Energy for Sustainable Development 83 (2024) 101572

Contents lists available at ScienceDirect

Energy for Sustainable Development

journal homepage: www.journals.elsevier.com/energy-for-sustainable-development

Load profiles of residential off-grid solar systems on the Navajo Nation

Henry Louie [a,*], Scott O'Shea [b], Stanley Atcitty [c], Derrick Terry [d], Darrick Lee [e], Peter Romine [e]

[a] Seattle University, United States
[b] Lunar Energy, United States
[c] Sandia National Laboratories, United States
[d] Navajo Tribal Utility Authority, United States
[e] Navajo Technical University, United States

ARTICLE INFO

Keywords:
K-means clustering
Load profile
Off-grid
Solar
Navajo Nation

ABSTRACT

Standalone off-grid electrical systems, no matter where they are deployed or for what user class, are designed based upon the load they are expected to serve. State-of-the-art computerized off-grid system design tools require the user to specify the expected load profile, that is, how the power consumption changes throughout the day. Often, this is at an hourly resolution, and some characterization of the distribution of power around the average values may be required. Specifying realistic and reasonable load profiles is a barrier to the appropriate design of standalone systems. This research extends previous studies on daily energy consumption of residential solar-powered off-grid systems on the Navajo Nation to provide hourly load profiles, statistical characteristics, and probabilistic models. The data analyzed come from 90 homes over a two-year period. K-means clustering is used to identify prototypical normalized load profiles when the data are grouped by year, season, weekday, and weekend. Eight parametric probability density functions are fit to the grouped data at an hourly resolution. Their fit to the data is evaluated using the Cramér-von Mises (CvM) statistic. The results show that the load profiles tend to be night-peaking and that Log Normal and Gumbel distributions can reasonably model variation in the data. The load profiles and probabilistic models can be used in off-grid design software and to synthesize load profiles for design and future research.

Introduction

The Navajo Nation is a sovereign nation located in the southwestern region of the United States. Its territory spans approximately 70,000 km^2. More than 160,000 people live within its land (U.S. Census Bureau, 2020). Navajo Nation is home to exceptional mineral, energy, and solar resources, yet over 10,000 homes within it are without grid-connected electricity (Gallucci, 2019). The burden of this form of energy poverty on human development and quality of life are numerous and well-documented, affecting income, health, and educational outcomes (Asghar et al., 2022; Franco et al., 2017; International Energy Agency, 2014; Kanagawa & Nakata, 2008; Sarkodie & Adams, 2020a, 2020b; Sovacool & Ryan, 2016). It is also a notable form of distributional energy injustice, where the un-electrified homes are unduly burdened with the impacts of an electricity system that they do not directly benefit from (Sovacool et al., 2016).

The Navajo Tribal Utility Authority (NTUA) has made progress in eroding the electrification deficit through various programs, including grid extension and by offering off-grid solar systems (Navajo Tribal Utility Authority, 2022). Grid extension generally is the least-cost option for homes that are in relatively dense clusters and are close to the existing grid. However, consistent with their pastoral tradition, many homes on the Navajo Nation are separated from each other by large distances and at a low density. Some homes are as far as 70 km from the grid. With grid extension costs typically greater than USD$40,000 per kilometer, off-grid systems offer a viable alternative for those homes far from the grid.

NTUA has been implementing off-grid systems on the Navajo Nation since the 1990's (Begay, 2018). More recently, partially supported by funds from the U.S. government CARES Act (Coronavirus Aid, Relief, and Economic Security Act), NTUA has installed over 450, 3.8 kW off-grid solar systems on the Navajo Nation (2020 CARES Act Final Report for Navajo Nation Leadership, 2021).

Data acquisition systems were integrated into the off-grid solar

* Corresponding author.
E-mail addresses: hlouie@ieee.org (H. Louie), osheas@seattleu.edu (S. O'Shea), satcitt@sandia.gov (S. Atcitty), derrickt@ntua.com (D. Terry), dlee@navajotech.edu (D. Lee), promine@navajotech.edu (P. Romine).

https://doi.org/10.1016/j.esd.2024.101572
Received 14 August 2024; Received in revised form 27 September 2024; Accepted 30 September 2024
Available online 2 November 2024

systems, which enabled the creation of a rich data set of electricity consumption of off-grid homes on the Navajo Nation. The first analysis of this data set was done in (Louie et al., 2023) and focused on the daily energy consumption characteristics of 127 of the homes. This analysis was prioritized because average daily consumption, along with the solar resource, are two key inputs to the design of an off-grid system, particularly if a standards-based methodology is applied (Louie, 2018). The analysis showed that the average AC-side consumption was 2.78 kWh per day, but with a wide range of variation between homes. Forty percent of the total energy consumed was attributed to just 20 % of the homes (Louie et al., 2023). The daily consumption tended to follow seasonal patterns, which peaked in the summer and exhibited a 10 % decline in year-over-year consumption (Louie et al., 2023).

In (Louie et al., 2023), the authors identified the data set as having important value at sub-daily timescales by using it to create prototypical hourly load profiles. These load profiles are needed in numerical design approaches, including those using software such as HOMER Pro (Bekele & Palm, 2010; Ma et al., 2014; UL, 2023; Xendee, 2024). Some software also requires parameters that characterize the stochastic nature of the load (UL, 2023). This research extends the analyses in (Louie et al., 2023) by identifying prototypical hourly load profiles through K-means clustering and evaluating eight parametric probabilistic models to characterize the stochasticity of the load.

The primary contributions of this research are the detailed presentation of the load profiles and parametric probabilistic models derived from data from 90 of NTUA's off-grid systems. These load profiles can be used in off-grid system design of homes in similar contextual and irradiation conditions, and by researchers in need of realistic, representative load profiles. Due to the unique circumstances of the Navajo Nation, the intent of this research is not to show how the load profiles are generalizable or applicable to other contexts, but rather, it supports the view that off-grid users on Tribal Lands have electricity-use characteristics that are different than documented elsewhere (Louie et al., 2023).

A load profile, in its simplest form, describes the typical power consumption of a facility or load center over the course of a 24-hour period. It may represent the consumption of a single user or an aggregation of users. Load profiles of mini-grid and other off-grid system users in Sub-Saharan Africa have been presented in literature (Williams et al., 2017; Yoder & Williams, 2020). But, as described in (Louie et al., 2023), there is evidence that consumption of off-grid users in the Sub-Saharan context is strikingly different than on the Navajo Nation.

Researchers have developed approaches to synthesizing realistic load profiles. One example is LoadProGen (Mandelli et al., 2016, 2017). While these approaches are useful, they require some knowledge of the users, for example the appliance ownership or geographical location. An alternative method to constructing load profiles is to employ a data-driven approach, relying on historical load data from similar users. In (Blodgett et al., 2017), a data-driven approach was shown to reduce average daily consumption errors by 70 % over other approaches. The research presented in this paper supports a data-driven approach for load profile creation. The contributions of this paper are: 1) it considers a novel data set – individual domestic users on a Tribal Land; 2) it identifies prototypical annual, seasonal, weekday, and weekend load profiles, and 3) it evaluates parametric probabilistic models of hourly load.

This paper is arranged as follows. The Data set description section provides background information on the data considered in this work. The Methodology section describes the methodology for developing the load profiles. Representative load profiles are provided in the following section. Statistical analysis of the load profiles is presented in the Statistical analysis section. Parametric probability distributions are fit to the data and evaluated in the Probabilistic models section. Discussion of the load profiles and additional insight are provided in the Discussion section. The Conclusions and future work Section summarizes the work and describes the next steps for this research.

Data set description

This paper considers off-grid residential solar systems installed on the Navajo Nation by NTUA in late 2020. An image of the type of off-grid system considered in this work is shown in Fig. 1. The system consists of a 3.8 kW bi-facial solar array arranged into two sub-arrays each with a charge controller; 16 gel lead acid batteries each rated at 12 V, 183 Ah (20-hour rate), and arranged into a 48 V battery bank with nominal capacity of 35.1 kWh; and a 8 kW inverter. The off-grid systems are each outfitted with a cellular-connected data acquisition system by Samsara (Samsara, Inc, 2020). This paper considers 90 of the homes with off-grid systems. These homes were selected because their data records span two years and they generally had high-quality data.

Although many system quantities are measured, this work is primarily concerned with measurements of real power, expressed in watts, and measured at the AC-side of the inverter. The data considered spanned from 1 January 2021 to 31 December 2022. The data for each day began at 0:00 UTC, which was then adjusted to Mountain Standard Time (−6). No adjustments for daylight saving time were made.

The data were automatically sampled, but the sampling was done at irregular intervals, typically ranging from every 5 to 30 s. The data are first arranged into a minutely time-series by averaging all the samples recorded within the same minute according to their timestamp. Missing values were excluded from the averaging. Any minute without data was marked as such. A ten-minute data set was then created from the minutely data by averaging. Again, missing values were excluded from the averaging, and any ten-minute interval without data was marked as such. Linear interpolation was applied to the remaining data to fill-in any missing data in the 10-minute data set. If the first or last data point for a day is missing, then linear interpolation was not applied. Instead, the day was excluded from the data set.

Any day that was missing more than 18 ten-minute (three hours total) intervals of data or was excluded from the data set. Any day that was flagged as having an outage was also excluded. Outages were usually caused by the inverter disconnecting the load—and data acquisition system—when the battery voltage fell below a temperature-dependent threshold. An outage was assumed to have occurred if the minutely battery voltage fell below 47.1 V and was immediately followed by an interruption of data lasting for at least 15 min. Other times, data were missing due to communication hardware malfunction.

Fig. 1. Image of the type of off-grid solar system considered in this work. The 8 kW inverter, charge controllers, 48 V battery bank and data acquisition system are in the enclosure in the foreground. The bi-facial 3.8 kW PV array tilted at 35 degrees in the background (Louie et al., 2023).

Methodology

This research considers normalized hourly load profiles of the data. This section describes the methodology for creating these load profiles and for associating them into clusters based upon their shape. Several overlapping subsets of the data are considered. These are referred to as "All Years" (data from both 2021 and 2022), "2022", "2021", "Winter" (1 Jan.–31 Mar. of 2021 and 2022), "Spring"(1 Apr.–30 Jun. of 2021 and 2022), "Summer" (1 Jul.–30 Sep. of 2021 and 2022), "Fall"(1 Oct.–31 Dec. of 2021 and 2022), "Weekday" and "Weekend" (weekday and weekend days from 2021 and 2022). These subsets of data are referred to as "Groups". The number of days of data considered for each Group after preprocessing is shown in the rightmost column of Table 1. The process for creating a normalized load profile for each system in a Group is displayed in Fig. 2.

The days are indexed numerically so that d is 1 for 1 January 2021, and d is 730 for 31 December 2022. The day indices that belong to the Group under consideration is denoted as $\mathcal{G}$. The index of the last day in the Group is d_{last}. Starting with the first system k, each day is considered one at a time. If the day d is in $\mathcal{G}$, then the ten-minute data for that day are normalized using the ℓ-1 norm. In this way, the sum of the 10-minute normalized load profile for day d and system k is exactly equal to 1.

This process repeats for each day in $\mathcal{G}$ through d_{last}. For example, in the All Years Group, up to 730 10-minute normalized load profiles are created for system k. These load profiles are then averaged to create a single 144-element 10-minute average normalized load profile. This load profile is then downsampled by summing the six 10-minute values that correspond to each hour for all hours, resulting in a single normalized hourly load profile. The sum of the 24-elements in this load profile is exactly equal to 1. The elements of the normalized hourly load profile can be interpreted as the proportion of that day's total energy consumption that was consumed during a given hour, on average.

The process repeats for the next system until all 90 systems have been considered for the Group, nominally yielding 90 separate normalized hourly load profiles (note that a particular system may not have any valid data in a certain Group, and so fewer than 90 load profiles may be generated for that Group). Then, the next Group is considered. Hereafter for the sake of concision, normalized hourly load profiles are referred to simply as "load profiles".

The load profiles within each Group exhibit a variety of patterns. Simple averaging of all the load profiles in a Group masks this variation. To better represent the typical load profile patterns within each Group, a clustering algorithm was used. Specifically, SciKit-Learn's implementation of the Lloyd's K-means algorithm was used with 30 different initializations per Group (Lloyd, 1982; Pedregosa et al., 2011).

The K-means algorithm partitions the load profiles into M clusters for each Group. Each load profile "belongs" to exactly one cluster. Each cluster has one "centroid". Each centroid can be thought of as a 24-element vector whose elements correspond to an hour of the day and whose values are equal to the average load profile of their cluster. The centroids are hereafter referred to as "prototype" or "prototypical" load profiles as they are representative of the load profiles belonging to their cluster for a given Group.

Table 1
Percent and number of days of data per group.

Group	Cluster					Total days
	1	2	3	4	5	
	(%)	(%)	(%)	(%)	(%)	
All years	40.0	14.7	2.5	21.1	21.6	33,249
2022	26.7	40.9	5.5	11.2	15.7	19,391
2021	35.3	19.4	8.4	13.6	23.2	13,583
Winter	20.2	34.5	15.8	9.0	20.4	6445
Spring	39.4	4.4	22.8	12.1	21.2	10,013
Summer	16.6	27.4	35.7	7.5	12.8	9569
Fall	18.0	18.4	13.4	34.7	15.4	6654
Weekday	3.2	32.4	36.8	23.6	4.2	14,622
Weekend	40.1	16.1	17.4	24.8	1.5	9543

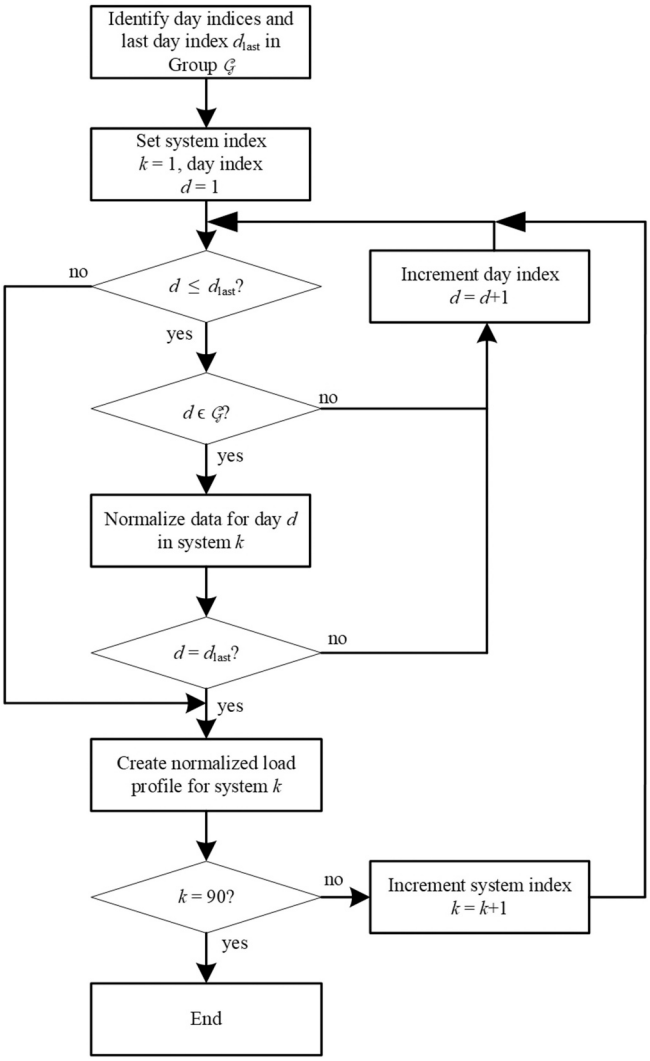

Fig. 2. Block diagram of the normalized load profile creation process for a given Group.

The number of clusters M produced by the algorithm for each Group is specified by the user. Determining M is not always straightforward with unsupervised machine learning techniques such as K-means. The approach used in this work is to consider the silhouette score (Rousseeuw, 1987). Silhouette scores range from −1 to 1 with larger positive numbers typically indicating better clustering. Silhouette scores were calculated for each group for a number of clusters ranging from 3 to 15. Although the number of clusters corresponding to the greatest silhouette score was not the same for each Group, it was most often achieved at or near five clusters. For the sake of clarity and consistency, five clusters were used for each Group. The percentage of the total days of data corresponding to each cluster and Group are summarized in Table 1.

Prototypical load profiles

This section presents the prototype load profiles for each Group of data based on the methodology previously described. Recall that a prototype load profile for a cluster is the average of the load profiles belonging to that cluster. The prototypical load profiles tell the story of a typical day's consumption. In the design of off-grid systems powered by

a solar array, the energy consumed before sunrise and after sunset is especially important, as this influences the size of the battery bank. Generally, it is advantageous for the evening and overnight consumption to be low. Also of interest is the peak of the load profile, as this influences the size of the inverter.

All Years

Examined first are the load profiles consisting of the two year range of data. The prototype load profiles for each of the five clusters are shown in Fig. 3. Also shown are the actual load profiles that were associated with each cluster. Here, *N* refers to the number of systems whose load profiles were associated with each cluster. Generally, the actual load profiles match the prototype load profiles but, as can be expected with K-means clustering, none are an exact fit. With some exceptions, the hourly consumption generally ranged from three to 6 % of the daily energy.

The 33 homes in Cluster 1 had relatively consistent consumption throughout the day with a peak in the evening around 20:00. Evening peaks were also evident in Clusters 2 and 5. Clusters 3 exhibited a day-peaking characteristic. However, only four homes had load profiles belonging to this cluster. Cluster 4 exhibited a consistent daytime load, slightly increasing until 20:00. With the exception of Cluster 2, the prototype load profiles exhibit a rise beginning at approximately 6:00, when presumably the occupants wake up and begin using appliances.

Yearly

Comparing the years 2021 and 2022 show how, if at all, the load profiles evolved over time. Note that as described in (Louie et al., 2023), the consumption in 2022 was approximately 10 % less than in 2021. The prototypical load profiles for 2021 and 2022 are shown in Figs. 4 and 5,

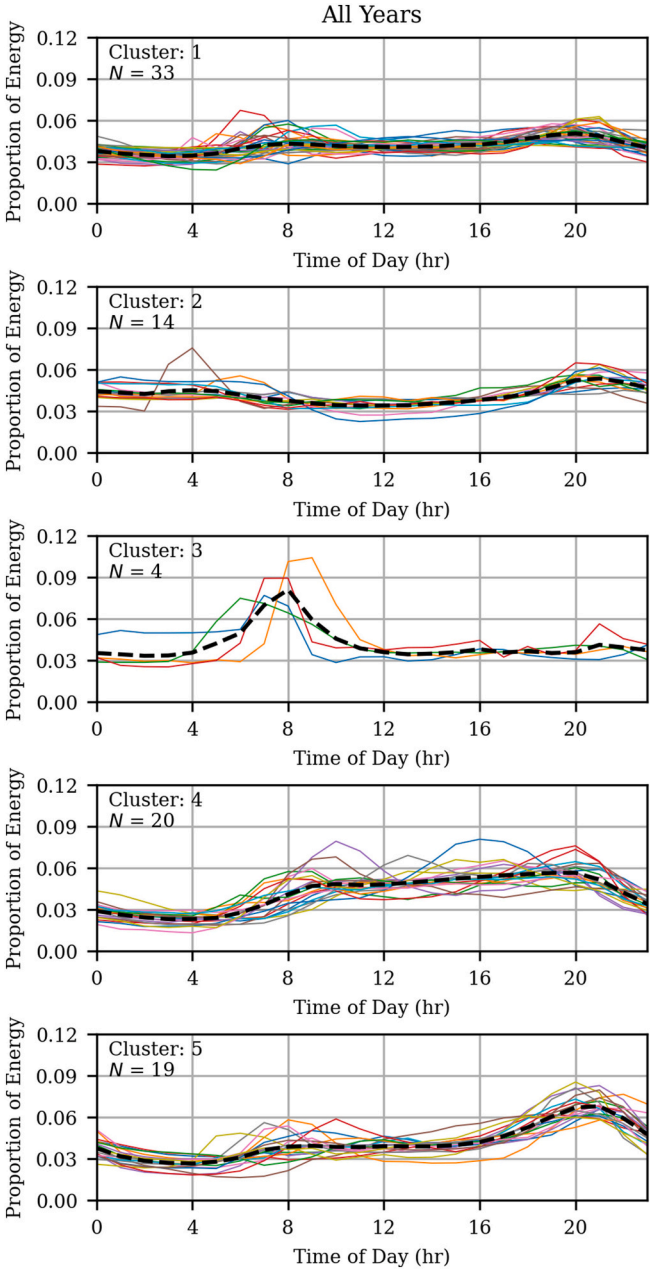

Fig. 3. Load profiles (thin lines) and prototype load profile (dashed lines) of each cluster for the All Years Group.

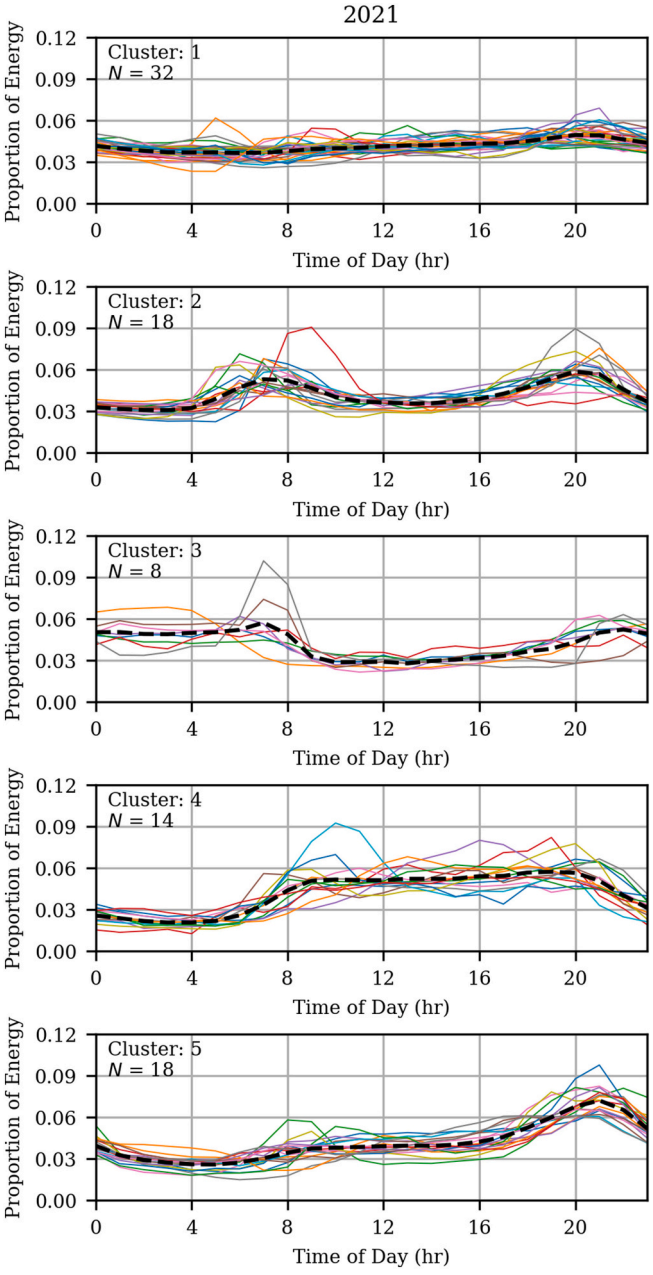

Fig. 4. Load profiles (thin lines) and prototype load profile (dashed lines) of each cluster for the 2021 Group.

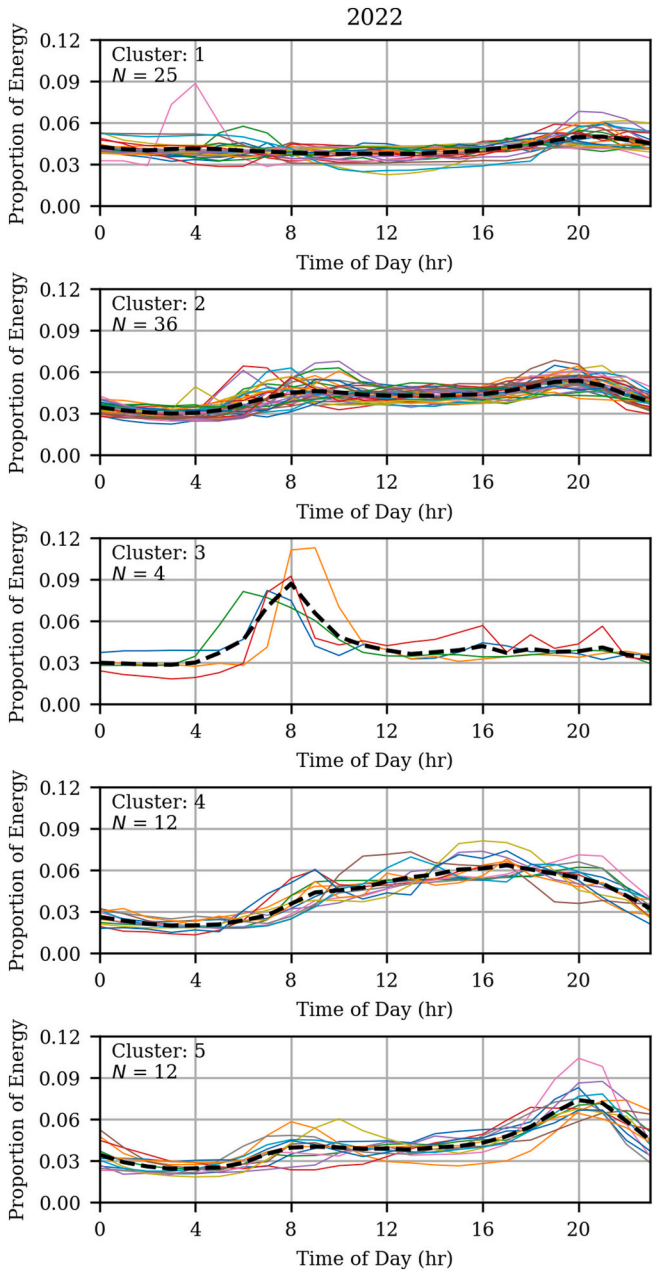

Fig. 5. Load profiles (thin lines) and prototype load profile (dashed lines) of each cluster for the 2022 Group.

respectively. Keep in mind that the homes assigned to a certain cluster in 2021 do not necessarily correspond to the same numbered cluster in 2022. For both years, only a small number of homes exhibited morning- or day-peaking behavior (Cluster 1 and 2 in both figures). The majority of the homes saw peaks occurring in the evening, around 20:00. A notable difference is Cluster 3 in 2021 showed a pronounced dip in consumption from approximately 9:00 to 15:00 which is not represented by any of the clusters in 2022. Overall, the general shapes of the load profiles do not appear to be substantially evolving from one year to the next.

However, the load profiles of many of the individual homes did change from 2021 to 2022. This is shown in Fig. 6. The circles on the left represent the five clusters of the 2021 Group; the five circles on the right represent the 2022 Group. The area of the circles are proportional to the number of homes associated with each cluster. The circles and wedges are color-coded based on the 2021 clustering. For example, the 32

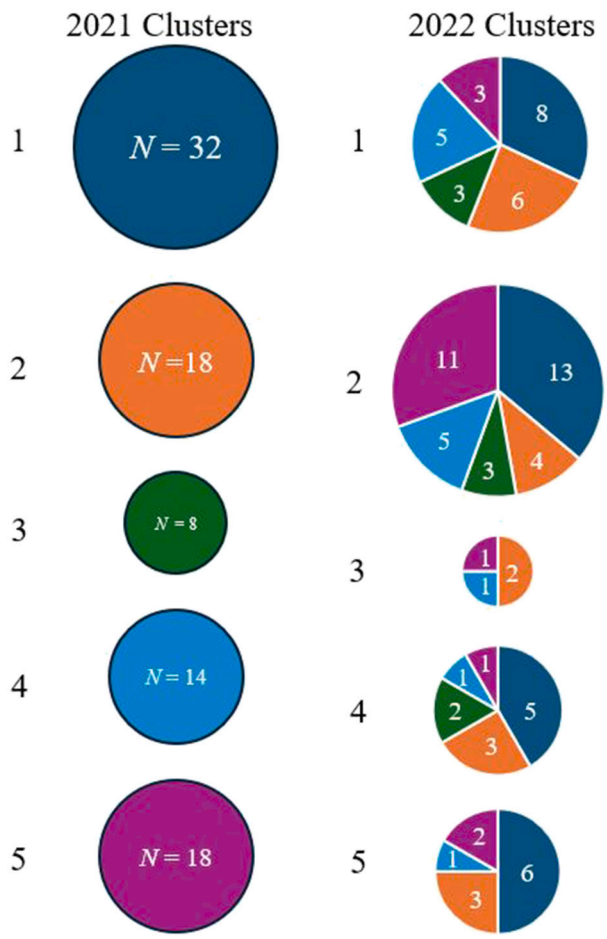

Fig. 6. Cluster membership composition from 2021 clustering to 2022 clustering.

homes that were clustered with Cluster 1 in 2021—represented by the dark blue color—scattered into different clusters in 2022, with 8 belonging to Cluster 1, 13 belonging to Cluster 2, none belonging to Cluster 3, and so on. While one might expect the most of homes associated together in the same cluster in 2021 to also be associated together in 2022—even if the cluster number itself was different—this was generally not the case. Rather, the homes in a given cluster in 2021 tended to disperse in the 2022 clustering. The only 2021 cluster where more than half of the homes stayed grouped together was Cluster 5, in which 11 of the 14 homes moved to Cluster 2 in the 2022 Group. Inspecting the load profiles for Cluster 5 in the 2021 Group and Cluster 2 in the 2022 Group, shows that in general these homes reduced their evening peak somewhat.

Quarterly load profiles

Considered next are quarterly (seasonal) load profiles. In (Louie et al., 2023), a seasonal component to average daily energy consumption as noted, with increased consumption in the summer months. The data for both years are combined for each quarter. The resulting prototype load profiles are shown in Fig. 7. Note that a small number of systems did not have valid data remaining after the cleaning process for given Group. Therefore, the total number of load profiles in each Group does not always equal 90. The load profiles during the summer months (July through September) generally exhibit more consumption during the daytime, whereas the fall (October through December) and winter (January through March) months often have more pronounced evening peaks.

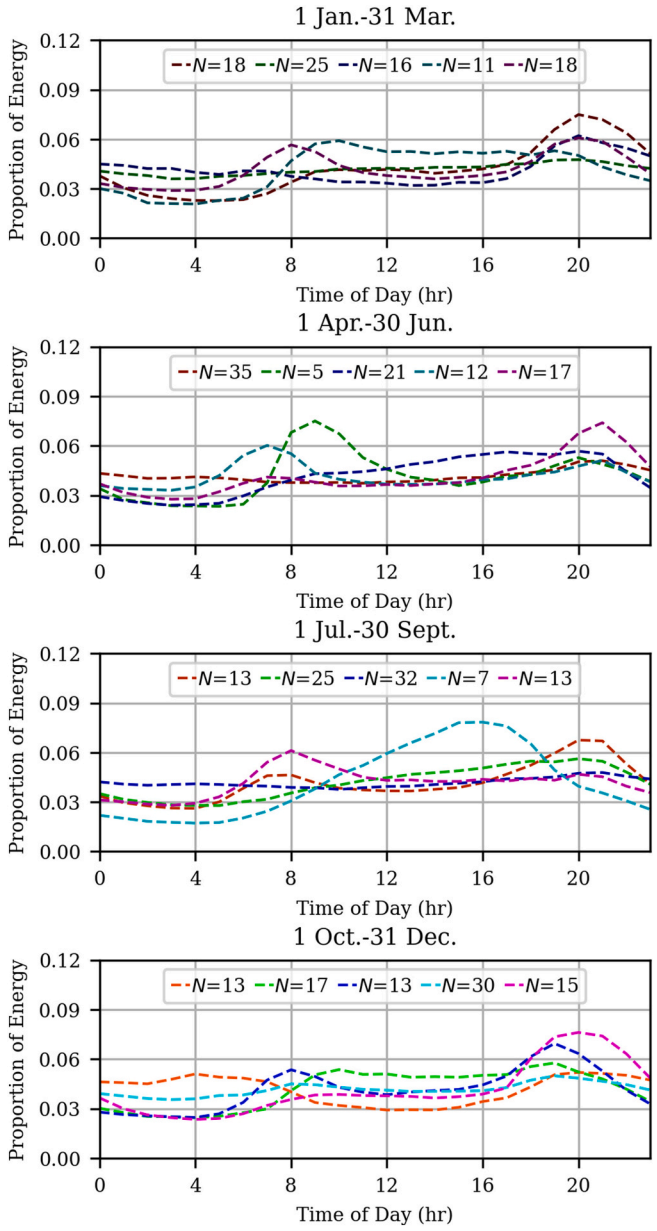

Fig. 7. Prototype load profiles for the Quarterly (seasonal) Group.

Weekday and Weekend

It is common for the load of most users to be different during the weekdays and weekends. Here, the weekend is considered to start at 18:00 Mountain Standard Time on Friday and end on 17:50 on Sunday. In this way, evening behavior on days preceding a typical day off of work are captured. Fig. 8 shows the clustered load profiles for weekdays and weekends, respectively. Data for 2021 and 2022 are combined. The profiles do not exhibit substantial differences. The predominance of night-peaking load remains. Although, the daytime consumption is somewhat greater on the weekend days. This suggests that homes are similarly occupied during weekends and weekdays and appliances—which can be consistent with anecdotal observations of irregular, inconsistent, or self-employment.

Representative load profiles

Although the prototype profiles provide information on typical patterns of average consumption, the statistical details of the variation

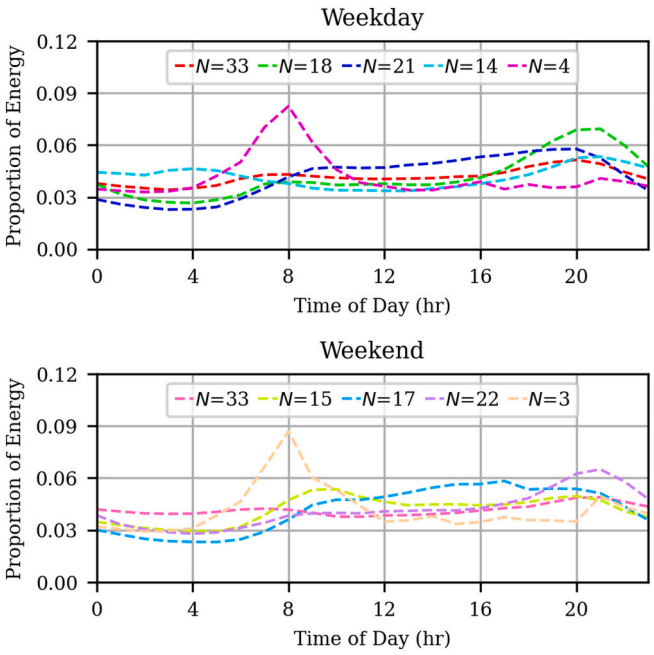

Fig. 8. Prototype load profile (dashed lines) of each cluster for the Weekday and Weekend Groups.

within each hour is often also of interest. To illustrate the variation, actual load for five specific homes – each corresponding to a different Cluster for the All Years group— are provided in Fig. 9. The specific homes chosen are those that had the lowest average mean squared error between their load profile and the corresponding prototype load profile.

The load are expressed as box plots for each hourly interval. The whiskers extend to 1.5 times the inter-quartile range. Outliers are displayed as small "+" symbols. Shown in this way, the variation that occurs in each interval is apparent. In all cases, consumption two to three times the median, or more, was observed. Of particular note is the variation for Cluster 3, where on some days over 20 % of the day's total energy was consumed in just 1 h. Moreover, the distribution is generally asymmetric, with outliers that are greater than the median appearing more frequently and severe than those less than the median. This suggests that a Normal (Gaussian) model of consumption may be insufficient to characterize the observed distribution of load.

Statistical analysis

The load profiles presented in the previous sections show the average load for each hour for each data set. As was shown in Fig. 9, there can be substantial variation in consumption during the same hour on different days, even in the same home. This variation is described by the statistical moments of standard deviation σ, skewness γ, and kurtosis β. The moments are provided in Table 2 for the All Years group for Cluster 1. Also shown are the 0.10 and 0.90 quantiles, denoted as Q(10) and Q(90), respectively. These quantiles are useful in understanding the extremes of consumption that occurred each hour. The statistics for the remaining clusters are found in the Appendix.

Although the statistics are different for each cluster, hour, and group, some general trends emerge. Of particular note is that all the hours exhibit a positive skewness, indicating that the distributions are asymmetric with a tendency for more extreme positive values than negative. This is consistent with the observations from Fig. 9. The kertosis for almost all hours indicate that the distributions are leptokurtic. A leptokurtic distribution is one that generally has "flatter" tails, producing more frequent outliers than a Normal distribution.

The statistical moments are useful in suggesting which parametric

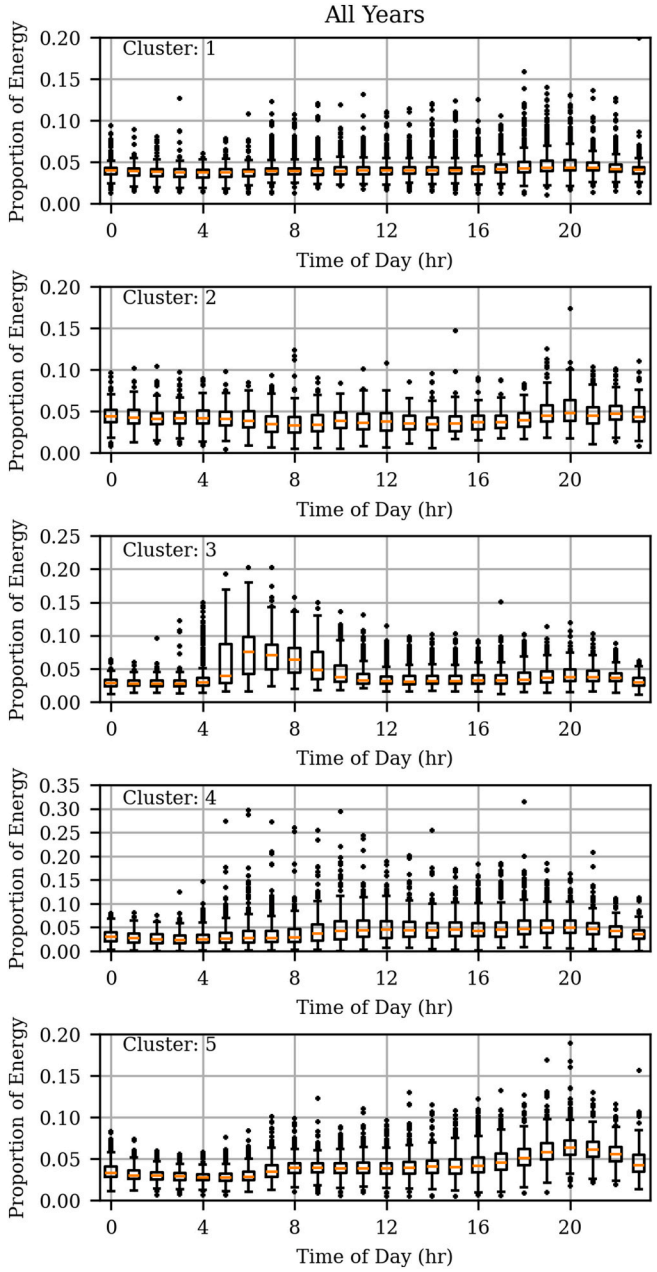

Fig. 9. Box plots of hourly load for homes whose load profiles that most closely match prototype load profiles for the All Years Group.

distributions are reasonable candidates to fit to the data, as discussed in the following sections.

Probabilistic models

This section evaluates parametric distributions as models of the load for each Group, cluster, and hour. Eight common distributions are considered: Beta, Gamma, Gumbel, Log Normal, Nakagami, Normal, Rayleigh and Weibull. These were selected based on inspection of the statistical moments and histograms of the data.

Cramér-von Mises statistic

The parameters for each distribution were determined using the maximum likelihood estimation procedure. This step was implemented using the SciPy library for Python (Virtanen et al., 2020). The fit of each distribution was evaluated using the Cramér-von Mises (CvM) statistic. The CvM statistic was selected because unlike the χ statistic, the CvM it is strictly objective. Briefly, the CvM statistic compares the CDF of the considered distribution to the empirical cumulative distribution function (eCDF) of the considered load data for a given hour and cluster.

The CvM criterion is:

$$w^2 = \int_{-\infty}^{\infty} (F(x) - \widehat{F}(x))^2 dF(x) \tag{1}$$

where $F(x)$ is the CDF of the considered distribution and $\widehat{F}(x)$ is the eCDF of the data for a specific hour of a load profile cluster (Berg, 2009; Genest et al., 2009). The CvM test statistic (T) is practically implemented as:

$$T = nw^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(\frac{2i-1}{2n} - F(x_i) \right)^2 \tag{2}$$

where x_i are the data sorted in ascending order, and n is the number of data for a considered Group, cluster, and hour (D'Agostino & Stephens, 1986). With this formulation, a smaller value of T indicates a closer fit of the distribution to the data. The CvM statistic was computed for each Group, cluster, hour and for each of the distributions considered.

It is helpful to have a visual sense of the distributions' fit to the data. Examples of fit distributions for two different hours for Cluster 1 of the All Years group are provided in Figs. 10 and 11.

The figures show the histogram of the data along with the distributions with the smallest (best) and largest (worst) CvM test statistics T. Fig. 10 is an example where the best fitting distribution (Log Normal) resulted in a low CvM statistic. The dashed line reasonably approximates the histogram of the actual data. For comparison, also shown is the worst fitting distribution for this data—the Rayleigh distribution. Fig. 11, on the other hand, was the hour with the overall worst fitting distribution for this cluster. The best fitting distribution was a Gumbel with a CvM statistic of 1.63. While certainly not as close of a fit as the Log Normal distribution in Fig. 10 is, it is at least reasonable. The worst fitting distribution was the Normal.

The performance of the distributions for the All Years Group are summarized in Table 3. The table shows the average CvM statistic for each distribution when averaged across all five clusters and all hours. The Log Normal distribution has the lowest average CvM statistic, and thus can be concluded as the best fitting by that measure. The Gumbel,

Table 2
Statistical moments for Cluster 1: All Years Group.

Hour	Mean	Std.	Skew.	Kurt.	Q(10)	Q(90)
0	0.038	0.014	2.132	19.403	0.023	0.053
1	0.036	0.013	1.326	8.429	0.022	0.050
2	0.035	0.013	1.613	12.235	0.020	0.049
3	0.034	0.013	1.240	5.706	0.020	0.047
4	0.035	0.015	3.302	39.200	0.019	0.049
5	0.037	0.019	4.536	48.673	0.019	0.053
6	0.040	0.024	4.628	47.043	0.020	0.064
7	0.042	0.024	3.183	20.162	0.021	0.070
8	0.043	0.025	2.978	17.365	0.022	0.070
9	0.043	0.023	2.918	16.064	0.023	0.069
10	0.042	0.022	3.642	32.676	0.023	0.064
11	0.041	0.021	3.081	20.651	0.022	0.062
12	0.041	0.020	3.087	19.785	0.022	0.060
13	0.041	0.020	3.202	21.366	0.022	0.059
14	0.041	0.020	2.976	18.937	0.022	0.059
15	0.042	0.021	3.370	27.898	0.022	0.061
16	0.042	0.020	2.948	21.049	0.022	0.062
17	0.044	0.023	3.818	35.663	0.023	0.065
18	0.047	0.023	3.385	32.928	0.025	0.070
19	0.049	0.022	2.749	21.924	0.028	0.074
20	0.051	0.021	2.107	12.668	0.030	0.075
21	0.049	0.020	2.103	12.587	0.030	0.073
22	0.045	0.017	1.843	10.006	0.027	0.065
23	0.041	0.016	2.511	20.991	0.025	0.058

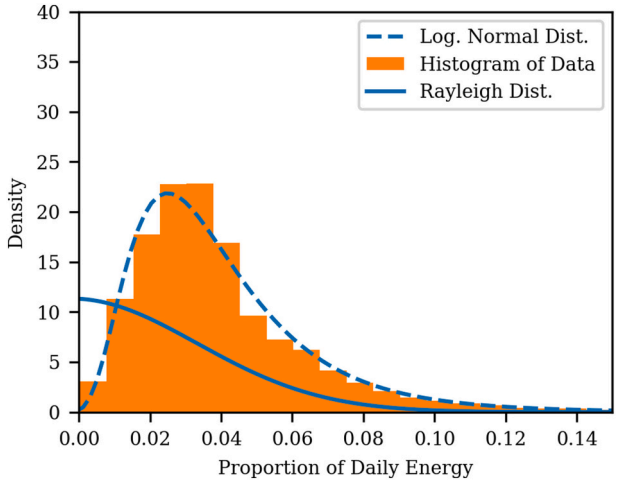

Fig. 10. Best fitting ($T = 0.17$) and worst fitting ($T = 1.47$) distributions to data for cluster 3, hour 7 in the All Years group.

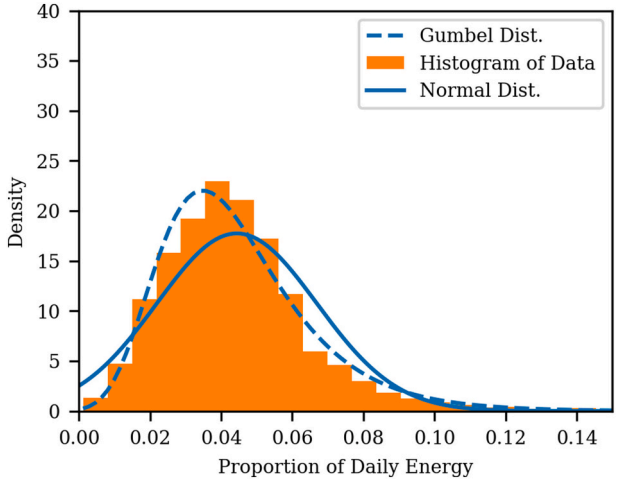

Fig. 11. Best fitting ($T = 1.63$) and worst fitting ($T = 7.9$) distributions to data for cluster 3, hour 15 in the All Years group.

Table 3
Fit summary All Years Group.

Distribution	Avg. CvM Statistic	Best fit
Log Normal	2.8	42.3 % (457)
Gumbel	3.1	46.0 %(497)
Beta	3.7	5.4 %(58)
Gamma	3.9	2.8 % (30)
Nakagami	8.0	0.6 % (7)
Weibull	8.6	1.1 % (12)
Normal	13.8	1.1 % (12)
Rayleigh	13.8	0.6 % (7)

Beta, and Gamma distributions also achieved similarly low CvM averages. Note that the Normal distribution was the second worst performing, only behind the Rayleigh distribution.

Another meaningful way of comparing the distributions is by the number of hours for each cluster and year that a distribution had the lowest CvM score among the other distributions. This is summarized in rightmost column in Table 3. Note that because there are five clusters with hourly granularity and there are nine Groups, there are $5 \times 24 \times 9 = 1080$ opportunities for a distribution to have the lowest CvM. By this measure, the Gumbel is the best-performing. It was the best-fit distribution in 46 % of the hours, with the Log Normal distribution being the best fit in 42.3 % of the hours. The Normal distribution was the best-fitting in just 1.1 % of the hours. Overall, the analysis shows that if a single distribution is to be used to model the variation of load in a given hour, the Log Normal or Gumbel distributions should be chosen.

Cramér-von Mises test

The CvM statistics are now used in the non-parametric Cramér-von Mises hypothesis test (Dytham, 2011; Mooney, 1997). Let $F^*(y, \boldsymbol{z})$ be the best-fitting distribution with parameters $\boldsymbol{z}$ for a given Group, cluster, and hour. The normalized load data for the given Group, cluster, and hour are $\boldsymbol{x} = [x_1, \ldots, x_n]$. Let the CvM statistic of $F^*(y, \boldsymbol{z})$ when evaluated against $\boldsymbol{x}$ using (2) be $T^*(\boldsymbol{x})$. The null hypothesis H_0 is that the data $\boldsymbol{x}$ are from $F^*(y, \boldsymbol{z})$. From a statistical viewpoint, in a well-fitting distribution the null hypothesis is accepted.

The first step is to map $T^*(\boldsymbol{x})$ to its p-value. This is accomplished using a Monte Carlo approach. A Monte Carlo approach is necessitated because the distribution of the CvM statistics—known as the "sampling distribution"–depends on $F^*(y, \boldsymbol{z})$ and the number of data points n. Determining closed form representations of the sampling distribution is not always tractable, and so the Monte Carlo approach is warranted (Mohr, 1990; Mooney, 1997).

Random samples $\widetilde{\boldsymbol{x}}_{\boldsymbol{1}} = [\widetilde{x}_{1,1}, \cdots, \widetilde{x}_{1,n}]$ are drawn from $F^*(y, \boldsymbol{z})$. The CvM statistic, $T_1(\widetilde{\boldsymbol{x}}_{\boldsymbol{1}})$ is computed as in (2) using $\widetilde{\boldsymbol{x}}_{\boldsymbol{1}}$ instead of $\boldsymbol{x}$. The process repeats to generate n CvM statistics $\boldsymbol{T}(\widetilde{\boldsymbol{x}}) = [T_1(\widetilde{\boldsymbol{x}}_{\boldsymbol{1}}), \ldots, T_n(\widetilde{\boldsymbol{x}}_{\boldsymbol{n}})]$.

The sampling distribution is approximated by $\boldsymbol{T}(\widetilde{\boldsymbol{x}})$, and maps $T^*(\boldsymbol{x})$ to a p-value (Mohr, 1990; Mooney, 1997). The p-value is the proportion of $\boldsymbol{T}(\widetilde{\boldsymbol{x}})$ that is greater than or equal to $T^*(\boldsymbol{x})$ (recall that larger values of T indicate a worse fit). In other words, the p-value can be thought of as estimate of the probability that $F^*(y, \boldsymbol{z})$ will fit the distribution of the normalized load data $\boldsymbol{x}$ better than the distribution of random samples $\widetilde{\boldsymbol{x}}_{\boldsymbol{j}}$. In this formulation, a larger p-value indicates stronger evidence that the distribution fits the data well.

As an example, Cluster 2 for the Summer Group had $n = 2624$ hourly data points for hour 4:00. The best-fitting distribution $F^*(y, \boldsymbol{z})$ is Log Normal with location, scale, and shape parameters of $\boldsymbol{z} = -0.014$, 0.041, 0.290 and corresponding CvM statistic $T^*(\boldsymbol{x}) = 0.082$. To determine the p-value, 2624 random samples $\widetilde{\boldsymbol{x}}_{\boldsymbol{1}}$ were drawn from the same Log Normal distribution $F^*(y, \boldsymbol{z})$ and the corresponding CvM statistic $T_1(\widetilde{\boldsymbol{x}}_{\boldsymbol{1}})$ was computed. The Monte Carlo process is repeated to generate 2624 CvM statistics (the sampling distribution), as shown in Fig. 12. A total of 1738 out of the 2624 elements in $\boldsymbol{T}(\widetilde{\boldsymbol{x}})$ are greater than or equal

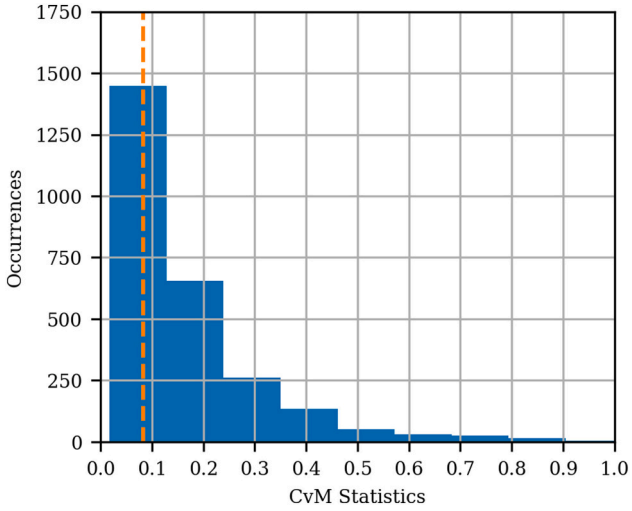

Fig. 12. Histogram of CvM statistics for Cluster 2 of the Summer Group generated by Monte Carlo simulation. The vertical dashed line marks the CvM statistic of the fit Log Normal distribution.

to $T^*(\boldsymbol{x})$, corresponding to a p-value of 0.66. With this large p-value, there is insufficient reason to reject the null hypothesis, and it can be reasonably concluded that the load for hour 4:00 is produced by a stochastic process following a Log Normal distribution with parameters −0.014, 0.041, 0.290. A smaller p-value such as 0.01 would indicate that it is unlikely that the data actually are from $F^*(y, \boldsymbol{z})$.

As may be expected, the best-fitting distributions exhibited a wide range of p-values, which are summarized in the cumulative distribution function (CDF) plot in Fig. 13. Usually, a level of significance α is subjectively selected as a threshold for rejecting or accepting the null hypothesis. In this research the p-value itself is more useful for interpreting the fit of the distribution, rather than if the null hypothesis is accepted or rejected. After all, the goal is to identify the best-fitting distribution, not evaluate hypotheses of the fit. Still, for completeness and to more conveniently summarize the results, a value of $\alpha = 0.20$ is selected. Approximately 25 % of the 1080 best-fit distributions resulted in p-values greater than or equal to 0.20, indicating that these distributions fit the data well from a statistical viewpoint. The number of hours where this occurred varied unevenly depending the Group and cluster, as summarized in Table 5. Distributions with p-values greater than or equal to 0.20 are marked in bold in the tables in the Appendix.

Synthesis of load

The load profiles and parametric models have an important application in synthesizing a time-series of load. The general process for doing so is described in the following. To begin, the average daily load must be specified by the user. One approach to doing so is to consult the average daily load for off-grid homes on the Navajo Nation described in (Louie et al., 2023). Other approaches of estimating the average load are described in (Louie, 2018). For each hour of the day, consult the load profile of choice and multiply the average daily load by the load profile mean. Using the average daily load of 2.78 kWh/day in (Louie et al., 2023), and the All Years Group load profile cluster 0 is used. Consulting Table 4, for hour 0:00, is seen that the Log Normal distribution best fits the data. The location, scale, and shape parameters for this hour are −0.023, 0.598, and 0.216, respectively. The load value can be sampled from this distribution, and then multiplied by 2.78 kWh to model the hourly consumption. This approach is repeated using the corresponding distribution and parameters for the remaining hours.

This approach does not capture the inter-hour relationships of the load, but this form of timeseries modeling is not often done in off-grid design programs. Timeseries models can be developed as a natural extension of the research presented in this paper.

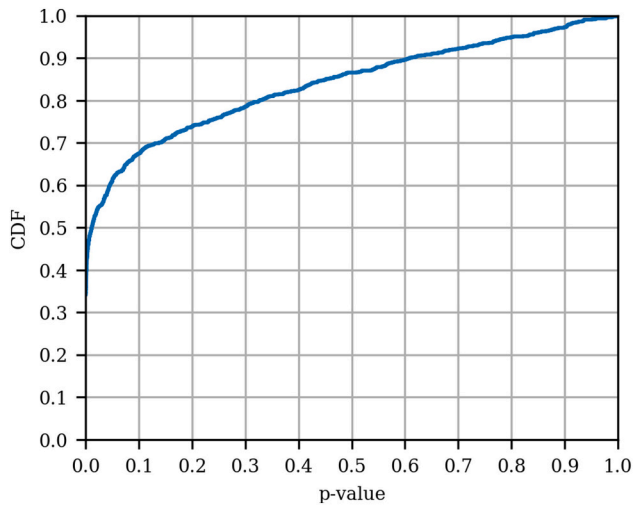

Fig. 13. Cumulative distribution function of p-values for the best-fitting distributions for all hours, clusters, and Groups.

Table 4
Best fit distribution and parameters for All Years Group Cluster 1.

Time	Distribution	Parameter
0	Log Normal	−0.023, 0.0598, 0.216
1	Log Normal	−0.032, 0.0675, 0.181
2	Log Normal	−0.028, 0.0622, 0.195
3	Log Normal	−0.031, 0.0638, 0.189
4	Log Normal	−0.015, 0.0476, 0.279
5	Gumbel	0.029, 0.0124
6	Gumbel	0.031, 0.0143
7	Gumbel	0.033, 0.0152
8	Gumbel	0.034, 0.0155
9	Gumbel	0.034, 0.0150
10	Gumbel	0.033, 0.0144
11	Gumbel	0.033, 0.0140
12	Gumbel	0.033, 0.0139
13	Gumbel	0.033, 0.0141
14	Gumbel	0.033, 0.0142
15	Gumbel	0.034, 0.0145
16	Gumbel	0.034, 0.0146
17	Gumbel	0.035, 0.0154
18	Gumbel	0.038, 0.0160
19	Log Normal	−0.017, 0.0638, 0.294
20	Gumbel	0.042, 0.0161
21	Log Normal	−0.019, 0.0655, 0.268
22	Log Normal	−0.019, 0.0618, 0.249
23	Log Normal	−0.030, 0.0697, 0.202

Table 5
Number of hours with p-values greater than or equal to $\alpha = 0.20$.

Group	Cluster				
	1	2	3	4	5
All Years	0	0	6	4	5
2022	0	0	5	1	0
2021	4	18	0	3	14
Winter	6	0	0	5	5
Spring	0	23	11	8	16
Summer	18	9	0	20	7
Fall	0	10	22	0	13
Weekday	0	10	22	0	13
Weekend	0	2	9	5	15

Discussion

It is interesting to compare the load profiles for these off-grid systems to those reported elsewhere. The load profiles on the Navajo Nation generally exhibit night-peaking characteristics. This is consistent with residential off-grid systems reported in mini-grids in Sub-Saharan Africa (Williams et al., 2017). However, the load profiles on the Navajo Nation generally have higher load during the day than in Sub-Saharan Africa, where zero or nearly zero daytime consumption is common. This could be attributed to several possible and plausible reasons. Anecdotally, off-grid homes on the Navajo Nation seem to have more appliances than in Sub-Saharan Africa. The "stand-by" load of these appliances would contribute to greater daytime consumption. Also, most mini-grids in Sub-Saharan Africa have energy-based tariffs, which incentivizes homes to limit consumption. Unplugging appliances, which eliminates their stand-by consumption, is a common practice. The off-grid homes considered in this research paid a flat monthly fee, which is partially due to the added technical complexity of implementing metering and payment systems.

Although there was some variation in load profiles between the Groups, in general, the same characteristics were found in most clusters—an increase in load from 6:00 to 8:00, steady daytime consumption, and an evening peak around 18:00. This suggests that the statistics and probabilistic models for the All Years Group can be generally used. However, since some design approaches consider winter, summer, and weekday and weekend load, statistics for these are provided in the

Appendix.

The parametric distributions that had the overall best fit were the Log Normal and Gumbel. Therefore, a reasonable approach would be to use it as the default model of variation of consumption.

As previously described, a data-driven approach for off-grid system design can significantly reduce load estimation error, reducing capital costs or improving reliability. This research demonstrates how high-resolution measured load data can be analyzed and used to generate load profiles to achieve that aim. As more load data becomes available for off-grid systems in a variety of contexts, it is worthwhile to discuss the general considerations and limitations of this approach.

First, it is emphasized that load data alone is insufficient to develop meaningful load profiles for a data-driven load estimation approach. Contextual information is extremely important in understanding how the data should and should not be used. As mentioned, one should not expect load profiles based on data from the Navajo Nation to serve as a reasonable estimate of load in rural Zambia. In particular, the tariffs charged for energy, willingness and ability to pay, availability and types of appliances used, and general socio-economic and environmental conditions all influence consumption. The developed load profiles should only be used in similar contexts.

Second, the technology itself may influence load profiles. The inverter and battery ratings impose an upper limit on consumption and the shape of the load profile. Ideally, the load profile would represent the unconstrained or true demand for electricity. This is not the case when there are outages caused by energy shortages or equipment malfunction. Including days where an outage occurs alters the load profile—the consumption is zero during an outage—and therefore does not represent the true demand. The average load profile will be artificially lower during times prone to outages. On the one hand, excluding outage days runs the risk of eliminating days where consumption was perhaps higher than normal. In this paper, outage days were excluded as they were infrequent in the homes considered. It is less straightforward how to handle situations when there are a high number of outages, for example more than 20 % of the days. One approach would be to develop a separate load profile for the outage days, and backfill the times when an outage occurs with the load profile from days in which outages did not occur. Since most energy-constrained outages occur late in the evening when the load tends to be constant, an alternative approach would be to use linear interpolation across the outage period.

Third, the scope of the data is important. At least one year of data should be collected to capture seasonal variation. The minimum sampling rate should be one hour as this is the resolution used by most design software.

Fourth, when data from multiple homes are available, it is important to identify prototypical load profiles as was done in this work before any data aggregation. Engineering judgement is needed to determine if differences exhibited in the prototypical load profiles identified by the clustering technique are salient enough to warrant disaggregation of the data. Keep in mind that when several prototypical load profiles are identified, the practical challenge of deciding which to use in load estimation for a new home is introduced. Capturing demographic and contextual information can help in this. For example, if a certain prototypical load profile is found to primarily occur in homes of retired elders, then it is reasonable to apply that load profile to design an off-grid system for home occupied by retired elders. Similarly, splitting the data temporally—bi-annually, seasonally, monthly, and weekday/weekend—should be explored. This is particularly important in locations with large seasonal variation in insolation.

Lastly, the average or median values are generally not sufficient to represent the load. As was shown in this data set, there can be wide variation in load for any given hour of a load profile. Understanding the extremes—especially high consumption levels—will lead to a more appropriate system design.

Conclusions and future work

This research presented normalized load profiles and associated statistical moments and probabilistic models derived from high-resolution consumption data from 90 off-grid homes on the Navajo Nation. The normalized load profile describes the average proportion of a day's total energy consumption that was consumed during each hour. A K-means clustering approach identified five prototypical normalized load profiles for each grouping of data: All Years, 2022, 2021, Winter, Spring, Summer, Fall, Weekday, and Weekend. Despite some differences in the prototypical load profiles, several common, general features emerged: the load profiles almost all peaked near 18:00, and most had low overnight that started to increase at approximately 6:00. Year-to-year comparison showed that it was common for the load profiles of the individual homes to change, albeit modestly, from one year to the next. These load profiles support a data-driven approach to off-grid energy production system design, and can directly be incorporated into off-grid system design software such as HOMER Pro. Statistical moments were computed for the load profiles, and variation that of consumption that occurs within each hour was examined. Eight parametric distributions were evaluated for the suitableness in modeling this variation. In general, the Log Normal and Gumbel were found to be the best performing when evaluated by the Cramér-von Mises statistic. An important conclusion is that the usual assumption of Normal variation around the mean is not appropriate. The distributions exhibited a range of p-values, and approximately 25 % passed a hypothesis test based on a level of significance $\alpha = 0.20$.

The data and analyses presented in this paper can enable further research in the domain of off-grid system design and analysis. Future research avenues include developing timeseries models of the load that capture inter-hour relationships of consumption, and combing the load profiles with solar data to evaluate the sizing of off-grid systems on the Navajo Nation or similar locations, and aggregating the load profiles to study the viability of minigrids vis-á-vis standalone off-grid systems. In addition, further research is needed to explain why the load profiles exhibit the characteristics described in this paper.

When considering the development of load profiles based on data, an important open research question is the benefit gained by using highly nuanced load profiles in the design phase. For example, what is lost by using a single hourly load profile to represent an entire year versus developing load profiles for each month or by accurately modeling the variability? The more nuance and detail included increases the data collection and processing burden that might not be offset by meaningful improvements in design.

CRediT authorship contribution statement

Henry Louie: Writing – review & editing, Writing – original draft, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Scott O'Shea:** Formal analysis. **Stanley Atcitty:** Writing – review & editing, Methodology, Conceptualization. **Derrick Terry:** Resources, Data curation. **Darrick Lee:** Writing – review & editing, Conceptualization. **Peter Romine:** Resources, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Co-author Derrick Terry is an employee of NTUA, which provided access to the data. However, the raw data was accessed by a different author, and Mr. Terry did not conduct data analysis nor did he direct or dictate the conclusions from the analyses presented in this paper.

Acknowledgements

This paper is based upon work supported by the National Science Foundation under Grant # 2137027/2137028. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. The data considered in this work was provided by Navajo Tribal Utility Authority. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Appendix A

Statistical quantities and best-fit distribution parameters are provided in the Appendix for prototype load profiles of selected groups of data. The Groups included are the All Years, Winter, Summer, Weekday, and Weekend. The 2021 and 2022 Groups are not included as they were generally similar; the Spring and Fall Groups are not included as it is the seasons with more extreme climate that are generally of interest in off-grid solar design.

Tables 6–10 provide the mean and standard deviation of the prototype load profiles for each hour of the selected Groups of data. Higher order moments are generally not needed in off-grid design software, and so are excluded. Tables 11–20 provide the parameters of the best-fitting distribution for each hour of each cluster for each of the selected Groups of data. The parameters are ordered as: location, scale, and shape parameters (for Beta distributions, the shape parameters are ordered as a, b). Distributions whose corresponding p-values are greater than or equal to 0.20 are marked in bold. Consult SciPy (Virtanen et al., 2020) for formulations of the distributions.

Table 6
Normalized load profiles for All Years Group.

	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
Time	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
0	0.038	0.014	0.045	0.016	0.036	0.020	0.029	0.016	0.037	0.018
1	0.036	0.013	0.044	0.016	0.037	0.021	0.026	0.015	0.031	0.015
2	0.035	0.013	0.043	0.016	0.036	0.017	0.024	0.014	0.028	0.013
3	0.034	0.013	0.044	0.022	0.036	0.018	0.023	0.013	0.027	0.013
4	0.035	0.015	0.045	0.024	0.040	0.023	0.023	0.014	0.026	0.013
5	0.037	0.019	0.044	0.020	0.053	0.032	0.024	0.016	0.028	0.017
6	0.040	0.024	0.042	0.019	0.063	0.035	0.028	0.021	0.031	0.019
7	0.042	0.024	0.039	0.019	0.073	0.033	0.034	0.025	0.037	0.024
8	0.043	0.025	0.038	0.022	0.069	0.037	0.041	0.029	0.039	0.025
9	0.043	0.023	0.036	0.021	0.050	0.034	0.047	0.032	0.040	0.025
10	0.042	0.022	0.034	0.021	0.040	0.026	0.049	0.033	0.039	0.023
11	0.041	0.021	0.034	0.020	0.036	0.022	0.048	0.032	0.038	0.022
12	0.041	0.020	0.034	0.020	0.035	0.020	0.048	0.029	0.039	0.023
13	0.041	0.020	0.034	0.019	0.033	0.018	0.048	0.028	0.039	0.023
14	0.041	0.020	0.036	0.021	0.034	0.018	0.049	0.029	0.039	0.021
15	0.042	0.021	0.037	0.021	0.035	0.020	0.051	0.030	0.040	0.021
16	0.042	0.020	0.039	0.024	0.037	0.022	0.053	0.031	0.042	0.023
17	0.044	0.023	0.040	0.025	0.037	0.022	0.054	0.032	0.046	0.026
18	0.047	0.023	0.043	0.024	0.036	0.018	0.056	0.032	0.053	0.029
19	0.049	0.022	0.047	0.024	0.036	0.018	0.057	0.029	0.060	0.029
20	0.051	0.021	0.052	0.022	0.037	0.018	0.058	0.029	0.067	0.028
21	0.049	0.020	0.054	0.022	0.037	0.017	0.053	0.025	0.068	0.027
22	0.045	0.017	0.051	0.019	0.037	0.017	0.042	0.021	0.059	0.025
23	0.041	0.016	0.047	0.016	0.035	0.018	0.034	0.018	0.047	0.022

Table 7
Normalized load profiles for Summer Group.

	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
Time	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
0	0.034	0.016	0.035	0.014	0.043	0.013	0.023	0.016	0.031	0.014
1	0.030	0.012	0.032	0.013	0.041	0.013	0.021	0.015	0.030	0.014
2	0.028	0.011	0.030	0.012	0.040	0.012	0.019	0.014	0.029	0.013
3	0.026	0.011	0.029	0.012	0.041	0.017	0.018	0.013	0.028	0.012
4	0.026	0.011	0.029	0.013	0.041	0.017	0.018	0.012	0.030	0.017
5	0.031	0.025	0.028	0.013	0.041	0.017	0.018	0.013	0.035	0.023
6	0.039	0.031	0.030	0.015	0.040	0.016	0.021	0.018	0.044	0.029
7	0.046	0.028	0.031	0.016	0.040	0.018	0.024	0.017	0.054	0.033
8	0.047	0.028	0.035	0.018	0.039	0.017	0.031	0.023	0.057	0.035
9	0.042	0.024	0.038	0.019	0.038	0.016	0.037	0.029	0.053	0.036
10	0.038	0.023	0.040	0.018	0.037	0.016	0.045	0.033	0.049	0.034
11	0.037	0.023	0.043	0.020	0.038	0.016	0.051	0.036	0.045	0.029
12	0.037	0.021	0.044	0.020	0.039	0.015	0.058	0.039	0.043	0.025
13	0.037	0.019	0.046	0.021	0.040	0.015	0.065	0.042	0.042	0.024

(continued on next page)

Table 7 (*continued*)

	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
Time	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
14	0.038	0.020	0.047	0.021	0.040	0.016	0.071	0.042	0.042	0.024
15	0.039	0.019	0.049	0.022	0.041	0.016	0.078	0.045	0.042	0.023
16	0.042	0.021	0.051	0.024	0.042	0.016	0.078	0.048	0.043	0.025
17	0.047	0.026	0.053	0.028	0.044	0.017	0.076	0.050	0.044	0.027
18	0.052	0.027	0.055	0.029	0.044	0.017	0.065	0.050	0.045	0.025
19	0.059	0.029	0.055	0.026	0.045	0.016	0.049	0.035	0.045	0.024
20	0.067	0.028	0.057	0.023	0.047	0.016	0.040	0.026	0.049	0.025
21	0.066	0.027	0.055	0.022	0.048	0.016	0.036	0.022	0.046	0.022
22	0.053	0.023	0.048	0.019	0.046	0.014	0.032	0.021	0.040	0.018
23	0.041	0.021	0.040	0.016	0.044	0.014	0.027	0.019	0.035	0.018

Table 8
Normalized load profiles for Winter Group.

	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
Time	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
0	0.037	0.020	0.039	0.016	0.045	0.018	0.029	0.020	0.034	0.014
1	0.030	0.016	0.037	0.015	0.042	0.017	0.026	0.019	0.031	0.013
2	0.026	0.013	0.036	0.015	0.041	0.017	0.022	0.014	0.030	0.014
3	0.024	0.012	0.034	0.013	0.039	0.018	0.022	0.014	0.030	0.013
4	0.023	0.012	0.035	0.015	0.038	0.016	0.022	0.015	0.030	0.016
5	0.023	0.012	0.038	0.020	0.037	0.017	0.024	0.023	0.032	0.017
6	0.023	0.014	0.041	0.023	0.040	0.021	0.025	0.020	0.037	0.022
7	0.026	0.019	0.043	0.023	0.040	0.024	0.030	0.027	0.047	0.028
8	0.033	0.024	0.043	0.022	0.040	0.026	0.040	0.036	0.052	0.031
9	0.042	0.027	0.043	0.024	0.039	0.025	0.052	0.038	0.049	0.028
10	0.043	0.025	0.040	0.021	0.037	0.024	0.061	0.047	0.045	0.026
11	0.042	0.024	0.040	0.020	0.036	0.026	0.058	0.045	0.040	0.022
12	0.042	0.024	0.040	0.021	0.036	0.028	0.052	0.038	0.039	0.022
13	0.042	0.025	0.039	0.020	0.036	0.026	0.050	0.033	0.039	0.022
14	0.041	0.023	0.040	0.021	0.036	0.025	0.047	0.035	0.038	0.021
15	0.042	0.022	0.040	0.022	0.037	0.023	0.048	0.030	0.039	0.022
16	0.044	0.025	0.041	0.020	0.037	0.032	0.049	0.034	0.039	0.021
17	0.047	0.026	0.042	0.025	0.039	0.030	0.050	0.034	0.042	0.024
18	0.054	0.028	0.045	0.025	0.042	0.023	0.052	0.038	0.049	0.025
19	0.065	0.029	0.051	0.024	0.052	0.025	0.057	0.036	0.057	0.025
20	0.072	0.029	0.053	0.023	0.056	0.025	0.056	0.033	0.059	0.024
21	0.070	0.027	0.050	0.021	0.056	0.022	0.049	0.026	0.056	0.022
22	0.061	0.027	0.047	0.018	0.052	0.021	0.043	0.024	0.048	0.020
23	0.048	0.024	0.042	0.017	0.048	0.021	0.036	0.022	0.039	0.017

Table 9
Normalized load profiles for Weekday Group.

	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
Time	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
0	0.038	0.014	0.037	0.019	0.029	0.015	0.045	0.016	0.035	0.020
1	0.036	0.013	0.031	0.015	0.026	0.014	0.044	0.017	0.036	0.021
2	0.035	0.013	0.028	0.013	0.024	0.013	0.043	0.016	0.036	0.017
3	0.034	0.013	0.027	0.013	0.023	0.013	0.045	0.023	0.036	0.018
4	0.035	0.016	0.027	0.014	0.023	0.014	0.046	0.025	0.040	0.023
5	0.037	0.020	0.029	0.018	0.025	0.017	0.045	0.021	0.052	0.032
6	0.041	0.024	0.031	0.020	0.029	0.023	0.042	0.019	0.064	0.035
7	0.043	0.025	0.037	0.025	0.035	0.026	0.039	0.019	0.073	0.034
8	0.043	0.024	0.039	0.025	0.042	0.029	0.038	0.022	0.070	0.038
9	0.042	0.023	0.038	0.024	0.047	0.031	0.035	0.020	0.051	0.034
10	0.041	0.021	0.037	0.022	0.048	0.032	0.034	0.020	0.040	0.026
11	0.040	0.020	0.037	0.021	0.047	0.033	0.033	0.020	0.037	0.023
12	0.040	0.020	0.038	0.022	0.046	0.028	0.033	0.020	0.035	0.020
13	0.041	0.020	0.037	0.022	0.047	0.027	0.033	0.018	0.033	0.018
14	0.041	0.020	0.037	0.021	0.048	0.027	0.035	0.020	0.033	0.017
15	0.041	0.020	0.039	0.021	0.050	0.028	0.036	0.020	0.036	0.021
16	0.042	0.020	0.041	0.023	0.052	0.030	0.038	0.024	0.038	0.024
17	0.044	0.024	0.046	0.027	0.054	0.032	0.040	0.023	0.036	0.020
18	0.047	0.023	0.054	0.031	0.056	0.031	0.043	0.024	0.037	0.018
19	0.050	0.022	0.062	0.031	0.059	0.029	0.047	0.025	0.036	0.017

(*continued on next page*)

Table 9 (*continued*)

	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
Time	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
20	0.052	0.021	0.069	0.029	0.059	0.028	0.052	0.023	0.037	0.017
21	0.050	0.019	0.070	0.027	0.054	0.025	0.054	0.022	0.038	0.017
22	0.045	0.017	0.060	0.026	0.043	0.021	0.051	0.019	0.037	0.017
23	0.041	0.015	0.048	0.023	0.034	0.017	0.047	0.017	0.034	0.018

Table 10
Normalized load profiles for Weekend Group.

	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
Time	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
0	0.041	0.015	0.034	0.015	0.030	0.016	0.038	0.018	0.030	0.008
1	0.040	0.014	0.033	0.015	0.027	0.015	0.033	0.015	0.029	0.008
2	0.039	0.014	0.031	0.014	0.025	0.014	0.030	0.014	0.029	0.007
3	0.039	0.015	0.030	0.014	0.023	0.013	0.028	0.013	0.030	0.011
4	0.039	0.017	0.029	0.014	0.023	0.013	0.027	0.013	0.035	0.021
5	0.040	0.018	0.029	0.014	0.022	0.012	0.028	0.015	0.055	0.036
6	0.042	0.021	0.032	0.020	0.024	0.015	0.031	0.023	0.066	0.038
7	0.042	0.022	0.038	0.025	0.028	0.020	0.035	0.023	0.070	0.027
8	0.042	0.023	0.047	0.031	0.036	0.024	0.039	0.027	0.068	0.031
9	0.040	0.021	0.054	0.034	0.045	0.030	0.040	0.024	0.059	0.032
10	0.038	0.019	0.054	0.037	0.048	0.029	0.040	0.022	0.047	0.024
11	0.038	0.018	0.050	0.032	0.048	0.026	0.039	0.021	0.039	0.021
12	0.039	0.020	0.047	0.030	0.049	0.028	0.041	0.022	0.035	0.012
13	0.039	0.020	0.044	0.027	0.051	0.030	0.041	0.022	0.035	0.013
14	0.040	0.020	0.045	0.029	0.053	0.031	0.042	0.020	0.036	0.013
15	0.040	0.020	0.044	0.027	0.055	0.031	0.041	0.021	0.035	0.012
16	0.042	0.022	0.044	0.025	0.055	0.030	0.043	0.022	0.035	0.012
17	0.043	0.025	0.045	0.025	0.057	0.031	0.046	0.023	0.038	0.017
18	0.044	0.022	0.046	0.026	0.055	0.033	0.049	0.025	0.038	0.014
19	0.046	0.020	0.049	0.025	0.056	0.025	0.055	0.027	0.039	0.014
20	0.049	0.020	0.050	0.025	0.056	0.028	0.063	0.027	0.039	0.013
21	0.049	0.019	0.048	0.023	0.053	0.025	0.065	0.027	0.041	0.016
22	0.046	0.017	0.041	0.017	0.045	0.022	0.058	0.023	0.039	0.012
23	0.043	0.015	0.037	0.017	0.036	0.019	0.047	0.022	0.033	0.012

Table 11
Best fit parameters for All Years Group.

	Cluster 2		Cluster 3	
Time	Dist.	Parameters	Dist.	Parameters
0	LogN	−0.024, 0.0670, 0.224	LogN	0.005, 0.0273, 0.539
1	LogN	−0.021, 0.0625, 0.240	LogN	0.003, 0.0299, 0.505
2	LogN	−0.024, 0.0651, 0.228	LogN	−0.003, 0.0363, 0.426
3	Gum.	0.036, 0.0138	LogN	−0.004, 0.0370, 0.416
4	Gum.	0.036, 0.0142	LogN	−0.004, 0.0395, 0.478
5	Gum.	0.036, 0.0140	Gam.	0.000, 0.0182, 2.896
6	Gum.	0.034, 0.0136	**Weibull**	0.000, 0.0710, 1.912
7	Gum.	0.031, 0.0134	**LogN**	−0.070, 0.1392, 0.227
8	Gum.	0.030, 0.0137	**Beta**	−0.003, 7.8767, 3.827, 415
9	Gum.	0.028, 0.0139	LogN	−0.004, 0.0454, 0.611
10	Gum.	0.026, 0.0134	LogN	−0.004, 0.0384, 0.540
11	Gum.	0.026, 0.0134	LogN	−0.004, 0.0358, 0.473
12	Gum.	0.026, 0.0138	LogN	−0.002, 0.0333, 0.462
13	Gum.	0.026, 0.0136	Gum.	0.026, 0.0120
14	Gum.	0.027, 0.0142	Gum.	0.027, 0.0123
15	Gum.	0.028, 0.0143	Gum.	0.027, 0.0131
16	Gum.	0.030, 0.0151	LogN	−0.005, 0.0380, 0.442
17	Gum.	0.031, 0.0159	Gum.	0.028, 0.0137
18	Gum.	0.034, 0.0158	Gum.	0.028, 0.0134
19	Gum.	0.037, 0.0165	**Gum.**	0.029, 0.0130
20	Gum.	0.042, 0.0171	**Gum.**	0.029, 0.0129
21	LogN	−0.019, 0.0705, 0.276	Gum.	0.030, 0.0128
22	LogN	−0.022, 0.0704, 0.247	Gum.	0.030, 0.0119
23	Beta	−0.008, 2.8437, 11.43, 582	**LogN**	0.004, 0.0271, 0.505

Table 12
Best fit parameters for All Years continued.

	Cluster 4		Cluster 5	
Time	Dist.	Parameters	Dist.	Parameters
0	**LogN**	−0.010, 0.0365, 0.386	**LogN**	−0.006, 0.0396, 0.404
1	**LogN**	−0.009, 0.0333, 0.389	**Gum.**	0.025, 0.0112
2	**Gum.**	0.019, 0.0102	**Gum.**	0.022, 0.0101
3	Gum.	0.018, 0.0096	**Gum.**	0.021, 0.0095
4	Gum.	0.017, 0.0098	Gum.	0.021, 0.0096
5	LogN	−0.005, 0.0257, 0.480	Gum.	0.021, 0.0107
6	LogN	−0.002, 0.0252, 0.600	LogN	−0.003, 0.0300, 0.500
7	LogN	−0.003, 0.0307, 0.614	LogN	−0.004, 0.0348, 0.542
8	LogN	−0.005, 0.0399, 0.585	LogN	−0.006, 0.0390, 0.514
9	LogN	−0.008, 0.0480, 0.519	Gum.	0.029, 0.0169
10	LogN	−0.008, 0.0494, 0.507	Gum.	0.029, 0.0162
11	Gum.	0.035, 0.0206	Gum.	0.029, 0.0157
12	Gum.	0.036, 0.0198	Gum.	0.029, 0.0157
13	Gum.	0.036, 0.0199	Gum.	0.030, 0.0155
14	Gum.	0.037, 0.0202	Gum.	0.030, 0.0154
15	Gum.	0.038, 0.0209	Gum.	0.031, 0.0156
16	Gum.	0.040, 0.0215	Gum.	0.032, 0.0165
17	Gum.	0.041, 0.0223	Gum.	0.036, 0.0181
18	Gum.	0.043, 0.0226	LogN	−0.012, 0.0600, 0.400
19	LogN	−0.034, 0.0872, 0.304	Gum.	0.047, 0.0224
20	LogN	−0.041, 0.0953, 0.278	LogN	−0.048, 0.112, 0.242
21	Beta	−0.034, 5.14, 12.324721	LogN	−0.064, 0.130, 0.200
22	Beta	−0.008, 6.88, 5.529757	Beta	−0.015, 4.69, 9.00, 559
23	**Beta**	−0.002, 7.4202, 3.986824	**Gam.**	0.000, 0.0102, 4.617

Table 13
Best fit parameters for Summer Group.

	Cluster 1		Cluster 2		Cluster 3	
Time	Dist.	Parameters	Dist.	Parameters	Dist.	Parameters
0	Gum.	0.027, 0.011	**Beta**	−0.003, 4.654, 7.170, 871	LogN	−0.010, 0.051, 0.242
1	**Gum.**	0.024, 0.009	**Beta**	−0.001, 4.611, 6.868, 937	LogN	−0.013, 0.053, 0.222
2	**Gum.**	0.023, 0.008	**Beta**	−0.002, 2.477, 7.029, 528	LogN	−0.015, 0.054, 0.215
3	**Gum.**	0.022, 0.008	**LogN**	−0.013, 0.040, 0.286	Gum.	0.035, 0.011
4	**Gum.**	0.021, 0.008	**LogN**	−0.014, 0.041, 0.290	Gum.	0.035, 0.011
5	LogN	0.004, 0.022, 0.601	**Gum.**	0.023, 0.010	Gum.	0.034, 0.011
6	LogN	0.005, 0.027, 0.673	**Gum.**	0.024, 0.011	Gum.	0.033, 0.011
7	**LogN**	0.001, 0.038, 0.575	Gum.	0.025, 0.012	Gum.	0.033, 0.012
8	**LogN**	−0.002, 0.042, 0.530	Gum.	0.027, 0.013	Gum.	0.032, 0.012
9	**LogN**	−0.004, 0.041, 0.492	LogN	−0.017, 0.052, 0.323	Gum.	0.031, 0.012
10	**LogN**	−0.002, 0.035, 0.518	LogN	−0.033, 0.070, 0.251	LogN	−0.014, 0.050, 0.267
11	**LogN**	−0.001, 0.033, 0.525	Beta	−0.009, 4.793, 7.279668	LogN	−0.017, 0.053, 0.257
12	**LogN**	−0.003, 0.036, 0.460	LogN	−0.033, 0.075, 0.245	LogN	−0.020, 0.057, 0.240
13	**LogN**	−0.005, 0.038, 0.426	LogN	−0.026, 0.070, 0.274	LogN	−0.027, 0.065, 0.211
14	**LogN**	−0.004, 0.038, 0.429	LogN	−0.026, 0.070, 0.276	LogN	−0.024, 0.063, 0.224
15	**LogN**	−0.004, 0.039, 0.427	LogN	−0.027, 0.073, 0.273	LogN	−0.024, 0.064, 0.226
16	**LogN**	−0.005, 0.043, 0.420	Gum.	0.040, 0.018	LogN	−0.021, 0.062, 0.240
17	**Beta**	0.006, 9.200, 3.002, 671	Gum.	0.042, 0.019	LogN	−0.017, 0.059, 0.263
18	**Gam.**	0.005, 0.014, 3.339	Gum.	0.044, 0.020	LogN	−0.014, 0.056, 0.268
19	**Gum.**	0.046, 0.021	Gum.	0.044, 0.019	LogN	−0.016, 0.059, 0.248
20	LogN	−0.035, 0.098, 0.267	LogN	−0.030, 0.084, 0.254	LogN	−0.020, 0.066, 0.226
21	Beta	−0.004, 7.234, 7.074, 725	LogN	−0.031, 0.083, 0.244	LogN	−0.022, 0.069, 0.218
22	**LogN**	−0.014, 0.063, 0.337	**LogN**	−0.032, 0.078, 0.230	LogN	−0.018, 0.062, 0.214
23	LogN	−0.001, 0.038, 0.445	**LogN**	−0.019, 0.057, 0.267	LogN	−0.014, 0.057, 0.227

Table 14
Best fit parameters for Summer Group continued.

	Cluster 4		Cluster 5	
Time	Dist.	Parameters	Dist.	Parameters
0	LogN	−0.002, 0.021, 0.639	**Gum.**	0.025, 0.011
1	LogN	−0.002, 0.019, 0.637	**Gum.**	0.024, 0.010
2	**LogN**	−0.001, 0.016, 0.668	**Gum.**	0.023, 0.010
3	**Gam.**	0.001, 0.009, 1.826	**LogN**	−0.010, 0.036, 0.312
4	**Gam.**	0.001, 0.009, 1.761	Gum.	0.023, 0.011
5	**Gam.**	0.001, 0.010, 1.733	LogN	0.001, 0.028, 0.599
6	**Gam.**	0.001, 0.012, 1.618	LogN	−0.001, 0.037, 0.633
7	**Beta**	0.000, 8.15, 2.012, 694	Gam.	0.002, 0.022, 2.363
8	**Beta**	0.001, 17.3, 1.657, 971	LogN	−0.003, 0.051, 0.566
9	**Gam.**	0.002, 0.023, 1.557	LogN	0.006, 0.037, 0.690
10	Beta	0.001, 23.3, 1.707, 897	LogN	0.004, 0.037, 0.635
11	**Naka.**	0.004, 0.059, 0.535	LogN	0.001, 0.037, 0.562
12	**Naka.**	0.004, 0.067, 0.569	LogN	−0.001, 0.039, 0.502
13	**Naka.**	0.003, 0.075, 0.663	LogN	−0.001, 0.038, 0.493
14	**Weibull**	0.002, 0.076, 1.687	LogN	−0.003, 0.040, 0.461
15	**Weibull**	0.005, 0.081, 1.637	LogN	−0.001, 0.038, 0.476
16	**Weibull**	0.002, 0.085, 1.660	LogN	0.000, 0.039, 0.492
17	**Beta**	−0.003, 30.8, 2.542, 989	LogN	0.000, 0.039, 0.498
18	**Gam.**	0.000, 0.036, 1.782	**Gum.**	0.035, 0.017
19	**Gum.**	0.034, 0.025	Gum.	0.035, 0.017
20	**Gum.**	0.029, 0.019	Gum.	0.038, 0.018
21	**Beta**	−0.004, 12.1, 3.19, 949	**LogN**	−0.019, 0.061, 0.318
22	**Gam.**	0.000, 0.014, 2.30	LogN	−0.016, 0.053, 0.308
23	Beta	0.000, 4.493, 2.004, 338	**Gum.**	0.028, 0.012

Table 15
Best fit parameters for Winter Group.

	Cluster 1		Cluster 2		Cluster 3	
Time	Dist.	Parameters	Dist.	Parameters	Dist.	Parameters
0	Beta	0.003, 9.58, 2.84, 798	LogN	−0.023, 0.060, 0.246	LogN	−0.037, 0.080, 0.207
1	**LogN**	−0.005, 0.032, 0.448	LogN	−0.029, 0.064, 0.214,	LogN	−0.045, 0.086, 0.187
2	**LogN**	−0.006, 0.030, 0.390	LogN	−0.029, 0.063, 0.216,	Norm.	0.041, 0.017
3	**Gum.**	0.018, 0.009	Beta	−0.018, 3.278, 15.2, 941	Beta	−0.014, 4.75, 10.4, 922
4	Gum.	0.018, 0.009	LogN	−0.020, 0.054, 0.259,	Norm.	0.038, 0.016
5	**Gum.**	0.017, 0.009	Gum.	0.030, 0.014,	LogN	−0.031, 0.066, 0.244
6	Gum.	0.017, 0.010	Gum.	0.032, 0.015,	Gum.	0.031, 0.015
7	LogN	−0.004, 0.026, 0.545	Gum.	0.034, 0.016,	Gum.	0.031, 0.016
8	Gam.	0.001, 0.015, 2.235	Gum.	0.034, 0.015,	Gum.	0.030, 0.016
9	**Gam.**	0.001, 0.017, 2.383	Gum.	0.034, 0.015	Gum.	0.029, 0.016
10	**Gum.**	0.033, 0.018	Gum.	0.032, 0.014	Gum.	0.027, 0.016
11	Gum.	0.032, 0.018	Gum.	0.032, 0.014	Gum.	0.026, 0.015
12	Gum.	0.032, 0.017	Gum.	0.032, 0.014	Gum.	0.026, 0.016
13	Gum.	0.032, 0.017	Gum.	0.032, 0.014	Gum.	0.026, 0.015
14	Beta	0.000, 5.411, 3.430, 453	Gum.	0.032, 0.014,	Gum.	0.027, 0.015
15	Rayleigh	0.001, 0.033	Gum.	0.032, 0.015,	Gum.	0.027, 0.015
16	Gum.	0.033, 0.019	LogN	−0.021, 0.059, 0.289,	Gum.	0.027, 0.016
17	Gum.	0.035, 0.020	Gum.	0.034, 0.015,	LogN	−0.004, 0.037, 0.526
18	LogN	−0.028, 0.078, 0.327	Gum.	0.036, 0.016,	Gum.	0.032, 0.017
19	LogN	−0.083, 0.146, 0.192	Gum.	0.041, 0.017,	LogN	−0.031, 0.079, 0.293
20	LogN	−0.092, 0.162, 0.171	Gum.	0.042, 0.017,	LogN	−0.043, 0.097, 0.236
21	LogN	−0.092, 0.160, 0.163	Gum.	0.041, 0.016,	LogN	−0.041, 0.095, 0.226
22	LogN	−0.071, 0.129, 0.200	Gum.	0.039, 0.015,	LogN	−0.037, 0.087, 0.232
23	Naka.	0.004, 0.050, 0.942	LogN	−0.021, 0.061,	LogN	−0.026, 0.072, 0.262

Table 16
Best fit parameters for Winter Group continued.

	Cluster 4		Cluster 5	
Time	Dist.	Parameters	Dist.	Parameters
0	**Gum.**	0.021, 0.013	LogN	−0.016, 0.049, 0.266
1	Gum.	0.019, 0.011	LogN	−0.021, 0.051, 0.233
2	**Gum.**	0.016, 0.010	LogN	−0.020, 0.049, 0.246
3	Gum.	0.016, 0.010,	LogN	−0.022, 0.050, 0.245
4	Gum.	0.016, 0.010,	LogN	−0.014, 0.043, 0.300
5	LogN	−0.002, 0.022, 0.615	Gum.	0.025, 0.012
6	LogN	−0.003, 0.023, 0.596	Gum.	0.028, 0.014
7	LogN	−0.002, 0.024, 0.743	LogN	−0.008, 0.049, 0.467
8	**Gamma**	0.000, 0.025, 1.607	LogN	−0.008, 0.053, 0.471
9	**Gum.**	0.036, 0.025	LogN	−0.009, 0.053, 0.440
10	Gum.	0.043, 0.029	LogN	−0.007, 0.046, 0.460
11	Gum.	0.041, 0.026	Gum.	0.031, 0.015
12	Gum.	0.037, 0.024	Gum.	0.030, 0.015
13	Gum.	0.036, 0.022	Gum.	0.030, 0.015
14	Gum.	0.034, 0.022	Gum.	0.029, 0.014
15	Gum.	0.035, 0.022	Gum.	0.030, 0.015
16	Gum.	0.035, 0.022	Gum.	0.031, 0.015
17	Gum.	0.036, 0.023	Gum.	0.032, 0.017
18	Gum.	0.038, 0.023	**Gum.**	0.038, 0.019
19	Gum.	0.043, 0.025	**LogN**	−0.025, 0.079, 0.289
20	Ray.	−0.003, 0.048	**LogN**	−0.034, 0.090, 0.248
21	Norm.	0.049, 0.026	**LogN**	−0.030, 0.083, 0.251
22	Ray.	−0.002, 0.037	**Gum.**	0.039, 0.015
23	**Gum.**	0.026, 0.017	Gum.	0.032, 0.013

Table 17
Best fit parameters for Weekday Group.

	Cluster 1		Cluster 2		Cluster 3	
Time	Dist.	Parameters	Dist.	Parameters	Dist.	Parameters
0	LogN	−0.023, 0.060, 0.215	**LogN**	−0.005, 0.038, 0.426	**LogN**	−0.011, 0.037, 0.375
1	LogN	−0.033, 0.068, 0.179	**LogN**	−0.006, 0.035, 0.393	**Gum.**	0.020, 0.011
2	Beta	−0.011, 2.404, 13.604, 692	**LogN**	−0.010, 0.036, 0.338	Gum.	0.019, 0.010
3	LogN	−0.031, 0.064, 0.188	**Gum.**	0.021, 0.010	Gum.	0.018, 0.009
4	LogN	−0.013, 0.046, 0.296	**Gum.**	0.021, 0.010	Gum.	0.018, 0.010
5	Gum.	0.030, 0.013	LogN	−0.003, 0.028, 0.483	LogN	−0.004, 0.025, 0.492
6	Gum.	0.032, 0.014	LogN	−0.004, 0.031, 0.503	LogN	−0.002, 0.026, 0.613
7	Gum.	0.033, 0.015	LogN	−0.003, 0.034, 0.563	LogN	−0.004, 0.033, 0.606
8	LogN	−0.005, 0.043, 0.429	LogN	−0.004, 0.037, 0.535	**LogN**	−0.007, 0.042, 0.544
9	Gum.	0.033, 0.015	Gum.	0.028, 0.016	LogN	−0.008, 0.048, 0.505
10	Gum.	0.033, 0.014	Gum.	0.028, 0.016	LogN	−0.007, 0.048, 0.511
11	Gum.	0.032, 0.014	Gum.	0.028, 0.015	Gum.	0.034, 0.020
12	Gum.	0.032, 0.014	Gum.	0.029, 0.015	Gum.	0.035, 0.019
13	Gum.	0.033, 0.014	Gum.	0.028, 0.015	Gum.	0.036, 0.020
14	Gum.	0.033, 0.014	Gum.	0.028, 0.015	Gum.	0.036, 0.020
15	Gum.	0.033, 0.014	Gum.	0.030, 0.015	Gum.	0.037, 0.020
16	Gum.	0.034, 0.015	**Gum.**	0.031, 0.017	Gum.	0.039, 0.021
17	Gum.	0.035, 0.015	Gum.	0.035, 0.018	Gum.	0.041, 0.022
18	Gum.	0.038, 0.016	Gum.	0.041, 0.021	Gum.	0.043, 0.023
19	Gum.	0.041, 0.016	**Gum.**	0.049, 0.023	LogN	−0.045, 0.101, 0.268
20	Gum.	0.042, 0.016	**Beta**	−0.013, 8.809, 8.106, 864	LogN	−0.054, 0.110, 0.240
21	Gum.	0.041, 0.015	Beta	−0.020, 6.794, 10.62, 799	LogN	−0.076, 0.128, 0.188
22	LogN	−0.016, 0.059, 0.265	**LogN**	−0.040, 0.097, 0.250	Naka.	0.000, 0.048, 1.132
23	LogN	−0.020, 0.059, 0.236	**Gam.**	0.000, 0.011, 4.397	**Beta**	−0.004, 4.716, 4.763, 592

Table 18
Best fit parameters for Weekday Group continued.

	Cluster 4		Cluster 5	
Time	Dist.	Parameters	Dist.	Parameters
0	LogN	−0.025, 0.067, 0.223–8888	LogN	0.003, 0.028, 0.503
1	LogN	−0.002, 0.019, 0.637	LogN	0.002, 0.030, 0.499
2	**LogN**	−0.001, 0.016, 0.668	LogN	−0.004, 0.037, 0.414
3	**Gam.**	0.001, 0.009, 1.826	LogN	−0.004, 0.037, 0.419
4	**Gam.**	0.001, 0.009, 1.761	LogN	−0.003, 0.038, 0.499
5	**Gam.**	0.001, 0.010, 1.733	LogN	−0.005, 0.050, 0.552
6	**Gam.**	0.001, 0.012, 1.618	**Weibull**	0.000, 0.072, 1.932
7	**Beta**	0.000, 8.15, 2.012, 694	**Beta**	−0.021, 9.902, 7.571, 788
8	**Beta**	0.001, 17.3, 1.657, 971	**Beta**	0.001, 13.785, 3.256, 647
9	**Gam.**	0.002, 0.023, 1.557	LogN	−0.002, 0.043, 0.642
10	Beta	0.001, 23.3, 1.707, 897	LogN	−0.003, 0.037, 0.553
11	**Naka.**	0.004, 0.059, 0.535	LogN	−0.003, 0.035, 0.486
12	**Naka.**	0.004, 0.067, 0.569	LogN	−0.002, 0.033, 0.469
13	**Naka.**	0.003, 0.075, 0.663	Gum.	0.026, 0.012
14	**Weibull**	0.002, 0.076, 1.687	Gum.	0.026, 0.012
15	**Weibull**	0.005, 0.081, 1.637	Gum.	0.027, 0.013
16	**Weibull**	0.002, 0.085, 1.660	LogN	−0.004, 0.037, 0.468
17	**Beta**	−0.003, 30.8, 2.542, 989	Gum.	0.028, 0.013
18	**Gam.**	0.000, 0.036, 1.782	Gum.	0.029, 0.013
19	**Gum.**	0.034, 0.025,	**Gum.**	0.028, 0.013
20	**Gum.**	0.029, 0.019,	**Gum.**	0.029, 0.013
21	**Beta**	−0.004, 12.1, 3.19, 949	Gum.	0.030, 0.013
22	**Gam.**	0.000, 0.014, 2.30	Gum.	0.030, 0.012
23	Beta	0.000, 4.493, 2.004, 338	LogN	0.004, 0.027, 0.505

Table 19
Best fit parameters for Weekend Group.

	Cluster 1		Cluster 2		Cluster 3	
Time	Dist.	Parameters	Dist.	Parameters	Dist.	Parameters
0	LogN	−0.016, 0.056, 0.244	LogN	−0.039, 0.072, 0.200	**Gam.**	0.000, 0.009, 3.477
1	LogN	−0.017, 0.056 0.237	LogN	−0.032, 0.063, 0.221	**Gam.**	0.000, 0.008, 3.273
2	Beta	−0.002, 1.98 8.95, 417	LogN	−0.027, 0.056, 0.238	**Gum.**	0.019, 0.010
3	LogN	−0.013, 0.050 0.265	LogN	−0.027, 0.055, 0.241	**LogN**	−0.009, 0.030, 0.372
4	Gum.	0.032, 0.012	**LogN**	−0.023, 0.051, 0.266	LogN	−0.008, 0.028, 0.399
5	Gum.	0.033, 0.013	Naka.	0.001, 0.031, 1.111	LogN	−0.008, 0.028, 0.401
6	Gum.	0.033, 0.014	**Gum.**	0.024, 0.013	LogN	−0.003, 0.024, 0.501
7	Gum.	0.034, 0.014	Gum.	0.028, 0.016	**LogN**	−0.003, 0.027, 0.562
8	Gum.	0.033, 0.014	LogN	−0.007, 0.047, 0.500	**LogN**	−0.006, 0.037, 0.525
9	Gum.	0.032, 0.014	LogN	−0.008, 0.055, 0.495	**Gum.**	0.033, 0.020
10	Gum.	0.031, 0.013	LogN	−0.003, 0.049, 0.554	Gum.	0.036, 0.020
11	LogN	−0.017, 0.053 0.296	LogN	−0.005, 0.048, 0.504	Gum.	0.037, 0.020
12	Gum.	0.031, 0.014	Gum.	0.035, 0.019	Gum.	0.037, 0.020
13	Gum.	0.031, 0.014	Gum.	0.033, 0.018	Gum.	0.039, 0.021
14	Gum.	0.032, 0.014	Gum.	0.034, 0.018	Gum.	0.040, 0.021
15	Gum.	0.032, 0.014	Gum.	0.034, 0.018	Gum.	0.042, 0.022
16	Gum.	0.033, 0.015	Gum.	0.033, 0.017	Gum.	0.043, 0.021
17	Gum.	0.034, 0.015	Gum.	0.034, 0.018	Gum.	0.043, 0.022
18	Gum.	0.035, 0.015	Gum.	0.036, 0.018	Gum.	0.042, 0.022
19	LogN	−0.020, 0.064 0.279	Gum.	0.038, 0.019	LogN	−0.061, 0.114, 0.213
20	LogN	−0.023, 0.069 0.259	Gum.	0.039, 0.019	LogN	−0.042, 0.095, 0.269
21	LogN	−0.020, 0.067 0.258	Beta	−0.010, 6.662, 6.662767	Beta	−0.024, 6.979, 9.652, 868
22	LogN	−0.020, 0.064 0.239	Beta	−0.021, 3.195, 12.797644	**LogN**	−0.031, 0.073, 0.284
23	LogN	−0.017, 0.059 0.237	LogN	−0.027, 0.062, 0.258	**Weibull**	0.001, 0.039, 1.904

Table 20
Best fit parameters for Weekend Group continued.

	Cluster 4		Cluster 5	
Time	Dist.	Parameters	Dist.	Parameters
0	**Gum.**	0.030, 0.013	**LogN**	−0.005, 0.033, 0.218
1	**Gum.**	0.026, 0.011	**Gum.**	0.026, 0.006
2	**Gum.**	0.024, 0.010	**LogN**	−0.009, 0.037, 0.180
3	LogN	−0.011, 0.037, 0.310	**Gum.**	0.026, 0.007
4	**Gum.**	0.022, 0.010	LogN	0.012, 0.018, 0.690
5	Gum.	0.022, 0.011	LogN	0.012, 0.030, 0.858
6	LogN	−0.003, 0.030, 0.496	Weibull	0.013, 0.058, 1.417
7	LogN	−0.006, 0.036, 0.487	**Norm.**	0.070, 0.027
8	LogN	−0.006, 0.039, 0.517	**Weibull**	0.024, 0.049, 1.479
9	Gum.	0.030, 0.016	LogN	−0.003, 0.056, 0.479
10	Gum.	0.031, 0.016	LogN	−0.002, 0.044, 0.432
11	Gum.	0.030, 0.015	Gum.	0.032, 0.011
12	Gum.	0.031, 0.016	Gum.	0.030, 0.009
13	Gum.	0.032, 0.015	**Gum.**	0.030, 0.010
14	Gum.	0.033, 0.015	**Gum.**	0.030, 0.010
15	Gum.	0.033, 0.015	Gum.	0.030, 0.009
16	Gum.	0.034, 0.016	**Gum.**	0.029, 0.009
17	Gum.	0.036, 0.017	Gum.	0.031, 0.011
18	Gum.	0.038, 0.018	**Gum.**	0.031, 0.011
19	Gum.	0.044, 0.020	**Gum.**	0.033, 0.011
20	Gum.	0.050, 0.022	**Gum.**	0.033, 0.011
21	LogN	−0.037, 0.099, 0.251	**Gum.**	0.034, 0.011
22	LogN	−0.048, 0.104, 0.215	**LogN**	−0.004, 0.041, 0.270
23	**Gum.**	0.038, 0.017	**Gum.**	0.028, 0.008

References

2020 CARES Act Final Report for Navajo Nation Leadership. (2021). *Technical Report*. Fort Defiance, AZ: Navajo Tribal Utility Authority.

Asghar, N., Amjad, M. A., Rehman, H. ur, Munir, M., & Alhajj, R. (2022). Achieving sustainable development resilience: Poverty reduction through affordable access to electricity in developing economies. *Journal of Cleaner Production, 376*, Article 134040.

Begay, S. K. (2018). Navajo residential solar energy access as a global model. *The Electricity Journal, 31*, 9–15.

Bekele, G., & Palm, B. (2010). Feasibility study for a standalone solar–wind-based hybrid energy system for application in Ethiopia. *Applied Energy, 87*, 487–495.

Berg, D. (2009). Copula goodness-of-fit testing: An overview and power comparison. *European Journal of Finance, 7-8*, 675–701.

Blodgett, C., Dauenhauer, P., Louie, H., & Kickham, L. (2017). Accuracy of energy-use surveys in predicting rural mini-grid user consumption. *Energy for Sustainable Development, 41*, 88–105.

D'Agostino, R. B., & Stephens, M. A. (1986). *Goodness-of-fit Techniques*. New York, NY: Marcel Dekker.

Dytham, C. (2011). *Choosing and using statistics: A biologist's guide* (4th ed.). Chichester, UK: Wiley–Blackwell.

Franco, A., Shaker, M., Kalubi, D., & Hostettler, S. (2017). A review of sustainable energy access and technologies for healthcare facilities in the Global South. *Sustainable Energy Technologies and Assessments, 22*, 92–105.

Gallucci, M. (2019). Plugging in the Navajo Nation: A pilot program brought grid power to 200 homes, but there are 15,000 to go. *IEEE Spectrum, 56*, 6–7.

Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and power study. *Insurance: Mathematics and economics, 44*, 199–213.

International Energy Agency, Africa energy outlook 2014, https://www.iea.org/reports/africa-energy-outlook-2014, 2014. Accessed: Dec. 2022.

Kanagawa, M., & Nakata, T. (2008). Assessment of access to electricity and the socio-economic impacts in rural areas of developing countries. *Energy Policy, 36*, 2016–2029.

Lloyd, S. (1982). Least squares quantization in pcm. *IEEE Transactions on Information Theory, 28*, 129–137.

Louie, H. (2018). *Off-Grid Electrical Systems in Developing Countries*. Cham, Switzerland: SpringerNature.

Louie, H., Atcitty, S., Terry, D., Lee, D., & Romine, P. (2023). Daily electrical energy consumption characteristics and design implications for off-grid homes on the navajo nation. *Energy for Sustainable Development, 73*, 315–325.

Ma, T., Yang, H., & Lu, L. (2014). A feasibility study of a stand-alone hybrid solar–wind–battery system for a remote island. *Applied Energy, 121*, 149–158.

Mandelli, S., Brivio, C., Moncecchi, M., Riva, F., Bonamini, G., & Merlo, M. (2017). Novel loadprogen procedure for micro-grid design in emerging country scenarios: application to energy storage sizing, Energy Procedia. In *, 135. 11th International Renewable Energy Storage Conference, IRES 2017, 14–16 March 2017, Düsseldorf, Germany* (pp. 367–378).

Mandelli, S., Merlo, M., & Colombo, E. (2016). Novel procedure to formulate load profiles for off-grid rural areas. *Energy for Sustainable Development, 31*, 130–142.

L. B. Mohr, Understanding significance Testing, volume 73 of *Quantitative Applications in Social Sciences,* Sage Publications, Thousand Oaks, CA, 1990.

C. Mooney, Monte Carlo Simulation, volume 116 of *Quantitative Applications in Social Sciences,* Sage Publications, Thousand Oaks, CA, 1997.

Navajo Tribal Utility Authority, About us, https://www.ntua.com/about-us.html, 2022. Accessed: Dec. 2022.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65.

Samsara, Inc., Equipment monitoring, https://www.samsara.com/products/equipment-monitoring/, 2020. Accessed: Dec. 2022.

Sarkodie, S. A., & Adams, S. (2020a). Electricity access and income inequality in South Africa: Evidence from Bayesian and NARDL analyses. *Energy Strategy Reviews, 29*, Article 100480.

Sarkodie, S. A., & Adams, S. (2020b). Electricity access, human development index, governance and income inequality in Sub-Saharan Africa. *Energy Reports, 6*, 455–466.

Sovacool, B. K., Heffron, R. J., McCauley, D., & Goldthau, A. (2016). Energy decisions reframed as justice and ethical concerns. *Nature Energy, 1*.

Sovacool, B. K., & Ryan, S. E. (2016). The geography of energy and education: Leaders, laggards, and lessons for achieving primary and secondary school electrification. *Renewable and Sustainable Energy Reviews, 58*, 107–123.

U.S. Census Bureau, Total population in navajo nation reservation and off-reservation trust land, https://data.census.gov/all?g=2500000US2430, 2020. Accessed: May 2024.

UL, HOMER software, https://www.homerenergy.com, 2023. Accessed: Dec. 2023.

P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods 17 (2020) 261–272.

Williams, N. J., Jaramillo, P., Cornell, B., Lyons-Galante, I., & Wynn, E. (2017). Load characteristics of East African microgrids, in. *IEEE PES PowerAfrica, 2017*, 236–241. https://doi.org/10.1109/PowerAfrica.2017.7991230

Xendee, Xendee, https://www.xendee.com, 2024. Accessed: Apr. 2024.

Yoder, E., & Williams, N. J. (2020). Load profile prediction using customer characteristics, in. *IEEE PES/IAS PowerAfrica, 2020*, 1–5. https://doi.org/10.1109/PowerAfrica49420.2020.9219811