

Logarithmic Resilience Risk Metrics That Address the Huge Variations in Blackout Cost

Arslan Ahmad¹, Graduate Student Member, IEEE, and Ian Dobson², Life Fellow, IEEE

Abstract—Resilience risk metrics must address the customer cost of the largest blackouts of greatest impact. However, there are huge variations in blackout cost in observed distribution utility data that make it impractical to properly estimate the mean large blackout cost and the corresponding risk. These problems are caused by the heavy tail observed in the distribution of customer costs. To solve these problems, we propose resilience metrics that describe large blackout risk using the mean of the logarithm of the cost of large-cost blackouts, the slope index of the heavy tail, and the frequency of large-cost blackouts.

Index Terms—Resilience metric, power distribution reliability.

I. EVENTS, COST, AND METRICS

FOR a resilience analysis, we extract from historical outage data recorded by utilities events in which outages bunch up and overlap [1], [2]. Alternatively, the events could be simulated with power system models [3]. The events we observe range in size from isolated single outages to large events with hundreds of outages caused by extreme weather. The customer costs for each event can be estimated based on factors such as the customer hours without power and the types of customers. Let n_{customer} be the number of customers served in a distribution system. Then the cost of the i th event in the distribution system normalized by the number of customers served is

$$c_i = \frac{\text{customer cost for event } i}{n_{\text{customer}}} \quad (1)$$

Suppose there are n events with normalized costs¹

$$C_{\text{all}} = \{c_1, c_2, \dots, c_n\}. \quad (2)$$

For a resilience analysis we are most interested in events with large cost, so we set a threshold c_{large} and define the costs of the n_{large} large cost events as

$$C_{\text{large}} = \{\text{costs } c \text{ in } C_{\text{all}} \text{ such that } c \geq c_{\text{large}}\} \quad (3)$$

Received 17 May 2025; revised 11 August 2025; accepted 14 September 2025. Date of publication 19 September 2025; date of current version 22 October 2025. This work was supported in part by NSF under Grant 2153163 and Grant 2429602, in part by Argonne National Laboratory, in part by US DOE EERE under Grant DE-EE0010724, and in part by PSERC Project S110. Paper no. PESL-00166-2025. (Corresponding author: Ian Dobson.)

The authors are with the Department Electrical & Computer Engineering, Iowa State University, Ames Iowa, IA 50011 USA (e-mail: dobson@iastate.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TPWRS.2025.3612225>.

Digital Object Identifier 10.1109/TPWRS.2025.3612225

¹Customer costs are assumed positive by neglecting events with zero cost.

If the number of years of outage data is n_{year} , the average annual rate of events with positive customer cost is $E_{\text{rate}} = n/n_{\text{year}}$. Then we easily obtain empirical estimates of the probability p_{large} of large cost events and its associated annual rate f_{large} and recurrence interval RI_{large} :

$$p_{\text{large}} = \frac{n_{\text{large}}}{n}; f_{\text{large}} = p_{\text{large}} E_{\text{rate}}; RI_{\text{large}} = f_{\text{large}}^{-1} \quad (4)$$

For this letter we choose c_{large} so that $p_{\text{large}} = 0.1$ or so that $p_{\text{large}} = 0.05$. The choice of c_{large} is discussed at the end of section III. c_{large} is kept fixed once chosen to allow comparisons of large event risk over time.

Now we quantify the risk of large cost events. The risk of large events is sometimes defined as the probability of a large event multiplied by the mean cost of large events. However, the large event costs vary over several orders of magnitude. To address this, and to avoid the severe difficulties caused by attempting to estimate the mean cost that are discussed in Section III, we instead transform the large cost data by applying a logarithm before taking the mean. That is, we define the Average Log Event Cost metric ALEC for large cost events as

$$\text{ALEC} = \frac{1}{n_{\text{large}}} \sum_{c \in C_{\text{large}}} \log_{10} c \quad (5)$$

The use of the logarithm to base ten shows that ALEC indicates the mean order of magnitude of the large event costs. For example, $\text{ALEC} = 2$ indicates costs of mean order of magnitude 2, or 10^2 . Note that 10^{ALEC} is the geometric mean of the large event costs.

We also define the Annual Log Cost Resilience Index ALCRI for large cost events, which usefully combines frequency and ALEC, as

$$\text{ALCRI} = \frac{1}{n_{\text{year}}} \sum_{c \in C_{\text{large}}} \log_{10} c = f_{\text{large}} \text{ALEC} \quad (6)$$

It is well-known in reliability that there is huge variability in blackout sizes and correspondingly large variations in the SAIDI reliability metric if extreme events are included. Indeed, the distribution of daily SAIDI typically follows a heavy-tailed lognormal distribution, and major event days are defined and often excluded from the calculations if their daily SAIDI is in the tail of the distribution [9]. Instead of excluding the tail events as needed for consistent year-to-year estimates of reliability, here we address resilience by characterizing the tail of the distribution of event cost with metrics.

TABLE I
UTILITY DATA AND METRICS

	Utility-1	Utility-2	Utility-3	Utility-4	Utility-5
n	5716	2706	3830	7000	6485
n_{year}	6	17.4	11	10	11
k	370.2	228.2	323.0	339.9	339.9
c_{maxobs}	\$5063	\$1095	\$3523	\$1954	\$88
c_{large}	\$1.05	\$0.62	\$2.40	\$1.10	\$2.15
n_{large}	572	271	384	701	325
p_{large}	0.1	0.1	0.1	0.1	0.05
f_{large}	96	16	35	70	30
α	0.83	0.73	0.75	1.00	1.47
CI_{α}	(.76,.89)	(.65,.82)	(.68,.83)	(.93,1.1)	(1.3,1.6)
ALEC	0.55	0.38	0.96	0.48	0.63
CI_{ALEC}	(.50,.59)	(.31,.46)	(.90,1.0)	(.44,.51)	(.60,.66)
RSE_{ALEC}	0.040	0.094	0.031	0.034	0.026
RSE_{pb}	1.44	1.70	1.05	1.46	0.66
RSE_{LNb}	1.38	1.53	0.65	0.89	0.13

all costs in 2022 USD; f_{large} in per year.

The Area Index of Resilience AIR metric [10] also averages the logarithm of customer hours to reduce its variability, and Carreras [11] discusses the corresponding heavy tail problem for blackout sizes in transmission systems. In fields other than power system resilience engineering, it is usual to fit the statistics of extreme events with heavy-tailed distributions and to estimate the tail index α with Hill estimators [6], [7], [8].

II. UTILITY DATA PROCESSING

We process the outage data of five different Northeastern and Midwestern US utilities. The outage data includes the start time, end time, and number of customers affected by each outage. Only the unplanned distribution system outages are selected from the data and analyzed. Outage events are extracted from the outage data after data cleaning (details in [4]), and the cost of each event is calculated by multiplying the total customer hours lost in that event by a factor k , where k is the average cost per customer per hour. k accounts for the widely varying costs of residential, commercial, and industrial loads by assuming that each event has the proportion of these loads outaged that is the same as the proportion of these loads in the entire system [4]. Table I gives data and metrics for each utility. We gratefully acknowledge Muscatine Power and Water for generously providing the Utility-2 outage data.

III. RISK AND CONSEQUENCES OF HEAVY TAILS

Risk associates probabilities with costs of events [5], and can be described by a probability distribution as a cost exceedance function $\bar{F}_C(c) = P[C > c]$, which is the probability of the event cost C exceeding the value c as c varies.² Fig. 1 shows the exceedance function $\bar{F}_C$ of per-event customer cost for Utility-3 on a log-log plot. Fig. 1 shows an approximate straight-line behavior of slope magnitude $\alpha = 0.75$ in the tail for $c \geq c_{\text{large}} = 2.4$. That is, the tail in Fig. 1 has the approximate

²The cost exceedance function is also known as the survival function or complementary cumulative distribution function (CCDF) or risk curve.

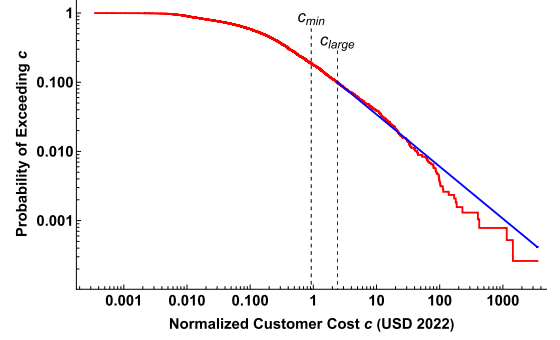


Fig. 1. Exceedance curve of normalized customer cost from utility data of Utility-3, along with fitted Pareto distribution.

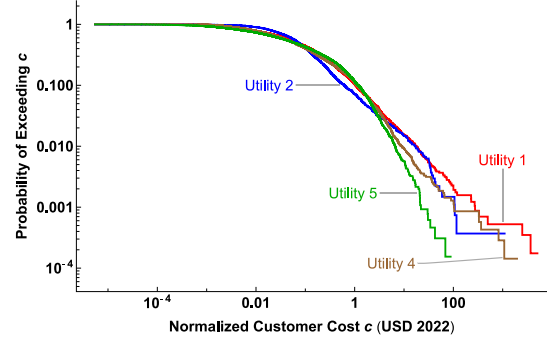


Fig. 2. Exceedance curves of normalized customer cost for utilities 1, 2, 4, 5.

power-law behavior³ described by a Pareto distribution P with minimum value c_{large} and tail slope index α :

$$\bar{F}_P(c) = (c/c_{\text{large}})^{-\alpha}, \quad c \geq c_{\text{large}} \quad (7)$$

Formally testing the tails for all five utilities indicates that linear approximations to the tails on the log-log plots are reasonable.⁴ Fig. 2 shows the empirical exceedance curves for utilities 1, 2, 4, 5.

If $\alpha \leq 1$, the mean of Pareto distribution (7) is infinite. If $1 < \alpha \leq 2$, the mean is $(\alpha c_{\text{large}})/(\alpha - 1)$, but the standard deviation is infinite. The relative standard error of an estimate of the mean with M samples is infinite for $\alpha \leq 2$. This means that even when $1 < \alpha \leq 2$, the mean large blackout cost is hard to accurately estimate. Table I shows that $\alpha < 2$ for all the utilities analyzed in this letter.

In contrast, the logarithm of the data in the Pareto tail $\log_{10} P$ for $c \geq c_{\text{large}}$ has a shifted-exponential distribution with rate

³To verify the straight-line behavior on the log-log plot, take the log of (7) to obtain $\log \bar{F}(c) = -\alpha \log c + \alpha \log c_{\text{large}}$.

⁴The plausibility of a Pareto (straight-line) tail is not rejected for utilities 1, 2, and 5 at $p = 0.1$, using Clauset's goodness-of-fit test [6], and the plausibility of a truncated lognormal fit is not rejected at $p=0.1$ for utilities 1, 2, 3, and 4. The truncated lognormal fit is quadratic, but with a large enough standard deviation parameter σ so that the quadratic coefficient $-1/(2\sigma^2)$ is small [7, section 1.2.3] and the linear approximation is acceptable. A null hypothesis H_0 of Pareto fit above c_{large} is not rejected at $p = 0.1$ for utilities 1, 2, 5, and H_0 of truncated lognormal fit above c_{large} is not rejected at $p = 0.1$ for all utilities. A likelihood ratio test between Pareto and lognormal above c_{large} shows insufficient evidence to prefer one over the other at $p = 0.1$.

parameter $\alpha \ln 10$ and shift parameter $\log_{10} c_{\text{large}}$:

$$\bar{F}_{\text{exp}}(x) = e^{-(\alpha \ln 10)(x - \log_{10} c_{\text{large}})}, \quad x \geq \log_{10} c_{\text{large}} \quad (8)$$

If the cost data follows the Pareto tail (7) for $c \geq c_{\text{large}}$, then $\text{ALEC} = \log_{10} c_{\text{large}} + (\alpha \ln 10)^{-1}$ is the mean of the log transformed data (8). The relative standard error of an estimate of ALEC with M samples is then $\text{RSE}_{\text{ALEC}} = [(1 + \alpha \ln c_{\text{large}})\sqrt{M}]^{-1}$. While the heavy Pareto tail (7) has an infinite mean, the mean (ALEC) of the log-transformed tail (8) follows the usual statistics of light-tailed distributions for all values of $\alpha > 0$.

The tail index α is the absolute slope of the tail above c_{large} on the log-log plot of the exceedance curve, so that α describes a linear trend for the observed risk⁵. α can be estimated using the maximum likelihood Hill estimator [7], [8], and related to ALEC according to:

$$\alpha = \left[\frac{1}{n_{\text{large}}} \sum_{c \in C_{\text{large}}} \ln \frac{c}{c_{\text{large}}} \right]^{-1} = \frac{(\ln 10)^{-1}}{\text{ALEC} - \log_{10} c_{\text{large}}} \quad (9)$$

The probability p_{large} and the absolute slope α describe the large cost portion of the exceedance curve [4], so this is also the case for p_{large} and ALEC. That is, p_{large} and α describe the trend for the large blackouts, and reducing p_{large} or ALEC improves this trend and reduces the risk of blackouts. If the large blackout cost trend in Fig. 1 is quite reasonably extrapolated linearly beyond the maximum observed cost (as explained in the next section), then reducing p_{large} or ALEC also reduces the risk of large blackouts beyond the largest observed.

Now we discuss choosing the c_{large} threshold for large cost events. First, c_{large} should be chosen in the approximately linear upper region of the log-log plot of the exceedance curve. The linear region can be determined by $c \geq c_{\text{min}}$, where c_{min} is estimated using Clauset's method [6]. Second, there is a tradeoff: a smaller c_{large} includes more large event data, giving a less variable estimate of α and ALEC, whereas a larger c_{large} better describes the trend indicated by the very largest cost events, especially if there could be some change in slope for the very largest cost events. For this letter, we choose c_{large} corresponding to $p_{\text{large}} = 0.1$ or 0.05 with the result that c_{large} is above but near c_{min} to minimize the variability of the estimates of α and ALEC. Determining a suitable value of c_{large} , or, more precisely, the percentile corresponding to c_{large} , is a well-known delicate issue in estimating α with the Hill estimator (9) [7], [8].

IV. EXTRAPOLATING TO LARGEST POSSIBLE COST

We extrapolate the slope of the exceedance curve beyond the maximum observed cost c_{maxobs} so that the likely performance of the metrics can be evaluated while including the unobserved future blackouts of highest risk. The extrapolation of the trend is somewhat uncertain, but the extrapolation nevertheless seems more reasonable than ignoring the even more extreme events indicated by the observed trend—it seems unlikely that the

largest cost blackout has already been observed in the data. On the other hand, in (7) the cost has no upper bound, and that is also unrealistic. This section considers two ways to extrapolate the observed data up to a maximum possible blackout cost.

There is some uncertainty about the maximum possible blackout cost. To obtain a rough estimate, we suppose that the maximum possible blackout cost $c_{\text{max}} = 744k$ corresponds to the customer cost of a total blackout lasting one month.

To extrapolate the exceedance curve tail beyond c_{maxobs} , we fit two different distributions to the observed costs exceeding c_{large} . Both distributions have large but finite mean and variance. A variant P_b of the Pareto distribution (7) with upper bound c_{max} is given by $\bar{F}_{P_b}(c) = (c^{-\alpha} - c_{\text{max}}^{-\alpha}) / (c_{\text{large}}^{-\alpha} - c_{\text{max}}^{-\alpha})$ for $c_{\text{large}} < c \leq c_{\text{max}}$. A bounded variant LN_b of the lognormal distribution conditions the standard lognormal distribution to have $c_{\text{large}} < c \leq c_{\text{max}}$. We calculate the relative standard error of estimates of the mean of P_b and LN_b with M samples as $\text{RSE} = (\text{standard deviation}) / ((\text{mean})\sqrt{M})$.

We also extrapolate to estimate the variability of ALEC: Assuming the Pareto tail (7), which extrapolates beyond c_{maxobs} , ALEC is the mean of (8), and we calculate the relative standard error of estimates of ALEC, RSE_{ALEC} , using (8).⁶

V. RESULTS

Table I shows all the results. The empirical customer cost exceedance curves have an approximately linear trend for the larger cost events on a log-log plot. In particular, utilities 1, 2, 4 have very heavy tails of absolute slope $\alpha < 1$, utility 3 has $\alpha \approx 1$, and utility 5 has a heavy tail with $1 < \alpha < 2$. These heavy tails indicate that large cost events show high risk and great variability, and that there is no typical large cost event. The less severe and more resilient case of utility 5 can likely be attributed to it being primarily urban. ALEC and the associated slope index α can be practically estimated with reasonable variability, as shown by their 95% confidence intervals CI_{ALEC} and CI_{α} for events observed over periods from 6 to 17 years.

However, there are severe difficulties in estimating the mean costs of the large events without taking the logarithm. We extrapolated the exceedance curve beyond the maximum observed cost in two ways (P_b and LN_b) to a maximum possible blackout. This gives large, finite means, but with high variability in estimates of the means: The relative standard errors RSE_{P_b} and RSE_{LN_b} are at least an order of magnitude larger than RSE_{ALEC} for utilities 1, 2, 3, 4, and at least 5 times larger than RSE_{ALEC} for utility 5. Therefore, utilities 1, 2, 3, 4 require at least two orders of magnitude more observed events to achieve the same variability of estimates of the mean as ALEC, showing that estimating the means of the untransformed data with reasonable accuracy requires impractically large amounts of data. Utility 5 requires an order of magnitude more observed events.

VI. CONCLUSION

We quantify risk with a customer cost exceedance curve for five distribution utilities by processing their detailed outage

⁵A larger absolute slope α reduces large blackouts. α does not depend on the multiplicative scaling of the costs or k or $n_{\text{customers}}$.

⁶Upper bounding (8) with c_{max} makes no practical difference.

data into events and assuming that the event customer cost is proportional to the customer hours lost. The empirical customer cost exceedance curves have heavy tails, implying that the risk of large-cost events is very high despite their rarity. It is impractical to estimate the mean cost of large events for our utility data, because the available amount of data is insufficient to get a workable estimate of the mean. Moreover, there is no representative or typical large-cost event. These conclusions imply that the usual practices of computing metrics involving or optimizing means for extreme event risk can be invalid. For example, Value at Risk remains valid [3], but Conditional Value at Risk, which involves taking a mean value, cannot be applied. While we are optimistic that the heavy tails will be observed more universally so that these conclusions apply more broadly, the heaviness of the tails in the data of other utilities must be individually checked.

However, we do suggest metrics that can describe the extreme event risk by characterizing the tail of the cost exceedance curve with the probability p_{large} of exceeding a large event cost and either the Average Log Event Cost ALEC metric or the slope index α . α can be calculated with a Hill estimator, and ALEC is directly related to α . p_{large} can also be converted to the annual frequency of large cost events f_{large} , and f_{large} can be multiplied by ALEC to yield the Annual Log Cost Resilience Index ALCRI. These metrics should be used instead of relying on mean values of extreme event cost to get practical and meaningful results. ALEC is easy to explain and calculate: take the average of the logarithms of the costs of the large cost events. Similarly, ALCRI is an annual sum of the logarithms of the costs of the large cost events. ALEC and ALCRI work since taking the logarithm of heavy-tailed data gives light-tailed data to which the usual mean calculations apply.

Heavy tails and high variability are traditionally acknowledged in distribution systems when calculating reliability metrics by excluding the heavy tails using major event days. However, there is an increasing need for metrics to quantify the risk of extreme weather events in the heavy tails in order to monitor and mitigate resilience. This quantification currently relies on metrics such as Conditional Value at Risk and other mean values

that only work for sufficiently light tails. We show five cases of distribution utilities in which the heavy tails and huge variations of event costs cause problems with these metrics. To address this, we propose resilience metrics that can practically describe and take account of the heavy tails. That is, we address a problem of extreme variability in quantifying the risk of extreme events that occurs in some observed distribution systems' data, and suggest resilience metrics that solve this problem.

ACKNOWLEDGMENT

The views expressed herein do not necessarily represent the views of US DOE or the United States Government.

REFERENCES

- [1] A. M. Stanković et al., "Methods for analysis and quantification of power system resilience," *IEEE Trans. Power Syst.*, vol. 38, no. 5, pp. 4774–4787, Sep. 2023.
- [2] N. K. Carrington, I. Dobson, and Z. Wang, "Extracting resilience metrics from distribution utility data using outage and restore process statistics," *IEEE Trans. Power Syst.*, vol. 36, no. 2, pp. 5814–5823, Nov. 2021.
- [3] S. Poudel, A. Dubey, and A. Bose, "Risk-based probabilistic quantification of power distribution system operational resilience," *IEEE Syst. J.*, vol. 14, no. 3, pp. 3506–3517, Sep. 2020.
- [4] A. Ahmad and I. Dobson, "Quantifying distribution system resilience from utility data: Large event risk and benefits of investments," in *Proc. CIRED Workshop Resilience Electric Distribution Syst.*, Chicago, IL, USA, Nov. 2024, pp. 118–122.
- [5] S. Kaplan and B. J. Garrick, "On the quantitative definition of risk," *Risk Anal.*, vol. 1, no. 1, pp. 11–27, 1981.
- [6] A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data," *SIAM Rev.*, vol. 51, pp. 661–703, 2009.
- [7] J. Nair, A. Wierman, and B. Zwart, *The Fundamentals of Heavy Tails: Properties, Emergence, and Estimation*. New York, NY, USA: Cambridge Univ. Press, 2022.
- [8] S. I. Resnick, *Heavy-Tail Phenomena; Probabilistic and Statistical Modeling*. New York, NY, USA: Springer, 2007.
- [9] *IEEE Guide for Electric Power Distribution Reliability Indices*, IEEE Standard 1366-2022, 2022.
- [10] S. Pandey et al., "Resiliency metrics quantifying emergency response in a distribution system," *IEEE Power Energy Soc. Gen. Meeting*, Austin, TX, USA, pp. 1–5, 2025.
- [11] 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.