jul. 2009.

[76] Connecticut Light & Power. *Outage Map.* [Online]. Available: http://www.cl-p.com/outage/outagemap.aspx

[77] "2012 Electric reliability performance report," Dept. Public Ser. State New York, Office Elect., Gas, Water, Tech. Rep., Jun. 2013.

[78] S. Ertekin, C. Rudin, and T. H. McCormick, "Reactive point processes: A new approach to predicting power failures in underground electrical systems," *Ann. Appl. Statist.*, vol. 9, no. 1, pp. 122–144, 2015.

[79] R. D. Christie, "Statistical classification of major event days in distribution system reliability," *IEEE Trans. Power Del.*, vol. 18, no. 4, pp. 1336–1341, Oct. 2003.

[80] T. D. O'Rourke, S.-S. Jeon, R. T. Eguchi, and C. K. Huyck, "Advanced GIS for loss estimation and rapid post-earthquake assessment of building damage," in *Proc. Res. Prog. Accomplishments (MCEER)*, 2001, pp. 157–164.

[81] P. H. Larsen, K. H. LaCommare, J. H. Eto, and J. L. Sweeney, "Assessing changes in the reliability of the U.S. electric power system," Lawrence Berkeley Nat. Lab., Berkeley, CA, USA, Tech. Rep. LBNL188741, Aug. 2015.

[82] M. E. J. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemp. Phys.*, vol. 46, pp. 323–351, Dec. 2005.

[83] D. Divan, G. A. Luckjiff, W. E. Brumsickle, J. Freeborg, and A. Bhadkamkar, "A grid information resource for nationwide real-time power monitoring," *IEEE Trans. Ind. Appl.*, vol. 40, no. 2, pp. 699–705, Mar. 2004.

[84] S. Karnouskos, O. Terzidis, and P. Karnouskos, "An advanced metering infrastructure for future energy networks," in *New Technologies, Mobility and Security*, H. Labiod and M. Badra, Eds. Amsterdam, The Netherlands: Springer-Verlag, 2007, pp. 597–606.

[85] M. Uddin, A. Kuh, A. Kavcic, T. Tanaka, and D. P. Mandic, "Grid monitoring: Bounds on performances of sensor placement algorithms," in *Proc. 3rd Int. Conf. Smart Grids, Green Commun. IT Energy-Aware Technol. (ENERGY)*, Lisbon, Portugal, Mar. 2013, pp. 89–95.

[86] S. Meliopoulos, V. Madani, D. Novosel, and G. Cokkinides, "Synchrophasor measurement accuracy characterization," North American Synchrophasor Initiative Perform. & Standards Task Team (CERTS), Tech. Rep., 2010.

[87] J. Peppanen, M. J. Reno, M. Thakkar, S. Grijalva, and R. G. Harley, "Leveraging AMI data for distribution system model calibration and situational awareness," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 2050–2059, Jul. 2015.

[88] R. E. Brown, S. Gupta, R. D. Christie, S. S. Venkata, and R. Fletcher, "Distribution system reliability assessment: Momentary interruptions and storms," *IEEE Trans. Power Del.*, vol. 12, no. 4, pp. 1569–1575, Oct. 1997.

[89] D. A. Reed, "Electric utility distribution analysis for extreme winds," *J. Wind Eng. Ind. Aerodyn.*, vol. 96, no. 1, pp. 123–140, 2008.

[90] Q. Yan, T. Dokic, and M. Kezunovic, "Predicting impact of weather caused blackouts on electricity customers based on risk assessment," in *Proc. IEEE Power Energy Soc. General Meeting*, Boston, MA, USA, Jul. 2016, pp. 1–5.

[91] M. Bruneau *et al.*, "A framework to quantitatively assess and enhance the seismic resilience of communities," *Earthquake Spectra*, vol. 19, no. 4, pp. 733–752, Nov. 2003.

[92] S. E. Chang and M. Shinozuka, "Measuring improvements in the disaster resilience of communities," *Earthquake Spectra*, vol. 20, no. 3, pp. 739–755, 2004.

[93] R. Arghandeh, A. von Meier, L. Mehrmanesh, and L. Mili, "On the definition of cyber-physical resilience in power systems," *Renew. Sustain. Energy Rev.*, vol. 58, pp. 1060–1069, May 2016.

[94] J. P. Watson *et al.*, "Conceptual framework for developing resilience metrics for the electricity, oil, and gas sectors in the United States," Sandia Nat. Lab., New Mexico, CA, USA, Sandia Rep. SAND2014-18019, Sep. 2015.

ABOUT THE AUTHORS

Chuanyi Ji (Senior Member, IEEE) received the B.S. degree (honors) from Tsinghua University, Beijing, China, in 1983, the M.S. degree from the University of Pennsylvania, Philadelphia, PA, USA, in 1986, and the Ph.D. degree from the California Institute of Technology, Pasadena, CA, USA, in 1992, all in electrical engineering.

She was an Assistant and Associate Professor at the Rensselaer Polytechnic Institute, Troy, NY, USA, from 1991 to 2001. She was a visitor/consultant at Bell Labs Lucent, Murray Hill, NJ, USA, in 1999, and a visiting faculty member at the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2000. She is an Associate Professor with the Georgia Institute of Technology, Atlanta, GA, USA, which she joined in 2001. Her research is in large-scale networks, machine learning, and data analytics.

Dr. Ji received several awards, including a CAREER award from the National Science Foundation (NSF) and an Early CAREER award from the Rensselaer Polytechnic Institute.

Yun Wei received the B.S. degree from Peking University, Beijing, China, in 2010 and the Ph.D. degree from the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA, in 2017.

He is a quantitative researcher with the General Welfare Group/Cooperfund. His current research interests focus on time-series data analytics, machine learning, and advanced algorithmic modeling such as fractal/scaling algorithms and wave pattern models. His recent research includes management and control of heterogeneous and large networks using machine learning techniques, network resilience analysis, and development of analytical approaches using large-scale real data.

H. Vincent Poor (Fellow, IEEE) received the Ph.D. degree in electrical engineering and computer science from Princeton University, Princeton, NJ, USA, in 1977.

From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign, Urbana, IL, USA. Since 1990 he has been on the faculty at Princeton University, where he is the Michael Henry Strater University Professor of Electrical Engineering. During 2006–2016, he served as Dean of Princeton's School of Engineering and Applied Science. He has also held visiting appointments at several other institutions, most recently at the University of California Berkeley and Cambridge University. His research interests are in the areas of information theory and signal processing, and their applications in wireless networks and related fields. Among his publications in these areas is the recent book *Information Theoretic Security and Privacy of Information Systems* (Cambridge, U.K.: Cambridge Univ. Press, 2017).

Dr. Poor is a Member of the National Academy of Engineering and the National Academy of Sciences, and is a Foreign Member of the Royal Society. In 1990, he served as President of the IEEE Information Theory Society, and in 2004–2007 as the Editor-in-Chief of the IEEE Transactions on Information Theory. Recent recognition of his work includes the 2016 John Fritz Medal, the 2017 IEEE Alexander Graham Bell Medal, Honorary Professorships at Peking University and Tsinghua University, both conferred in 2016, a D.Sc. honoris causa from Syracuse University in 2017, and election as a Foreign Member of the National Academy of Engineering of Korea and an Honorary Member of the National Academy of Sciences of Korea, both in 2017.