Electricity consumption and load profile segmentation analysis for rural microgrid customers in Tanzania

Nathaniel J. Williams and Paulina Jaramillo
Department of Engineering and Public Policy
Carnegie Mellon University
Pittsburgh, PA USA

Kieran Campbell, Brian Musanga, Isaiah Lyons-Galante
Grid Research and Analytics
PowerGen Renewable Energy
Nairobi, Kenya

***Abstract*- Understanding electricity consumption behavior of different groups of customers is important to ensure appropriate microgrid designs. This paper assesses segmentation among microgrid customers in Tanzania according to electricity consumption behavior. Using the k-means clustering algorithm, 821 customers are placed into clusters by mean daily electricity consumption and mean normalized load profiles. We divided customers into three distinct segments by daily consumption level: < 140, 140-450, and > 450Wh. We also found five distinct load profile clusters. Clusters with high daytime electricity use tend to be commercial customers that consume large amounts of electricity. Households exhibit large evening peaks and consume relatively small amounts of electricity. Four of five load profiles show distinct evening peaks in consumption and one shows high night time electricity use. Microgrid developers can use these clusters and customer segments either to optimize the load profile for a power system by mixing certain customer types, or optimize the power system design to a given customer base.**

I. Introduction

Lack of access to electricity is a barrier to social and economic development for communities throughout East Africa. The rural electrification rate in Tanzania, for example, is only 17% [1]. At a country level, low electrification rates limit social and economic development that could lift millions out of poverty. Off-grid microgrids, small electricity networks that are able to operate autonomously, have emerged as a promising solution for addressing this energy access shoftfall in the near term. Microgrids can be deployed quickly and have a lower cost than extension of the main grid to these remote areas [2]-[8]. If or when the main grid arrives, microgrids can interconnect to it, creating a decentralized network that can aggregate loads and generation capacity, while maintaining the ability to operate as isolated systems when necessary. A number of African countries have recognized the value of off-grid solutions like microgrids, incorporating them into their national electrification plans. Tanzania is one such country that has implemented policies and programs supportive of the microgrid sector [9].

As a leading microgrid developer and operator, PowerGen Renewable Energy has built hundreds of power systems throughout East Africa since 2011, including more than 40 solar-powered microgrids in Kenya and Tanzania over the past four years. In doing so, they have collected a large quantity of data on their customers, both before their homes have been electrified through in-person surveys, and afterwards through data collected with smart meters. Understanding the profile and behaviors of microgrid customers is important for a number of reasons. Because microgrids in this context are typically small-scale and serve low-income customers, it is critical for project economics that microgrid developers appropriately size the system to meet expected demand—oversize and there is excess capital spend, undersize and there will need to be a costly system expansion later on [10], [11]. Additionally, gaining a better understanding of its microgrid customers allows microgrid operators to provide high quality and tailored customer service. Of particular interest is how customer characteristics translate into or predict consumption levels, consumption patterns and, thus expenditures on power.

This paper seeks to identify customer segmentation based on aggregate electricity consumption and temporal consumption patterns using clustering algorithms on mean customer daily electricity consumption and normalized load profiles. Using these clusters, we then examine the characteristics of customers that tend to fall into each of these catogories. This information will aid microgrid developers in estimating customer-level and aggregate electricity consumption patterns on their systems prior to construction. This work builds on previous work examining aggregate site level load characteristics on microgrids in East Africa [12]. Similar load profile clustering work in the microgrid context has been found in media reports [13] but such prior work did not delve into the characteristics of customers in each cluster or how load profiles relate to aggregate electricity consumption. Similar techniques to the ones we use have been applied in developed country contexts [14]-[16].

II. Data Collection Methods

At each site developed, PowerGen collects surveys from each potential customer to identify which to connect to the microgrid based on their location and characteristics. For those customers that are then connected to power, data is then collected on their behavior, including payments made and energy consumed hourly. Together these datasets can provide

978-1-5386-4163-7/18/$31.00 ©2018 IEEE

insights into how customer classifications (e.g., home, bar, restaurant, etc.) relate to consumption levels (average load) and patterns (load profile). These insights can inform how microgrid developers select sites and customers, design power systems, and then manage customer relationships once the site is operational.

A. Customer Application Surveys

After a site is selected for microgrid installation, the surveying team returns to site to collect customer applications for microgrid connections. The goal of customer application process is to identify customers to connect to the microgrid and inform the design of the reticulation network to support those customers. Key information collected in the customer application includes GPS coordinates, customer demographics, current electricity use, and potential electricity use. The surveys are conducted in person with a mobile phone application, and the results are then downloaded and aggregated into a customer database.

Understanding the composition of the customer base is imperative to success as a customer-oriented microgrid utility. The customers are broken down into four classifications: Homes, Businesses, Home & Business, and Public Premises. Homes, which are residential users, make up approximately two-thirds of the customer base in Tanzania, while the other three categories make up the remaining third. Fig. 1 shows the distribution of 832 customers across these categories.

Each customer category has a different profile in terms of consumption and expenditure. Home customers use about half as much power per day as the typical business, ~90 Wh vs. ~180 Wh. Customers at the low end of the Home category use power on a limited basis for lighting or phone charging only, while those higher up use a range of appliances including small TVs, subwoofers, fridges, etc. Most customers in the Home category are self-employed in agriculture as subsistence or small scale commercial farmers, as seen in Fig. 2. As a result, their income fluctuates seasonally with the harvest cycles.

Customers in the "Business" or "Home & Business" categories also use power for basic uses such as lighting but can offer a range of services that consume larger amounts of power. For example, the largest Business customers are serving the community with entertainment (e.g., a video hall or bar/hotel), which requires multiple TVs, speakers, and freezers. Other Business customers operate productive load equipment such as maize millers and welders. These customers have an outsized impact on total system consumption and consequently, on overall revenue collections. Fig. 3 shows the distribution of business types found in this sample of microgrids in Tanzania. Decisions about which customers to connect are ultimately made on the basis of connection cost and expected electricity demand.

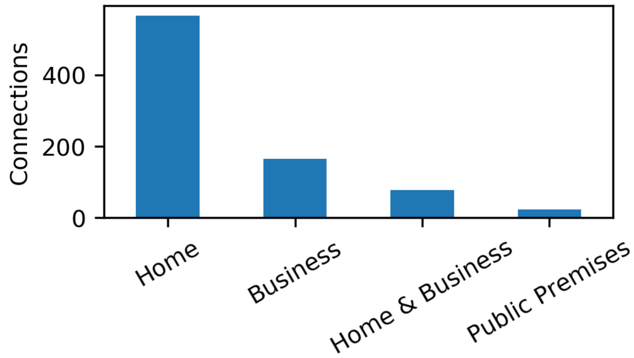

Fig. 1. Distribution of PowerGen Tanzania customer types

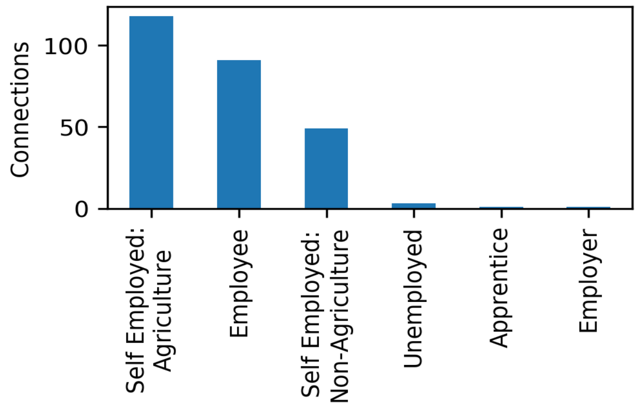

Fig. 2. Distribution of PowerGen Tanzania customers by employment

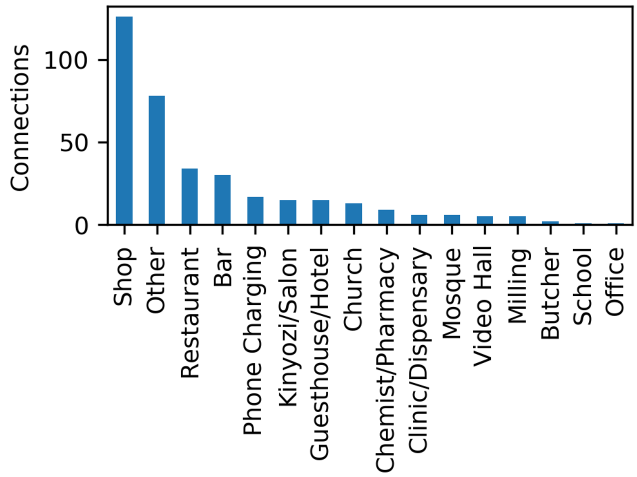

Fig. 3. Distribution of PowerGen Tanzania customers by business type

B. Electricity Consumption Data

PowerGen uses smart meter technology in its microgrids to collect energy payments from customers and disburse electricity for their consumption. "Pay-as-you-go"-integrated smart meters allow for the grid operator to manage customer connections automatically according to each customer's payment balance. Customers can make payments via mobile money or through a local agent, and their power will be switched on automatically. If a customer consumes their entire energy credit without topping up again, their meter will automatically switch off. This functionality is ubiquitous among solar home system (SHS) companies.

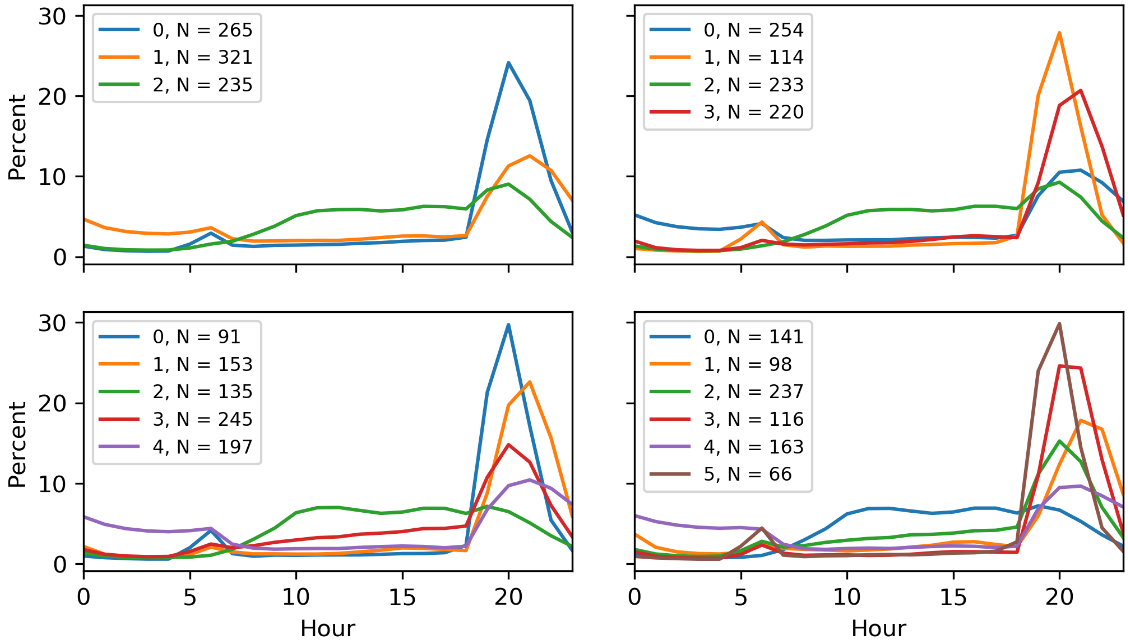

Fig. 4. Load profile cluster centroids for different numbers of clusters

In addition to recording payments, the smart meters collect granular electricity consumption data for each customer. Each microgrid contains a central server in the power system that aggregates the hourly usage values from smart meters throughout the system. These consumption data are then transmitted to a server. These data can then be monitored in real time to remotely troubleshoot issues, and analyzed over longer periods to garner insights into customer consumption behaviors.

III. Analysis Methods

This paper examines the energy consumption patterns of microgrid customers using the data described in the previous section. We break consumption behavior into two components: quantity consumed and diurnal consumption patterns. Using the k-means clustering algorithm [17], we categorize customers by their mean load profile (normalized to mean daily electricity consumption) and by their mean daily electricity consumption.

A. Load profile cluster analysis

Normalized mean load profiles for each meter were constructed by taking the mean electricity consumption for each hour of the day over all days for which the customer has a complete set of data (24 hourly values). Profiles are then normalized by dividing each hourly mean by the mean daily electricity consumption so that the integral of each profile is one. Prior to the computing load profiles, certain data points and customers were filtered out. Gaps in data often indicate that communication between the central server and the metering system was lost. When this occurs, all consumption between the time the connection was lost and reconnection is aggregated into the next data transmission. This means that the hourly values after these gaps do not represent actual consumption during that time step. These points are therefore filtered out. Furthermore, in order to ensure that profiles are representative of average consumption, customers with less than 30 full days of consumption data are not included in this analysis. A small number of customers that have more than 30 days of data show zero electricity consumption in the vast majority of days for which records exist, resulting in load profiles with large spikes at the few periods where they did use energy. These outliers distort the clustering analysis that was performed and, therefore, customers that show no electricity use in more than 25% of days for which data are available are excluded from the cluster analysis. After filtering, there were 821 customers included in the analysis.

In order to identify typical consumption patterns among customers, the normalized mean load profiles, consisting of 24 hourly values, were clustered using the k-means algorithm, implemented using the Python package, Scikit-Learn [18]. The algorithm was run on several numbers of cluster as described in the results section.

B. Mean daily consumption clustering

A further cluster analysis was performed by grouping customers by mean daily electricity consumption. The data were cleaned in the same way as the load profile data. Only days with complete hourly records were summed and included in the calculation of daily mean consumption. The k-means algorithm is then run on this single feature. By grouping customers into both load profile and mean daily consumption, we characterize customers both by magnitude and pattern of consumption. Using the demographic data collected in the customer application surveys, we then characterize the types of customers that fall into each load profile and consumption category.

IV. Results

A. Load profile cluster analysis results

Fig. 4 shows the load profile cluster centroids resutling from running the k-means algorithm on different numbers of clusters. Determining the 'correct' number of clusters to use is not always straightforward. One method is to look for an elbow in the plot of the k-means score versus the number of clusters. The k-means score is the sum of the squared distance from each observation to its cluster centroid. As the number of clusters increase, the score will naturally decrease; however, it is expected that the rate of decrease will drop once the 'correct' number of clusters is reached. Fig. 5 plots the k-means score against the number of clusters. The curve appears to level off between four and six clusters. The upper left plot in Fig. 4 shows the cluster centroids created using three clusters. All of the profiles have evening peaks with cluster zero having a small morning peak, cluster one having high night-time use and cluster two showing strong electricity use during the day. Adding a fourth cluster results in a cluster with a strong evening peak and a very small morning peak. With a fifth cluster, a cluster emerges with strong daytime use and no distinct evening peak. With six clusters, the additional cluster appears to differentiate between the size and precise timing of the evening peak for the clusters but no new clusters that are significantly qualitatively different. We therefore settle on five clusters as describing the distinct patterns in electricity consumption profiles among customers. Fig. 6 plots the final load profile centroids for the five clusters and Table I summarizes the qualitative characteristics of these five clusters.

TABLE I
DESCRIPTION OF LOAD PROFILES

Cluster	Description
0	Large evening peak with a morning peak
1	Large evening peak with small morning bump
2	Large daytime use with no evening peak
3	Growing use during day with evening peak
4	Evening peak with continued night time use

B. Mean daily consumption cluster analysis results

To complement the load profile cluster analysis, we performed a cluster analysis on mean daily consumption. Fig. 7 shows the distribution of mean daily consumption across customers split into three, four and five clusters presented as a swarm plot as a function of log consumption. The width of the swarm is an indication of the density of customers at each consumption level. Fig. 7 doesn't show any distinct clusters for mean daily consumption. Customers seem to fall along a continum with the majority of customers consuming between 10 and 100 Wh per day. The divisions between clusters for mean daily consumption will therefore be somewhat arbitrary, but the clustering method provides a quantitative method for establishing the cutoffs. With three clusters for mean daily consumption, only three customers fall in the highest use cluster. Adding a fourth cluster pulls higher consumption users from the low use cluster into an intermediate category and a few customers from the mid consumption cluster. The fifth cluster for mean daily consumption further segments the higher consumption users. We decided to use the five cluster groupings for mean daily consumption, but combined the three highest clusters into a 'high' use cluster. This results in a cutoff of 140 Wh/day between low and medium use customers and 450 Wh/day between medium and high consumption

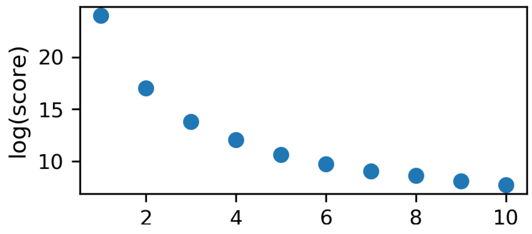

Fig. 5. k-means score as a function of number of clusters

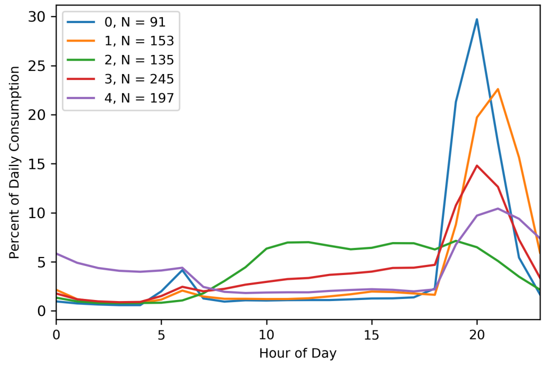

Fig. 6. Load profile cluster centroids with five clusters

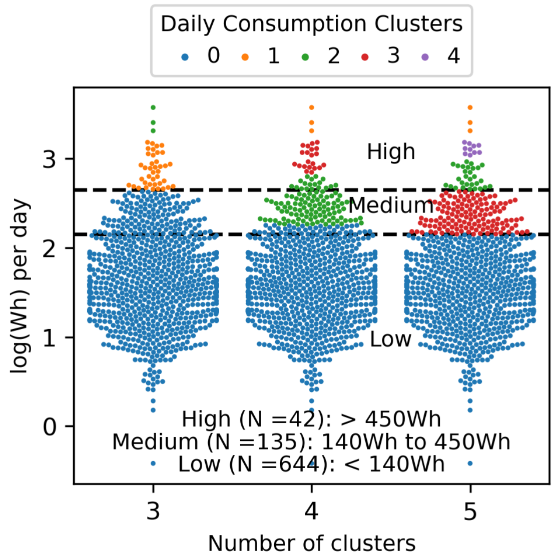

Fig. 7. Clustering of customers by mean daily consumption by customers

customers. In total, 42 customers fall into the high consumption category, 135 customers in the medium consumption category, and 644 customers in the low use category.

C. *Cluster demographics*

Using these load profile and consumption profiles, we can now describe the characteristics of customers that fall into the intersection of these clusters. Fig. 8 classifies customers by connection type (Business, Home, Home & Business and Public Premises) in each load profile/consumption cluster pair. High consumption customers fall primarily in load profile clusters two and three, both of which show higher daytime electricity use than the other clusters. High use customers also tend to fall in the 'Business' and 'Home & Business' categories. Medium use customers also fall frequently into load profile clusters two and three, as well as cluster four with high night time use. Customers in load profile clusters zero and one, which consume most of their energy during the evening peak, tend to be homes with low energy consumption. The low consumption customers are spread more evenly across load profile clusters and are primarily homes. There are a large number of homes in cluster four (evening peak with night time use) with low consumption. This may be due to security lights that operate overnight and, as a result of low overall consumption, make up a large portion of total energy use.

Fig. 9 breaks down business customers into specific business types. Only business types with a sample size greater than ten have been included in the figure. The stacked bar charts represent the percentage of businesses falling into each load profile cluster for each business type. Each bar is labeled by the number of businesses represented by the bar. Shops and bars dominate the high consumption category but are also among the most common businesses on the microgrids. The load profile cluster seems to depend more strongly on the level of consumption than the type of business. With the exception of cluster zero, each load profile type is present for at least one customer in each business type category. Shops are the most common type of business and there is no dominant load profile type. This may be because the category is generic and likely contains shops that are diverse in nature. The next most common business types are bars and restaurants. Bars are well represented in load profile clusters two and three which show stronger electricity use throughout the day. Hair salons/Kinyozis also fall frequently into these clusters. Restaurants are distributed across a number of load profile types. Average load profiles for several business types and homes are shown in Fig. 10. These profiles were constructed by averaging the normalized mean load profiles for all customers in each category. Bars show steadily increasing consumption through the day followed by an evening peak. Hair salons show strong day time use with a moderate evening peak. All profiles show clear peaks in the evening hours except for churches which have an irregular multimodal profile. Closer examination of individual church load profiles show a

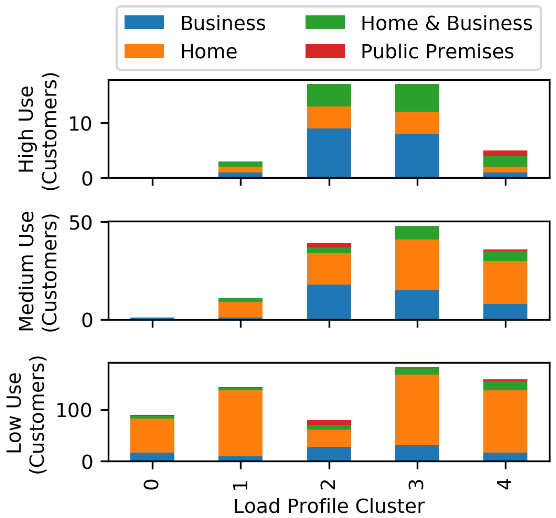

Fig. 8. Distribution of customers by type across load profile and consumption clusters

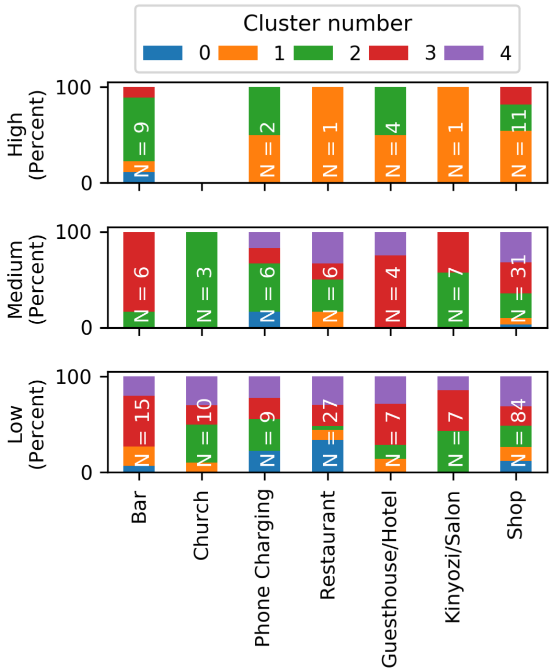

Fig. 9. Distribution of customers by load profile cluster across consumption cluster and business type

wide variety of irregular load profiles, often with mutliple peaks throughout the day.

V. Conclusions

This paper has identified five distinct load profiles across 821 microgrid customers in Tanzania. These customers were further segmented by level of consumption into low, medium, and high demand categories. Four of five low profile clusters show strong evening peaks. The exception, profile two, has strong daytime electricity use, continuing through the early evening, and is most prevalent among medium and high consumption business customers. Customers with strong

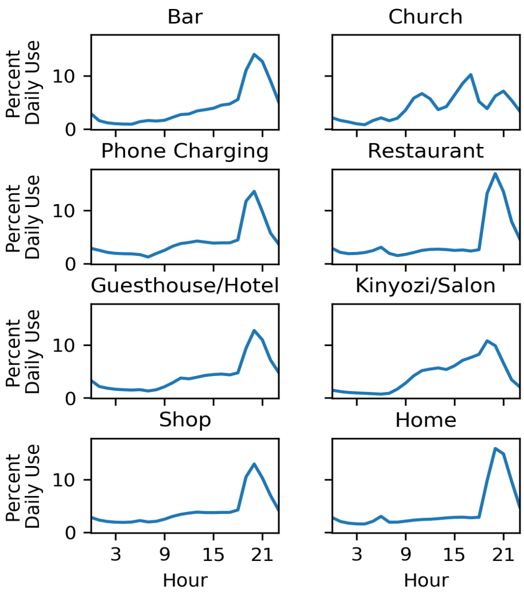

Fig. 10. Mean normalized load profiles for different customer

daytime use are valuable for a system because they tend to be high consumers and use power during the off-peak hours corresponding to solar generation, which helps to spread electricity consumption throughout the day and reduce storage requirements. Profile three also shows stronger day time energy use with a smaller evening peak relative to profiles zero, one, and four. Businesses that often fall in this category include bars, kinyozi/salons, phone charging stations, and shops. Households are generally low consumption users with strong evening peaks and low daytime electricity use.

Predicting customer electricity consumption and load profiles prior to microgrid construction is important for system design and assessing project financial feasibility. Conversely, it is also important for developers to identify customers that will lead to a financially sustainable system. This work is a first step towards assisting microgrid developers in better forecasting customer consumptive behavior. Future work will apply more sophisticated classification models to classify potential new customers by load profile according to their demographic profile. Furthermore, by better understanding the characteristics of customers with different consumption patterns, microgrid operators can design interventions to help graduate low consumption customers into higher use categories.

Acknowledgment

The authors gratefully acknowledge the support of PowerGen Renewable Energy and their willingness to share the data that made this work possible. The authors are also grateful for the financial support provided by USAID Power Africa and the National Renewable Energy Laboratory.

References

[1] International Energy Agency, "Energy Access Database." International Energy Agency, 2017.

[2] N. U. Blum, R. S. Wakeling, and T. S. Schmidt, "Rural electrification through village grids—assessing the cost competitiveness of isolated renewable energy technologies in Indonesia," *Renewable and Sustainable Energy Reviews*, vol. 22, pp. 482–496, Mar. 2013.

[3] S. Szabó, K. Bódis, T. Huld, and M. Moner-Girona, "Energy solutions in rural Africa: mapping electrification costs of distributed solar and diesel generation versus grid extension," *Environ. Res. Lett.*, vol. 6, p. 034002, Jul. 2011.

[4] M. Moner-Girona, K. Bódis, B. Korgo, T. Huld, I. Kougias, I. Pinedo-Pascua, F. Monforti-Ferrario, and S. Szabó, "Mapping the least-cost option for rural electrification in Burkina Faso," European Commission Joint Research Centre, 2017.

[5] D. Mentis, M. Welsch, F. F. Nerini, O. Broad, M. I. Howells, M. Bazilian, and H. Rogner, "A GIS-based approach for electrification planning—A case study on Nigeria," *Energy for Sustainable Development*, vol. 29, no. C, pp. 142–150, Dec. 2015.

[6] N. Moksnes, A. Korkovelos, D. Mentis, and M. I. Howells, "Electrification pathways for Kenya–linking spatial electrification analysis and medium to long term energy planning," *Environ. Res. Lett.*, vol. 12, no. 9, pp. 095008–14, Sep. 2017.

[7] D. Mentis, M. Andersson, M. I. Howells, H. Rogner, S. Siyal, O. Broad, A. Korkovelos, and M. Bazilian, "The benefits of geospatial planning in energy access - A case study on Ethiopia," *Applied Geography*, vol. 72, no. C, pp. 1–13, Jul. 2016.

[8] F. F. Nerini, O. Broad, D. Mentis, M. Welsch, M. Bazilian, and M. I. Howells, "A cost comparison of technology approaches for improving access to electricity services," *Energy*, vol. 95, no. C, pp. 255–265, Jan. 2016.

[9] L. A. Odarno, E. Sawe, M. Swai, M. J. J. Katyega, and A. Lee, "Accelerating Mini-Grid Deployment in Sub-Saharan Africa," World Resoures Institute, Oct. 2017.

[10] N. J. Williams, P. Jaramillo, and J. Taneja, "An investment risk assessment of microgrid utilities for rural electrification using the stochastic techno-economic microgrid model: A case study in Rwanda," *Energy for Sustainable Development*, vol. 42, no. C, pp. 87–96, Feb. 2018.

[11] N. J. Williams, P. Jaramillo, and J. Taneja, "PV-array Sizing in Hybrid Diesel/PV/Battery Microgrids under Uncertainty," IEEE PES Power Africa Conference, Livingstone, Zambia, 2016, pp. 189–193.

[12] N. J. Williams, P. Jaramillo, B. Cornell, I. Lyons-Galante, and E. Wynn, "Load Characteristics of East African Microgrids," IEEE PES Power Africa Conference, Accra, Ghana, 2017, pp. 236–241.

[13] J. Deign, "What Smart Meters Tell Us About Rural Microgrid Use in Emerging Markets," *Green Tech Media*, 15-Sep-2017. [Online]. Available: https://www.greentechmedia.com/articles/read/what-smart-meters-are-telling-us-about-microgrid-use. [Accessed: 28-Jan-2018].

[14] G. J. Tsekouras, P. B. Kotoulas, C. D. Tsirekis, E. N. Dialynas, and N. D. Hatziargyriou, "A pattern recognition methodology for evaluation of load profiles and typical days of large electricity customers," *Electric Power Systems Research*, vol. 78, no. 9, pp. 1494–1510, Sep. 2008.

[15] T. Räsänen, D. Voukantsis, H. Niska, K. Karatzas, and M. Kolehmainen, "Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data," *Applied Energy*, vol. 87, no. 11, pp. 3538–3545, Nov. 2010.

[16] G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping," *Energy*, vol. 42, no. 1, pp. 68–80, Jun. 2012.

[17] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. Berlin: Springer, 2008.

[18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, and M. Perrot, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, Oct. 2011.