

2020

Can the Markovian influence graph simulate cascading resilience from historical outage data?

Kai Zhou

Iowa State University, kzhou@iastate.edu

Ian Dobson

Iowa State University, dobson@iastate.edu

Zhaoyu Wang

Iowa State University, wzy@iastate.edu

Follow this and additional works at: https://lib.dr.iastate.edu/ece_conf



Part of the [Power and Energy Commons](#)

Recommended Citation

Zhou, Kai; Dobson, Ian; and Wang, Zhaoyu, "Can the Markovian influence graph simulate cascading resilience from historical outage data?" (2020). *Electrical and Computer Engineering Conference Papers, Posters and Presentations*. 91.

https://lib.dr.iastate.edu/ece_conf/91

This Conference Proceeding is brought to you for free and open access by the Electrical and Computer Engineering at Iowa State University Digital Repository. It has been accepted for inclusion in Electrical and Computer Engineering Conference Papers, Posters and Presentations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Can the Markovian influence graph simulate cascading resilience from historical outage data?

Abstract

It is challenging to simulate the cascading line outages that can follow initial damage to the electric power transmission system from extreme events. Instead of model-based simulation, we propose using a Markovian influence graph driven by historical utility data to sample the cascades. The sampling method encompasses the rare, large cascades that contribute greatly to the blackout risk. This suggested new approach contributes a high-level simulation of cascading line outages that is driven by standard utility data.

Disciplines

Power and Energy

Comments

This is a pre-print of the proceeding Zhou, Kai, Ian Dobson, and Zhaoyu Wang. "Can the Markovian influence graph simulate cascading resilience from historical outage data?" (2020). Posted with permission.

Can the Markovian influence graph simulate cascading resilience from historical outage data?

Kai Zhou Ian Dobson Zhaoyu Wang

Department of Electrical and Computer Engineering, Iowa State University, Ames IA, USA

Abstract—It is challenging to simulate the cascading line outages that can follow initial damage to the electric power transmission system from extreme events. Instead of model-based simulation, we propose using a Markovian influence graph driven by historical utility data to sample the cascades. The sampling method encompasses the rare, large cascades that contribute greatly to the blackout risk. This suggested new approach contributes a high-level simulation of cascading line outages that is driven by standard utility data.

I. INTRODUCTION

Historical transmission line outage data that is routinely collected by utilities can be processed and grouped into cascades of outages [1]. These processed cascades are the observational bedrock for the study of cascading failure, since they occurred in practice. However, if one assumes some initial outages and seeks to predict the probabilistic extent of further cascading, ranging from no further outages to blackouts, the historical cascades are limiting: the particular initial outages may not occur in the historical cascades, and even if they do occur once or twice, it is only one or two samples of the possible cascading outcomes. To address this problem, we propose using the Markovian influence graph to describe the statistics of the historical outage data, and then sample from the Markovian influence graph to simulate the consequences of some assumed initial outages. This gives a high-level and flexible statistical model of cascading that can be driven by standard utility data. In suggesting this approach, we are motivated by the resilience problem of estimating the cascading that can follow damage to the power transmission system in an extreme event [2]–[4].

Extreme events such as storm, fire, or earthquake can damage multiple power system components. Then further power system components can outage in a cascade. Usually the cascading only outages components without damaging them, but the cascading does make the blackout more widespread and impactful, and can seriously hinder the subsequent recovery from the event. There is considerable expertise modeling probabilistic power system component failure under extreme conditions of wind, flooding, icing, earthquake and fire. However, the cascading phase of resilience is much less well

We thank Benjamin Carreras of BACV Solutions, Oak Ridge TN, David E. Newman of University of Alaska-Fairbanks, and José-Miguel Reynolds-Baredo of Universidad Carlos III de Madrid for providing the OPA results used to determine load shed. We thank Bonneville Power Administration for making publicly available the outage data that made this paper possible. The analysis and any conclusions are strictly those of the authors and not of BPA. We thank anonymous reviewers for their comments improving the paper. This work was supported by USA NSF grants 1609080 and 1735354.

characterized. Given the initially damaged components, one can simulate the cascading using a model-based simulation. While useful, simulation only captures a limited subset of approximated cascading mechanisms. The alternative that we suggest and explore in this paper uses a Markovian influence graph driven by historical utility data to generate samples of the cascaded transmission lines. Throughout the paper we are interested in properly sampling from the largest cascades since these dominate the risk [5], because straightforward sampling does not work well for the larger cascades.

There is a large literature on model-based simulation of cascading (reviewed in [6], [7]), and substantial literature on influence or interaction graphs and fault chains driven by simulated data (reviewed in [8], [9]). To our knowledge, the only previous work on influence graphs driven by real data is [8], [10].

The need for higher-level statistical simulation of cascading arose from broader studies of the multiple phases of resilience. For example, Romero [2] optimized investments to improve resilience to earthquakes, and discussed but did not model the cascading phase of resilience. Recently Kelly-Gorham [3], [4] proposed a high-level statistical method driven largely by observed statistics called CRISP to quantify power transmission system overall resilience in all its phases. CRISP models the cascading phase of resilience by sampling from a probability distribution of the total number of lines out based on historical data. Then, in [4], given the number of lines out, the lines outaged in a cascade are chosen in accordance with an observed probability distribution [11] of network distance between cascaded line outages.

In this paper, given the initial outages, we aim to more accurately capture the spatial and interdependence statistics of which lines outage by simulating the recently developed Markovian influence graph driven by utility data. We also discuss possibilities for determining the load shed from the samples of cascaded lines and hence the risk of a widespread blackout caused by the extreme event and the cascading.

II. HISTORICAL DATA AND INFLUENCE GRAPH

This section summarizes the historical outage data and the Markovian influence graph used in this paper; all detail is referred to [8], [11].

Detailed transmission line outage data are routinely collected by utilities worldwide, such as in North America's Transmission Availability Data System (TADS). Here, as an illustration, we use some publicly available historical line outage data [12] recorded over fourteen years by a utility in

the WECC interconnection of the USA. The data includes the forced line outages, names of the outaged lines, and their start time to the nearest minute. There are 10942 forced outages and over 500 different transmission lines outaging in the data with rated voltages ranging from 69 kV to 500 kV.

The line outages in the historical utility data are grouped into cascades and then divided into generations in each cascade using a simple method based on the outage timing that is described in [11]. This data processing combines into 6687 cascades a variety of dependent outages, including the initiating outages, cascading interactions within the network, and outages occurring in faster succession with a common cause such as bad weather. Thus the extreme conditions of interest in this paper are included in the data and its processing.

The cascades of outages divided into generations are then processed to form the Markovian influence graph. The Markovian influence graph is a Markov chain with discrete time steps and a discrete state space. Each new time step corresponds to a new generation of cascading. Each observed line outage (or a set of line outages that occur nearly simultaneously) forms a state of the Markov chain, and statistics of the pairwise interaction of states in successive generations are extracted. For example, given a specific line outage in a cascade generation, the probabilities of transition to other specific line outages in the next generation of the cascade are estimated from the transitions that occurred in the data. That is, the estimated transition probabilities represent the influence between successive line outages. At each state, there is a probability of the cascade stopping, which is represented as a transition to a special state with no line outages that is indicated by the empty set of outages $\{\}$. After the first transition to $\{\}$, the influence graph remains stopped at $\{\}$ for all subsequent generations.

A simple 3-line toy example of forming the Markovian influence graph from outage data is shown in Figure 1. The Markov chain indicated by the influence graph at the bottom of Figure 1, models the statistics of outage data at the top of Figure 1. This example uses a constant transition matrix for ease of illustration, but the practical Markovian influence graph has varying transition matrices for different generations. This is because the transitions vary, especially between the first transition and subsequent transitions and because particular care is taken in forming the influence graph that the overall stopping probabilities at each generation match those in the data. The formation of the influence graph is somewhat intricate [8] and is designed to leverage the sparse data as much as feasible, particularly for the higher cascade generations that are rarer in the observed data.

The Markovian influence graph resulting from the historical outage data is a Markov chain with 1094 states. 543 of the states are single line outages, and the rest of the states have several line outages that occurred nearly simultaneously.

The Markovian influence graph in [8] exploits the asymptotic properties of the Markov chain to find the transmission lines most involved in large blackouts. In this paper we sample from the Markov chain to find cascades consistent with the

assumed initial failures and the statistics of how cascades spread on the network.

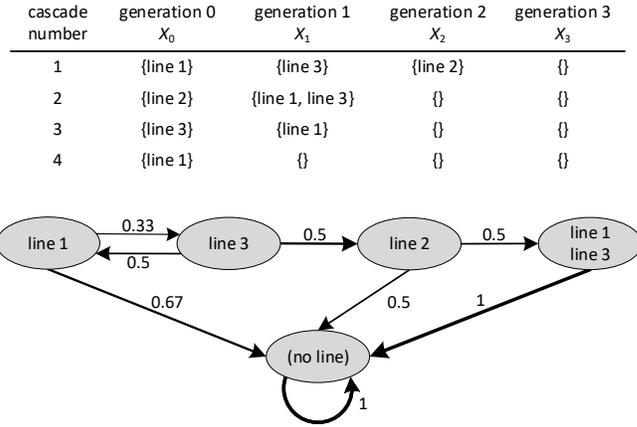


Fig. 1. The influence graph (bottom graph) formed from the line outage data (top table) for a 3-line system. Reproduced from [8] under CC BY 4.0 license.

III. SAMPLING CASCADES WITH THE INFLUENCE GRAPH

Let Y_0 be the set of initially failed lines that are damaged by the extreme event. We express Y_0 as a disjoint union of m Markov chain states:

$$Y_0 = x_0^{(1)} \cup x_0^{(2)} \cup \dots \cup x_0^{(m)} \quad (1)$$

Consider the state $x_0^{(r)}$ in Y_0 with $1 \leq r \leq m$. Let the r th Markov chain starting from state $x_0^{(r)}$ but subsequently avoiding any initially failed states be $X_0^{(r)}, X_1^{(r)}, \dots$. That is, $P[X_0^{(r)} = x_0^{(r)}] = 1$ and $P[X_k^{(r)} \in Y_0] = 0$ for $k > 0$. (The transition matrix for the Markov chain $X_0^{(r)}, X_1^{(r)}, \dots$ is easily obtained by preventing transitions to states in Y_0 by deleting the columns of the transition matrix corresponding to states in Y_0 and renormalizing.)

We write $|x|$ for the number of line outages in state x . The number of lines out in the r th chain is

$$N^{(r)} = \sum_{k=0}^{\infty} |X_k^{(r)}| \quad (2)$$

and the total number of lines out is

$$N = \sum_{r=1}^m N^{(r)} \quad (3)$$

Note that (2) and (3) neglect any repeats of lines out within or between chains.

A. Simulating the influence graph

We first describe a straightforward but inferior way to do the simulation. For the r th chain we need to simulate $X_0^{(r)}, X_1^{(r)}, \dots$ from its starting state $x_0^{(r)}$ until it stops. That is, the simulation produces a series of states $x_0^{(r)}, x_1^{(r)}, x_2^{(r)}, \dots$ until it stops by transitioning to the empty state $\{\}$. Suppose state $x_j^{(r)}$ is produced at step j . Then the next state is

produced as follows: Let $e_j^{(r)}$ be the row vector with a one at the index of state $x_j^{(r)}$ and zeros elsewhere. Let P_k be the transition matrix from generation k to $k+1$. Then $e_j^{(r)} P_k$ is a probability distribution over the states not in Y_0 ¹. Sample from this probability distribution to obtain the state $x_{k+1}^{(r)}$. Thus $x_0^{(r)}, x_1^{(r)}, x_2^{(r)}, \dots$ are produced.

The problem with this straightforward way to do the simulation is that it will mainly sample the frequent short cascades with few line outages, so that a huge number of samples is needed to accurately estimate the longer cascades. An advantage of the influence graph is that it can be easily modified to sample more uniformly over the range of the possible cascades by manipulating the cascade stopping probabilities. Instead of allowing chains to stop by themselves, the stopping is inhibited until a maximum number of cascade generations g_{\max} is simulated, and then the chain stops. At each generation before g_{\max} , the line outages are recorded, and, although the chain does not stop, the probability that the state would have stopped is recorded. This gives many samples of the number of line outages for each of the generations $0, 1, 2, \dots, g_{\max}$, and these samples range from a small to a large number of line outages. And the probability of stopping at each intermediate length cascade can be calculated.

We now give the details of this improved simulation. Suppose the r th chain is simulated and is at state $x_k^{(r)}$ at generation $k < g_{\max}$. When the simulation samples from the probability distribution to obtain the next state $x_{k+1}^{(r)}$, it is easy to prohibit the choice $x_{k+1}^{(r)} = \{\}$ that would stop the chain. (This is equivalent to zeroing the probability of transition to $\{\}$ and renormalizing the probabilities of the other transitions.) It is also straightforward to record $x_k^{(r)}$ (which contains the lines outaged in generation k), and the probability $\sigma_k^{(r)} = \text{P}[\text{transition from } x_k^{(r)} \text{ to } \{\}]$ that the chain stops when the state is $x_k^{(r)}$. $\sigma_k^{(r)}$ is the entry in the first column of the transition matrix P_k corresponding to $x_k^{(r)}$. The probability that the r th chain has exactly k generations is

$$q_k^{(r)} = (1 - \sigma_0^{(r)})(1 - \sigma_1^{(r)}) \dots (1 - \sigma_{k-1}^{(r)})\sigma_k^{(r)} \quad (4)$$

More precisely, we have simulated (realized) one particular sequence of states $x_1^{(r)}, x_2^{(r)}, \dots, x_k^{(r)}$ that avoid stopping. Now, conditioned on the states that do happen occurring in this sequence, we compute in the Markov chain that does not avoid stopping the probability of stopping at generation k with (4).

We indicate the first simulation of the r th chain by the superscript $(r; 1)$. We perform the first simulation of the r th chain up to generation g_{\max} and extract results for each generation $k \leq g_{\max}$. For generation k , the total number of lines out is

$$n_k^{(r;1)} = \sum_{j=0}^k |x_j^{(r;1)}| \quad (5)$$

¹During the simulation, however, we allow lines not in Y_0 to outage again in successive generations except the next generation.

and the probability of $n_k^{(r;1)}$ lines out is equal to $q_k^{(r;1)}$, since the number of lines out increases at each non-stopping generation. Repeating the simulation of the r th chain t times for the same initial state $x_0^{(r)}$ gives different sequences $x_0^{(r;s)}, x_1^{(r;s)}, x_2^{(r;s)}, \dots$ for $s = 1, 2, \dots, t$, generating many samples of the number of line outages $n_k^{(r;s)}$ and their probabilities $q_k^{(r;s)}$ for $s = 1, 2, \dots, t$ and $k = 0, 1, \dots, g_{\max}$. All these results are combined to give the distribution of the number of line outages $N^{(r)}$ in the simulations of the r th chain:

$$\text{P}[N^{(r)} = v] = \frac{1}{t} \sum_{k=0}^{g_{\max}} \sum_{s=1}^t I[n_k^{(r;s)} = v] q_k^{(r;s)} \quad (6)$$

where the indicator function $I[\cdot]$ limits the sums in (6) to the results giving v line outages. Thus (6) is the average of all the probabilities corresponding to the t possible occurrences of v line outages in the simulations of the r th chain.

Then, according to (3) and assuming the chains are independent, we evaluate the distribution of the total number of lines out N by convolving the distributions $N^{(1)}, N^{(2)}, \dots, N^{(m)}$. The convolution is done by multiplying probability generating functions:

$$\text{E}[z^N] = \text{E}[z^{(\sum_{r=1}^m N^{(r)})}] = \prod_{r=1}^m \text{E}[z^{N^{(r)}}] \quad (7)$$

The coefficient of z^v in $\text{E}[z^N]$ is the probability $\text{P}[N = v]$.

IV. PROBABILITY DISTRIBUTION OF LOAD SHED

The load shedding of a cascade is denoted as L . We want to estimate f_L , the probability distribution of load shed. We do this by conditioning on the number of line outages.

The number of line outages N ranges from ℓ_0 to ℓ_{\max} , where ℓ_0 = number of lines in Y_0 . We partition the range of N into K disjoint bins B_1, B_2, \dots, B_K so that

$$\{\ell_0, \ell_0 + 1, \ell_0 + 2, \dots, \ell_{\max}\} = B_1 \cup B_2 \cup \dots \cup B_K \quad (8)$$

We use the following subsections to obtain $f_{L|N \in B_\kappa}$ the distribution of the load shed given that the number of lines out are in bin B_κ . The bins (8) are chosen large enough so that there is sufficient data in each bin to be able to approximate $f_{L|N \in B_\kappa}$.

From the distribution of N provided in section III, we can easily evaluate the bin probabilities:

$$b_\kappa = \text{P}[N \in B_\kappa] = \sum_{v \in B_\kappa} \text{P}[N = v] \quad (9)$$

The idea is to evaluate the distribution of load shed f_L by conditioning on the number of lines in the bins:

$$f_L = \sum_{\kappa=1}^K b_\kappa f_{L|N \in B_\kappa} \quad (10)$$

We now use the OPA simulation to approximate $f_{L|N \in B_\kappa}$.

A. Load shed given the number of lines out

Given that the number of lines N outaged after cascading are in bin B_κ , we want to obtain the distribution of load shed $f_{L|N \in B_\kappa}$. We use a probability distribution of load shed because we are trying to estimate the risk of a future extreme event, and the power system loading condition, generator dispatch and maintenance status for a future event are uncertain and variable. This variability will produce different load sheds for the same line outages, or the same number of line outages.

The OPA model [13]–[16] has been validated to approximate well the observed bulk statistics of blackouts of the WECC [17], [18]. Here, noting that our historical data is from part of WECC, we use OPA results on a 1553 bus model of WECC to generate the conditional distributions of load shed $f_{L|N \in B_\kappa}$, $\kappa = 1, 2, 3, \dots, 12$. Note that OPA is a long-term simulation that samples from a variety of grid loading conditions. The OPA results consist of 58 903 cascades. Each cascade yields the load shed and the number of lines out.

The OPA results are easily sorted into the bins B_1, B_2, \dots, B_{12} according to the number of lines outaged in each cascade. Bin B_κ has κ line outages for $1 \leq \kappa \leq 11$, and bin B_{12} has 12 or more line outages. Each bin has at least 83 data points. The empirical distribution for load shed in bin B_κ is fit with the lognormal distribution $f_{L|N \in B_\kappa}$. Figure 2 shows three of these fits. The data points that have a fraction of load shed less than 0.01 are excluded. The mean μ and standard deviation σ of the lognormal distributions for the 12 bins are shown in Table I. The mean and standard deviation increase as the number of line outages in cascades increase, as expected. Moreover, the Kolmogorov-Smirnov test for each bin’s fitting has a p-value at least 0.1, so these fits are statistically significant.

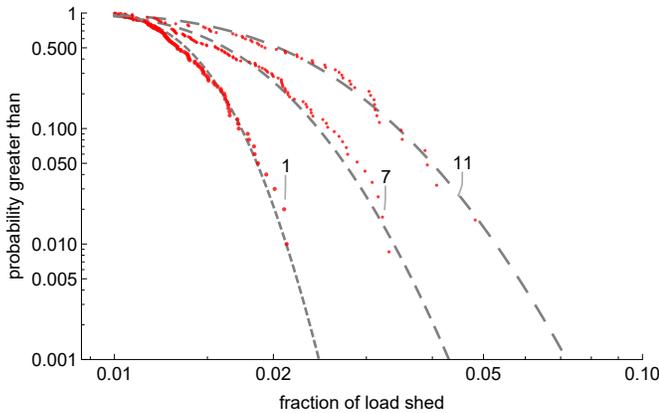


Fig. 2. The survival functions of the distribution of fraction of load shed for cascades with 1, 7, or 11 line outages.

V. RESULTS

A. Simulation of line outages

We use the improved sampling of section III-A to sample cascades from the Markovian influence graph formed from the utility data in section II. Specifically, starting with assumed 3

TABLE I
PARAMETERS OF THE LOGNORMAL DISTRIBUTIONS OF THE LOAD SHED GIVEN THE NUMBER OF LINE OUTAGES IN THE κ TH BIN

κ	1	2	3	4	5	6	7	8	9	10	11	12
μ	-4.29	-4.26	-4.27	-4.22	-4.21	-4.18	-4.12	-4.18	-4.05	-3.99	-3.89	-3.75
σ	0.19	0.22	0.23	0.26	0.30	0.32	0.33	0.31	0.32	0.33	0.41	0.42

initial outages, we simulate 100 cascades up to $g_{\max} = 100$ generations. Since the simulation also records data for each cascade stopping at any generation before 100 generations, this is equivalent to simulating 10 000 cascades, in which 100 cascades have 1 generation, 100 have 2 generations, and so on.

To contrast the improved sampling with straightforward sampling, we also simulate 10 000 cascades with the same initial outages using the straightforward sampling method of simply simulating until the cascade stops, with no special control of the stopping. The two simulations have close execution times. Figure 3 shows that the survival functions match except for some variability in the tail due to limited samples from the straightforward sampling. With the same simulation time, the improved sampling has two benefits: it has smaller standard deviations and generates more large cascade samples. For example, the standard deviation of the probability that cascades have more than 30 line outages is 0.00004 for improved sampling, and 0.0004 for straightforward sampling. As the number of line outages increases, this advantage is even more significant. The straightforward sampling focuses on the small cascades and does not sample enough large cascades to accurately estimate the large cascades. In contrast, the improved sampling samples uniformly across a full range of cascade sizes to better estimate a longer tail. The Markovian influence graph flexibly allows this improved sampling, addressing the straightforward sampling problem common in the literature of inherently undersampling large cascades.

Although in this paper we only estimate the distribution of the number of lines out, there is a wealth of detailed information in the simulated cascades that could be useful.

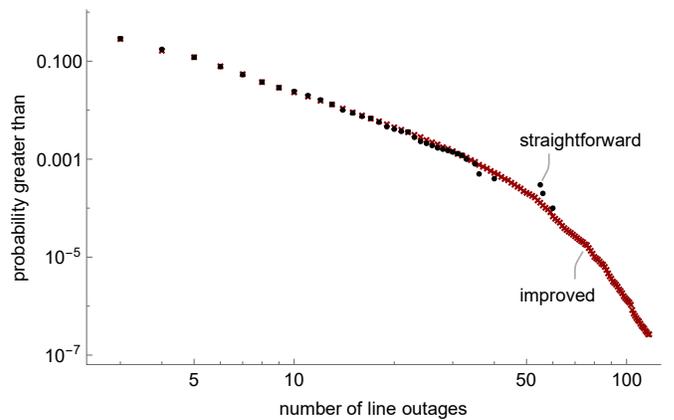


Fig. 3. The survival function of the number of line outages N given 3 initial outages using the improved sampling (red crosses) and the straightforward sampling (black dots).

B. Distribution of load shed

After estimating distribution of the number of line outages N , we proceed to estimate the distribution of load shed using the method described in Section IV. Subsection IV-A calculates the conditional lognormal load shed distributions $f_{L|N \in B_1}, f_{L|N \in B_2}, \dots, f_{L|N \in B_{12}}$. The distribution of N gives the bin weights b_1, b_2, \dots, b_{12} according to (9). Then the distribution of load shed f_L is the mixture of $f_{L|N \in B_1}, f_{L|N \in B_2}, \dots, f_{L|N \in B_{12}}$ weighted by b_1, b_2, \dots, b_{12} as described by (10).

Figure 4 shows the survival function of the distribution of load shed f_L given 3 initial outages (red solid curve). In Figure 4, we also vary the number of initial outages to simulate different initial line damage scenarios. As the number of initial outages increases, the probability of large load shed increases.

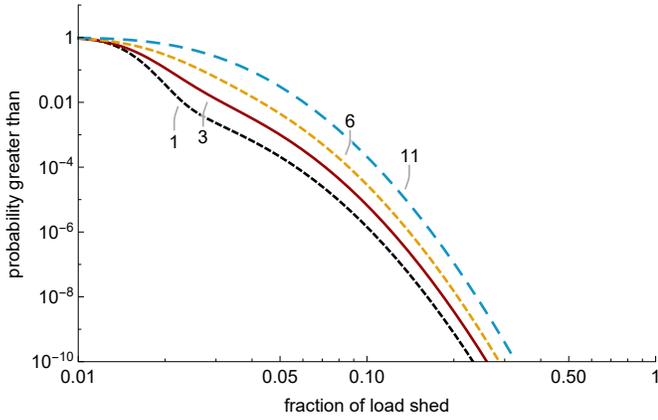


Fig. 4. Survival functions of load shed with 1, 3, 6, or 11 initial outages.

VI. COMPARING SIMULATION DRIVEN BY HISTORICAL DATA WITH MODEL-BASED SIMULATION

This section describes and contrasts the strengths and weaknesses of model-based simulation and simulation of the Markovian influence graph driven by historical data.

Realism: A major limitation is that model-based simulations are practically constrained to approximate a limited subset of cascading mechanisms. The Markovian influence graph driven by historical data uses the statistics of real cascades, which encompass all the cascading mechanisms encountered in the historical period. It produces many cascades not observed in the real cascades. However, the Markovian influence graph does not describe the pairwise interactions between outages that could happen but that did not happen in the historical period. The power grid slowly changes over the historical period as components age and upgrades to the grid and operational procedures are made. The Markovian influence graph pools together all the interactions in the power system over the historical period. For example, if an interaction was mitigated half way during the historical period, it still contributes a possible interaction to the Markovian influence graph.

Validation: When used to predict cascades, model-based simulation can produce cascades that are often judged to be credible, but most model-based simulations are not yet validated against historical cascade data. (One of the exceptions is the OPA simulation used in subsection IV-A, which is validated with WECC data in [17], [18].) An appropriate validation is reproducing the form of cascade statistics, and there is notable progress towards this goal [19], [20]. On the other hand, the Markovian influence graph describes the historical statistics of successive line outages, and reproduces the statistics of numbers of lines out [8], so important aspects of validation are inherent or already checked. There are additional assumptions in the processing of the historical data and the influence graph formulation, and some of these issues are discussed below. However, the more subtle validations of the Markovian influence graph would seem to require more elaborate approaches to statistical validation.

Sampling of grid conditions: Another requirement, which is not always satisfied in the literature, is that model-based simulations should sample appropriately from a range of grid operating conditions [21]. Historical data inherently samples all the actual grid conditions encountered over the observed time, and this is often an appropriate sampling.

Sampling of cascades: Predicted cascading is inherently probabilistic due to the many interactions and protection actions that involve thresholding in an uncertain environment. Note that even “deterministic” model-based simulations can sample from cascade possibilities by randomizing the grid conditions. As regards sampling technique, the Markovian influence graph easily allows computing the rarer but riskier long cascades while tracking the outcomes and probabilities of all the truncations of the long cascades, as explained in subsection III-A. A corresponding advantage in computing the largest cascades can be achieved for model-based simulation using splitting [22] or other methods.

Markov assumption: The Markovian influence graph only describes the statistics of successive Markov states in the historical cascades. Each Markov state is a specific line outage or set of line outages. The issue is the extent to which one can assume that knowing the state in a cascade generation is sufficient to approximate the statistics of which state is in the next generation. This is a pragmatic but fairly strong assumption.

Limited data: The Markovian influence graph is formed from historical data, which is limited in extent, especially for the higher cascade generations. This limitation can be partially mitigated [8], but not eliminated. In practice the higher generations are combined together in some ways to get sufficient data. Model-based simulations can, if not too detailed, produce larger amounts of cascading data.

Commonality between cascades: In this paper, the Markovian influence graph describes the statistics of all types of cascades, but some of these may not be the cascades of interest. That is, there is an assumption that the same set of probabilistic cascading interactions tend to occur for all cascades. In particular, statistical patterns in small cascades are

to some extent extrapolated to large cascades. It is certainly possible to restrict the historical data to the subset of cascades of interest if the subset is large enough, but there is the tradeoff that as data set becomes smaller, estimation becomes more uncertain. In model-based simulation it seems easier to restrict the cascades simulated, but the challenges of validation for the restricted subset of cascades remain.

VII. COMBINING HISTORICAL AND SIMULATED DATA

We observe that the Markovian influence graph could be used to combine historical cascading data from utilities with cascading data produced by model-based simulation. One method is to simply combine the cascades from both data sets and then form the Markovian influence graph from the combined data set. This method would weight the historical and model-based cascades according to their respective amounts of data. If a different weighting is desired; for example, an equal weight to historical and model-based cascades despite differing amounts of data, then this can be easily achieved by forming two separate Markovian influence graphs and taking their weighted average.

In particular, suppose that the historical cascades give a Markovian influence graph with transition matrices P_k^{hist} , $k = 1, 2, \dots$ and the model-based cascades give a Markovian influence graph with transition matrices P_k^{model} , $k = 1, 2, \dots$. We form the union of the states of the historical and model-based influence graphs and add rows and columns of zeros to the transition matrices to expand them so that they describe transitions between the union of the states. This gives transition matrices $P_k^{\text{histexpand}}$ and $P_k^{\text{modelexpand}}$, $k = 1, 2, \dots$. Then, if the desired weight on the historical data is w^{hist} with $0 < w^{\text{hist}} < 1$, then the weighted average Markovian influence graph has transition matrices

$$w^{\text{hist}} P_k^{\text{histexpand}} + (1 - w^{\text{hist}}) P_k^{\text{modelexpand}}, \quad k = 1, 2, \dots$$

This construction relies on the fact that the weighted average of stochastic matrices are stochastic matrices. Combining the two influence graphs as described above is straightforward, assuming that the components that outage in the cascades are described in the same way and in the same format.

VIII. CONCLUSION

This paper suggests a new form of cascading simulation driven by the detailed transmission line outage data that is routinely collected by utilities. This historical outage data is first processed into cascades and generations within cascades, and then used to form the Markovian influence graph that describes the statistics of outages in successive cascade generations as a Markov chain. Some initial line outages are assumed, and in this paper these are the lines damaged by some extreme events, such as weather, fire, icing, or earthquake. Our immediate aim is to simulate and quantify the cascading of line outages after the initial damage. The Markovian influence graph is sampled to produce the simulated cascades. The simulated cascades are statistically similar to but more variable than the cascades in the historical data. The Markovian influence graph easily

allows improved sampling that is more uniform across all sizes of cascades, and this gives better estimates of the large cascades that are rare but significant for cascade risk.

The Markovian influence graph produces cascades of specific line outages but no direct estimates of load shed. We show one way to estimate load shed by using a model-based simulation, OPA, to evaluate the probability distribution of load shed conditioned on the number of line outages. The distribution of load shed is then a weighted sum of these conditional distributions, with the weights determined by the line outage statistics produced by the Markovian influence graph. Other methods of estimating load shed can be developed and compared in future work. The combined result of the Markovian influence graph cascading simulation and the load shed estimation is probability distributions of load shed for choices of specific initial lines damaged by the extreme event. We also explain how the Markovian influence graph can describe a weighted combination of cascading data from historical data and from the model-based simulation.

We critically examine the strengths and limitations of model-based cascading simulation versus cascading simulation driven by historical data via the Markovian influence graph. While more work is needed to develop and test the Markovian influence graph simulation, it seems to be a viable alternative to model-based simulation. It is particularly exciting that the Markovian influence graph simulation has strengths and weaknesses that complement those of model-based simulation, so that future decisions about cascading resilience could be informed by two different and complementary approaches to cascading simulation.

REFERENCES

- [1] I. Dobson, "Estimating the propagation and extent of cascading line outages from utility data with a branching process," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 2146–2155, Nov. 2012.
- [2] N. Romero, L. K. Nozick, I. Dobson, N. Xu, and D. A. Jones, "Seismic retrofit for electric power systems," *Earthquake Spectra*, vol. 31, no. 2, pp. 1157–1176, May 2015.
- [3] M. R. Kelly-Gorham, P. Hines, and I. Dobson, "Using historical utility outage data to compute overall transmission grid resilience," in *Mod. Electr. Power Syst. Conf.*, Wroclaw Poland, September 2019.
- [4] M. R. Kelly-Gorham *et al.*, "Using utility outage statistics to quantify improvements in bulk power system resilience," in *PSCC*, Portugal, Aug. 2020, and to appear in *Electric Power Systems Research*.
- [5] B. A. Carreras, D. E. Newman, and I. Dobson, "North American blackout time series statistics and implications for blackout risk," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4406–4414, Nov. 2016.
- [6] R. Baldick *et al.*, "Initial review of methods for cascading failure analysis in electric power transmission systems," in *IEEE PES General Meeting*, Pittsburgh, PA, USA, Jul. 2008.
- [7] M. Vaiman *et al.*, "Risk assessment of cascading outages: methodologies and challenges," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 631–641, May 2012.
- [8] K. Zhou, I. Dobson, Z. Wang, A. Roitershtein, and A. P. Ghosh, "A Markovian influence graph formed from utility line outage data to mitigate cascading," to appear in *IEEE Trans. Power Syst.*, 2020, also preprint arXiv:1902.00686.
- [9] U. Nakarmi *et al.*, "Interaction graphs for reliability analysis of power grids: A survey," *Energies*, vol. 13, 2020, paper 2219.
- [10] K. Zhou, I. Dobson, P. Hines, and Z. Wang, "Can an influence graph driven by outage data determine transmission line upgrades that mitigate cascading blackouts?" in *PMAPS*, Boise, ID, USA, Jun. 2018.

- [11] I. Dobson *et al.*, “Obtaining statistics of cascading line outages spreading in an electric transmission network from standard utility data,” *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4831–4841, Nov. 2016.
- [12] BPA transmission services operations & reliability website. [Online]. Available: <https://transmission.bpa.gov/Business/Operations/Outages>
- [13] B. A. Carreras *et al.*, “Complex dynamics of blackouts in power transmission systems,” *Chaos*, vol. 14, no. 3, pp. 643–652, 2004.
- [14] 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*, vol. 17, no. 2, p. 026103, 2007.
- [15] H. Ren, I. Dobson, and B. A. Carreras, “Long-term effect of the n-1 criterion on cascading line outages in an evolving power transmission grid,” *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1217–1225, 2008.
- [16] S. Mei, X. Zhang, and M. Cao, *Power grid complexity*. Beijing, China: Tsinghua University Press, Springer, 2011.
- [17] B. A. Carreras *et al.*, “Validating OPA with WECC data,” in *Proc. 46th Hawaii Int. Conf. System Sciences*, Maui, HI, USA, Jan. 2013.
- [18] —, “Validating the OPA cascading blackout model on a 19402 bus transmission network with both mesh and tree structures,” in *Proc. 52th Hawaii Int. Conf. System Sciences*, Maui, HI, USA, Jan. 2019.
- [19] E. Ciapessoni *et al.*, “Benchmarking quasi-steady state cascading outage analysis methodologies,” in *PMAFS*, Boise, ID, USA, Jun. 2018.
- [20] J. Bialek *et al.*, “Benchmarking and validation of cascading failure analysis tools,” *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4887–4900, Nov. 2016.
- [21] I. Dobson and D. E. Newman, “Cascading blackout overall structure and some implications for sampling and mitigation,” *Int. J. Electr. Power Energy Syst.*, vol. 86, pp. 29–32, Mar. 2017.
- [22] J. Kim, J. A. Bucklew, and I. Dobson, “Splitting method for speedy simulation of cascading blackouts,” *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3010–3017, 2012.