Understanding long-term energy use in off-grid solar home systems in sub-Saharan Africa

R. Perriment [a] , V. Mergulhão [b] , V. Kumtepeli [a] , P. Parikh [b] , M.D. McCulloch [a] ,
D.A. Howey [a] ,*

[a] Department of Engineering Science, University of Oxford, Oxford, OX1 3PJ, United Kingdom
[b] Engineering for International Development Centre, Bartlett School of Sustainable Construction, University College London, 1-19 Torrington Place, London, WC1E 7HB, United Kingdom

ARTICLE INFO

ABSTRACT

Solar home systems provide low-cost electricity for rural off-grid communities. As access to them increases, more long-term data becomes available on how these systems are used throughout their lifetime. This work analyses a dataset of 1,000 systems across sub-Saharan Africa. Dynamic time warping clustering was applied to the load demand data from the systems, identifying five distinct archetypal daily load profiles and their occurrence across the dataset. Temporal analysis reveals a general decline in daily energy consumption over time, with 77% of households reducing their usage compared to the start of ownership. On average, there is a 33% decrease in daily consumption by the end of the second year compared to the peak demand, which occurs on the 96th day. Combining the load demand analysis with payment data shows that this decrease in energy consumption is observed even in households that are not experiencing economic hardship, indicating there are reasons beyond financial constraints for decreasing energy use once energy access is obtained.

Introduction

Sustainable Development Goal 7 (SDG7) highlights the need for affordable, reliable, and sustainable energy access for all. Meeting this pressing challenge requires a dual focus on transitioning to renewable energy sources while simultaneously ensuring energy access for the 666.4 million people without electricity worldwide, 84.8% of whom live in sub-Saharan Africa (Guterres, 2025). Abundant solar resources and dramatically decreasing costs of photovoltaic panels make solar-based technology solutions a prime contender for sustainable electrification in sub-Saharan Africa.

While there has been significant progress towards universal electrification, the rate of global electrification has slowed in recent years due to the increasing difficulty of providing access to the remaining unelectrified population, who often live in remote, isolated areas where grid extension is not feasible (Almeshqab & Ustun, 2019). The popularity of mini-grids to serve these regions is growing, but they require large upfront investments, which in turn require guaranteed sufficient demand (Narayan et al., 2020). On the other hand, solar home systems (SHSs) are a viable alternative with lower upfront costs for providing Tier 1 and 2 energy access. (Tier 1 energy access includes electricity for lighting, radio and phone charging with a minimum of 12 Wh/day, while Tier 2 access includes all Tier 1 functions plus

televisions and fans with a minimum of 200 Wh/day.) The simplicity and lower upfront costs of SHSs make them an attractive choice—across Africa, 74% of off-grid energy investments between 2010 and 2020 were in SHSs (Irena, 2022).

While the roll-out of smart meters in high-income countries (HICs) has led to an influx of data on electricity consumption, there is a lack of data on electricity consumption in low- and lower-middle-income countries (LMICs). This is particularly true for newly electrified households where there was no previous access to electricity. In this paper we analyse a large SHS dataset from the Africa-based energy supplier BBOXX, applying clustering techniques to understand typical load profiles in newly electrified households. The payment structure used by BBOXX for their SHS is also very different to the standard price-per-kWh format used in HICs. Instead, a pay-as-you-go (PAYG) system is used; the amount each household pays is fixed per day of using the system, and based on the number of appliances the system powers. The PAYG approach allows users to top-up usage when they desire, in their local currency, and enables them to stop paying on certain days if they wish. However, when payment stops, electricity supply is also stopped.

Rural off-grid household load data is important for many stakeholders, from system designers to aid sizing and design choices, to local

---

governments to aid policy decisions, to researchers trying to improve off-grid systems and their use. We analyse the energy consumption of over 1,000 households with SHS across sub-Saharan Africa. Initially, we cluster the time series data to obtain representative load profiles. We then analyse the changes in household energy use over time by looking at shifts from cluster to cluster, and compare this with payment data to better understand the drivers for energy behaviour.

Literature review

Quantifying the electrical load requirements of rural off-grid households is essential for understanding individuals' needs and ensuring appropriate design and maintenance choices are made. Before access to big field datasets was possible, the common method of estimating load demand was consumer use surveys, where households were asked questions on anticipated appliance ownership and times of use (Boait, Advani, & Gammon, 2015; Boait et al., 2017; Mandelli, Merlo, & Colombo, 2016; Pandyaswargo et al., 2020; Sandwell et al., 2016). However, surveys have since been proven to be an inaccurate indicator of load demand, both in terms of magnitude and time of use, with errors up to four times greater than actual demand data (Allee, Williams, Davis, & Jaramillo, 2021; Blodgett, Dauenhauer, Louie, & Kickham, 2017; Gelchu, Ehnberg, & Ahlgren, 2023; Hartvigsson & Ahlgren, 2018).

A data-driven approach is established as more accurate in literature; however, the lack of available field data is a challenge (Bisaga, Puźniak-Holford, Grealish, Baker-Brian, & Parikh, 2017). In some cases, a limited amount of load demand data is available and this can be used alongside other techniques. Allee et al. (2021) combine a data-driven approach with survey results to predict the load demand of 1,378 Tanzanian mini-grid customers, demonstrating that the use of surveys can be beneficial when supplemented by measured data. Few et al. (2022) combine mini-grid load demand data with information on climatic conditions and household level characteristics, such as appliance ownership, to simulate household level load profiles. Both cases show that additional information, beyond stand-alone load-demand data, further increases load demand estimation capabilities. Yet, reliable load demand data is still the essential starting block.

Greater granularity of the load-demand data enables useful insights for system design and maintenance, but sometimes aggregated data is all that is available. Data can be aggregated on a household level (i.e. for multiple households) (Bhattacharyya & Palit, 2021) or a temporal level (i.e. multiple households for a single time period) (Heeten et al., 2017), but in both instances, the results are diminished compared to using raw granular data, and this can negatively impact design and maintenance choices (Jurasz, Guezgouz, Campana, & Kies, 2022).

Where ample data is available, clustering can be a useful tool for visualisation and understanding. In the case of load demand clustering, k-centred methods are a common choice due to simple implementation (Yilmaz, Chambers, & Patel, 2019). Some authors have made efforts to compare the effectiveness of different clustering methods (McLoughlin, Duffy, & Conlon, 2015; Rajabi et al., 2020; Wang et al., 2015), but different results are found in each case. This highlights the importance of having a clear goal for the clustering process, ensuring that the algorithms and distance metrics produce relevant clusters (Hennig, 2015; Von Luxburg, Williamson, & Guyon, 2012).

Dynamic time warping (DTW) is an elastic distance measure for comparing time series data that has a one-to-many mapping and is robust to shifts in time (Sakoe & Chiba, 1978). This allows a shape-based comparison, as proposed in Lin, Li, Tian, Fu, and Li (2019), without being dependent on dimensionality reduction. The benefits of using DTW over Euclidean distance in load profile clustering are beginning to be recognised in literature (Aghabozorgi, Seyed Shirkhorshidi, & Ying Wah, 2015; Ausmus et al., 2020; Begum, Ulanova, Wang, & Keogh, 2015; Ratanamahatana & Keogh, 2005). However, due to its quadratic computational complexity, it is not commonly implemented (Rajabi

et al., 2020). Teeraratkul, O'Neilly, and Lallz (2016) and Ausmus et al. (2020) clustered household load profiles using DTW with k-medoids and k-means respectively. Both found significant improvement compared to using Euclidean distance, and k-medoids has the advantage of using a time series from the dataset as the representative cluster centre, which can improve interpretability. However, use of k-centred methods in the cluster assignment stage makes the algorithm vulnerable to getting stuck in local optima because the iterative method to solve the k-medoids and k-means problem does not guarantee global optimality. Kumtepeli, Perriment, and Howey (2024) instead formulated the k-medoids problem as a mixed integer program (MIP), allowing global optimality within the cluster assignment while using DTW for the distance metric. MIP is beneficial for alleviating the risk of getting stuck in local optima, however using MIP on very large datasets (with respect to number of time series) is computationally expensive and infeasible.

Clustering load profiles from LMICs is much less common than load data for HICs in literature because less data is available. Williams, Jaramillo, Cornell, Lyons-Galante, and Wynn (2017) characterised load profiles from eleven mini-grids across Africa finding trends in consumption on daily and monthly time scales. However, no clustering was done, and household level data was not made available, so comprehension of household trends is not possible. Lorenzoni et al. (2020) conducted hierarchical clustering on the load demand data of 61 mini-grids in LMICs, finding clusters of aggregated household data. The key differentiation between the clusters was the peaks in load demand. The authors also found that as the systems aged, and the energy connection tier increased, the use profile became flatter. Lukuyu, Shiran, Kennedy, Urpelainen, and Taneja (2023) used k-means to cluster daily load demand profiles of SHS customers in East Africa over a year. However, they used a monthly mean to represent the daily load profile for each household, which loses the granularity of day-to-day variability and gives smoothed profiles for the results.

In addition to understanding day-to-day energy use, it is important to consider long-term behaviour, particularly in the context of energy access for previously unelectrified households. In these situations, behaviour is more unpredictable (Muhumuza, Zacharopoulos, Mondol, Smyth, & Pugsley, 2018; Riva, Tognollo, Gardumi, & Colombo, 2018). Therefore, long-term load-demand analysis is imperative to gain insights into how systems deployed in LMICs are used. It is commonly thought that energy demands will increase as energy access is obtained (Muhumuza et al., 2018; Opiyo, 2020; Riva et al., 2018), however, this conjecture needs to be supported by real-world data (Riva, Sanvito, Tonini, & Colombo, 2019). Bisaga and Parikh (2018) compared energy use of households across Rwanda, splitting them into groups depending on the length of time they had been using systems. They found that households that had systems for over a year generally used less energy each day than those who had obtained systems within the last 6 months, despite the more established households having generally more appliances. Kizilcec, Spataru, Lipani, and Parikh (2022) also conducted work on Rwandan SHS energy use, finding a reduction in electricity demand over the first year of use for the majority of households in their dataset. The decrease in electricity consumption was even more pronounced in households that owned a TV.

Beyond the first year of ownership, Dominguez, Orehounig, and Carmeliet (2021) found a general increase in monthly electricity expenditure after initial adoption, along with an increase in appliance ownership, but this plateaued at the 2–3 year mark. Conversely, Masselus et al. (2024) found that rural households in Rwanda with grid connections decreased their electricity consumption over the first ten years of connection. This analysis includes monthly electrical consumption data from 147,074 rural households, focusing on 174 households with survey data. The surveys found that household appliances largely consisted of lighting and entertainment devices, but economically productive appliances were a rarity. Louie, Atcitty, Terry, Lee, and Romine (2023) analysed the load demand of off-grid households in the Navajo
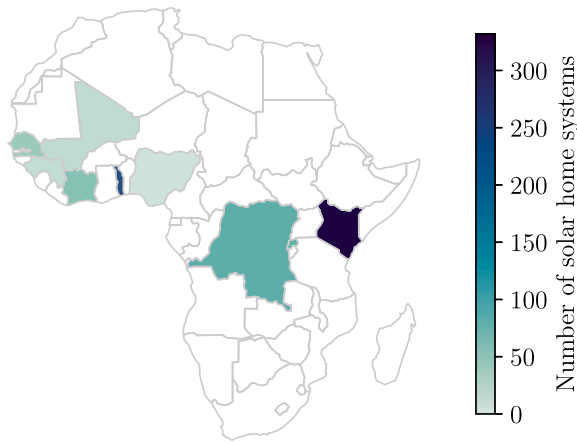
Fig. 1. Geographical distribution of the downselected SHSs used in the clustering process. The countries coloured white hosted no systems in our dataset.

Nation over 2 years and found a large variation both between and within households, the latter being impacted by seasonal changes. Additionally, energy use in the second year was on average 10% lower than in the first year, highlighting longer-term temporal load shifts.

The overall picture for long-term temporal changes in energy use remains unclear—increased consumption is anticipated but there are also datasets indicating the opposite. Further information is required to understand underlying causes. The barrier to rural electrification is often economic (Blimpo, Postepska, & Xu, 2020; Kizilcec & Parikh, 2020), but there is a research gap in understanding the link between diverse PAYG payment data (Mergulhão, Capra, Voglitsis, & Parikh, 2023) and energy use. Existing studies establish a link between energy use and payment, but focus mainly on payment behaviour impacts (Guajardo, 2019) and very coarse-grained energy consumption data (Lukuyu et al., 2023).

This work makes several contributions: firstly, a large-scale analysis of daily SHS load-demand profiles is undertaken, elucidating patterns in real-world data; secondly, progression of energy use over several years is considered, rather than observing a single snapshot in time; finally, payment and energy data are combined, enabling insight into the role of economic factors. This research also provides access to the measured data for others to use, supporting the collective efforts of researchers and industry to achieve SDG7.

Energy use dataset and analysis

The dataset analysed in this paper was provided by BBOXX, and consists of time series from 1,000 SHSs across Africa, each with a 50 W solar panel, a 12 V, 17 Ah lead–acid battery, and various DC loads, including lighting, phone charging, fans, radios and TVs. These systems were downselected from the wider BBOXX database using stratified sampling based on system lifetime so that a range of systems were represented (including new adopters and more established users). To enable longer-term analysis, all SHSs younger than 1 year were excluded, and the oldest system in the dataset is 1230 days old. The dataset covers SHSs installed between September 2018 and December 2020, with consumption data available up to January 2022. The geographical distribution of systems in sub-Saharan Africa is shown in Fig. 1, with most located in Kenya, Togo, the Democratic Republic of the Congo and Rwanda. The time for each system (normally stored as UTC) was corrected to local time in each country to ensure comparable solar and usage conditions.

The time, output current, and output voltage were used for load-demand analysis, with non-uniform sampling, i.e., measurements were recorded whenever a change above a certain threshold was detected.

This was at second resolution, with a 10-minute backstop recording taken if no change was detected within that time. The following method was used to convert these measurements to hourly load demands:

1. Calculate instantaneous power $P(t)$ at each time t from measured voltage $V(t)$ and load current $I_{\text{out}}(t)$,

$$P(t) = V(t) \times I_{\text{out}}(t). \tag{1}$$

2. Maintain the previous value until the next time stamp (i.e. zero-order hold).

3. Determine the energy consumed in one hour by integrating the power over hourly intervals, $\Delta t = 1$ h,

$$E(t) = \int_{t-\Delta t}^{t} P(\tau)d\tau. \tag{2}$$

4. For each household, construct a vector of hourly energy consumption values, E, where each element is the energy consumed in that hour (t_{end} represents the final time step in the time series data), i.e.,

$$E_k = E(k \cdot \Delta t) \quad \text{where} \quad k \in \left[1, \frac{t_{\text{end}}}{\Delta t}\right]. \tag{3}$$

For the 1,000 systems considered, there are a total of 647,021 daily load profiles. Since many of these are very similar, we downsampled a subset from the total set to improve computational efficiency. Stratified sampling was used, dividing the dataset into subgroups based on daily energy consumption. First, the distribution of daily energy consumption across the entire dataset was found, then time series were randomly selected so as to keep the same distribution of daily energy consumption in the sampled subset. The process was applied twice: Initially, for each household, ten representative days were selected, giving 10,000 time series in total. This ensures that every household and the overall spread of energy consumption are represented. Then, we applied the same method to reduce this set down to 2,000 time series in total. These steps together ensure we have both a representatively diverse dataset and the option to apply an algorithm to find a global optimum in the clustering process (see next section). After the cluster centres were found, each of the 647,021 daily load profiles were assigned to their nearest cluster according to DTW distance. The end result is that each household can then be represented by a chronological sequence of days, where each day is labelled with the cluster that is the closest match to the load profile of that day.

The BBOXX dataset also contains customer payment data that has previously been analysed extensively (Mergulhão et al., 2023). For this work, the relevant payment information is the time series of remaining credit for each SHS. The PAYG approach uses fixed daily rates, where customers purchase access to electricity for a given day, rather than access to a metered amount of energy in a certain time period. The contracts vary both by location and load: household energy limits are restricted based on the number of appliances owned, which relates to the customer tariff. To allow comparison between households, we convert the credit in local currency to the number of days of electricity use remaining. This removes the need for conversion between currencies and payment amounts. Once a household runs out of credit, the system enters a 'disabled' state where discharging is not permitted until further payment—this is deemed an economic outage (Ferrall, Callaway, & Kammen, 2022). The clustering exercise previously described ignored this condition (i.e., clustered systems may or may not be in this stage), but we later consider it when analysing the results.

The PAYG model charges a fixed daily rate determined by the number of appliances connected, which defines the household's tariff. Each tariff corresponds to a technical energy ceiling, set through adjusted cut-off voltages, and once activated for the day, customers

can use electricity freely up to this limit. Payment therefore governs whether electricity is available on a given day, not how much is used once a system is active. Financial improvement can increase consumption through additional appliances or tariff upgrades, whereas financial hardship will likely result in skipping payment days rather than reducing use on paid days.

We also calculated another payment metric, the utilisation rate, defined as the number of days a household is not in the economic outage state divided by the number of days the household has owned the system. For example, a household with a utilisation rate of 0.9 experienced economic outages on 10% of total days since SHS purchase.

Clustering method

This work uses our DTW-MIP clustering algorithm, fully described in Kumtepeli et al. (2024). A brief overview of the approach is given here for readability and completeness. The DTW distance (cost) between individual elements of two time series x and y of length n and m (indexed by i and j), respectively, is given by

$$c_{i,j} = (x_i - y_j)^2 + \min \begin{cases} c_{i-1,j-1} \\ c_{i-1,j} \\ c_{i,j-1} \end{cases} \quad (4)$$

where the final element $c_{n,m}$ is the total DTW distance between series x and y. This problem can be solved using dynamic programming. Here we chose not to use a warping window to restrict the parts of one time series that can be mapped to another (Sakoe & Chiba, 1978). This allows long-range similarities to be found between peaks in energy consumption, i.e., whether they occur at night or in the morning. Dynamic time warping already penalises warping because more warping means more pairwise comparisons that are embedded in the final cost. Therefore, the hard constraint of a window is not necessary.

The DTW distance between each time series is stored in a symmetric matrix $D_{p \times p}$, where p is the total number of time series. A binary square matrix $A_{p \times p}$ is used for MIP cluster assignment, where $A_{ij} = 1$ if time series j is in cluster with centroid i. There are k clusters (5b), and each time series must be in one cluster only (5c). Clusters can be found by solving the optimisation problem

$$\min_A \sum_i \sum_j D_{ij} \times A_{ij} \quad (5a)$$

subject to

$$\sum_{i=1}^{p} A_{ii} = k, \quad (5b)$$

$$\sum_{i=1}^{p} A_{ij} = 1 \qquad \forall j \in [1, p], \quad (5c)$$

$$A_{ij} \leq A_{ii} \qquad \forall i, j \in [1, p]. \quad (5d)$$

To choose the value of k, the algorithm was repeated for 2 to 8 clusters. We then compared the costs and silhouette scores (Shahapure & Nicholas, 2020) to select a final value $k = 5$.

Results and discussion

The results of this study are divided into three sections. The first presents representative daily load demand profiles identified through clustering. The second focuses on long-term changes in household energy use, first exploring consumption trends and then long-term trends in cluster allocation over the entire dataset. Finally, we consider links between payment data and electricity use.

Daily clustering

Fig. 2 depicts the centroids of the five clusters resulting from the DTW-MIP clustering exercise, and heatmaps for each cluster. The centroids are the individual time series at the median of each cluster. The clusters are archetypally named by the dominant load profile characteristic within the cluster; however, they are not exclusive categories, and there is some overlap in usage levels and times between clusters. It is also helpful to examine all the time series within each cluster to gain further insight into discrepancies between and within clusters. The heatmaps present each time series in each cluster as a row, with lighter blues representing low energy consumption and darker blues indicating high consumption. Overall, the results confirm the general agreement between the cluster representatives and their associated time series.

Table 1 shows the allocation of time series across clusters within the subset, along with the results from the extrapolation to the whole dataset. The validity of the extrapolation is generally acceptable, with a larger proportion of time series in the *economic outage* subset of *low use*, but there is a reasonable consistency between the full dataset and subset.

Table 2 shows characteristics of the time series within each cluster, analysed on the whole dataset, including the peaks and information on appliances and time of energy use. A peak within a time series is defined as having double the amplitude of the mean energy consumption of the time series, at least two hours apart from another distinct peak (to reduce the risk of fluctuations across the threshold counting as two peaks). Additionally, to prevent profiles with low consumption returning many peaks, a peak must have a minimum amplitude of 10 W. The appliance power reflects the sum of the rated power of all appliances owned by a household, but these are not necessarily switched on or drawing power simultaneously or continuously.

The *low use* cluster stands out as distinct from the other clusters, and signals that either there is an economic outage, or there is virtually no usage despite payment. To differentiate between these two scenarios, a new subset was created, classifying *low use* time series as *economic outage* if the credit balance was also zero. It is notable that 24% of days in the dataset belong to *economic outage* and *low use* combined. Of these, 75% (i.e., 18%/24%) correspond to economic outages, but there is still a significant proportion of days where the systems are paid for, but not used.

The *moderate use* cluster is large and has the next lowest mean daily consumption. Typical moderate use (heatmap, Fig. 2) is characterised by consistent low use throughout the day, with more energy consumed in the evening and a slight increase in morning use. The evening use spans between 15:00 and 21:00 and therefore crosses sunset at around 18:00, but drops to almost no consumption after midnight. The peak demand is fairly low and few households with days in this cluster have TVs, which is reflected in the low mean appliance power. Alternatively, households with TVs and higher power appliances are not using all them on days when they are in this cluster.

Nighttime use is another large cluster, with a slightly larger mean daily consumption than *moderate use*, however there is significant overlap in the daily energy consumption between these two clusters. For days in *nighttime use* the proportion of households owning TVs is low, much like for *moderate use*, and there is a low appliance power of 55 W. The characteristic that distinguishes a time series from being in *moderate use* vs. *nighttime use*, is that *nighttime use* profiles consume most of their energy overnight between 18:00 and 06:00, whereas in the *moderate use* case electricity is consumed in the late afternoon and evening.

The final two clusters, *double peak* and *single peak*, represent higher consumption (households are likely to have a TV), with the double peak case slightly higher than the single. And, of course, the number of peaks is different between the two clusters—as seen in Table 2 and Fig. 2. A final distinction between these two higher consumption clusters is that
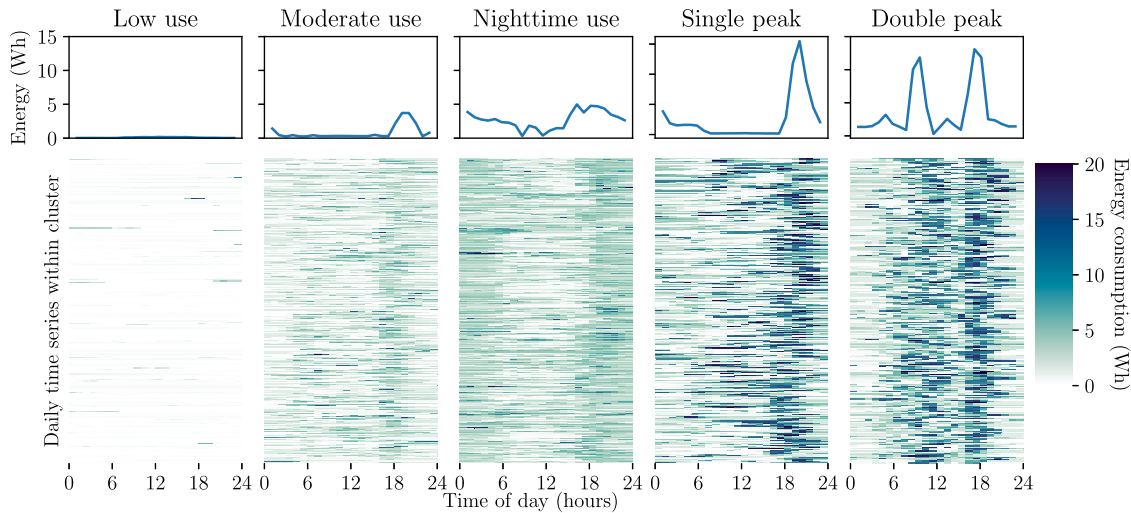
Fig. 2. The 5 representative cluster centres from the DTW-MIP clustering of daily SHS load demand profiles, with the heatmaps for all time series within each cluster below. Each heatmap row represents a 24-hour load demand within that cluster. The clusters can be archetypally named: *low use*, *moderate use*, *nighttime use*, *single peak* and *double peak*.

Table 1

Comparison of clustering results from subset vs. extrapolation to full dataset.

Cluster	Clustering subset			Full dataset		
	Count	Proportion of subset	Mean (Wh/day)	Count	Proportion of dataset	Mean (Wh/day)
Economic outage	–	–	–	118 258	0.18	0
Low use	259	0.13	4	39 040	0.06	12
Moderate use	583	0.29	42	154 765	0.24	39
Nighttime use	573	0.29	69	181 696	0.28	73
Single peak	339	0.17	102	86 362	0.14	102
Double peak	246	0.12	110	66 900	0.10	122

Table 2

Mean characteristics of all clustered time series.

Cluster	Daily peaks	TV ownership (%)	Weekday use (%)	Daytime use (%)	Appliance power (W)
Economic outage	0	49	72	N/A	73
Low use	0	48	72	50	71
Moderate use	0.2	17	72	50	50
Nighttime use	0.1	22	71	41	55
Single peak	1.1	84	71	46	96
Double peak	1.9	92	71	59	97
Total population	0.5	43	71	48	68

there is a higher proportion of daytime use in *double peak*, caused by the additional morning peak alongside the evening peak.

Finally, we include the proportion of days in each cluster that are weekdays (Table 2), showing that there is no significant difference between clusters—in all cases the weekday proportion is around 72%, i.e., 5/7. This differs from HICs where there is often a noticeable difference between weekday and weekend consumption (Trotta, 2020).

In summary, there is significant heterogeneity in household energy use despite the identical technical capabilities of each system (e.g., same sized battery and solar panel). The causes of this might be financial or due to the diverse energy needs of different households on different days, caused by diversity in activities. There could be a case for supplying different intensity users with differently sized systems.

Long-term general consumption trend

Fig. 3 shows an initial increase in the mean daily consumption for all households, peaking on the 96th day. This is likely to be a 'settling in period' where households also purchase appliances and become familiar with the system. Over the first six months, daily electricity consumption increases for 58% of customers, but by the end of the first

year, only 43% of customers continue to increase consumption. Beyond the first year, 77% of users are decreasing daily consumption. This decrease is particularly prominent in the first 400 days, and then diverges depending on household energy use. The mean energy consumption over two years shows a 33% reduction from its peak, dropping from 66 Wh/day to 44 Wh/day. This is contrary to prior literature, where the expectation is that consumption will continue to increase after a system is acquired, since users will obtain more appliances and find increasingly energy-intensive activities to engage with Dominguez et al. (2021). Instead, our results show that individual SHS users follow a long-term consumption trend similar to that witnessed for individual grid-connected users (Masselus et al., 2024; Mugyenyi et al., 2025).

Mean consumption in high-use households begins to plateau beyond 400 days. When economic outages are removed the decrease in consumption is slightly stronger, especially beyond 700 days. Another important observation is that when economic outages are ignored, the consumption in high-use households continues to increase until around 250 days, i.e., the apparent decrease in the mean consumption of high-use households after the peak around day 96 is actually caused by a high proportion of economic outages.

Table 3
Household mean characteristics separated by dominant cluster—where they spend the majority of time.

Dominant cluster	Homogeneity (%)	Utilisation rate (%)	TV ownership (%)	Flexible appliances	Appliance power (W)
Economic outage	51	48	62	1.4	82
Low use	42	84	53	1.1	72
Moderate use	56	89	8	0.7	44
Nighttime use	61	90	19	0.7	57
Single peak	46	93	91	1.9	102
Double peak	48	92	98	1.8	99
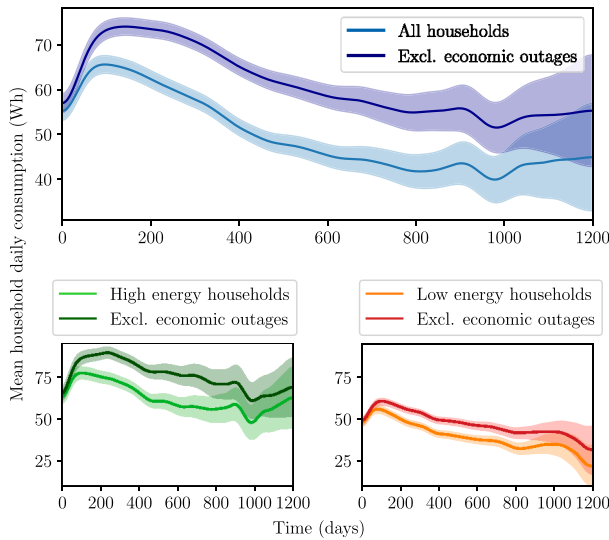
Fig. 3. Top: Mean daily household electricity consumption initially increases, but then decreases. The confidence interval becomes wider over time as the number of households in the sample decreases. Bottom: High-use households (> 50 W of appliances), left, vs. low-use (≤ 50 W of appliances), right.

Fig. 4. Spread of usage over clusters: for each household with a certain dominant cluster (rows), the proportion of time spent in other clusters is indicated by the columns.
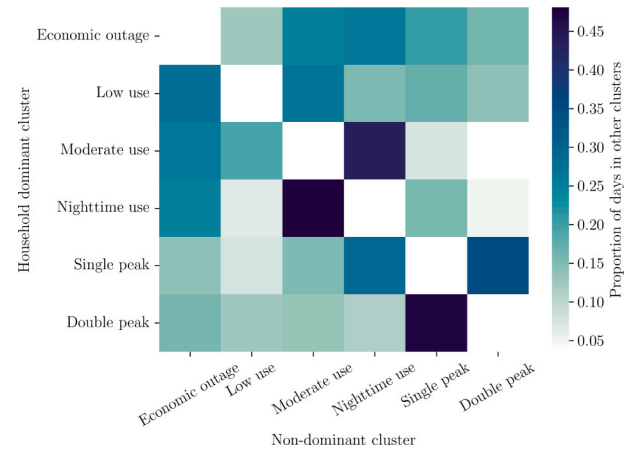
In low-use households, the mean relative decrease by the end of the first year is even greater (24%) than in high-use households (12%). The decrease continues, although at a lower rate, until around 800 days. The difference between the mean consumption with and without economic outages in low-use households grows quickly in the first part of ownership and then gradually increases over time.

A gradual decrease in electricity use can lead to oversized systems. This is undesirable because the full solar generation potential is not realised (Bhatti & Williams, 2021; Soltowski, Campos-Gaona, Strachan, & Anaya-Lara, 2019), batteries are not cycled optimally, shortening lifetime (Narayan et al., 2019; Perriment, Kumtepeli, McCulloch, & Howey, 2023), and customers pay for a larger system than needed, increasing their electricity cost. This is particularly detrimental because the cost of energy is a significant barrier to electrification (Kizilcec & Parikh, 2020; Kyere, Dongying, Bampoe, Kumah, & Asante, 2024).

Cluster allocation trend

We now analyse long-term trends in cluster allocation for individual households over time. An interesting finding is the lack of homogeneity within households, where homogeneity is defined as the fraction of days that a household's demand is in its dominant cluster (the cluster the household spends the most time in). Table 3 shows that the mean homogeneity for all households is only 55%.

Higher consumption clusters (*single peak* and *double peak*) exhibit lower homogeneity. Table 3 also shows that households with days in higher consumption clusters have greater appliance power, with the majority owning a TV. Therefore, households with days in higher consumption dominant clusters have the potential to be in all clusters, however, the majority of households with other dominant clusters

do not consume larger amounts of energy due to lack of appliances. Additionally, in the higher consumption clusters, households have more flexible loads. Base households have lights and phone chargers and these are relatively inflexible because lighting is needed when it is dark and phone charging when batteries are depleted. Here, we define anything other than lighting and phone charging to be a flexible load. In the higher consumption clusters, households have appliances like TVs and radios which they may not use every day, leading to more frequent day-to-day changes in electricity consumption.

Fig. 4 depicts how households, split into their dominant clusters, consume energy when not following the dominant cluster behaviour. This further demonstrates that households dominated by higher consumption can also have usage days in all other clusters, whereas it is very uncommon for *moderate use* dominated households to have days in the *single peak* and *double peak* clusters.

Economic outage is observed in all households, but is particularly prevalent in the three lower consumption dominated cases (*low use*, *moderate use* and *nighttime use*). This is logical, because customers who initially pay less have access to smaller amounts of available energy, and fewer appliances, and may be more likely to face financial hardship or be less engaged with their system—both of which lead to missed payments and economic outages.

Fig. 4 also shows that clusters with similar average consumption are more closely linked within households (e.g., *moderate use* and *nighttime use*, similarly *single peak* and *double peak*). Therefore, while household energy consumption is clearly non-uniform, it is more common for households to vary their consumption behaviour within similar magnitudes of use or to not use the system at all and not pay.

Fig. 5 shows the proportion of households in each cluster on each day. The most important finding is the sharp increase in *economic outage* over the first year. This is the primary explanation for reduced electricity consumption, as previously seen in Fig. 3, and shows that customers either cannot afford to keep up with payments or are choosing not to spend their money on SHS credit. Once *economic outage* begins to plateau, *low use* starts to gradually increase. The *low use* cluster
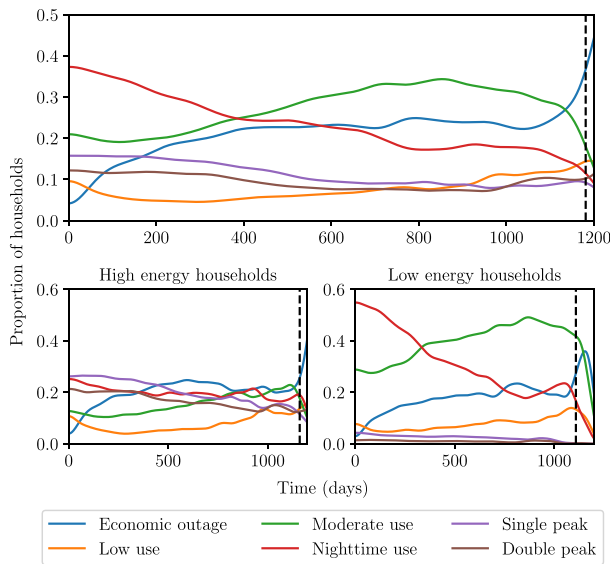
Fig. 5. Top: Proportion of households in each cluster, every day. Bottom: High-use households (>50 W of appliances), left, and low-use (≤50 W of appliances), right. Black dashed line indicates when number of households drops below 10.

Fig. 6. Proportion of households in each cluster over time, split into households with a utilisation rate over 90% and under 90% at time of analysis. The black dashed line indicates when the number of households in the sample goes below 10.

shows when households are choosing to use the system minimally regardless of economic considerations. The increase therefore shows disengagement with the system for non-economic reasons.

Previously, Fig. 4, a link between *nighttime use* and *moderate use* was observed. This is reinforced in Fig. 5, especially in the low-use subplot, where *nighttime use* significantly reduces over the time of analysis, while *moderate use* increases. We hypothesise that this is the driver for the decreasing use in low-use households shown in Fig. 3. Given their already low consumption, a shift from *nighttime use* to *moderate use* has a substantial impact. For instance, households might initially use lighting throughout the night for security purposes, but over time, reduce this to evening hours only. To fully understand the reasons behind this behavioural change, further research including surveys is necessary.

The high energy household subplot in Fig. 5 shows a similar trend of consistently decreasing *single peak* and *double peak* with increase in *moderate use* and *low use*. Thus, households reduce peak consumption over time and reduce the use of flexible loads.

Utilisation rate and electricity use

Fig. 5 illustrates a rise in *economic outage* over time. However, a deeper understanding of household behaviour can be gained by examining utilisation rates. Fig. 6 presents the time variation in the proportion of households within each cluster segmented by utilisation rate. There are 530 households in the dataset with a utilisation rate of 0.9 or above, but only 121 households with a utilisation rate of less than 0.5.

The observed increase in *economic outage* is primarily driven by households with lower utilisation rates, which, by definition, is less evident in households with higher utilisation rates. In the low utilisation rate group, *nighttime use* significantly reduces, particularly in the first year; *low use* and *moderate use* fluctuate but stay relatively constant, whereas *single peak* and *double peak* steadily decrease. In low utilisation households, there is not a significant increase in any cluster aside from *economic outage*, so these households are clearly changing their energy behaviour for financial reasons—either an affordability issue or a lack of desire to spend money on the system.

For households with higher utilisation rate, the proportion of *nighttime use* continually decreases but at a steadier rate than in lower
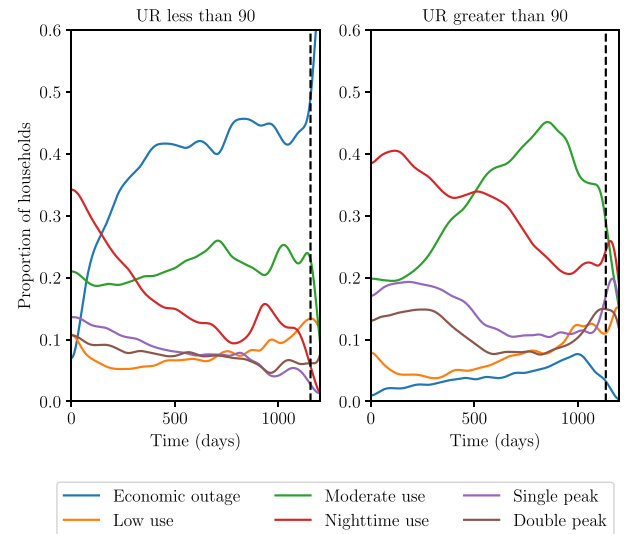
utilisation households. Higher utilisation rate households are more likely to be in *single peak* and *double peak* clusters than lower utilisation rate households. The mean proportion of households in these two high-consumption clusters across the whole time period is 19% for low utilisation rate households and 32% for high utilisation households. Thus, households that pay more consistently are more likely to be in the higher use clusters. However, beyond the first year, *single peak* and *double peak* rates decrease, while *low use* and *moderate use* increase. This demonstrates how households are reducing their electricity consumption even when *economic outage* is negligible, which further builds on the previously stated hypothesis that households decrease electricity consumption even while still paying for and engaging with the SHS.

Overall, splitting the households by utilisation rates highlights several key issues:

- Energy consumption patterns vary with utilisation rate.
- *Economic outage* increases significantly over the first year of ownership in low utilisation households. In high utilisation households, the small proportion of outages increases gradually over time, peaking at the 3-year mark.
- There is a shift to lower consumption clusters over time and it is more pronounced in households with utilisation rate over 90%.
- Higher consumption clusters are more common in higher utilisation rate households.

As described in Section "Energy use dataset and analysis", PAYG payments determine daily access to energy, rather than giving access to a certain amount of energy. Therefore, financial hardship affects the number of days that a product is activated in a certain time period rather than the energy consumption on paid days. Behavioural changes in electricity use are likely to be the primary drivers of the decreasing consumption.

Conclusions

This work has analysed a large dataset of energy use in 1,000 SHSs across sub-Saharan Africa. Time series clustering identified five distinct daily usage profiles, archetypally defined as *low use, moderate use, nighttime use, single peak* and *double peak*, in order of increasing mean energy use. The representation of these across the dataset shows
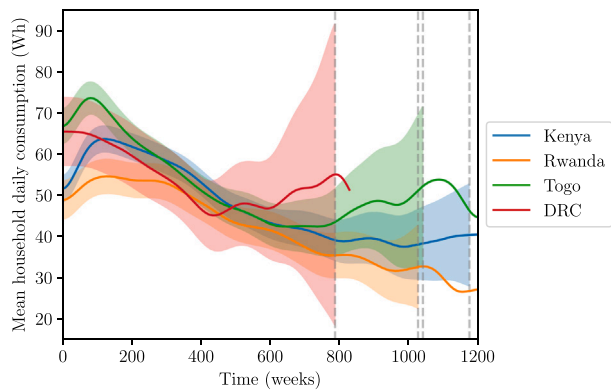
Fig. A.1. Mean daily energy consumption by country, with confidence intervals. The vertical lines at the ends of the confidence intervals indicate when there are fewer than 10 households for that country.

that systems are only fully utilised, i.e., their energy usage is in the *single peak* or *double peak* cluster, on 24% of days—the same proportion of days that the systems are in the *economic outage* and *low use* clusters. The majority of time, the systems provide basic supply of *moderate use* and *nighttime use* electricity. Households frequently change, from day to day, how they use their systems, with cluster homogeneity for households averaging only 55%.

Analysis of trends in household behaviour over a period of up to 1,200 days revealed a general decrease in energy consumption among both high and low consumption households. While economic outages are not uncommon under the PAYG scheme, they are not the sole reason for decreasing energy consumption, indicating that changes in energy behaviour extend beyond financial considerations. The decrease in energy consumption when not in the economic outage state is characterised by a decrease in *nighttime* energy use, and reduction of days with large evening peaks. Utilisation rates are slightly higher for households that predominantly exhibit *single peak* and *double peak* profiles, but even these households experience economic outages.

Overall, the systems are often underused and therefore oversized, leading to knock-on effects on costs for households and suppliers. The significant proportion of economic outages suggests that costs remain a limitation for many households, warranting further investigation into how to maximise energy provision from each SHS in order to reduce the overall cost of energy. There is a need for SHS designs and pricing structures to better align with the actual energy requirements of households. Adaptable solutions may be key for long term adoption, and this may include downsizing systems over time in some cases.

Further work should include surveys of households to understand the underlying causes of the non-financial decrease in energy consumption. Moreover, continuous study of energy behaviour as households gain access to electricity is essential for better understanding their needs and how these evolve over time.

CRediT authorship contribution statement

R. Perriment: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **V. Mergulhão:** Writing – review & editing, Investigation, Formal analysis, Data curation. **V. Kumtepeli:** Writing – review & editing, Supervision, Software, Methodology. **P. Parikh:** Writing – review & editing, Supervision. **M.D. McCulloch:** Writing – review & editing, Supervision, Conceptualization. **D.A. Howey:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Rebecca Perriment reports financial support from BBOXX and the Engineering and Physical Sciences Research Council. Vasco Mergulhao reports financial support from BBOXX. Volkan Kumtepeli reports financial support from the Engineering and Physical Sciences Research Council. Priti Parikh reports financial support from the Royal Academy of Engineering. David Howey is a co-founder of Brill Power. Malcolm McCulloch declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank BBOXX for project funding and access to SHS data, and EPSRC for project funding (ref. EP/W027321/1). Prof. Parikh is grateful to the Royal Academy of Engineering, United Kingdom for part funding.

Appendix A. Country level consumption trends

The observed decline in energy consumption occurred across several countries in the dataset. Fig. A.1 shows the consumption trends per country for the four countries with the largest number of households in the dataset. There are minor differences in the magnitude of mean daily consumption by country, as well as some differences in rate of decline. However the overall decreasing consumption is consistent across the countries, particularly in the first 600 days. This indicates that, in the rural off-grid context, the decreasing electricity consumption trend is a common feature rather than being driven by country-specific economic or contextual factors. Beyond the first 600 days, the mean daily energy consumption in Rwanda and Kenya continues gradually decreasing, whereas in DRC and Togo it plateaus or slightly increases. However, in the latter cases, the uncertainty is larger due to greater variance and fewer samples.

Data availability

The dataset associated with this paper is available at doi: 10.5287/ora-6r4vapx6b.

References

Aghabozorgi, S., Seyed Shirkhorshidi, A., & Ying Wah, T. (2015). Time-series clustering - A decade review. *Information Systems*, *53*, 16–38. http://dx.doi.org/10.1016/j.is.2015.04.007.

Allee, A., Williams, N. J., Davis, A., & Jaramillo, P. (2021). Predicting initial electricity demand in off-grid Tanzanian communities using customer survey data and machine learning models. *Energy for Sustainable Development*, *62*, 56–66. http://dx.doi.org/10.1016/j.esd.2021.03.008.

Almeshqab, F., & Ustun, T. S. (2019). Lessons learned from rural electrification initiatives in developing countries: Insights for technical, social, financial and public policy aspects. In *Renewable and sustainable energy reviews*: Vol. *102*, (pp. 35–53). Elsevier Ltd, http://dx.doi.org/10.1016/j.rser.2018.11.035.

Ausmus, J. R., Sen, P. K., Wu, T., Adhikari, U., Zhang, Y., & Krishnan, V. (2020). Improving the accuracy of clustering electric utility net load data using dynamic time warping. In *Proceedings of the IEEE power engineering society transmission and distribution conference*: Vol. *2020-October*, Institute of Electrical and Electronics Engineers Inc., http://dx.doi.org/10.1109/TD39804.2020.9299915.

Begum, N., Ulanova, L., Wang, J., & Keogh, E. (2015). Accelerating dynamic time warping clustering with a novel admissible pruning strategy. In *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining*: Vol. *2015-August*, (pp. 49–58). Association for Computing Machinery, http://dx.doi.org/10.1145/2783258.2783286.

Bhattacharyya, S. C., & Palit, D. (2021). A critical review of literature on the nexus between central grid and off-grid solutions for expanding access to electricity in Sub-Saharan Africa and South Asia. In *Renewable and sustainable energy reviews*: Vol. *141*, Elsevier Ltd, http://dx.doi.org/10.1016/j.rser.2021.110792.

Bhatti, S., & Williams, A. (2021). Estimation of surplus energy in off-grid solar home systems. *Renewable Energy and Environmental Sustainability*, 6, 25. http://dx.doi.org/10.1051/rees/2021024.

Bisaga, I., & Parikh, P. (2018). To climb or not to climb? Investigating energy use behaviour among Solar Home System adopters through energy ladder and social practice lens. *Energy Research and Social Science*, 44, 293–303. http://dx.doi.org/10.1016/j.erss.2018.05.019.

Bisaga, I., Puźniak-Holford, N., Grealish, A., Baker-Brian, C., & Parikh, P. (2017). Scalable off-grid energy services enabled by IoT: A case study of BBOXX SMART Solar. *Energy Policy*, 109, 199–207. http://dx.doi.org/10.1016/j.enpol.2017.07.004.

Blimpo, M. P., Postepska, A., & Xu, Y. (2020). Why is household electricity uptake low in Sub-Saharan Africa? *World Development*, 133, http://dx.doi.org/10.1016/j.worlddev.2020.105002.

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. http://dx.doi.org/10.1016/j.esd.2017.08.002.

Boait, P., Advani, V., & Gammon, R. (2015). Estimation of demand diversity and daily demand profile for off-grid electrification in developing countries. *Energy for Sustainable Development*, 29, 135–141. http://dx.doi.org/10.1016/j.esd.2015.10.009.

Boait, P., Gammon, R., Advani, V., Wade, N., Greenwood, D., & Davison, P. (2017). ESCoBox: A set of tools for mini-grid sustainability in the developing world. *Sustainability (Switzerland)*, 9(5), http://dx.doi.org/10.3390/su9050738.

Den Heeten, T., Narayan, N., Diehl, J. C., Verschelling, J., Silvester, S., Popovic-Gerber, J., et al. (2017). Understanding the present and the future electricity needs: Consequences for design of future Solar Home Systems for off-grid rural electrification. In *Proceedings of the 25th conference on the domestic use of energy* (pp. 8–15). Institute of Electrical and Electronics Engineers Inc., http://dx.doi.org/10.23919/DUE.2017.7931816.

Dominguez, C., Orehounig, K., & Carmeliet, J. (2021). Understanding the path towards a clean energy transition and post-electrification patterns of rural households. *Energy for Sustainable Development*, 61, 46–64. http://dx.doi.org/10.1016/j.esd.2021.01.002.

Ferrall, I., Callaway, D., & Kammen, D. M. (2022). Measuring the reliability of SDG 7: The reasons, timing, and fairness of outage distribution for household electricity access solutions. *Environmental Research Communications*, 4(5), http://dx.doi.org/10.1088/2515-7620/ac6939.

Few, S., Barton, J., Sandwell, P., Mori, R., Kulkarni, P., Thomson, M., et al. (2022). Electricity demand in populations gaining access: Impact of rurality and climatic conditions, and implications for microgrid design. *Energy for Sustainable Development*, 66, 151–164. http://dx.doi.org/10.1016/j.esd.2021.11.008.

Gelchu, M. A., Ehnberg, J., & Ahlgren, E. O. (2023). *Comparison of electricity load estimation methods in rural mini-grids: Case study in Ethiopia* (pp. 1–5). IEEE, http://dx.doi.org/10.1109/powerafrica57932.2023.10363276.

Guajardo, J. A. (2019). How do usage and payment behavior interact in rent-to-own business models? Evidence from developing economies. *Production and Operations Management*, 28(11), 2808–2822. http://dx.doi.org/10.1111/poms.13067.

Guterres, A. (2025). *The sustainable development goals report 2025*. United Nations publication, issued by the Department of Economic and Social Affairs, URL https://unstats.un.org/sdgs/report/2025/.

Hartvigsson, E., & Ahlgren, E. O. (2018). Comparison of load profiles in a mini-grid: Assessment of performance metrics using measured and interview-based data. *Energy for Sustainable Development*, 43, 186–195. http://dx.doi.org/10.1016/j.esd.2018.01.009.

Hennig, C. (2015). What are the true clusters? *Pattern Recognition Letters*, 64, 53–62. http://dx.doi.org/10.1016/j.patrec.2015.04.009.

Irena (2022). Renewable energy market analysis: Africa and its regions. URL www.afdb.org.

Jurasz, J., Guezgouz, M., Campana, P. E., & Kies, A. (2022). On the impact of load profile data on the optimization results of off-grid energy systems. *Renewable and Sustainable Energy Reviews*, 159, http://dx.doi.org/10.1016/j.rser.2022.112199.

Kizilcec, V., & Parikh, P. (2020). Solar home systems: A comprehensive literature review for Sub-Saharan Africa. In *Energy for sustainable development*: *Vol. 58*, (pp. 78–89). Elsevier B.V., http://dx.doi.org/10.1016/j.esd.2020.07.010.

Kizilcec, V., Spataru, C., Lipani, A., & Parikh, P. (2022). Forecasting solar home system customers' electricity usage with a 3D convolutional neural network to improve energy access. *Energies*, 15(3), http://dx.doi.org/10.3390/en15030857.

Kumtepeli, V., Perriment, R., & Howey, D. A. (2024). DTW-C++: Fast dynamic time warping and clustering of time series data. *Journal of Open Source Software*, 9(101), 6881. http://dx.doi.org/10.21105/joss.06881.

Kyere, F., Dongying, S., Bampoe, G. D., Kumah, N. Y. G., & Asante, D. (2024). Decoding the shift: Assessing household energy transition and unravelling the reasons for resistance or adoption of solar photovoltaic. *Technological Forecasting and Social Change*, 198, http://dx.doi.org/10.1016/j.techfore.2023.123030.

Lin, S., Li, F., Tian, E., Fu, Y., & Li, D. (2019). Clustering load profiles for demand response applications. *IEEE Transactions on Smart Grid*, 10(2), 1599–1607. http://dx.doi.org/10.1109/TSG.2017.2773573.

Lorenzoni, L., Cherubini, P., Fioriti, D., Poli, D., Micangeli, A., & Giglioli, R. (2020). Classification and modeling of load profiles of isolated mini-grids in developing countries: A data-driven approach. *Energy for Sustainable Development*, 59, 208–225. http://dx.doi.org/10.1016/j.esd.2020.10.001.

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. http://dx.doi.org/10.1016/j.esd.2023.02.012.

Lukuyu, J., Shiran, M., Kennedy, R., Urpelainen, J., & Taneja, J. (2023). Purchasing power: Examining customer profiles and patterns for decentralized electricity systems in East Africa. *Energy Policy*, 172, http://dx.doi.org/10.1016/j.enpol.2022.113331.

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. http://dx.doi.org/10.1016/j.esd.2016.01.005.

Masselus, L., Ankel-Peters, J., Gonzalez Sutil, G., Modi, V., Mugyenyi, J., Munyehirwe, A., et al. (2024). *10 years after: Long-term adoption of electricity in rural Rwanda*: *Working paper* http://dx.doi.org/10.4419/96973261.

McLoughlin, F., Duffy, A., & Conlon, M. (2015). A clustering approach to domestic electricity load profile characterisation using smart metering data. *Applied Energy*, 141, 190–199. http://dx.doi.org/10.1016/j.apenergy.2014.12.039.

Mergulhão, V. P., Capra, L., Voglitsis, K., & Parikh, P. (2023). How do they pay as they go?: Learning payment patterns from solar home system users data in Rwanda and Kenya. *Energy for Sustainable Development*, 76, http://dx.doi.org/10.1016/j.esd.2023.101295.

Mugyenyi, J., Muhwezi, B., Fobi, S., Massa, C., Taneja, J., Williams, N. J., et al. (2025). Post-connection electricity demand and pricing in newly electrified households: Insights from a large-scale dataset in Rwanda. *Energy Policy*, 198, http://dx.doi.org/10.1016/j.enpol.2024.114449.

Muhumuza, R., Zacharopoulos, A., Mondol, J. D., Smyth, M., & Pugsley, A. (2018). Energy consumption levels and technical approaches for supporting development of alternative energy technologies for rural sectors of developing countries. In *Renewable and sustainable energy reviews*: *Vol. 97*, (pp. 90–102). Elsevier Ltd, http://dx.doi.org/10.1016/j.rser.2018.08.021.

Narayan, N., Chamseddine, A., Vega-Garita, V., Qin, Z., Popovic-Gerber, J., Bauer, P., et al. (2019). Exploring the boundaries of Solar Home Systems (SHS) for off-grid electrification: Optimal SHS sizing for the multi-tier framework for household electricity access. *Applied Energy*, 240, 907–917. http://dx.doi.org/10.1016/j.apenergy.2019.02.053.

Narayan, N., Vega-Garita, V., Qin, Z., Popovic-Gerber, J., Bauer, P., & Zeman, M. (2020). The long road to universal electrification: A critical look at present pathways and challenges. *Energies*, 13(3), http://dx.doi.org/10.3390/en13030508.

Opiyo, N. N. (2020). How basic access to electricity stimulates temporally increasing load demands by households in rural developing communities. *Energy for Sustainable Development*, 59, 97–106. http://dx.doi.org/10.1016/j.esd.2020.09.006.

Pandyaswargo, A. H., Ruan, M., Htwe, E., Hiratsuka, M., Wibowo, A. D., Nagai, Y., et al. (2020). Estimating the energy demand and growth in off-grid villages: Case studies from Myanmar, Indonesia, and Laos. *Energies*, 13(20), http://dx.doi.org/10.3390/en13205313.

Perriment, R., Kumtepeli, V., McCulloch, M., & Howey, D. (2023). *Lead-acid battery lifetime extension in solar home systems under different operating conditions* (pp. 1–5). IEEE, http://dx.doi.org/10.1109/powerafrica57932.2023.10363281.

Rajabi, A., Eskandari, M., Ghadi, M. J., Li, L., Zhang, J., & Siano, P. (2020). A comparative study of clustering techniques for electrical load pattern segmentation. *Renewable and Sustainable Energy Reviews*, 120, http://dx.doi.org/10.1016/j.rser.2019.109628.

Ratanamahatana, C. A., & Keogh, E. (2005). Three myths about dynamic time warping data mining. In *Proceedings of the 2005 SIAM international conference on data mining*. http://dx.doi.org/10.1137/1.9781611972757.50.

Riva, F., Sanvito, F. D., Tonini, F., & Colombo, E. (2019). Modelling long-term electricity load demand for rural electrification planning. In *2019 IEEE Milan PowerTech* (pp. 1–6). IEEE, http://dx.doi.org/10.1109/PTC.2019.8810727.

Riva, F., Tognollo, A., Gardumi, F., & Colombo, E. (2018). Long-term energy planning and demand forecast in remote areas of developing countries: Classification of case studies and insights from a modelling perspective. In *Energy strategy reviews*: *Vol. 20*, (pp. 71–89). Elsevier Ltd, http://dx.doi.org/10.1016/j.esr.2018.02.006.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(1), 43–49. http://dx.doi.org/10.1109/TASSP.1978.1163055.

Sandwell, P., Chambon, C., Saraogi, A., Chabenat, A., Mazur, M., Ekins-Daukes, N., et al. (2016). Analysis of energy access and impact of modern energy sources in unelectrified villages in Uttar Pradesh. *Energy for Sustainable Development*, 35, 67–79. http://dx.doi.org/10.1016/j.esd.2016.09.002.

Shahapure, K. R., & Nicholas, C. (2020). Cluster quality analysis using silhouette score. In *Proceedings - 2020 IEEE 7th international conference on data science and advanced analytics* (pp. 747–748). Institute of Electrical and Electronics Engineers Inc., http://dx.doi.org/10.1109/DSAA49011.2020.00096.

Soltowski, B., Campos-Gaona, D., Strachan, S., & Anaya-Lara, O. (2019). Bottom-up electrification introducing new smart grids architecture-concept based on feasibility studies conducted in Rwanda. *Energies*, 12(12), 2439. http://dx.doi.org/10.3390/en12122439.

Teeraratkul, T., O'Neilly, D., & Lallz, S. (2016). Condensed representation and individual prediction of consumer demand. In *2016 4th IEEE international conference on smart energy grid engineering* (pp. 11–16). Institute of Electrical and Electronics Engineers Inc., http://dx.doi.org/10.1109/SEGE.2016.7589192.

Trotta, G. (2020). An empirical analysis of domestic electricity load profiles: Who consumes how much and when? *Applied Energy*, *275*, http://dx.doi.org/10.1016/j.apenergy.2020.115399.

Von Luxburg, U., Williamson, R. C., & Guyon, I. (2012). Clustering: Science or art?. In *Proceedings of machine learning research*, *Proceedings of ICML workshop on unsupervised and transfer learning* (pp. 65–79). Bellevue, Washington, USA: URL https://proceedings.mlr.press/v27/luxburg12a.html.

Wang, Y., Chen, Q., Kang, C., Zhang, M., Wang, K., & Zhao, Y. (2015). Load profiling and its application to demand response: A review. *Tsinghua Science and Technology*, *20*(2), 117–129. http://dx.doi.org/10.1109/TST.2015.7085625.

Williams, N., Jaramillo, P., Cornell, B., Lyons-Galante, I., & Wynn, E. (2017). Load characteristics of east african microgrids. In *IEEE PES-IAS PowerAfrica*. IEEE, http://dx.doi.org/10.1109/PowerAfrica.2017.7991230.

Yilmaz, S., Chambers, J., & Patel, M. K. (2019). Comparison of clustering approaches for domestic electricity load profile characterisation - Implications for demand side management. *Energy*, *180*, 665–677. http://dx.doi.org/10.1016/j.energy.2019.05.124.