Bayesian estimates of transmission line outage rates that consider line dependencies

Kai Zhou, James R. Cruise, Chris J. Dent, Ian Dobson, Louis Wehenkel, Zhaoyu Wang, Amy L. Wilson

***Abstract*—Transmission line outage rates are fundamental to power system reliability analysis. Line outages are infrequent, occurring only about once a year, so outage data are limited. We propose a Bayesian hierarchical model that leverages line dependencies to better estimate outage rates of individual transmission lines from limited outage data. The Bayesian estimates have a lower standard deviation than estimating the outage rates simply by dividing the number of outages by the number of years of data, especially when the number of outages is small. The Bayesian model produces more accurate individual line outage rates, as well as estimates of the uncertainty of these rates. Better estimates of line outage rates can improve system risk assessment, outage prediction, and maintenance scheduling.**

I. Introduction

Transmission line outage rates are foundational for many reliability calculations, but in historical data the counts of outages for the more reliable lines are low, and estimated individual line outage rates are highly uncertain. There are several ways in which individual transmission lines are partially similar, including their length, rating, geographical location, and their proximity. We leverage these partial similarities with a Bayesian hierarchical method to improve the estimation of line outage rates from historical data.

The conventional method of estimating annual line outage rates divides the number of outages by the number of years of data. However, these estimates have a high variance when the data are insufficient. Indeed, in a year, many lines either do not fail or only fail once.

One pragmatic approach to mitigate the problem of limited outage counts is to group or pool lines together to get an estimate for the outage rate of that group. The lines can be grouped by area [1]–[3], or by line voltage rating. Lines in the same area experience similar weather conditions, and lines of the same rating have similar construction. However, the similarity between lines in these groups is only partial, variations of outage rates within the groups are neglected, and it is unwieldy to group lines according to multiple characteristics.

Transmission line outage rates are often supposed to be proportional to line length, and they are often quoted as rates per unit length [4], [5]. However, a line's outage rate is not strictly proportional to the line length because of substation and other effects, making the dependence on line length only a partial dependence. Indeed, our historical line outage data shows only a limited dependence on line length.

There is a middle ground between pooling lines in groups assuming perfect line dependencies within the group, and completely neglecting dependencies between lines by computing individual line outage rates in isolation. To exploit the partial dependencies of line outage rates, this paper proposes a Bayesian hierarchical method to estimate outage rates of individual transmission lines. In particular, our method can leverage the multiple partial dependencies in line length, rating, network proximity, and geographical area to give better outage rates of individual lines. This is done by explicitly modeling the dependence of outage rates on line length and rating and by using covariance kernels to model the dependencies between lines in close proximity. Our method can, therefore, learn about the outage rates of individual lines from lines close-by and with similar lengths and ratings. This means that where there is little data associated with a line (because the outage rate is small), our method can still estimate an outage rate for that line and its uncertainty. Also, by borrowing information from other lines, we can expect smaller uncertainties associated with estimates of outage rates, without assuming that all lines within a group have the same outage rate (as would be the case if we pooled the data).

The Bayesian hierarchical model proposed in this paper makes better estimates of outage rates. In particular, the proposed model:

- estimates annual outage rates for individual transmission lines more accurately by leveraging partial similarities between lines, including proximity, length, and rated voltage, especially when the annual outage counts are low or the data is limited. The estimates have lower standard deviation for given data, or the same standard deviation for less data.
- has performance better than the conventional method of simply dividing the number of outages by the number of years observed, especially when the data is limited.
- instead of pooling lines with one characteristic in common, gives a way to combine multiple partial similarities between lines.
- provides not only mean line outage rate, but also the uncertainty of this estimate.
- shows that line length and rated voltage correlate with line outage rate, but the correlation is not strong.

KZ,ID,ZW are with Iowa State University, Ames IA USA; dobson@iastate.edu. JRC is with Riverlane Research, UK. CJD and ALW are with University of Edinburgh, Scotland. LW is with University of Liège, Belgium.

We thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support during the programme Mathematics of Energy Systems where work on this paper was initiated. This work was supported by EPSRC grant EP/R014604/1.&#34. KZ,ID,ZW acknowledge support from USA NSF grants 1609080 and 1735354. LW acknowledges the support of F.R.S.-FNRS and the Simons Foundation. We gratefully thank Bonneville Power Administration for making publicly available the outage data that made this paper possible. The analysis and conclusions are strictly those of the authors and not of BPA.

- works using one standard line outage dataset routinely collected by transmission utilities worldwide.

These advances benefit applications of the line outage rates, and we start to explore these benefits for maintenance and reliability calculations in [6].

After reviewing the literature in Section II, Section III presents the historical outage data collected by a large utility and the modeling of the line dependencies. Sections IV and V present the Bayesian hierarchical model and the processing of the utility data. As we do not know the true outage rates from historical outage data, we use synthetic data to validate and evaluate the performance of the Bayesian hierarchical model in Section VI. Section VII concludes the paper.

II. Literature review

Bayesian approaches encode uncertainty in uncertain parameters such as outage rates as random variables. The Bayesian analysis aims to estimate a probability distribution for the uncertain parameters by incorporating all of our knowledge and accurately reflecting the uncertainty. Bayes theorem is used to combine data with prior distributions that describe initial knowledge of the uncertainty. The prior distributions are updated with the available data to give a posterior distribution that describes the uncertainty in the parameter values given all the available data. The mean or mode of the posterior distribution can be used to give a point estimate of the parameter. For further detail explaining Bayesian methods we suggest [7] as an introduction and [8] as a reference.

Bayesian methods are ideal for problems with limited data (such as estimation of outage rates), where it is necessary to use all the information available. Studies in ecology and social science have shown that when data are limited, Bayesian methods have less bias and are more robust than frequentist methods that consider parameters as fixed values [9], [10]. When lots of data are available, the data outweighs any effect of the prior distributions and a Bayesian method is less advantageous.

There is previous research predicting outage rates using Bayesian methods. Li [1] Iešmantis [4], and Moradkhani [5] present three Bayesian hierarchical models. All three hierarchical models have a Poisson distribution for outage counts, but how the outages are counted and lower levels of the model are different. Li [1] develops a hierarchical model to predict outage counts in a substation district given weather conditions, in which the log of the outage rate is a linear combination of weather factors. Iešmantis [4] presents a Poisson-Gamma random field model to estimate 230 kV transmission lines outage rates in a specified rectangular cell. The grid cells are introduced to model spatial dependence by constructing a correlation matrix in the Gamma field. The hierarchical model in Moradkhani [5] estimates failure rates of individual overhead distribution feeders, which are assumed to be independent of each other. To have an analytical form for the posterior distribution, conjugate priors are used, which results in a Gamma posterior distribution. Bayesian networks are also applied to estimate outage rates. Zhou [3] proposes a simple Bayesian network to predict weather-related outage rates given lightning and wind conditions over the whole system. Zhou compares the Bayesian network with a Poisson regression model and concludes that the Bayesian network is preferable. Yang [11] gives interval estimates of outage rates of individual transmission lines given weather conditions using a credal network with imprecise priors, which is an extension of Bayesian networks. Dunn [12] formulates a Bayesian hierarchical model for the total outage counts in a system. All components share the same failure rate derived from a fragility curve. In contrast to all the references above, our paper estimates outage rates of individual transmission lines using a Bayesian hierarchical model considering line dependencies.

Transmission line outages are correlated with each other in several ways. Lines in the power grid interconnect at substations, and some faults or substation arrangements may trip several lines simultaneously. Multiple line outages also occur due to protection schemes such as control protection groups and remedial action schemes. Moreover, lines in the same area experience similar weather conditions. There is some previous work on these dependencies. Li [1] uses the network adjacency matrix to model district dependencies. Similarly, Dokic [2] uses the weighted adjacency matrix to model substation dependencies. The difference between them is that [1] models the dependencies as a covariance matrix from the Bayesian perspective, while [2] uses an embedding method by learning vector representations of dependencies from a frequentist perspective. Iešmantas [4] models geographical dependencies between the outage rate per kilometer of 230 kV lines by making a rectangular grid of the area. Portions of lines in the same rectangle are assumed to have the same geographical influence, and the correlation between lines in different rectangles is assumed and modeled in the Gamma field. The main conclusion of [4] is that geographical correlation between line outage rates is present but weak. However, our method captures partial similarities between lines, including proximity, length, and rated voltage as a layer in the Bayesian hierarchical model.

Many researchers focus on predicting outage probabilities in a short term according to the weather condition [1]–[3], [11], [13]–[15]. [5], [11], [15] consider the data deficiency when building the outage rate model.

III. Exploring historical outage data and modeling line dependencies

Utilities routinely collect detailed outage data. For example, NERC's Transmission Availability Data System (TADS) collects outage data from North American utilities. Here, to illustrate our methods, we use some publicly available historical line outage data [16].

A. Historical outage data

The historical line outage data we use consists of transmission line outages recorded by a North American utility [16] for fourteen years since 1999. The data record forced and scheduled line outages, including the sending and receiving bus names of outaged lines, outage start and end times and dates, line attributes such as lengths, voltage ratings, districts in which a line is, and outage causes. Some lines cross several

TABLE I
ANNUAL OUTAGE COUNTS, LINE ATTRIBUTES, AND BAYESIAN ESTIMATES OF OUTAGE RATES AFTER 1ST, 7TH AND 14TH YEARS FOR 4 LINES

Line	Outage counts in different years														Line attributes			Annual outage rate		
ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Voltage(kV)	Length(mile)	District	1st	after 7th	after 14th
29	0	0	0	0	0	0	0	0	3	2	0	0	0	0	230	8.3	P	0.32	0.17	0.37
11	0	0	1	0	0	1	0	0	1	0	0	0	0	1	500	22.65	N	0.36	0.33	0.34
2	1	2	0	0	0	0	0	0	0	0	0	0	0	0	230	7.62	A	0.73	0.48	0.28
8	1	2	4	2	1	2	2	2	2	1	3	8	6	2	500	148.86	E	0.93	1.85	2.56

districts. There are 549 lines outaging in the data with rated voltages of 69, 115, 230, 287, 345, and 500 kV.

We neglect the scheduled outages and only consider the forced line outages. We also exclude the two 1000 kV HVDC lines, and momentary outages (outage duration does not exceed one minute). There are lines that failed once or twice in most of the years but suddenly failed, for example, ten times in one year. One common reason that a line could fail several times in a day is outages and reclosures for the same cause. So if a line fails several times in a day, we only count it once. Table I shows an example of the outage data.

B. Data exploration

We initially explore the line outage data using the conventional method of estimating annual line outage rates by dividing the number of outages by the number of years of data. We first pool all the line data together (i.e. treat as one homogeneous data set) to calculate the overall mean and standard deviation of outage rates, which are 0.6 and 0.7 outages per year, respectively. Next, we examine the individual conventional line outage rates. The mean variance-to-mean ratio of outage counts for each line is 1.2, which indicates that the outage counts show some overdispersion[1].

The power system network can be deduced directly from the outage data using the method of [17], and we show the conventional outage rates on the network in Figure 1 to visualize the spatial correlation. Close lines tend to have close colors, which indicates line dependencies from network proximity.

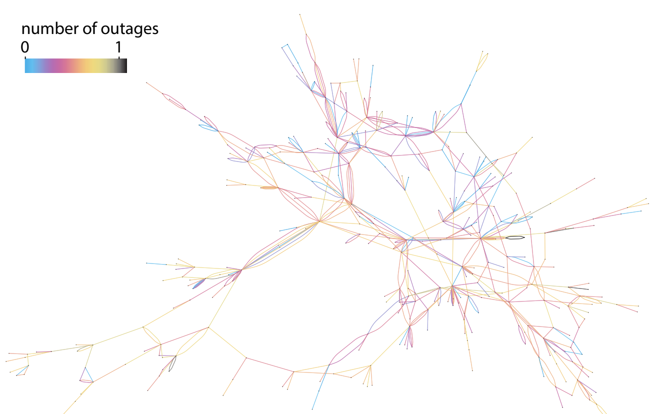

Fig. 1. The number of average annual forced outages over 14 years on network indicated by different colors (network layout is not geographic).

[1]Overdispersion means that the variance is larger than the mean. The Poisson distribution commonly used for count data does not apply when there is overdispersion because the Poisson mean and variance are equal.

C. Scaling line lengths and voltage ratings

The line lengths and voltage ratings are transformed and scaled so that their magnitudes and variations are scale-free and comparable. We do this in ways suggested by Gelman [18] for generic priors.

Line lengths in the vector $\boldsymbol{L}$ are first transformed by the natural logarithm to make the range of values less extreme, and then divided by the scale so that their variations are order of magnitude one:

$$\boldsymbol{x}_L = \frac{\ln \boldsymbol{L}}{\mathrm{scale}(\ln \boldsymbol{L})} \tag{1}$$

Here the scale of the sample in a vector $\boldsymbol{z}$ is estimated by the Mean Absolute Deviation, which is $\mathrm{scale}(\boldsymbol{z}) = \mathrm{median}(\boldsymbol{z} - \mathrm{median}(\boldsymbol{z}))$. Note that we use bold variables for vectors in this paper, and functions such as $\ln$ are applied element-wise so that $\ln \boldsymbol{L} = [\ln L_1, ..., \ln L_N]'$.

Similarly, the line voltage ratings $\boldsymbol{V}$ are first scaled by $\mathrm{SD}(\boldsymbol{V})$, the standard deviation of $\boldsymbol{V}$, and then divided by the scale:

$$\boldsymbol{x}_V = \frac{\boldsymbol{V}/\mathrm{SD}(\boldsymbol{V})}{\mathrm{scale}(\boldsymbol{V}/\mathrm{SD}(\boldsymbol{V}))} \tag{2}$$

It is usually considered that the line length and voltage rating have a positive correlation. Indeed, the BPA data shows this correlation, but it is a weak correlation: the Pearson correlation coefficient is 0.34 (0.12 for transformed lengths and voltage ratings).

D. Line proximity

The proximity of lines is quantified by the weighted sum of two kernels, which reflect two aspects of proximity. The first kernel is based on districts. Lines in the same district are more likely to experience the same weather conditions. Another kernel is based on network distance in terms of line length, which, to some extent, reflects both geographic proximity and the physical and engineering interactions in the power grid. We fit a linear regression model with correlated lines (described below) to support the form of the Bayesian hierarchical model and give guidance on setting priors.

1) Districts: There are 12 districts, and districts for each line are represented by a feature vector $\phi_{dis} \in \{0,1\}^{12}$ whose coordinates correspond to the districts, and are set to 1 for each district crossed by that line, and to 0 otherwise. The scalar product in this feature space thus counts the number of common districts crossed by two lines.

We define the district kernel as:

$$\boldsymbol{\Sigma}_1 = \exp\left(-||\phi_{dis}(i) - \phi_{dis}(j)||_2^2 - \mathbb{I}_{i \neq j}\right) \tag{3}$$

where $||\cdot||_2$ stands for the two-norm[2], and $\mathbb{I}_{i\neq j}$ is an indicator function. The reason why $\mathbb{I}_{i\neq j}$ is included is that a line is most similar to itself. The kernel $\mathbf{\Sigma}_1$ has the form of a correlation matrix since it is positive definite.

2) Network distance: The network distance between lines L_i and L_j along the network lines is defined as

$$d_{ij} = d(L_i, L_j) = \text{minimum length in miles of a network path joining midpoint of } L_i \text{ to midpoint of } L_j.$$

For example, the distance of line to itself is zero and the distance of a line to a neighboring line with at least one bus in common is half of the total length of the two lines.

Then we use the exponential kernel $\mathbf{\Sigma}_2$ which is

$$\mathbf{\Sigma}_2 = \exp[-2d(L_i, L_j)] \tag{4}$$

As $d(L_i, L_i) = 0$, the diagonal elements of $\mathbf{\Sigma}_2$ are one.

3) Combining the two kernels: The network proximity $\mathbf{\Sigma}$ is the weighted sum of above two kernels:

$$\mathbf{\Sigma} = w\mathbf{\Sigma}_1 + (1-w)\mathbf{\Sigma}_2, \tag{5}$$

where $0 < w < 1$. For example, if the two kernels are equally important, then $w = 0.5$.

We find the weights by fitting a linear regression model for the logarithm of average outage counts with $\boldsymbol{\beta}_0$ following a multivariate normal distribution to model correlation:

$$\ln \frac{\boldsymbol{N}}{t} = \boldsymbol{\beta}_0 + \beta_L \boldsymbol{x}_L + \beta_V \boldsymbol{x}_V, \tag{6}$$

$$\boldsymbol{\beta}_0 \sim \mathcal{N}(m\mathbf{1}, \sigma^2\mathbf{\Sigma}), \tag{7}$$

where $\boldsymbol{N}$ is a column vector whose entry N_i is the total number of counts in t years for line i, $\mathbf{1}$ is a column vector of ones, m, β_L, β_V are scalars, and

$$\sigma^2\mathbf{\Sigma} = \sigma^2(w\mathbf{\Sigma}_1 + (1-w)\mathbf{\Sigma}_2)) = \sigma_1^2\mathbf{\Sigma}_1 + \sigma_2^2\mathbf{\Sigma}_2. \tag{8}$$

For computation convenience, we decouple the dependencies between different lines in (8) by a coordinate transformation to diagonalize the covariance matrix $\sigma^2\mathbf{\Sigma}$. This transforms the multivariate normal random vector $\boldsymbol{\beta}_0$ in (7) into independent univariate normal random variables in the vector $\boldsymbol{\beta}_0'$. This decoupling facilitates the maximum likelihood calculation below. In particular, by simultaneous diagonalization [19, p.286], we find a matrix $\boldsymbol{Q}$ such that $\boldsymbol{Q}^T\mathbf{\Sigma}_1\boldsymbol{Q} = \boldsymbol{I}$ and $\boldsymbol{Q}^T\mathbf{\Sigma}_2\boldsymbol{Q} = \mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is a diagonal matrix. Define $\boldsymbol{\beta}_0' = \boldsymbol{Q}^T\boldsymbol{\beta}_0$, then

$$\begin{aligned}\boldsymbol{\beta}_0' &\sim \mathcal{N}(m\boldsymbol{Q}^T\mathbf{1}, \boldsymbol{Q}^T(\sigma_1^2\mathbf{\Sigma}_1 + \sigma_2^2\mathbf{\Sigma}_2)\boldsymbol{Q}) \\ &\sim \mathcal{N}(m\boldsymbol{Q}^T\mathbf{1}, \sigma_1^2\boldsymbol{I} + \sigma_2^2\mathbf{\Lambda}).\end{aligned} \tag{9}$$

We use Maximum Likelihood Estimation to estimate the parameters $\sigma_1^2, \sigma_2^2, m, \beta_L, \beta_V$ from the utility data. The log likelihood logL is

$$\begin{aligned}\boldsymbol{y} &= \boldsymbol{Q}^T(\ln \frac{\boldsymbol{N}}{t} - \beta_L\boldsymbol{x}_L - \beta_V\boldsymbol{x}_V) \\ \text{logL} &= \sum_i \ln f(y_i|m(\boldsymbol{Q}^T\mathbf{1})_i, \sigma_1^2 + \sigma_2^2\Lambda_i)\end{aligned} \tag{10}$$

[2]Since the vectors only have entries 0 or ±1, the one-norm is the same as the two-norm in this context.

where $\boldsymbol{y}$ is a column vector with ith entry y_i, $f(\cdot|\mu, \sigma^2)$ is the PDF of a normal distribution with mean μ and variance σ^2, $(\boldsymbol{Q}^T\mathbf{1})_i$ is the ith entry of $\boldsymbol{Q}^T\mathbf{1}$, and Λ_i stands for the ith diagonal entry of $\mathbf{\Lambda}$.

The maximum of logL in (10) is attained when $\sigma_1^2 = 0.45, \sigma_2^2 = 0.42$, $m = -1.5$ and $(\beta_L, \beta_V) = (0.13, 0.12)$. By normalizing σ_1^2 and σ_2^2, we have $w = \sigma_1^2/(\sigma_1^2 + \sigma_2^2) = 0.52$. The positive values of β_L and β_V indicate that longer lines or higher voltage lines tend to have higher outage rates, which is reasonable. These values shall give guidance on setting priors in Section IV.

We check the model assumptions by using the residual plot and QQ-plot as shown in Figure 2. $\boldsymbol{\beta}_0'$ has no correlation, so we focus on the transformed linear model, and Pearson residuals are used here as $\boldsymbol{\beta}_0'$ has heterogeneous variance. The Pearson residual is estimated by $\epsilon_i' = \epsilon_i/\sqrt{\sigma_1^2 + \sigma_2^2\Lambda_i}$, where the raw residuals are $\boldsymbol{\epsilon} = \boldsymbol{Q}^T(\ln \boldsymbol{N}/t - \beta_L\boldsymbol{x}_L - \beta_V\boldsymbol{x}_V - m\mathbf{1})$ There is no noticeable trend in the residual plot, and the QQ-plot shows that the Pearson residual follows the normal distribution.

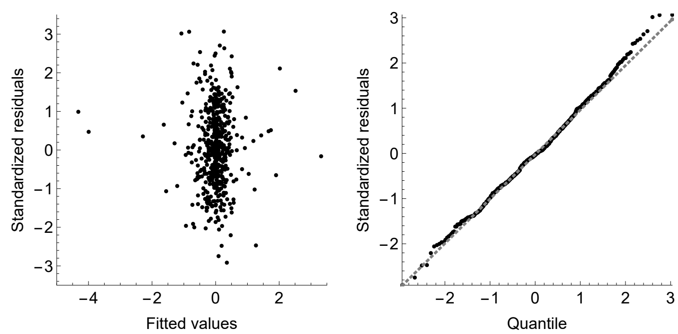

Fig. 2. Residual plot (left) and QQ-plot (right) for Pearson residuals.

IV. The Bayesian hierarchical model with line dependencies

We propose a Bayesian hierarchical model of outage counts incorporating line dependencies.

We assume that outage counts follow a Poisson distribution:

$$N_i \sim \text{Poisson}(\lambda_i t_i), \quad i = 1, ..., n \tag{11}$$

where N_i is the outage count for line i over t_i years, λ_i is the annual outage rate, and n is the number of lines.

We assume that the outage rates λ_i follow a Gamma distribution:

$$\lambda_i \sim \text{Gamma}(\alpha, \alpha/\mu_i), \quad i = 1, ..., n \tag{12}$$

The Gamma distribution is chosen for two reasons: It is a conjugate prior for the Poisson distribution. Moreover, the Gamma distribution mean is μ_i and its variance is μ_i^2/α. The variance of the Gamma distribution increases quadratically as the mean increases, which allows for the overdispersion observed in Section III-B.

The mean outage rate μ_i is modeled via a linear regression model with correlated lines. The linear regression model assumes the predicted variable is normally distributed, but μ_i

is positive and may have a large range of values, so μ_i is transformed by a log function [20, Sec. 3.6]:

$$\ln \boldsymbol{\mu} = \boldsymbol{\beta}_0 + \beta_L \boldsymbol{x}_L + \beta_V \boldsymbol{x}_V \tag{13}$$

where $\boldsymbol{\mu}$, $\boldsymbol{\beta}_0$ are column vectors.

$\boldsymbol{\beta}_0$ follows a multivariate normal distribution:

$$\boldsymbol{\beta}_0 \sim \mathcal{N}(m\mathbf{1}, \sigma^2(w\boldsymbol{\Sigma}_1 + (1-w)\boldsymbol{\Sigma}_2)) \tag{14}$$

Line dependencies are captured by the covariance matrix of this multivariate normal distribution, σ^2 is a scalar which controls the magnitude of the covariance and w controls the weights of the two kernels. The parameters α, β_V, β_V, m, σ^2 and w will be estimated using prior distributions in combination with the data as described below.

The prior distributions are:

$$\begin{aligned} \alpha &\sim \text{Half Normal}(0.7, 8^2) & \beta_L &\sim \text{Normal}(0.13, 5^2) \\ m &\sim \text{Normal}(-1.5, 5^2) & \beta_V &\sim \text{Normal}(0.12, 5^2) \\ \sigma^2 &\sim \text{Half Normal}(0, 0.5^2) & w &\sim \text{Beta}(1, 1) \end{aligned} \tag{15}$$

These priors are set to ensure that the parameters have a reasonable range and/or mean [3] when compared to our knowledge about the system and the model tested in Section III-B. As there is not much information about the standard deviations about these priors, we make these priors weakly informative. The detail is as follows.

The prior for α is a half-normal distribution with $\alpha > 0$. As discussed in Section III-B, the mean annual outage rate is 0.6, and the standard deviation is 0.7. This suggests the expected value of μ is 0.6, so the expected value of α would be $0.6^2/0.7^2 = 0.7$ (as $\mu_i^2/\alpha = \text{Var}\lambda_i$). The standard deviation of α is $\frac{(0.6+2\times0.7)^2}{0.7^2} - 0.6 \approx 8$ (the numerator is the maximum of μ in a typical range estimated by two times the standard deviation, $\frac{(0.6+2\times0.7)^2}{0.7^2}$ is the maximum of α).

Priors for m, β_L, β_V are normal distributions. The linear regression model in Section III-D suggests expected values for these parameters. $\boldsymbol{x}_L$ and $\boldsymbol{x}_V$ have range $[-10, 10]$ after scaling using method described in Section III-C, and we observe that the range of $\ln \boldsymbol{N}/t$ is $[-10, 10]$ conservatively. Therefore, we set the standard deviations of m, β_L, β_V to 5 so that 95% of the values lie in $[-10, 10]$ and they vary mostly in the same magnitude, which produces weakly informative priors.

σ^2 functions as a variance. The inverse-gamma prior is usually preferred since it is a conditional conjugate distribution. Gelman [21], however, does not recommend the inverse-gamma prior as the estimation of σ^2 would be sensitive to the parameters of inverse-gamma distribution when σ^2 is near zero. Thus, we let σ^2 have a half-normal prior. Section III-D shows that σ_1^2, σ_2^2 are about 0.5, so we set the standard deviation of σ^2 to 0.5 to make at least 95% of the values of σ^2 to lie in $[0, 1]$.

We give w a uniform prior as we know that w lies in $[0, 1]$ and the expectation of w is $0.52 \approx 0.5$ from Section III-D.

We now summarize the Bayesian hierarchical model. The Bayesian hierarchical model is specified by (11,12,13,14) together with the prior distribution of the parameters (15). Note that the partial dependencies between the lines are expressed in (13,14).

The model parameters, including the outage rates $\boldsymbol{\lambda}$, are

$$\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\beta}_0, \alpha, \beta_L, \beta_V, m, w) \tag{16}$$

The objective is to estimate the posterior distribution of the parameters $p(\boldsymbol{\theta}|\boldsymbol{N})$ that is informed by the line outage counts $\boldsymbol{N}$. By Bayes' theorem, the posterior distribution is

$$p(\boldsymbol{\theta}|\boldsymbol{N}) = \frac{p(\boldsymbol{N}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{N})} \tag{17}$$

Because normalization can be applied later, it is sufficient to calculate the unnormalized numerator of (17). We can exploit the dependencies in the hierarchical model (12,13,14) to get

$$p(\boldsymbol{N}|\boldsymbol{\theta}) = p(\boldsymbol{N}|\boldsymbol{\lambda}) = \prod_i p(N_i|\lambda_i) \tag{18}$$

$$p(\boldsymbol{\theta}) = \prod_i p(\lambda_i|\alpha, \mu_i) p(\boldsymbol{\mu}|\boldsymbol{\beta}_0, \beta_L, \beta_V) p(\boldsymbol{\beta}_0|m, w) \tag{19}$$

so that

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{N}) &\propto p(\boldsymbol{N}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \\ &\propto \prod_i p(N_i|\lambda_i) \prod_i p(\lambda_i|\alpha, \mu_i) p(\boldsymbol{\mu}|\boldsymbol{\beta}_0, \beta_L, \beta_V) \\ &\qquad \times p(\boldsymbol{\beta}_0|m, w)p(\alpha)p(\beta_L)p(\beta_V)p(m)p(w) \end{aligned} \tag{20}$$

V. Bayesian Processing of Real Data

The Bayesian hierarchical model described in the previous section is applied to the historical outage data.

A. Sampling posterior distributions using Stan

The posterior distributions (20) of the parameters (16) can be evaluated numerically by repeated sampling from the distribution with a Monte Carlo Markov Chain (MCMC) algorithm. MCMC is a class of algorithms for sampling from a probability distribution. We use the software Stan, which implements MCMC as Hamiltonian Monte Carlo (HMC) [22] with the algorithm adaptively tuned by the No-U-Turn Sampler (NUTS) [23]. Appendix A reproduces the algorithm of HMC with some explanatory comments and gives a detailed guide to the introductory and advanced literature on HMC.

We sample 2000 times, and the first 1000 samples are burn-in. Appendix B discusses technical details of model diagnostics and algorithm convergence. In this section, we focus on the result of the sampling.

B. Results of Bayesian estimates

We use the posterior mean as the point estimate of line outage rate because the posterior mean minimizes the Bayes risk in terms of squared error loss. Figure 3 shows the point estimates of line outage rates and their 95% credible intervals[4]. The mean outage rate of all lines is 0.74 outages per year,

[3] By saying that a range or mean is reasonable, we mean that the distribution of the prior has mean or range that is consistent with our prior knowledge, and it does not incorporate any further information.

[4] The credible interval is described by the multiplicative factor κ within which the outage rate λ_i can vary from the point estimate $\hat{\lambda}_i$ with 95% probability; that is, $P[\hat{\lambda}_i/\kappa \le \lambda_i \le \hat{\lambda}_i\kappa] = 95\%$.

and 82% of lines have rates less than 1 outage per year. There are two lines with very high outage rates. By inspecting the cause codes of these outages, one line outaged mainly because of foreign trouble (which is an external cause outside the power system, such as vehicles striking towers), while the other outaged mainly because of a remedial action scheme.

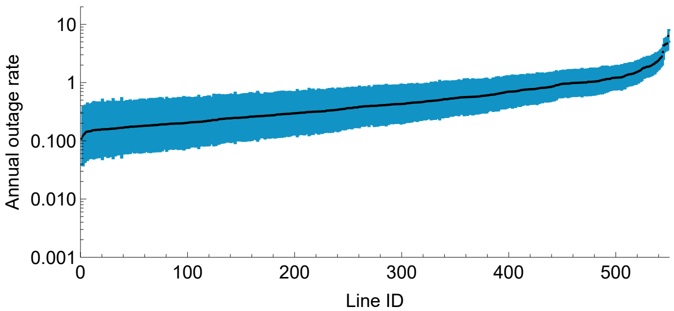

Fig. 3. Point estimates (black dots) and 95% credible intervals (blue bars) of annual outage rates. Lines are ordered by point estimates.

The values of β_L and β_V reveal the relationship between line lengths, voltage ratings, and outage rates. Figure 4 shows the posterior distributions of β_L and β_V and their correlation. The means of β_L and β_V are both 0.1. So the logarithm of the outage rate has a weakly positive correlation with transformed line length and transformed voltage rating. β_L and β_V have a very weak correlation, which is reasonable as $\boldsymbol{x}_L$ and $\boldsymbol{x}_V$ have a very weak correlation.

We use weakly informative priors in the Bayesian model. If we had access to previous studies in the region, or outage rates for other similar regions then these could be used to refine the priors. In this case we would expect the uncertainty in the outage rate estimates to be reduced.

We also test the sensitivity of the Bayesian model to the priors using 14-year data using two different sets of priors. The first case uses somewhat stronger informative priors. We reduce the standard deviation of the prior distributions of m, β_L, β_V from 5 to 1 and redo the calculations. In the second case, we randomly set parameters of priors by sampling from uniform distributions; then, we run the MCMC to estimate the posterior distributions. We compare the posterior mean and standard deviation of outage rates $\boldsymbol{\lambda}$ calculated using different priors, and find there is not much difference.

C. *Comparing the standard deviations of Bayesian and conventional estimates*

The Bayesian method produces a distribution of the outage rate, and it is straightforward to compute the standard deviation of this distribution. The conventional method estimates the outage rate with the sample mean. The standard deviation of the sample mean can be estimated as $s/\sqrt{n}$, where s is the sample standard deviation, and n is the sample size.

Figure 5 shows the ratio of the standard deviations of the Bayesian and conventional estimators. It shows that the standard deviation of the Bayesian estimator is typically smaller than the conventional estimator, especially when the data is limited to one year. The median ratio of standard deviations is

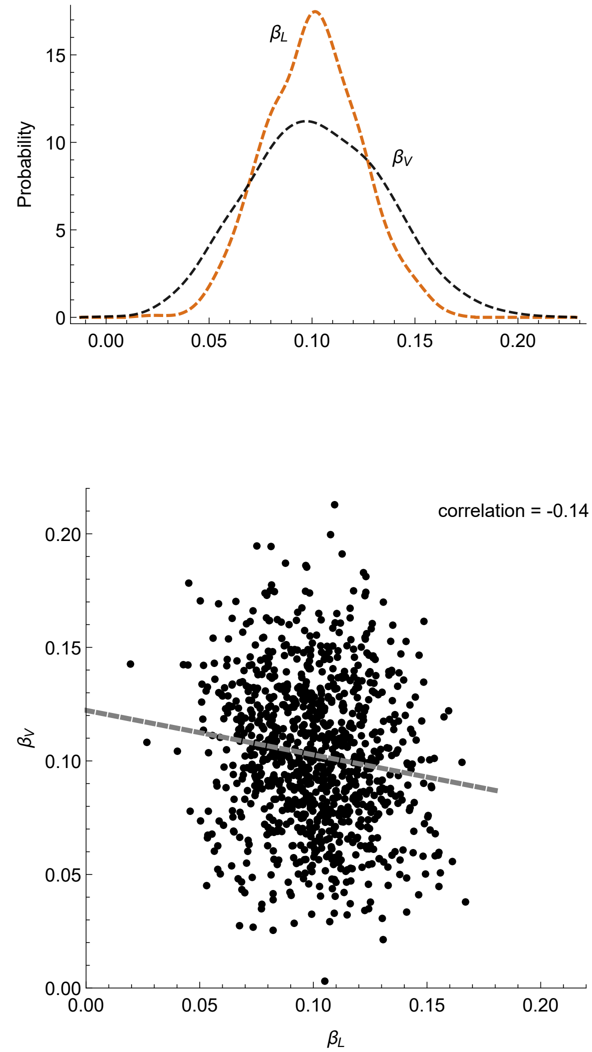

Fig. 4. Distributions of β_L and β_V (top) and their scatter plot and correlation (bottom).

0.66 for one year of data, while the median ratio is 0.93 for 14 years of data. Thus the Bayesian estimator typically achieves a lower standard deviation than the conventional one for limited data. Another way to present this finding is that given the same acceptable precision, the Bayesian method requires fewer data. Since the standard deviation is proportional to the square root of sample size, the Bayes estimator using one year of data achieves the same standard deviation as the conventional estimator using 2.30 years of data ($1/(0.66^2) = 2.30$). Similarly, the Bayesian estimator using 14 years of data achieves the same standard deviation as the conventional estimator using 16.2 years of data ($14/(0.93^2) = 16.2$).

D. *Performance on rarely outaged lines*

One advantage of the Bayesian method is that it provides a principled way of making line outage rates with no observed outages. The conventional estimate of outage rate is zero if a line has no outage in a year. However, it is more reasonable that the underlying outage rate of this line is a small value.

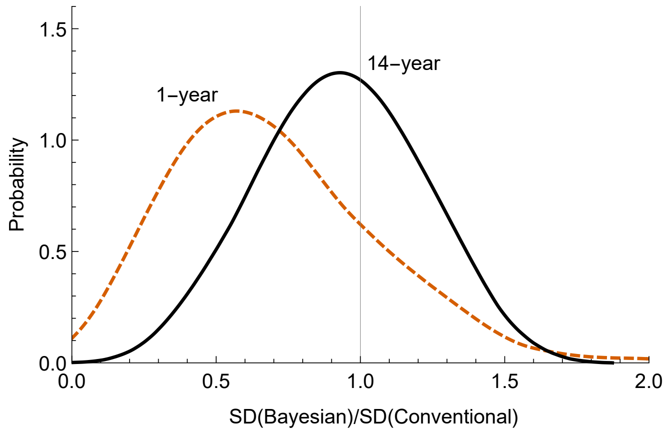

Fig. 5. Distributions of ratios of standard deviations of Bayesian estimator and conventional estimator using 1-year and 14-year data respectively. The ratio is SD(Bayesian)/SD(conventional).

Table I calculates 4 line outage rates with the data available after the 1st year, after the 7th year, and after the 14th year. In Table I, line 29 has no outage except in the 9th and 10th year. The Bayesian estimate of the outage rate of line 29 for the 1st year is 0.32, which is informed by correlations with other lines. By the 7th year, more years with no outages have been observed, so that the estimated outage rate decreases to 0.17. Line 29 outages several times in the 9th and 10th years, so its estimated rate over 14 years increases. There are also many zeros for lines 11 and 2, but the two outage rates vary differently as the distribution of zeros has different patterns. Most counts for line 11 are zeros, and single outages appear every several years. So we believe that the outage rate is roughly constant and small, which is captured by the Bayesian estimator. At the beginning, line 2 had several outages, and then it stops having outages. So this line has a decreasing outage rate. Line 8 is an example of a line with a high and increasing outage rate.

E. Validation of the Bayesian hierarchical model

Section III-D fits a linear regression model to the data, and Figure 2 shows that the assumptions for this regression model hold. This validates that the form of the Bayesian hierarchical model (particularly for (13), (14)) is reasonable.

As we do not know the true outage rates using real data, we generate synthetic data to further validate the Bayesian model in Section VI. That is, assuming that the real outage data follow the model detailed in III-D, we test that the Bayesian model accurately estimates the outage rates. As we have checked in Figure 2 that the model in III-D is a good fit to the real outage data, this is a reasonable method for validating the model when we do not have the true outage rates.

VI. Test Bayesian estimates on synthetic data

We build a generative model for synthetic datasets of arbitrary size, so the data are not limited in size, and the ground truth values are known. Then we test the Bayesian hierarchical model and the conventional estimates on the synthetic data. It turns out that the Bayesian hierarchical model predicts the outage rates well, and the Bayesian estimates compare favorably with the conventional method.

We also construct and test with synthetic data sets a Bayesian hierarchical model without correlations between the lines to evaluate the effect of line dependencies, which shows that modeling the dependencies reduces the variation of estimates.

A. The generative model for the synthetic data

In Section III-D, we fit a linear regression model with correlated lines. Based on this model, we generate outage counts according to the following model:

$$N_i \sim \mathrm{Poisson}(\lambda_i G) \tag{21}$$

$$G \sim \mathrm{Gamma}(a, a) \tag{22}$$

$$\ln \boldsymbol{\lambda} \sim \mathcal{N}(m\mathbf{1} + \beta_L \boldsymbol{x}_L + \beta_V \boldsymbol{x}_V, \boldsymbol{\Sigma}) \tag{23}$$

The parameters in (21–23) are assigned values according to the linear regression model with correlated lines. That is, $m = -1.5$, $\beta_L = 0.13, \beta_V = 0.12$, and $\boldsymbol{\Sigma} = 0.52\boldsymbol{\Sigma}_1 + 0.48\boldsymbol{\Sigma}_2$, which models the line dependencies.

Once we draw a sample from (23), the failure rate is known and fixed. So the variation of outage counts comes from the Poisson and Gamma distributions. In particular, using $EG = 1$, we derive from (21), (22) that the mean of N_i is the same as only using a Poisson distribution and that a controls the overdispersion:

$$\mathrm{E}N_i = \mathrm{E}[\mathrm{E}[N_i|G]] = \lambda_i \tag{24}$$

$$\mathrm{Var}N_i = \mathrm{E}[\mathrm{Var}[N_i|G]] + \mathrm{Var}[\mathrm{E}[N_i|G]] = \lambda_i + \lambda_i^2/a \tag{25}$$

The value of a is chosen so that the variance of the model matches the empirical variance calculated from the data. In particular, we find the quadratic that best fits the relationship between the empirical variance and mean to be $\sigma^2 = 0.14 + 0.54\mu + 0.53\mu^2$ (where σ^2 is the variance, μ is the mean). Since the coefficients of μ and μ^2 are close, we choose $a = 1$.

We generate three datasets with different sizes so that we have the equivalents of 1-year, 5-year, and 100-year data: 1) draw a sample of $\ln \boldsymbol{\lambda}$ from the multivariate normal distribution (23); 2) draw a sample of G from the Gamma distribution (22); 3) draw samples of N_i from the Poisson distribution (21) n times ($n \in \{1, 5, 100\}$). Thus, we obtain n annual outage counts for each line, and we know the true values of the outage rates $\boldsymbol{\lambda}$.

B. Comparing to the conventional estimates

The conventional estimates of outage rates are average outage counts per year. The conventional estimates and their standard deviations are obtained using Monte Carlo simulation: draw $B = 1000$ samples according to model (23), calculate the average count of each sample, and then calculate the standard deviation of the estimates.

We apply the Bayesian hierarchical model to synthetic datasets using MCMC with the same configuration as in Section IV, and use the mean of the posterior distribution as a point estimate.

1) Errors of point estimates: Figure 6 shows the distribution of errors of the Bayesian estimates and the conventional estimates (the estimation errors of the Bayesian method and conventional method have the same distribution for 100-year data, so that the plot is not shown). In general, the less the data, the wider the histogram. The error of the conventional estimates has two modes, and the probability of error near zero is lower for 1-year data. As the data size increases, the two modes merge into one. Moreover, for 1-year data, the standard deviation of the error is 0.6 for Bayesian estimates and 0.9 for conventional estimates; for 5-year data, the standard deviation is 0.3 for Bayesian estimates and 0.4 for conventional estimates. Therefore, the Bayesian estimates have a high chance of obtaining more accurate point estimates, especially when data is limited.

On the other hand, there is not much difference in the bias. Specifically, the bias is -0.007 for Bayesian estimates and -0.004 for conventional estimates using 1-year data, and the bias is 0.003 for both Bayesian estimates and conventional estimates using 5-year data.

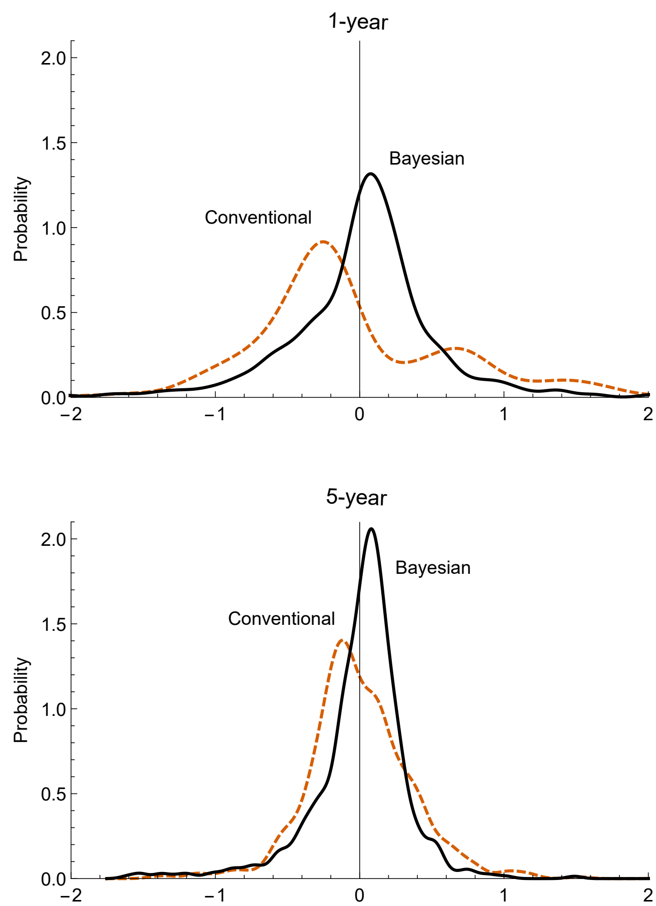

Fig. 6. Distributions of point estimation errors of Bayes estimates (posterior mean) and conventional estimates using 1-year and 5-year data.

2) Standard deviation: Figure 7 shows the distribution of the ratio of the standard deviation of the Bayesian estimator to that of the conventional estimator. The Bayesian estimator has a lower standard deviation when the data set is smaller. Specifically, the median of the ratio is 0.74 for 1-year data, 0.90 for 5-year data, and 0.99 for 100-year data.

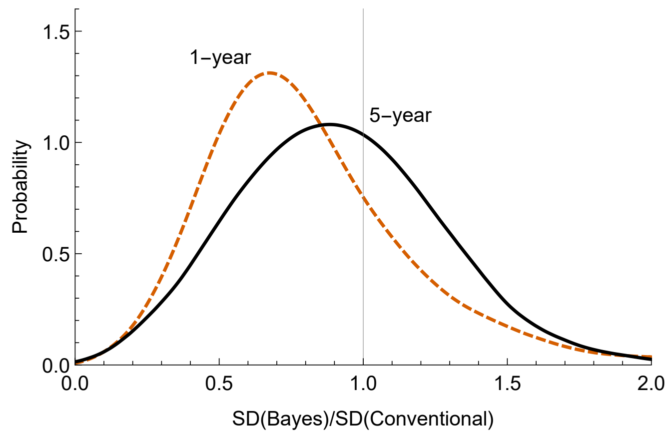

Fig. 7. Distributions of ratios of standard deviations of Bayes estimator and conventional estimator. The ratio is SD(Bayes)/SD(conventional).

3) Interval estimates: Figure 8 shows 95% credible intervals of the Bayes estimator using 1-year, 5-year, and 100-year data respectively. As the size of the dataset increases, we gain more information, and the width of the credible intervals decreases. Figure 8 also shows the true values of the outage rates as black dots. As expected with a 95% credible interval, approximately 5% of the true values lie outside the credible interval. The Bayesian point estimates (not indicated in Figure 8) lie in the center of the credible intervals and tend to be larger than the true values for low outage rates and smaller than the true values for high outage rates. This can be explained as the shrinkage towards the mean expected with Bayesian methods; see [7, Sec. 1.5].

C. Comparing to the Bayesian hierarchical model with independent lines

Previous work does not compute individual line outage rates while considering spatial dependencies between lines. We test the effect of the spatial dependence by removing it. The Bayesian hierarchical model with independent lines is:

$$N_i \sim \mathrm{Poisson}(\lambda_i t_i) \tag{26}$$
$$\lambda_i \sim \mathrm{Gamma}(\alpha, \alpha/\mu_i) \tag{27}$$
$$\ln \mu_i = \beta_0 + \beta_L x_{Li} + \beta_V x_{Vi} \tag{28}$$

The prior distributions are:

$$\begin{array}{ll} \alpha \sim \text{Half Normal}(0.7, 8^2) & \beta_L \sim \text{Normal}(0.13, 5^2) \\ \beta_0 \sim \text{Normal}(0, 1) & \beta_V \sim \text{Normal}(0.12, 5^2) \end{array} \tag{29}$$

We apply this restricted model to synthetic datasets using MCMC with the same configurations as in Section IV. The standard deviation when considering line dependencies is smaller than that without considering line dependencies. The medians of standard deviation ratios of this model to the conventional estimator for 100-year, 5-year, and 1-year data are 0.99, 0.93, and 0.89, which are greater than standard deviation ratios of the Bayesian model with line dependencies to the conventional estimator.

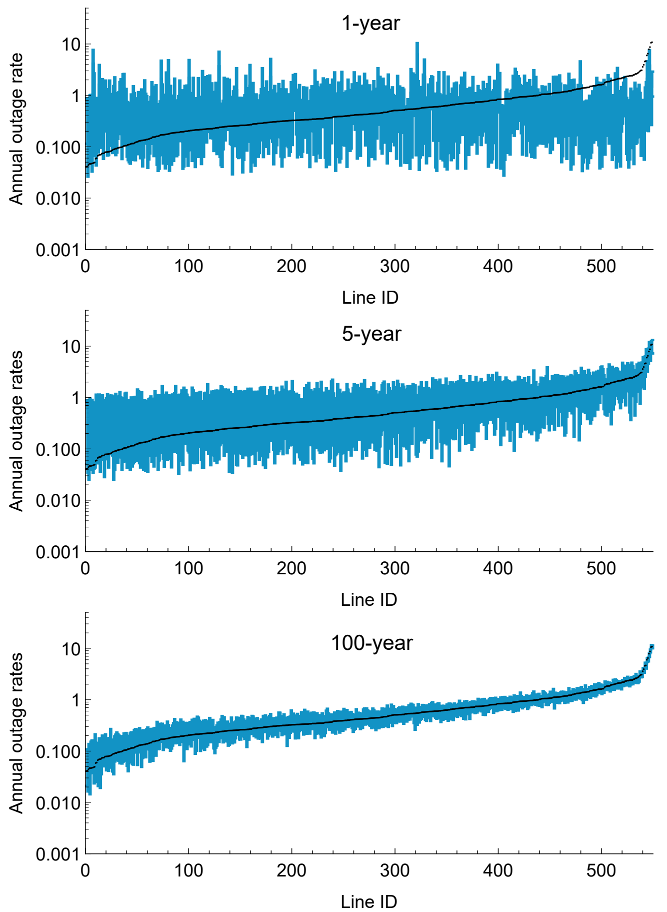

Fig. 8. 95% credible intervals of Bayesian estimates using 1-year, 5-year and 100-year data. Lines are ordered by outage rates (black dots).

VII. Discussion and Conclusion

We use a Bayesian hierarchical model to improve the estimation of annual outage rates for individual transmission lines. This Bayesian method incorporates several types of dependencies between lines and is applied to real outage data and tested with synthetic data. Particularly for the shorter observation periods with the lower outage counts, the Bayesian estimates perform better than the conventional estimates that simply divide the number of outages by the observation time: estimates of the individual line outage rates are more accurate, and the uncertainty of the estimates is reduced. Moreover, the comparison with a Bayesian model assuming spatially independent lines shows modeling line spatial dependencies reduces the standard deviation of estimates.

Our Bayesian hierarchical model offers an improvement over the conventional estimates for two reasons. Firstly, the Bayesian method can appropriately capture our prior knowledge of the parameter uncertainties with prior distributions. Secondly, because the model is hierarchical and models the dependence between lines, information about multiple partial commonalities can be appropriately shared across similar lines. These reasons imply that estimates can be improved for lines with no (or a small number of) outages.

Geographically close and neighboring lines experience similar weather conditions, may have a similar design, and share some physical and engineering interactions through the network. We model these line dependencies as a covariance matrix in the Bayesian hierarchical model. The covariance matrix is the weighted sum of two kernels that represent geographic district commonalities and network line proximity, respectively. The Bayesian model learns the weights of the two kernels from the outage data. Our modeling of these dependencies can be realized from a single utility outage dataset that is routinely collected, since the line district is recorded in the dataset, and the network can be readily deduced from the dataset [17]. Using only one dataset is advantageous since coordinating and combining different datasets is often arduous. However, it is conceivable that further advantage could be gained by including other factors such as average wind speed or altitude.

Previous work has often assumed that transmission line outage rates are proportional to line length [4] or grouped together lines of the same area [1]–[3]. We model these dependencies by linear factors in the outage rate, and the Bayesian model learns the weights for these factors. The results for our data are that individual line outage rates are only partially correlated with the line length or the voltage rating. Therefore, it is more reasonable to consider the outage rate for a whole line instead of the rate per mile.

The Bayesian method estimates the distribution of individual line outage rates. This is an advantage compared to methods that return point estimates, as a complete picture of the uncertainty around estimates is needed to make robust decisions about risk and maintenance. For example, if a line has a high point estimate outage rate that is very uncertain, it may be beneficial to wait to gather more information. If desired, any point or interval estimates can be easily obtained from the distribution, depending on the desired application of the outage rates. The quantification of the uncertainty of estimates is useful when the outage rates are used in other models and simulations. For example, a Monte Carlo simulation of transmission reliability can easily be modified to sample from the outage rate distribution to better capture the uncertainty in the estimated reliability.

We focus on overall line outage rates without considering different outage causes in this paper. However, the proposed Bayesian method can naturally be extended to investigate line outage rates for specific causes.

When data is limited, which is generally true for power system outage data, Bayesian estimates have smaller uncertainty than conventional estimates. Equivalently, with a specific acceptable standard deviation, the proposed Bayesian method needs less data than the conventional method. Thus, utilities can monitor individual line outage rates with fewer years of recording outages. There is a potential to more quickly identify lines with increasing outage rates and aging problems so that maintenance can be scheduled. For example, if utilities need two years of data using the conventional method to estimate line outage rates with a given uncertainty, they typically only need one year of data using the proposed Bayesian method to obtain an outage rate estimate that meets the same uncertainty requirement.

The general advantages of the hierarchical Bayesian method discussed above suggest benefits for various applications of line outage rates. We apply the hierarchical Bayesian method to start to explore and quantify these benefits in [6], which shows improved performance in detecting deterioration in line outage rates, quantifying the effect of storms, and a system reliability calculation.

Appendix A
Hamiltonian Monte Carlo

Hamiltonian Monte Carlo (HMC) is a sophisticated sampling algorithm combining ideas from Markov chains, rejection sampling, differential geometry, and numerical integration of Hamiltonian dynamics. This appendix reproduces the HMC Algorithm 1 from [23], briefly outlines how the algorithm works differently than other Markov chain Monte Carlo (MCMC) methods, and then recommends both tutorial and advanced references to HMC. Some general familiarity with MCMC is assumed.

Algorithm 1 Hamiltonian Monte Carlo

Given $\theta^0, \epsilon, L, \mathcal{L}, M$:
for $m = 1$ to M **do**
  Sample $r^0 \sim \mathcal{N}(0, I)$.
  Set $\theta^m \leftarrow \theta^{m-1}, \tilde{\theta} \leftarrow \theta^{m-1}, \tilde{r} \leftarrow r^0$.
  for $i = 1$ to L **do**
    Set $\tilde{\theta}, \tilde{r} \leftarrow \text{Leapfrog}(\tilde{\theta}, \tilde{r}, \epsilon)$.
  end for
  With probability $\alpha = \min\{1, \frac{\exp\{\mathcal{L}(\tilde{\theta}) - 0.5\tilde{r}\cdot\tilde{r}\}}{\exp\{\mathcal{L}(\theta^{m-1}) - 0.5 r^0 \cdot r^0\}}\}$,
    set $\theta^m \leftarrow \tilde{\theta}, r^m \leftarrow -\tilde{r}$.
end for
function Leapfrog(θ, r, ϵ)
$\tilde{r} \leftarrow r + (\epsilon/2)\nabla_\theta \mathcal{L}(\theta)$.
$\tilde{\theta} \leftarrow \theta + \epsilon\tilde{r}$.
$\tilde{r} \leftarrow r + (\epsilon/2)\nabla_\theta \mathcal{L}(\tilde{\theta})$.
return $\tilde{\theta}, \tilde{r}$

HMC has similar overall form as other Metropolis-Hastings Monte Carlo methods in that it proposes and probabilistically accepts successive samples of parameters to sample effectively from the posterior probability density. The successive samples are transitions in an ergodic Markov chain designed so that its final steady state distribution is the posterior probability density. gHowever, HMC samples differently than other methods in an enlarged space. In the notation of Algorithm 1, the parameter vector θ of "position" variables is augmented with a vector of "momentum" variables r to form an enlarged space of twice the dimension in which the successive samples are taken. The enlarged space enables Hamiltonian dynamics, where the "potential energy" $\mathcal{L}$ is the negative logarithm of the joint pdf of θ, and the "kinetic energy" is $\frac{1}{2} r \cdot r$.

Suppose the sampler is at (θ^0, r^0) in Algorithm 1. To propose a new sample at $(\tilde{\theta}, \tilde{r})$, the initial momentum r^0 is sampled from a Gaussian distribution, and then the Hamiltonian dynamics is integrated for L integration steps with integration step size ϵ. A symplectic leap-frog integrator that interleaves integration steps is used in order to preserve the Hamiltonian structure. Then the proposed sample is probabilistically accepted or rejected in a way similar to the Metropolis algorithm. Hoffman [23] proposed the No-U-Turn Sampler to avoid hand tuning the parameters L and ϵ controlling the integration.

To understand why HMC works, we refer readers to the approachable and intuitive expositions in [22] and [24, Cha.15] for expert explanations of the algorithm and to [23], [25]–[27] for further technical analysis. In particular, Betancourt discusses how HMC is "uniquely suited to the high-dimensional problems of applied interest." [22] and how HMC can tackle the correlations induced by hierarchical models [25]. The No-U-Turn Sampler has at least the same efficiency as a well-tuned HMC algorithm [25]. The convergence is usually checked by empirical diagnostic tools [27]. Also, we carefully set the initial values of the parameters to make the convergence faster by exploring the outage data in Section III.

Appendix B
Convergence of Sampling Algorithm

This appendix uses four methods to check the convergence of the Hamiltonian Monte Carlo algorithm used to sample the posterior distributions, including potential scale reduction factors, effective sample size diagnostics, and trace and autocorrelation plots. In addition, we check that the algorithm is not getting stuck in a local mode in the posterior distributions.

The Gelman-Rubin potential scale reduction factor diagnostic $\hat{R}$ is often used to check Markov chain Monte Carlo convergence [27]. $\hat{R}$ is defined as the ratio of the estimated pooled variance to the estimated within-chain variance (see [8, Sec. 11.4] for the equations of $\hat{R}$). Figure 9 plots the iterates of $\hat{R}$ for all parameters at increments of 20 iterations from four parallel Markov chains. Figure 9 shows that all $\hat{R}$s converge and are less than 1.1 after 400 iterations.

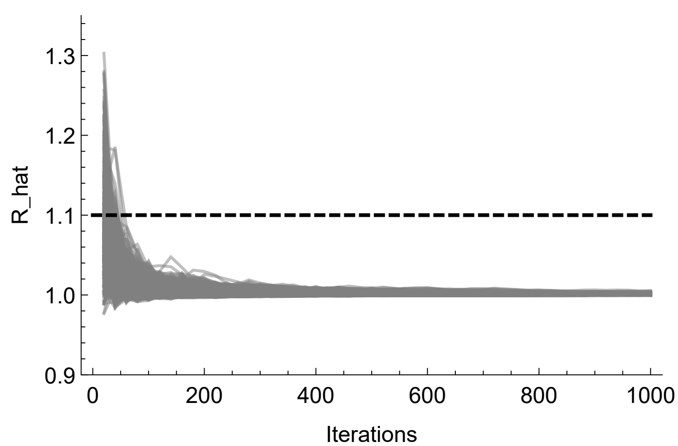

Fig. 9. Iterates of $\hat{R}$ for all parameters computed from four parallel Markov chains at increments of 20 iterations.

As suggested by Gelman [8], we also compute the effective sample size $\hat{n}_{eff}$, which is the equivalent number of independent samples that have the same standard error of the sample mean of the parameter as the Markov chain samples (see [8, Sec. 11.4] for the equations of $\hat{n}_{eff}$). It turns out that $\hat{n}_{eff}$s for all λs are greater than 100 per chain after 300 iterations, which shows that the estimates are reliable.

Graphical methods provide another way to check convergence. We make trace plots and autocorrelation function plots for each variable to check whether the chains are mixing and have large autocorrelation. It is not practical to show all the plots here. Instead, we randomly select four parameters to show the trace plots (Figure 10) and autocorrelation function plots (Figure 11). The two chains have mixed, and the autocorrelation decreases quickly and tends to zero.

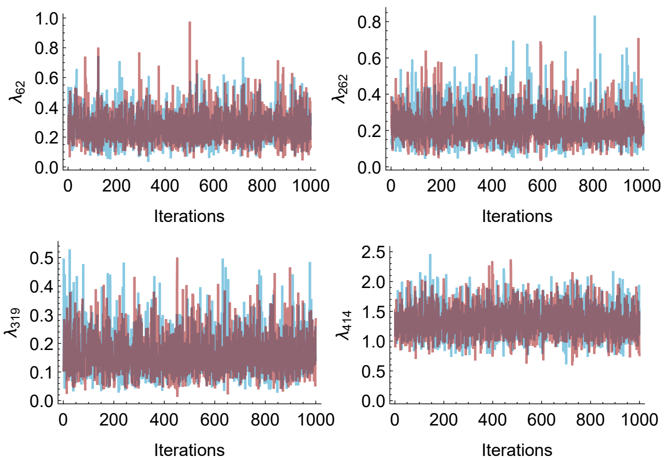

Fig. 10. Trace plots of two chains of four randomly selected λs.

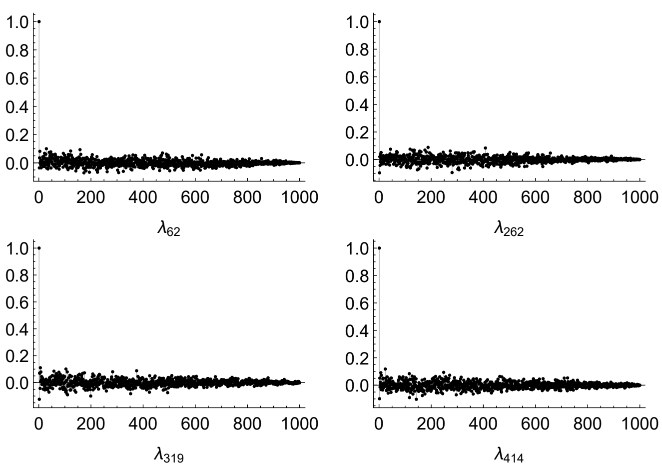

Fig. 11. Autocorrelation function plots of four randomly selected λs.

Based on the results of the four methods of checking convergence, we conclude that there is no evidence of non-convergence.

To check that the algorithm is not getting stuck in a local mode in the posterior probability distribution, we simulate two additional Markov Chains with random initial values sampled from a uniform distribution over the support of parameters. Each of these additional Markov Chains has 3000 iterations in which the first 2500 samples are burn-in and are thrown away. We compare the posterior distributions of all parameters estimated from the additional chains and the original chain with the initial values in the body of the paper, and we find no convergence issues. Moreover, as we are most interested in the outage rates λ, we implement a Kolmogorov-Smirnoff test on the corresponding distributions of outage rates of the two chains that start with random values. All the λs except two are judged to be from the same distribution with a significance level 0.01. And these two λs have close means (0.32 and 0.33, 0.14 and 0.15) and close standard deviations (0.14 and 0.13, 0.08 and 0.08).

References

[1] H. Li, L. A. Treinish, and J. R. M. Hosking, "A statistical model for risk management of electric outage forecasts," *IBM J. Res. Dev.*, vol. 54, no. 3, pp. 8:1–8:11, May 2010.

[2] T. Dokic et al., "Spatially aware ensemble-based learning to predict weather-related outages in transmission," in *Proc. 52th Hawaii Intl. Conf. System Science*, Maui, HI, USA, Jan. 2019.

[3] Y. Zhou, A. Pahwa, and S. Yang, "Modeling weather-related failures of overhead distribution lines," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1683–1690, Nov. 2006.

[4] T. Iešmantas and R. Alzbutas, "Bayesian spatial reliability model for power transmission network lines," *Electr. Power Syst. Res.*, vol. 173, pp. 214–219, 2019.

[5] A. Moradkhani *et al.*, "Failure rate estimation of overhead electric distribution lines considering data deficiency and population variability," *Int. Trans. Electr. Energ. Syst.*, vol. 25, no. 8, pp. 1452–1465, Apr. 2015.

[6] K. Zhou, J. R. Cruise, C. J. Dent, I. Dobson, L. Wehenkel, Z. Wang, and A. L. Wilson, "Applying Bayesian estimates of individual transmission line outage rates," to appear in PMAPS 2020.

[7] B. P. Carlin and T. A. Louis, *Bayesian methods for data analysis*. Boca Raton, FL, USA: CRC Press, 2008.

[8] A. Gelman et al., *Bayesian data analysis*. Boca Raton, FL, USA: CRC Press, 2013.

[9] M. Omlin and P. Reichert, "A comparison of techniques for the estimation of model prediction uncertainty," *Ecological modelling*, vol. 115, no. 1, pp. 45–59, Feb. 1999.

[10] D. Stegmueller, "How many countries for multilevel modeling? a comparison of frequentist and bayesian approaches," *American Journal of Political Science*, vol. 57, no. 3, pp. 748–761, Jul. 2013.

[11] M. Yang *et al.*, "Interval estimation for conditional failure rates of transmission lines with limited samples," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 2752–2763, Jul. 2018.

[12] L. N. Dunn *et al.*, "Bayesian hierarchical methods for modeling electrical grid component failures," *preprint arXiv:2001.07597*, 2020.

[13] R. Yao and K. Sun, "Toward simulation and risk assessment of weather-related outages," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4391–4400, Jul. 2019.

[14] K. Alvehag and L. Soder, "A reliability model for distribution systems incorporating seasonal variations in severe weather," *IEEE Trans. Power Del.*, vol. 26, no. 2, pp. 910–919, Apr 2011.

[15] Y. Wang *et al.*, "Evaluating weather influences on transmission line failure rate based on scarce fault records via a bi-layer clustering technique," *IET Gener. Transm. Distrib.*, vol. 13, no. 23, pp. 5305–5312, Nov. 2019.

[16] "Bonneville power administration transmission services operations & reliability website." https://transmission.bpa.gov/Business/Operations/Outages.

[17] 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.

[18] A. Gelman, "Prior choice recommendations." Github, https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations, (accessed Nov. 19, 2019).

[19] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[20] M. H. Kutner *et al.*, *Applied linear statistical models*. Boston, USA: McGraw-Hill Irwin, 2005.

[21] A. Gelman, "Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper)," *Bayesian analysis*, vol. 1, no. 3, pp. 515–534, 2006.

[22] M. Betancourt, "A conceptual introduction to Hamiltonian Monte Carlo," *arXiv preprint arXiv:1701.02434*, 2017.

[23] M. D. Hoffman and A. Gelman, "The No-U-Turn Sampler: adaptively setting path lengths in Hamiltonian Monte Carlo." *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1593–1623, 2014.

[24] B. Lambert, *A students guide to Bayesian statistics*. Sage, 2018.

[25] M. Betancourt and M. Girolami, "Hamiltonian Monte Carlo for hierarchical models," *Current trends in Bayesian methodology with applications*, vol. 79, no. 30, pp. 2–4, 2015.

[26] R. M. Neal *et al.*, "MCMC using Hamiltonian dynamics," *Handbook of Markov Chain Monte Carlo*, vol. 2, no. 11, p. 2, 2011.

[27] V. Roy, "Convergence diagnostics for Markov Chain Monte Carlo," *Annu. Rev. Stat. Appl.*, vol. 7, pp. 387–412, 2020.