Peak Load Estimation with the Generalized Extreme Value Distribution

Jonathan Lee
Mohini Bariya
Duncan Callaway

Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2022-60
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2022/EECS-2022-60.html

May 11, 2022

Peak Load Estimation with the Generalized Extreme Value Distribution

Jonathan T. Lee, *Student Member, IEEE,* Mohini Bariya, *Student Member, IEEE,* and Duncan Callaway *Member, IEEE*

***Abstract*—Estimating the coincident peak load of a group of loads is a critical task in power system planning and reliability analysis. Classical methods using coincidence and load factors have long been used, but leave a challenge for designers and modellers to determine appropriate factors to use and do not lend themselves to reliability analysis. This paper follows work that models peak load as a random variable, and contributes a parametric model that relates the probability distribution of peak load to average energy consumption using extreme value theory. This model allows designers to specify failure probabilities, and under some simple assumptions yields closed-form functions that can be used in planning models. The paper presents a procedure for fitting the model and discusses some modifications for tuning it to particular applications. Computational experiments on reference residential load data sets from Texas and London show the model predicts peak load with 2% median error on test data across a range of group size and failure probabilities. We find the performance degrades somewhat for small samples of more heterogenous loads, with a 13% median error on a set of 25 loads from New York with individual load factors as low as 0.02 and as high as 0.15.**

I. Introduction

MODERN power systems are being pushed to be more resilient, efficient, and flexible, with increasing attention to grid decentralization and distribution system planning. The cost-efficient design of microgrids and distribution networks, and the ability to model costs and reliability of different grid topologies, are important topics for both practitioners and researchers. Coincident peak load prediction at different points on distribution networks is a fundamental component of capacity sizing used in network modelling and planning. This paper explores the use of Extreme Value Theory (EVT) as a tool for stochastically estimating coincident peak load across different group sizes.

Beyond specific modelling problems, general theoretical insight into how the coincident peak load of a group of loads scales with the group size is important for understanding economies and diseconomies of scale in networked power systems. The positive economies of scale from relative coincident peak reduction by aggregating loads onto a common power source has been recognized since the early 20th century: because it is unlikely that all loads peak at exactly the same time, they can share a power supply with a total capacity less than if they each had individual supplies. However, this aggregation comes at the cost of a distribution network. Traditionally, the balance has been in favor of greater aggregations; however, reductions in the costs of smaller scale distributed energy resources, increasing costs from transmission and distribution networks due to wildfires and hardening against extreme weather events, and the resilience benefits from having self-sufficient sections of the grid call this into question. The ability to characterize this tradeoff depends directly on models for how coincident peak load scales with group size.

The greater availability of load data through automatic metering infrastructure presents an opportunity to improve methods for peak load estimation by enabling estimation procedures to be more data driven. This is a fertile research area that we discuss in our review in the next section, and it is almost certain that particular and perhaps more complex methods can be tailored to perform well in different contexts. In this paper, our objective is to develop a model that is transparent, general, easily applicable, insightful, and grounded in statistical theory. As such, we develop a compact parametric model with four parameters fit from data and two user-supplied inputs. Perhaps remarkably, our computational experiments show that it yields highly accurate estimates despite its simplicity, and we also provide some discussion on tuning it for specific contexts or objects.

The model, which we call the Extreme Value Load Estimator (EVLE), relates a certainty probability, which can be thought of as a predicted reliability or the complement of a failure probability, to a peak load value over a time interval. In other words, given a peak load value, it estimates the probability that over some time horizon a load will be observed that exceeds that value. Mathematically, this corresponds exactly to estimating the cumulative distribution function of the peak load. The model is also meant to be used in inverse for capacity sizing: given a desired probability, it estimates the corresponding load value, i.e. the capacity necessary for a conductor or a power source such that it has a desired probability of being overloaded. In this way, it is equivalent to estimating the value-at-risk of the maximum load.

The EVLE estimate depends on a prior assumption of the expectation of the load of the aggregated group over time, or equivalently, the total energy consumption of the group. This can also be framed as the number of customers each with a typical energy consumption. This assumption begs the question, "If the modeller can estimate the energy consumption, why do they not also know the peak?" One reason in planning problems is that an estimate of energy consumption can be generated from socioeconomic and environmental factors, e.g.

J.T. Lee and D. Callaway are with the Energy and Resources Group and the Department of Electrical Engineering and Computer Science, University of California, Berkeley.

M. Bariya is with the Department of Electrical Engineering and Computer Science, University of California, Berkeley

from home size or wealth, but such a model for peak load is less likely to be found. Another arises when modelling groups or aggregates of loads: the energy consumption of the group is additive, meaning it can be computed as the sum of each individual's energy consumption, but the peak load of the group cannot be computed from the individual peaks without additional assumptions or information. Theoretically, we should also expect that an empirically derived estimate of the mean of a signal will be more stable than an estimate of the maximum, and it has been observed that utilities generally predict peak load especially poorly [1]. The essential structural assumptions of the EVLE model are 1) scaling assumptions that relate the expectation and variance of the peak load to the mean, and 2) the form of the distribution of the peak load.

An example application of the EVLE model is to select an inverter size for a DC-coupled solar plus battery autonomous microgrid when the microgrid is expected to serve a community with an average energy consumption forecasted to be 200 kWh per day. Another is to size the conductor on feeder segments for a given network layout, where each segment serves e.g. 5, 10, 50 customers of different and known home sizes, where the modeller has access to the typical annual energy consumption for each of the home sizes. The EVLE model yields formulae similar to some longstanding methods in the literature, but provides a greater ability to control the reliability, or failure probability, of different designs.

II. Literature Review

We focus our review on peak load estimation at the distribution spatial scale and at the planning time scale, although we include some methodologically relevant work that studies the transmission scale or operational time scales. We first discuss classical "factor" based methods, modifications that have been proposed, and statistical analyses. Second, we survey models that directly address the relationship between peak power and energy across different numbers of customers. Third, we review recent work that uses extreme value theory to estimate peak load. We rely primarily on academic literature, engineering handbooks, and technical reports. The availability of actual utility practices vary regionally. In the U.S., [1], [2] detail peak load forecasting practices at the transmission scale, but otherwise we were unable to find good references covering modern utility practices at the distribution scale. Many of the papers we cite focusing on European countries include more information about the state-of-practice and industry guidelines, and we note these below. One important area we do not discuss is the use of simulations to estimate peak load. This is because when constructed in a "bottom-up" manner from behavioral models, this method is methodologically unrelated to the EVLE model; however, the EVLE method could be used in a "top-down" way to validate the results of bottom-up models.

The classical method for estimating coincident peak load for a group uses the *coincidence factor* (CF), or its reciprocal, the *diversity factor* (DF). The CF is the ratio of the peak aggregate load to the sum of the of peaks of the individual loads, so one can estimate the aggregate peak as the sum of individual peaks times the CF [3], [4]. If the individual peaks are not given directly, the *load factor* (LF) can be used to estimate each peak. The LF is the ratio of the mean load to the maximum. Note that we are referring to the maximum and mean over time, and the "aggregate" load refers to a time series given by summing over a group of loads at each moment in time. Some representative values for these factors are given in [3], which draws on empirical studies from the mid 20th century, and more recent empirical analyses of LFs can be found in [5], [6]. All of these show a substantial variation in LFs across customer type and among individual customers, and it is also well-recognized that the CF depends on the scale; i.e. it is lower for 100 customers than for 10. Reference [7] proposes an additional *contribution factor* by customer type to improve the accuracy of CFs. It is frequently noted in the literature that despite the availability of some empirical studies that give typical factors, it is difficult to determine the appropriate factors to use in any given context. They can in principal be fit from data, but additional statistical analysis beyond a standard regression is necessary to estimate probabilities of exceeding any given peak load. These considerations limit the usefulness of the factors.

A set of papers [8]–[10] examine the distributions of diversity and load factors of randomly sampled groups of customers. Reference [8] claims the DF is normally distributed, while later works [9] and [10] indicate it follows a Gamma distribution. Both [9] and [10] show the distribution of the DF depends on the group size. Their models do not take the group size into account or generalize across scale; they are fit to a particular group size. Additionally, when individual peaks are not known and the LF must be used, then the group peak depends on the ratio of random variables LF/DF and becomes more complicated than fitting the distribution of the peak directly with the energy usage or group size as a covariate. This issue is recognized in the recent study [11], which also finds that the expectation of the peak load has a linear dependence on the number of customers.

Models including the relationship between peak load and total energy consumption at distribution scales have been used for decades in Scandinavian power systems and are reviewed in [12]. In particular, the Velander method models the peak load as a linear combination of the annual energy consumption and its square root. In [13], the authors note there is little knowledge of the proper coefficients to use in this model. The model form can be derived by assuming the individual peak load is normal and i.i.d., and then a probability distribution can be constructed by assuming values for the mean and standard deviation as shown in [13]; however, the assumption of the normal distribution has not been compared against others. Fitting this model can be understood as a standard linear regression on energy consumption and its square root, and other regression techniques have been proposed. The study [14] uses a fuzzy linear regression model on energy consumption to obtain tighter bounds around the peak load, although the distribution of the residuals is not analyzed, and [15] uses a quantile regression to predict transmission-scale peak load in the U.S. from energy consumption and additional features. Neither [14] nor [15] consider the square root of energy consumption.

More recently, researchers have begun to explore prediction based on Extreme Value Theory (EVT). EVT studies the distribution of sample maxima based on the Extreme Value Theorem, which can be loosely thought of as an analog to the Central Limit Theorem. The theorem states that the only possible limiting distribution for the maximum of a sequence of random variables is the Generalized Extreme Value distribution (GEV). The GEV is classified according to three types, called the Reversed Weibull, Gumbel, and Frechèt, depending on its shape parameter ξ, with distinct tail behavior. A formal overview of EVT is given in [16] and a higher-level introduction including applications to electricity is found in [17]. This theory is naturally attractive for modelling maximum, or peak, loads. Reference [17] focuses on relatively short-term forecasting of peak load on a weekly time scale, finding that it follows the Reversed Weibull type with a sharp, bounded tail, and using EVT to generate point estimates of the upper bound. Reference [18] uses an EVT result to estimate, for a single load profile, the number of times the load exceeds a given threshold in a down-sampled period of time given behavior over a longer period. EVT predicts the threshold exceedances follow a Poisson Point Process (PPP). Most closely related to our work, [19] uses the PPP result to estimate the GEV parameters to predict the distribution of the peak load. Unlike our work, [19] focuses on transmission scale and predicting the future demand of a single substation, rather than load aggregation at the distribution scale. They include both the number of customers and total energy demand as covariates; however, they do not consider the square root of energy, which we find to be significant. Interestingly, they find solar photovoltaic penetration to be an insignificant covariate. Our study is the first to examine the ability of EVT to predict peak load across scales within the distribution system.

III. EVLE Model

In this section, we formally develop 1) a probabilistic model for the "block maximum" load over a period of time, and 2) a process for fitting the model to data. We emphasize that the model is of the empirical (also referred to as the observed, sample, or realized) maximum, and not the maximum possible value, because a system designer is concerned with the probability of actually observing a value higher than some threshold, and that the true maximum of a distribution may be unbounded while the empirical maximum can be computed with some probability.

Our central hypothesis is that the distribution of coincident peak load of a randomly selected group follows a GEV with a location and scale that depends on the expectation of the mean load of the group. The parameters of the model are fit from a set of empirical observations of sample groups of different sizes with varying empirical maxima and means. As a result the model can be used to predict the maximum load of groups of different size with different energy consumption.

A. Model Definition

Formally, we define a set of loads as $\mathcal{L}$, where $N := |\mathcal{L}|$ is the number of individual loads in the group. We denote the block maximum coincident load over a discrete interval $\mathcal{T} = \{1, \ldots, T\}$ as the random variable $\hat{M}_{\mathcal{L}}$, and estimate F, the cumulative distribution function (CDF) of $\hat{M}_{\mathcal{L}}$. To be precise, let the sample of load $n \in \{1, \ldots, N\}$ at time $t \in \mathcal{T}$ be $\mathcal{X}_{n,t}$. Then $\hat{M}_{\mathcal{L}} = \max_t \sum_n \mathcal{X}_{n,t}$ and $\hat{m}_{\mathcal{L}} = T^{-1} \sum_t \sum_n \mathcal{X}_{n,t}$, where $\hat{m}_{\mathcal{L}}$ is the block mean. We assume T is sufficiently large for F to converge to a GEV and also for $\hat{m}_{\mathcal{L}}$ to converge to the expectation of the group load over time, denoted $m_{\mathcal{L}}$. Any fit parameters will depend implicitly on the period length T, and also the sampling frequency, but we drop time from our notation for brevity, except in section III-C where we discuss extrapolation in time. In practice, we work with on the order of a year of data sampled sub-hourly.

We assume that the only dependence of F on the specific group $\mathcal{L}$ is captured by $m_{\mathcal{L}}$, and that a system designer has an estimate of $m_{\mathcal{L}}$ that is in general a probabilistic estimate with density $h_{\mathcal{L}}$ and support $\mathcal{H}_{\mathcal{L}}$. In this way the model is very general. In practice, most likely the designer will assume a point estimate m, (in which case $h_{\mathcal{L}}$ is a delta function at m, $h_{\mathcal{L}}(m) = 1$), but it allows the designer to use a more sophisticated estimate. Therefore, F is a *conditional* CDF, conditioned on the mean load $m_{\mathcal{L}}$.

To apply the model, a system designer must specify only a certainty probability $\phi \in (0, 1)$ in addition to $h_{\mathcal{L}}$. The model returns a value $M_{\mathcal{L},\phi}$ such that the probability of observing a maximum load $\hat{M}_{\mathcal{L}}$ higher than this value is $1 - \phi$. Given these inputs, the model can be stated as (1).

$$M_{\mathcal{L},\phi} = \{M : P(\hat{M}_{\mathcal{L}} \leq M) = \phi\} \tag{1}$$
$$P(\hat{M}_{\mathcal{L}} \leq M) = \int_{\mathcal{H}_{\mathcal{L}}} F(M|x) h_{\mathcal{L}}(x) dx$$
$$h_{\mathcal{L}}(m) = 1 \implies M_{\mathcal{L},\phi} = F^{-1}(\phi|m) \tag{2}$$

We assume F is approximately a GEV, parameterized by shape ξ with location α and scale $\beta > 0$. The GEV density $g_{\alpha,\beta,\xi}$ and CDF $G_{\alpha,\beta,\xi}$ are defined by (3)-(4) [16].

$$g_{\alpha,\beta,\xi}(x) = \frac{1}{\beta} t(x)^{\xi+1} e^{-t(x)} \tag{3}$$
$$G_{\alpha,\beta,\xi}(x) = e^{-t(x)} \tag{4}$$
$$t(x) = \begin{cases} e^{-\left(\frac{x-\alpha}{\beta}\right)} & \text{if } \xi = 0 \\ \left(1 + \xi\left(\frac{x-\alpha}{\beta}\right)\right)^{-1/\xi} & \text{if } \xi \neq 0 \end{cases} \tag{5}$$

When $\xi = 0$, the GEV reduces to the Gumbel distribution and has support $x \in (-\infty, \infty)$, otherwise the support is bounded by $\beta + \xi(x - \alpha) > 0$. When $\xi > 0$, the support is bounded to the left, and the GEV is heavy tailed to the right. When $\xi < 0$, the support is bounded to the right, and the GEV is light-tailed to the left. This was the behavior for electricity found in [17], [19] and is the typical case for maxima of random variables that are bounded on both sides, such as the uniform or Beta distributions, which is the natural assumption for electric loads. In our experiments, $\xi > 0$ was fit only in isolated degenerate cases, so we restrict consideration to $\xi \leq 0$ following [17], [19]. From here on out, we show formulae assuming $\xi \neq 0$ for brevity. In this case, the inverse CDF is given by:

$$G^{-1}_{\alpha,\beta,\xi}(\phi) = \alpha + \frac{\beta}{\xi}\left(\log^{-\xi}(\phi^{-1}) - 1\right) \quad (6)$$

We make the following two assumptions that relate the distributions parameters to the mean load: 1) the expectation of the peak load is linear with respect to the mean load *and* the square root of mean load, and 2) the variance is linear with respect to the mean load. These assumptions are theoretically justified for loads with similar individual statistics (see Appendix A for discussion). The assumptions are stated for coefficients a, b, and c in (7)-(8) with formulae for the expectation and variance of a GEV, from which expressions (9)-(10) for α and β conditional on m are derived. Γ denotes the Gamma function.

$$\mathrm{Exp} = \alpha + \beta\xi^{-1}(\Gamma(1-\xi)) := am + b\sqrt{m} \quad (7)$$

$$\mathrm{Var} = \beta^2\xi^{-2}\left(\Gamma(1-2\xi) - \Gamma(1-\xi)^2\right) := cm \quad (8)$$

$$\alpha = \alpha(m) = am + b\sqrt{m} - \beta(m)\xi^{-1}(\Gamma(1-\xi)) \quad (9)$$

$$\beta = \beta(m) = c|\xi|\left(\Gamma(1-2\xi) - \Gamma(1-\xi)^2\right)^{-1/2}\sqrt{m} \quad (10)$$

Defining a parameter vector $\theta = [\xi, a, b, c]$, it is straightforward to substitute (9)-(10) into (3)-(6) to obtain density $f(\cdot|m,\theta)$, CDF $F(\cdot|m,\theta)$, and inverse CDF $F^{-1}(\cdot|m,\theta)$ for $\hat{M}_{\mathcal{L}}$ conditional on the true mean load m and parameters θ. When $m_{\mathcal{L}}$ is a point estimate at m, then the simplification (2) holds, and the model can be written as:

$$M_{\mathcal{L},\phi} = k_1 m + k_{2,\phi}\sqrt{m} \;\text{ if } h_{\mathcal{L}}(m) = 1 \quad (11)$$

$$k_1 = a \quad (12)$$

$$k_{2,\phi} = b + \frac{\log^{-\xi}(\phi^{-1}) - 1 - \Gamma(1-\xi)}{\sqrt{\Gamma(1-2\xi) - \Gamma(1-\xi)^2}}c \quad (13)$$

It is noteworthy that this expression is the same form as the Velander formula [12], [13] for a given ϕ; however, this formulation allows the designer to specify the confidence ϕ in a meaningful way, thereby generalizing the Velander formula to include a reliability parameter. However, our model yields different equations to fit parameters to than standard regression fitting procedure and follows from different statistical assumptions.

B. Fitting the Model

In this section we show the two steps to training the model: 1) constructing aggregate load samples, and 2) fitting parameters with maximum-likelihood estimation (MLE). The main consideration with the sampling is the proportion of aggregations of different numbers of loads. For the MLE, we develop a search procedure with a locally convex subproblem that can be used to efficiently train the model on large data sets.

Each sample i selects a subset of load profiles $\mathcal{L}_i$ from a data set $\mathcal{L}$. The data set has N individual profiles and each sample has $N_i \leq N$. We first draw N_i from a distribution $P(N_i = k) = \binom{N}{k}2^{-N}$, where $k \in \{1, 2, \ldots, N\}$, such that the probability of choosing N_i loads to aggregate is proportional to the number of combinations of N_i loads in $\mathcal{L}$. Next, we randomly select N_i unique loads from $\mathcal{L}$ to obtain $\mathcal{L}_i$ and compute peak (M_i) and mean (m_i) load. This is repeated S times so that $\mathcal{L}_i$ sampled from $\mathcal{L}$ *with* replacement. We set S *a priori*, although a bootstrapping method could be used.

This sampling distribution for N_i is loosely "least-biased" in the case that predictions will be made for new groups similar to $\mathcal{L}$ without a prior assumption on the number of loads in the new group. In practice, users may have a large training data set, and benefit from restricting the sampling distribution to specific ranges of N_i that are of interest. For example, one could use a set of 1,000 loads to train separate models to predict for groups of 5 to 20 and 20 to 50.

Given the training samples, we use MLE to compute the parameters θ^* that maximize the probability of observing the output data $\{M_i\}$ given the input data $\{m_i\}$. This can be stated as (14), and is equivalent to (15) by assuming the samples are independent and transforming the objective to the log-likelihood.

$$\theta^* = \arg\max_{\theta} P(\{M_i\}|\{m_i\}, \theta) \quad (14)$$

$$= \arg\max_{\theta} \sum_{i=1}^{S} \log f(M_i|m_i, \theta) \quad (15)$$

The MLE problem includes the constraint $\beta(m_i) + \xi(M_i - \alpha(m_i)) > 0 \, \forall i$ enforcing the support of the distribution. As the objective is undefined for M_i outside of this constraint, it is necessary in practice to use a numerical solver that either enforces feasibility at every iteration or can preserve an estimate of the gradient when the objective is undefined.

The form of our model, where the location and shape parameters are linear functions of the covariates m and $\sqrt{m}$, is a generalization of those presented in [20]. It is common practice in the literature to solve the MLE problem for the GEV by solving its the first-order optimality conditions; however, [21] points out in an analysis of the consistency of the estimator that local maximizers exist, and it follows that a local solution to (14) does not converge to the true parameters. Because of this, we provide more details on our solution implementation for transparency.

We solve the problem numerically by starting with $\xi = 0$ (the Gumbel distribution) and solving the MLE problem for $[a, b, c]$, which can be shown to be convex under change of variables. We then search by decreasing ξ by a small fixed step-size of 0.01 and use the optimal $[a, b, c]$ from the previous step as an initial point to warm-start the MLE for the new ξ (projecting the initial $[a, b, c]$ to the feasible region if necessary). The search continues while $\xi > -0.5$, recording the likelihood and parameters at each step, and then concludes by selecting the parameters associated with the maximum likelihood. With this approach, we observed numerically stable results and solution times of a few seconds on a personal computer using the SLSQP algorithm with the SciPy optimization package in Python (`scipy.optimize.minimize`).

C. Extrapolation in Time

The EVLE model for the maximum load developed in Section III is fit over a a finite time interval $\mathcal{T}$ with T time steps, for example hourly samples over a year. In this section we describe how the model can be extrapolated to a longer time interval with potentially changing mean load, for example a 20 year horizon with load growth, when the model is fit over a subinterval with the same sampling frequency. In deriving this extrapolation, we assume the model is independent in each interval, conditional on the mean load in that interval. Care should be taken to validate this assumption in practice.

Denote the longer interval $\mathcal{T}'$ and the j'th of J subintervals as $\mathcal{T}_j$, so $\mathcal{T}' = \bigcup_j^J \mathcal{T}_j$. The model output over the longer interval $\mathcal{T}'$ is $M_{\mathcal{L},\phi,\mathcal{T}'}$ can be stated analogously to (1) as (16).

$$M_{\mathcal{L},\phi,\mathcal{T}'} = \{M : P(\hat{\mathcal{M}}_{\mathcal{L},\mathcal{T}'} \leq M) = \phi\} \tag{16}$$

Eq. (17) follows from the assumption that the model is independent in each interval, conditional on the mean load, and reduces to (18) if the density of the mean load in each interval j is a delta function at m_j.

$$P(\hat{\mathcal{M}}_{\mathcal{L},\mathcal{T}'} \leq M) = \prod_j^J \left(\int_{\mathcal{H}_{\mathcal{L}_j}} F_j(M|x) h_{\mathcal{L}_j}(x) dx \right) \tag{17}$$

$$h_{\mathcal{L}_j}(m_j) = 1 \Rightarrow P(\hat{\mathcal{M}}_{\mathcal{L},\mathcal{T}'} \leq M) = \prod_j^J F_j(M|m_j) \tag{18}$$

This further reduces to (19) when the model parameters $[\xi, a, b, c]$ are assumed identical in each sub interval, but the mean changes. This must be solved numerically. When the mean is constant, the model output can be written analytically as (20), and is equivalent to the model output for a single period with certainty probability $\phi^{\frac{1}{J}}$, which can be substituted for ϕ in (13) to obtain an expression for $M_{\mathcal{L},\phi^{\frac{1}{J}}}$.

$$F_j \equiv F \Rightarrow M_{\mathcal{L},\phi,\mathcal{T}'} = \{M : \prod_j^J F(M|m_j) = \phi\} \tag{19}$$

$$m_j \equiv m \Rightarrow M_{\mathcal{L},\phi,\mathcal{T}'} = F^{-1}(\phi^{\frac{1}{J}}|m) = M_{\mathcal{L},\phi^{\frac{1}{J}}} \tag{20}$$

IV. Computational Experiments

In this section, we perform computational experiments on publicly available datasets to validate the EVLE model and to compare it to alternatives. The first evaluates our scaling assumptions (7)-(8) against alternative linear and affine alternatives. The second compares the quality of fit of the EVLE model against alternative distributions to the GEV using the same scaling assumptions. The third compares examines how the model performance and fit parameters vary across the different load data sets.

We follow the methodology and nomenclature introduced in [22], of which the key aspects are defining independent variables for each experiment, varying confounding variables over a set of trials, and analyzing the distribution of performance metrics across trials for each independent variable. We evaluate performance by comparing the predicted distribution of the peak load to the observed peak load, normalizing across the mean load in three ways: 1) rescaling the empirical and predicted data to a standard GEV for visual tests, 2) introducing a prediction percent error metric, and 3) computing the Kullback-Leibler (KL) divergence (also called relative entropy) of the predicted and observed distributions.

TABLE I
Comparison of divergence on Texas Data

	Train	Test
Model A	5.71 (5.58, 5.78)	6.35 (5.88, 7.05)
Model B	6.22 (6.08, 6.27)	6.67 (6.36, 7.17)
Model C	6.23 (6.1, 6.28)	6.85 (6.4, 8.12)
EVLE	5.61 (5.48, 5.71)	6.09 (5.84, 6.76)
EVLEG	5.63 (5.49, 5.73)	5.95 (5.79, 6.31)

The rescaling procedure computes an empirical distribution $Z = z_i = (M_i - \alpha(m_i))/\beta(m_i)$. This Z can be compared visually and quantitatively against $G_{0,1,\xi}$. We define a prediction percent error ε_ϕ such that scaling all predicted peak loads in the sample by $1 - \varepsilon_\phi/100$ would make the predicted and observed failure rate equivalent:

$$\varepsilon_\phi : \frac{1}{S}\sum_{i=1}^{S} \mathcal{I}\left((1-\varepsilon_\phi/100)F^{-1}(\phi|m_i) > M_i\right) = \phi \tag{21}$$

$$\varepsilon := \frac{1}{S}\sum_{i=1}^{S} \left|\varepsilon_{\frac{i}{S}}\right| \tag{22}$$

Here, ε_ϕ is only defined for $\phi \in \{\frac{1}{S}, \frac{2}{S}, \ldots, 1\}$ and is computed as the largest value satisfying (21), $\mathcal{I}$ is the indicator function, and ε is the mean absolute percent error over all possible probabilities. In experiments where we use training and test data, we partition 50% of the data to each so that the number of loads is the same in both populations. In each trial, the partition is selected randomly as a confounding variable. We also use the KL-divergence from the fitted model to the empirical data as a performance metric. While ε_ϕ emphasizes the quality of the fit in matching the empirical percentiles, the divergence quantifies the similarity of the fitted and empirical densities. Divergence is defined on a pair of either discrete or continuous distributions; here we are interested in the divergence between the continuous fitted model and the discrete distribution of empirical samples from the true model. We therefore use a goodness of fit formulation of the divergence from [23] that is an estimator of the true divergence under finite empirical samples.

We use three residential data sets in the experiments, referred to as “Texas”, “New York”, and “London”. The Texas and New York data are from Pecan Street, Inc. [24] and the London data is from the Low Carbon London project as used in [10]. We processed the data by removing anomalously high values from the Pecan Street Data, removing negative values, and removing any load profiles with less than 90% data coverage. In constructing aggregate load profiles, we ignored any aggregate profiles with less than 95% data coverage averaged across all individuals in the group, and computed the aggregate mean as the sum of all individual means ignoring missing values. Table II gives the number of loads in each data set and their temporal sampling. Figure 1 shows the mean and max of individual loads. Note that both the U.S. data sets have higher mean loads, and some of the New York loads have extremely low load factors.

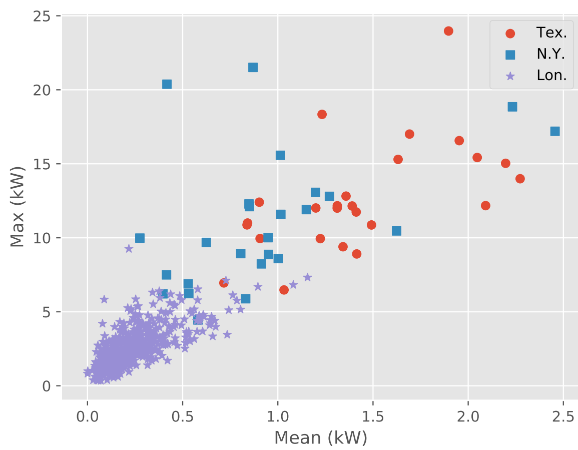

Fig. 1. Individual load characteristics

A. *Scaling Assumptions*

We compare the performance of the EVLE model to alternate models with similar, but theoretically unjustified, scaling assumptions. The comparisons show the efficacy of EVLE over other models in matching the empirical data, bolstering our claims from Section III on the relationship between the parameters of the peak load distribution and mean load. The three alternate scaling models are termed Model A, Model B, and Model C. The functional forms they assume for the peak load distribution parameters in terms of mean load are:

- Model A: $\text{Exp} = am + b,\ \text{Var} = cm$
- Model B: $\text{Exp} = am,\ \text{Var} = cm$
- Model C: $\text{Exp} = am,\ \text{Var} = c$

The fit of the models compared to EVLE is visualized in Fig. 2. The upper plot compares the empirical data with expectation of maximum load for a given mean load as predicted by each model. The lower plots compare the distribution of Z, the *rescaled* empirical data, with the fitted model distributions. If the model is accurate, we expect the distribution of Z to closely match the fitted distribution curve. Visually, the EVLE model fits the data better than Models A, B, or C. This is also evident in the kl-divergences from the fitted models to the data, recorded in Table I. The table includes divergence on training and test data, reporting the median followed by the 5th and 95th percentiles in parentheses across 30 trials. Notice that the EVLE divergence is lower than that of Model A, B, or C on both training and test data sets. Fig. 3 compares the absolute value of the prediction percent error ε_ϕ of the models across a range of failure probabilities. EVLE generally outperforms the alternate models across probabilities. The difference is very significant for high values of ϕ, implying that EVLE captures the tail of the empirical distribution the best. Altogether, these results are compelling evidence that the scaling assumptions of the EVLE model are sound.

B. *Comparison of Distributions*

A fundamental assumption of the EVLE model is that, for a given mean load, the corresponding maximum load follows

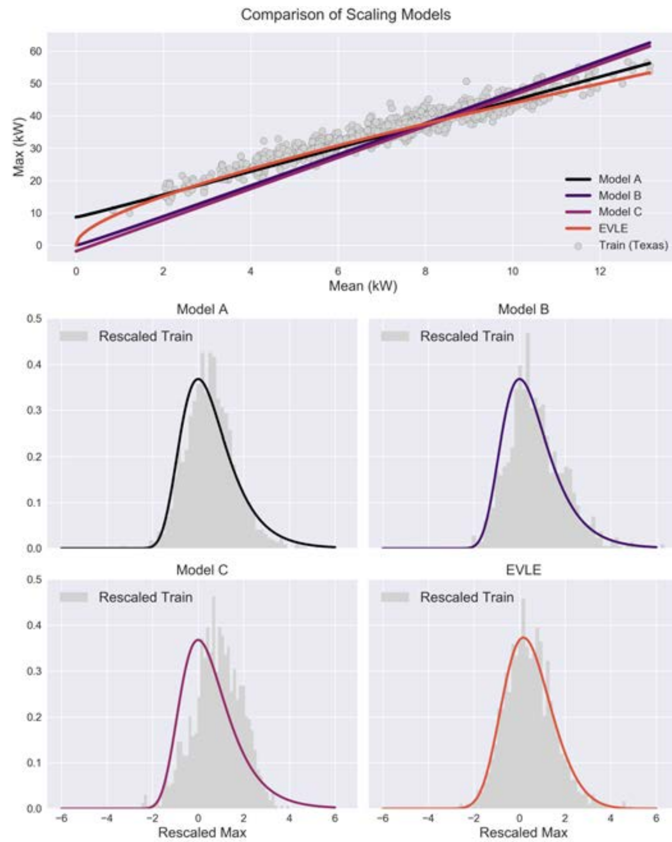

Fig. 2. Visualizing the efficacy of EVLE's scaling assumptions compared to three alternate models for a single, sample trial.

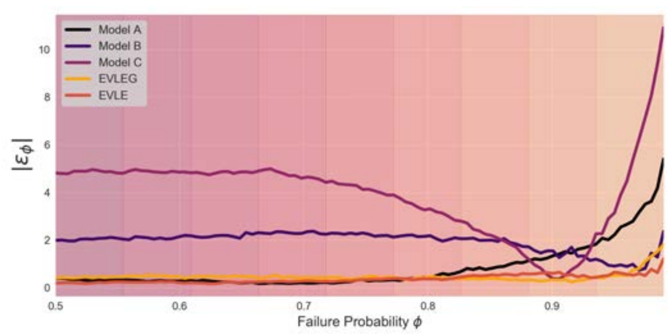

Fig. 3. Comparing the absolute value of the prediction percent error ε_ϕ over a range of failure probabilities ϕ between EVLE and alternate models on training data. A lower value implies better performance. Plot shows median across 30 trials.

a GEV distribution. This assumption is well motivated by the literature on extreme value theory. However, for further justification, we compare performance of the EVLE model to the *EVLEG* model, which makes identical scaling assumptions on the expectation and variance of maximum load, but models the distribution of the maximum load for a given mean load as Gaussian. Fig. 4 visualizes the fits of the EVLE and EVLEG. In the upper plots, empirical data is overlaid on color bands, where the color bands correspond directly to the bands in Fig. 2, and indicate the probability of observing a sample in that region according to each model. The lower plots compare the distribution of rescaled empirical data Z with the fitted

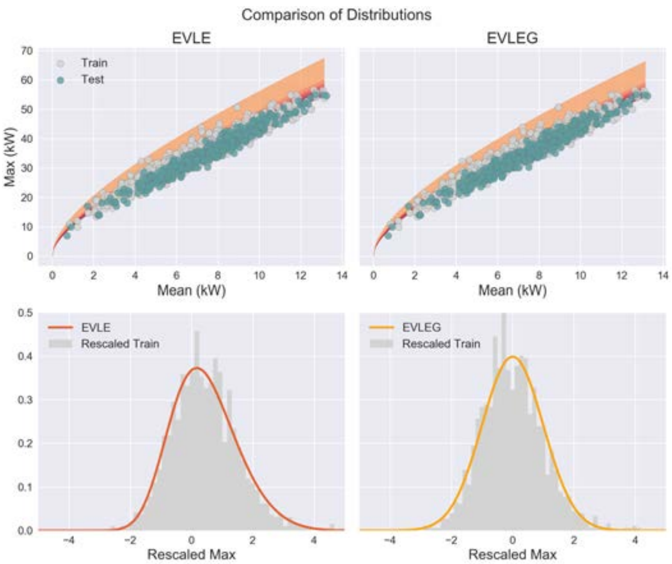

Fig. 4. Comparison of EVLE and EVLEG in describing the distribution of the maximum load for a single, sample trial.

distributions. Visually the models appear very close. In terms of kl-divergence too they have comparable performance: ELVE outperforms EVLEG on training, but not test data (Table I). A closer inspection of Fig. 4 suggests that EVLE does match the empirical distribution tails better than EVLEG: the scatter points appear to fit more comfortably within the EVLE probability bands and the asymmetric EVLE tails seem a better match to the empirical tail of the rescaled data. This assessment is corroborated by Fig. 3, in which EVLE does outperform EVLEG in terms of ε_ϕ at high failure probabilities. Therefore the GEV matches the empirical data distribution on the tails better than a Gaussian, as we would expect from extreme value theory.

C. Comparison of Data Sets

This experiment compares the EVLE model performance on the Texas, New York, and London data sets over 20 trials. The main results are in the bottom panel of Table II, which shows the median value of the parameters and error metric over the trials, followed by the 5th and 95th percentiles in parentheses. The much larger London data set yields the best performance, with a median error of 2.0% on test data over the trials. The Texas data yields similar median performance (2.4% median) but has a higher 95th percentile error.

The model performs notably worse on the New York data, which we attribute to it having especially high variation in the mean and max of individual loads as shown in Fig. 1. Here, the model fits a less than 1 (even equal to 0 in some trials) and yields higher b and c values. This can be explained by the dominance of the loads with low load factor in the small population (the scaling is highly nonlinear in this low mean load region), and shows a case where the data do not exhibit the scaling assumption of the model. Following the reasoning in Appendix A, this can be explained by the very high level of heterogeneity of the loads, and suggests that an extension to the model that considers different classes of loads could

TABLE II
EVLE Model Parameters and Performance By Data Set

	Texas	New York	London
N	25	25	624
Res.	1 min	1 min	30 min
Range	01/2018 - 01/2019	05/2019 - 11/2019	02/2012 - 01/2014
a	2.06 (1.82, 2.49)	0.61 (0.00, 1.25)	1.90 (1.73, 2.07)
b	7.72 (6.94, 8.36)	10.79 (8.96, 13.47)	2.00 (1.37, 2.88)
c	0.85 (0.64, 0.95)	1.86 (0.91, 2.38)	0.42 (0.38, 0.57)
ξ	-0.12 (-0.23, -0.07)	-0.07 (-0.17, -0.03)	-0.18 (-0.21, -0.06)
ε (train)	0.4 (0.2, 0.8)	1.1 (0.5, 2.5)	0.1 (0.1, 0.3)
ε (test)	2.4 (0.6, 9.3)	13.1 (1.9, 32.6)	2.0 (0.8, 5.2)

yield improved performance. However, the good performance London data, which contains some variation in load factor but is a larger data set, suggests the heterogeneity is less of an issue with a large sample size.

The parameters between the Texas and London data are similar, with a higher c value in the Texas data indicating a higher variance of the GEV, and consequently a greater sensitivity of the maximum load to the certainty probability ϕ.

V. Conclusion

In this paper, we develop the EVLE model, a method for peak load estimation using the Generalized Extreme Value (GEV) distribution to model the peak load as a random variable conditional on the average energy consumption. We show that a scaling model that assumes the expectation and variance of the peak load scale has both a theoretical intuition and performs well in computational experiments on groups of similar loads, and we show that the GEV distribution captures the tail of the distribution better than a Gaussian. We present a procedure for fitting the model from data, including the sampling strategy for constructing aggregate load samples that can be tuned for specific group sizes.

This paper contributes to a recent body of work applying Extreme Value Theory (EVT) to predicting peak electricity demand. Unlike classical factor based methods, EVT methods yield models for the distribution of peak load, which lends itself to reliability analysis and studying how system designs depend on reliability through the certainty probability ϕ. As a parametric approach with only four parameters that are fit from data, the model provides transparency and portability, which enables it to be included and interrogated in planning models. The model can take into account uncertainty in predicted energy consumption by accepting the mean load as an input that is a random variable. In this case, the model output is obtained by integrating with respect to the density of the mean load; however, when the mean load is a point estimate, the EVLE model simplifies to formula that can be evaluated directly (11). In principle, this model can be applied as an explicit risk-based constraint in an optimization-based planning model for sizing power system components, and this application is an important area for future work.

One of the computational experiments we conducted that compares the EVLE model on different data sets shows the model performs relatively poorly on data sets with a small number of individual loads with very heterogeneous individual load factors, specifically the New York meter data. We suspect this can be addressed by classifying loads and generalizing the model to include the number of loads from different classes as an input in addition to the mean load. Further developing this approach to include classification would be a welcome contribution, as would including the interaction of peak load with distributed generation. However, in the current form presented here, the model shows good prediction accuracy on test data when the individual loads have consistent load factors, as is the case with the London and Texas meter data. Most significantly, the model's accuracy is robust across different group sizes, mean loads, and failure probabilities.

Appendix A
Model Justification

We propose that the distribution of the maximum load conditioned on the mean load follows a GEV distribution: $\hat{M}_{\mathcal{L}} \sim g_{\alpha(m),\beta(m),\xi}$ where the GEV location and scale parameters are modelled as functions of the mean load according to (23) and (24) respectively.

$$\alpha(m) = am + b\sqrt{m} - \gamma\beta(m) \tag{23}$$
$$\beta(m) = c\sqrt{m} \tag{24}$$

Therefore, the claim of this model is that the expectation of the maximum load is linear in the mean and square root of the mean load, while the variance is a linear function of the square root of the mean load. To motivate these claims, we consider the simplified scenario where the individual loads are independent and identically distributed and sampled from a Gaussian Distribution:

$$\mathcal{X}_{n,t} \sim \mathcal{N}(\mu, \sigma^2) \tag{25}$$

The aggregate load over group $\mathcal{L}$ with size $N = |\mathcal{L}|$ at time t, denoted $\mathcal{X}_t^{(\mathcal{L})}$ is then normally distributed with mean $N\mu$ and variance $N\sigma^2$:

$$\mathcal{X}_t^{(\mathcal{L})} \sim \mathcal{N}(N\mu, N\sigma^2) \tag{26}$$
$$\Rightarrow m_{\mathcal{L}} = N\mu \tag{27}$$

We are interested in the maximum over T time points:

$$\hat{M}_{\mathcal{L}} = \max(\mathcal{X}_1^{(\mathcal{L})}, \ldots, \mathcal{X}_T^{(\mathcal{L})}) \tag{28}$$

There is no closed form solution for the distribution of $\hat{M}_{\mathcal{L}}$. However, the expectation of the maximum can be bounded above and below by functions of the variance and mean. From [25], if $Y \triangleq \max_{1\le i\le n} X_i$, where $X_i \sim \mathcal{N}(0, \sigma^2)$ are i.i.d., we have:

$$\frac{1}{\sqrt{\pi \log 2}}\sigma\sqrt{\log n} \le \mathbb{E}[Y] \le \sqrt{2}\sigma\sqrt{\log n}$$

If X_i is instead distributed according to $\mathcal{N}(\mu, \sigma^2)$—i.e. with arbitrary mean μ—we can incorporate this into the bounds using linearity of expectation: $\mathbb{E}[Y] = \mathbb{E}[Y - \mu] + \mu = \mathbb{E}[\max_{1\le i\le n}(X_i - \mu)]$. Note that $(X_i - \mu) \sim \mathcal{N}(0, \sigma^2)$. Therefore, the bounds become:

$$\mu + \frac{1}{\sqrt{\pi \log 2}}\sigma\sqrt{\log n} \le \mathbb{E}[Y] \le \mu + \sqrt{2}\sigma\sqrt{\log n} \tag{29}$$

We can now apply (29) to bounding $\hat{M}_{\mathcal{L}}$. Defining $\kappa \triangleq \sqrt{\log T}$, and plugging the distribution parameters of $\mathcal{X}_t^{(\mathcal{L})}$ in (26), we obtain:

$$N\mu + \frac{1}{\sqrt{\pi \log 2}}\sqrt{N}\sigma\kappa \le \mathbb{E}[\hat{M}_{\mathcal{L}}] \le N\mu + \sqrt{2}\sqrt{N}\sigma\kappa \tag{30}$$

Replacing N with $\frac{m_{\mathcal{L}}}{\mu}$ according to (27), allows the bounds of (30) to be expressed in terms of the aggregate mean. After some simplification, we obtain the final expression:

$$m_{\mathcal{L}} + \left(\frac{\sigma\kappa}{\sqrt{\mu\pi\log 2}}\right)\sqrt{m_{\mathcal{L}}} \le \mathbb{E}[\hat{M}_{\mathcal{L}}] \le m_{\mathcal{L}} + \left(\sqrt{\frac{2}{\mu}}\sigma\kappa\right)\sqrt{m_{\mathcal{L}}}$$

These bounds justify the form of (23).

We can standardize $\mathcal{X}_t^{(\mathcal{L})}$ to arise from a standard normal:

$$\mathcal{X}_t^{(\mathcal{L})} \sim \mathcal{N}(N\mu, N\sigma^2) \implies \frac{\mathcal{X}_t^{(\mathcal{L})} - N\mu}{\sqrt{N}\sigma} \sim \mathcal{N}(0, 1) \tag{31}$$

Suppose the maximum of T standard normals has variance $\sigma^2_{\max}$. Then, the variance of $\hat{M}_{\mathcal{L}} = \max(\mathcal{X}_1^{(\mathcal{L})}, \ldots, \mathcal{X}_T^{(\mathcal{L})})$ will be:

$$\mathrm{var}(\hat{M}_{\mathcal{L}}) = N\sigma^2\sigma^2_{\max} = \left(\frac{\sigma^2\sigma^2_{\max}}{\sqrt{\mu}}\right)m_{\mathcal{L}} \tag{32}$$

Therefore the variance of the maximum scales linearly in the mean aggregate load. The scale parameter is proportional to the square root of variance, which justifies the form of (24).

References

[1] J. P. Carvallo, P. H. Larsen, A. H. Sanstad, and C. A. Goldman, "Load forecasting in electric utility integrated resource planning," tech. rep., Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2017.

[2] T. Hong and M. Shahidehpour, "Load forecasting case study," *EISPC, US Department of Energy*, 2015.

[3] T. Gönen, *Electric Power Distribution Engineering*. CRC Press, 3rd ed., 2014.

[4] H. W. Beaty and D. G. Fink, *Standard Handbook for Electrical Engineers*. McGraw-Hill, 16th ed., 2012.

[5] T. Sen, *Electrical and Production Load Factors*. PhD thesis, Texas A & M University, 2010.

[6] J. R. New, M. B. Adams, E. D. Garrison, W. E. Copeland, B. Smith, and A. Campbell, "Nailing the peak: City-scale, building-specific load factor and contribution to a utility's hour of critical generation," tech. rep., Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 2019.

[7] R. Li, C. Gu, F. Li, G. Shaddick, and M. Dale, "Development of low voltage network templates—part ii: Peak load estimation by clusterwise regression," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3045–3052, 2014.

[8] J. Nazarko, R. Broadwater, and N. Tawalbeh, "Identification of statistical properties of diversity and conversion factors from load research data," in *MELECON'98. 9th Mediterranean Electrotechnical Conference. Proceedings (Cat. No. 98CH36056)*, vol. 1, pp. 217–220, IEEE, 1998.

[9] V. Chatlani, D. Tylavsky, D. Montgomery, and M. Dyer, "Statistical properties of diversity factors for probabilistic loading of distribution transformers," in *2007 39th North American Power Symposium*, pp. 555–561, IEEE, 2007.

[10] I. Konstantelos, M. Sun, and G. Strbac, “Quantifying demand diversity of households,” 2014.

[11] S. Bielecki, “Estimation of maximum loads of residential electricity users,” in *E3S Web of Conferences*, vol. 137, p. 01006, EDP Sciences, 2019.

[12] J. Dickert and P. Schegner, “Residential load models for network planning purposes,” in *2010 Modern Electric Power Systems*, pp. 1–6, IEEE, 2010.

[13] F. Provoost, M. van Lumig, F. Provoost, and A. com Michiel, “The use of smart meters to improve customer load models,” in *Electricity Distribution 2011. CIRED 2011. 21st International Conference and Exhibition on*, pp. 1–4, 2011.

[14] J. Nazarko and W. Zalewski, “The fuzzy regression approach to peak load estimation in power distribution systems,” *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 809–814, 1999.

[15] C. Gibbons and A. Faruqui, “Quantile regression for peak demand forecasting,” *Available at SSRN 2485657*, 2014.

[16] L. De Haan and A. Ferreira, *Extreme value theory: an introduction*. Springer Science & Business Media, 2007.

[17] M. Jacob, C. Neves, and D. Vukadinović Greetham, *Forecasting and Assessing Risk of Individual Electricity Peaks*. Springer Nature, 2020.

[18] J. Boano-Danquah, C. Sigauke, and K. A. Kyei, “Analysis of extreme peak loads using point processes: An application using south african data,” *IEEE Access*, vol. 8, pp. 146105–146115, 2020.

[19] Y. Li and B. Jones, “The use of extreme value theory for forecasting long-term substation maximum electricity demand,” *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 128–139, 2019.

[20] S. El Adlouni, T. B. Ouarda, X. Zhang, R. Roy, and B. Bobée, “Generalized maximum likelihood estimators for the nonstationary generalized extreme value model,” *Water Resources Research*, vol. 43, no. 3, 2007.

[21] C. Dombry *et al.*, “Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework,” *Bernoulli*, vol. 21, no. 1, pp. 420–436, 2015.

[22] J. D. Lara, J. T. Lee, D. S. Callaway, and B.-M. Hodge, “Computational experiment design for operations model simulation,” *Electric Power Systems Research*, vol. 189, p. 106680, 2020.

[23] J. Lequesne and P. Regnault, “vsgoftest: An package for goodness-of-fit testing based on kullback-leibler divergence,” *arXiv preprint arXiv:1806.07244*, 2018.

[24] Pecan Street, Inc., *Pecan Street Dataport*, 2020.

[25] G. Kamath, “Bounds on the expectation of the maximum of samples from a gaussian,” *http://www.gautamkamath.com/writings/gaussian*$_m ax.pdf$, 2015.