Energy Policy 123 (2018) 569–578

Contents lists available at ScienceDirect

Energy Policy

journal homepage: www.elsevier.com/locate/enpol

A longitudinal study of electricity consumption growth in Kenya

Simone Fobi[a], Varun Deshpande[b], Samson Ondiek[c], Vijay Modi[a], Jay Taneja[b,*]

[a] QSEL Lab – Department of Mechanical Engineering, Columbia University, New York, United States
[b] STIMA Lab – Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, United States
[c] Business Strategy, Kenya Power, Nairobi, Kenya

ARTICLE INFO

Keywords:
Electricity
Consumption
Growth
Urbanization
Kenya

ABSTRACT

During the past 5 years, electrification in Kenya has grown by more than 30% due primarily to increases in grid penetration and solar home systems. This represents a way forward for governments, international finance institutions, and entrepreneurs to address some of the challenges of energy access. However, little is understood about how consumption has evolved among these newly-electrified customers. In this paper, we address this by conducting a longitudinal analysis for 136k utility customers across Kenya over six years of electricity bills, uncovering critical trends in spatio-temporal evolution of electricity consumption. Our analysis reveals that recently-electrified customers are reaching their steady-state consumption more quickly than previous customers, that the steady-state is increasingly less, and that typical urban and peri-urban customers tend to consume 50% more electricity than rural customers. In addition we present implications for policymakers and electricity planners considering grid extension and distributed systems for improving electrification.

1. Introduction

Developing countries regularly make critical decisions on how to allocate precious public-sector resources to increase electricity access, often with little evidence. Governments, finance institutions, and entrepreneurs are exploring new pathways for electrification such as solar home systems and mini-grids, as well as redoubling investments in traditional grid extension, all in an effort to build sustainable institutions for delivering electricity services.

Grid extension efforts in Kenya have led to an up-tick in the percentage of population that has access to electricity at home; however, a less well-understood change is the evolution of consumption among these newly-electrified customers. Projecting future electricity consumption is difficult, underscored by the observation that projections tend to understate growth in electricity demand in the developing world (Wolfram et al., 2012). Plausible electrification strategies depend on analyzing existing customer data to predict the behavior of newly-connected customers.

Kenya is an example of a country that has vastly expanded its electrification – from 2010 to 2015, grid penetration has increased by 27%, more than doubling the number of customers on the centralized grid – see Fig. 1 (Kenya, 2016). In addition to the centralized grid, there are now upwards of 600,000 solar home systems deployed, which contribute another 5–6% in electrification (estimated using census figures (Kenya National Bureau of Statistics, 2009) and current population estimates (AfriPop, 2010)). Most of the grid connections from 2010 to 2016 were residential, and nationwide residential electricity consumption has increased at roughly 9% annually over the period. Despite these large gains, little is understood about how much electricity these new customers consume, and even less is known about how their consumption will change with time. This study seeks to address this question: *how much electricity do newly-connected electricity customers use, and how will that consumption evolve?*

To that end, we present a longitudinal study of electricity consumption growth in Kenya. This study is built upon a dataset of billing records from Kenya Power, the sole distribution utility in Kenya. The dataset includes monthly billing records over a six-year period, from 2010 through 2015, for a random sample from Kenya Power's customer database at the end of 2015. After cleaning and meta-data verification, the random sample amounts to roughly 136k residential customers. The scale and extent of the longitudinal dataset is heretofore unseen in the literature on electricity consumption for an African country. Further description of this dataset is provided in Section 3. To identify which customers in our randomly-sampled dataset are rural, we developed an algorithm for determining which areas of the country are urban, peri-urban, and rural based on a constrained clustering method – we describe this method and its relevance in Section 4 and Appendix A. Subsequently we show results for urban and rural consumption, where

* Corresponding author.
E-mail address: jtaneja@umass.edu (J. Taneja).

https://doi.org/10.1016/j.enpol.2018.08.065
Received 21 November 2017; Received in revised form 11 April 2018; Accepted 30 August 2018
Available online 01 October 2018

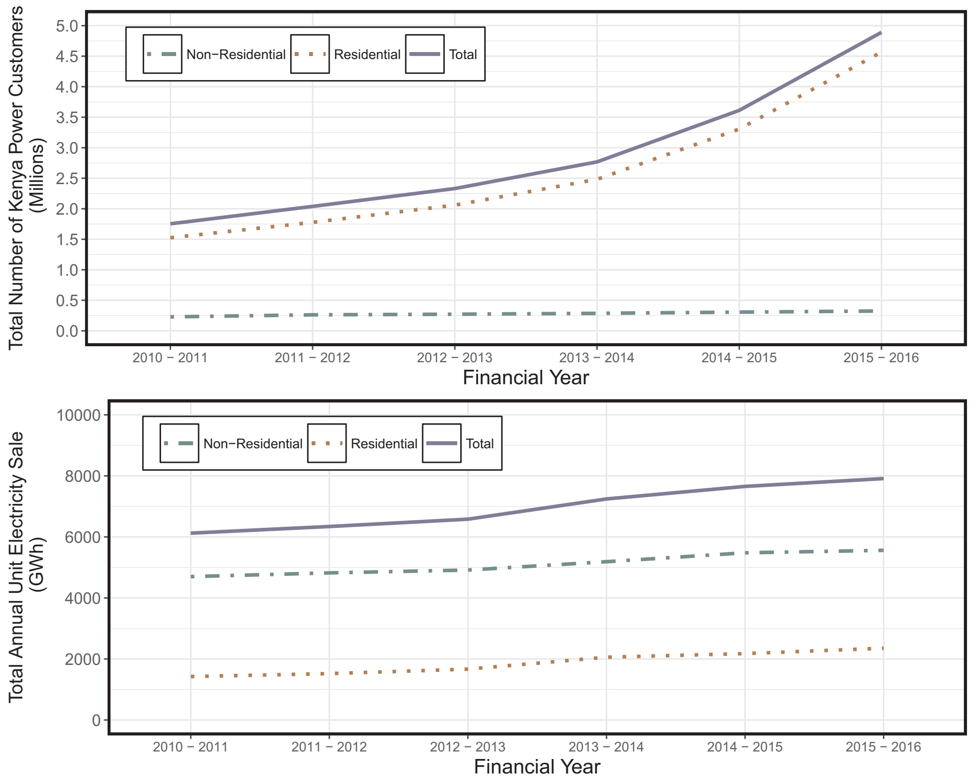


Fig. 1. Total number of customers and total electricity sales for Kenya Power between 2010 and 2016. Non-residential includes industrial, commercial, street lighting, and off-peak loads. Customer additions were mainly to the residential sector. Data are from Kenya Power annual reports (Kenya, 2016).

the urban results are a straightforward combination of both urban and peri-urban customers. In Section 5, we use the results of this method as well as other customer meta-data in order to segment our sample of customers and identify patterns of consumption growth among various groups. We conclude with implications of this study for policymakers and electricity planners, discussion of the limitations of our work, and next steps for research in the area.

2. Related work

Accurate electricity consumption estimates are important in designing electrical generation and delivery systems and meeting reliability requirements. A study in Malawi (Louie and Dauenhauer, 2016) uses off-grid data from 7 PV and battery systems to show the impact of incorrect load estimation on system cost and reliability. They found that system cost scaled proportionally with errors in consumption estimates, where over estimation led to significant increases in system cost of between USD 1.82 to USD 6.02 per watt-hour, while underestimating consumption eroded system reliability. This dichotomy between system cost and reliability emphasizes the need for data-driven approaches to understanding and predicting consumption, which can in turn yield more optimal system design.

In the case of residential electricity consumption, predictions are typically made by using multiple variables including socio-economic characteristics, appliance ownership, and living conditions. A literature review on the topic suggests that at least 62 variables potentially affect residential electricity usage (Jones et al., 2015). Other authors conclude that some important explanatory variables for household electricity consumption include appliance ownership, electricity tariffs, available income, and number of residents in the household (Villareal et al., 2016; Mensah et al., 2016; Esmaeilimoakher et al., 2016). While these analyses offer a deep-dive into electricity consumption patterns, they depend on expensive and time-consuming household surveys, rendering them difficult to scale with similar resolution to larger areas such as countries or regions.

Spatio-temporal analysis can provide insights to electricity consumption over large areas. Socio-economic and demographic variables such as population and income levels can be folded into such methods when studying electricity demand. For example, Amarala et al. (2005), Xie and Weng (2016), and Elvidge et al. (1997) demonstrate spatio-temporal analyses using satellite imagery to study population and energy dynamics in various regions. Results from these papers show a relationship between spatial dynamics, electricity consumption, and population. To explore the differences in electricity consumption due to urbanization, Xie and Weng (2016) use a pixel-based method to delineate urban, suburban and rural regions in China. A universal definition for urban regions was difficult to obtain and the Chinese administrative units “prefectural city” are a mix of both urban districts and rural counties. The authors use population adjusted nighttime lights to delineate urban areas. Land cover was then used to determine the optimal nighttime lights threshold for highly dense built-up areas in China. The obtained highly dense regions are labeled as the urban core while the difference between urban regions and urban core gives the suburban region. This definition of urbanization allows them to study differences in electricity patterns by urbanization levels.

Chévez et al. (2017) propose another approach for obtaining spatially homogeneous areas using k-means clustering algorithm. In this case, rather than using urban, suburban and rural as homogeneous areas, they define k clusters, where each cluster is a spatially homogeneous region. Homogeneity is defined by the authors as regions with similar electricity consumption. The algorithm classifies n users with M features into the k clusters. Given the number of clusters (k) defined a priori, the algorithm finds k clusters which minimize the euclidean distance as defined by sum of least squares. Initially, $k \times M$ values are chosen to represent cluster centroids. The authors compute the euclidean distance of each user from the initial centroids of the clusters and then assign the user to the cluster which yielded the smallest distance from the user. The process is repeated until users do not change clusters. Upon assigning users to clusters, the authors then study electricity consumption. Unlike the previous method proposed by Xie et al., the authors do not have to manually determine thresholds for each parameter to obtain spatially homogeneous regions. In our analysis, we propose a similar method for obtaining spatially homogeneous electricity consumptions where we define spatial homogeneity as levels of urbanization. We leverage the *k*-means clustering method, to obtain pixel-based urbanization levels as described in Section 4.1.

Electricity consumption over large areas can also be decomposed temporally to reveal unique consumption patterns. Multiple studies use cross-sectional or panel data to understand the spatial and temporal factors affecting electricity consumption. Sun et al. (2014) use 2013 cross-sectional data from the China Residential Energy Consumption Survey (CRECS) to study household consumption patterns. By performing single factor Analysis of Variance (ANOVA) and ordinary least square regression they are able to show that energy expenditure varies significantly with urbanization. Yan (2015) investigates the role of urbanization in provincial energy intensities in China. He uses a balanced panel dataset for 30 provinces from 2000 to 2012, to develop a temporal model used to understand the factors which drive energy intensity. The temporal model regressed urbanization levels for provinces in China along with other variables to estimate electricity intensity. Results showed that urbanization increased electricity intensities. In this paper, we take an alternative approach to empirical modeling as we do not have sufficient customer meta-data such as income levels and appliance ownership, to undertake a robust regression. In addition, the variables which we do have (urbanization and connection dates) are time-invariant. Thus by segmenting the study data through our approach we are still able to unpack the relative importance of urbanization and levels of experience on electricity consumption. Our segmentation approach puts less emphasis on regression coefficients which may be spurious due to serial correlation in models while highlighting both the underlying patterns and distributions within the study data.

3. Data description

Our study analyzes monthly electricity data of historical consumption in Kenya for residential customers from January 2010 through December 2015. The analysis first randomly samples customers from Kenya Power's customer database of about 4 million customers, at the end of 2015 and includes only residential customers with postpaid electricity meters.[1] This random sample consists of 152,752 customers. Using customer meta-data such as the meter GPS location and date of meter installation (connection), we remove customers with missing GPS location or installation date data. After this filtering our study dataset contains 135,579 customers. We use the bills of this study dataset for subsequent computations and analysis.

Each bill is provided as a series of components, according to a block tariff structure called the A0 (Residential) tariff. This tariff structure includes a combination of fixed and variable components; a description of these components is provided in Table 1. In addition to monthly units of electricity consumption (provided in kWh), each component also includes a bill amount (provided in Kenya Shillings – herein, KSh). In this study, we exclusively report on units of electricity consumption (*kWh*); discussion on the implications of this choice is provided in Section 6.2.

While most customers have bill data for all or nearly all months, there are some customers within this study dataset that have missing bill data, creating an unbalanced panel. Fig. 2 shows the months for which bill data are available for each customer, where customers are sorted by date of installation (connection). Each horizontal line represents a customer over time, with black indicating the presence of data for that given customer in a given month. Conversely, white represents the absence of data for the given customer in a given month. Please note that the sample is biased to the rate of growth in Kenya Power's customer base, and that we observe different epochs for each customer based on the relationship between their connection date and our study period (2010–2015). A small number of customers, seen in the topmost rows of the graph, have six years of billing data but are listed as having an installation date of March 1, 1995; we believe these customers originate prior to 1995, but have incorrect installation dates in our dataset. Based on our interaction with Kenya Power, installation dates for customers originating prior to 1995 were not recorded thus these customers were listed as having an installation date of March 1, 1995. We do not use data from these customers for determining customer consumption growth patterns.

The customers are spatially distributed across Kenya as seen in Fig. 3(a), where each dot represents a single electrical connection. For comparison, Fig. 3(b) shows the population density of Kenya, where each dot represents 100 people. Comparing customer locations to overall population density, there are heavy concentrations of both customers and people in the western, central, and coastal regions of Kenya. The electricity customer dataset is biased towards higher-

Table 1
Kenya Power residential (A0) tariff components. Note that the tariff description is as of the end of our study period; the tariff changed slightly on a couple of occasions during the study period. Energy Resource Commission (2014).

Component	Fixed/ Variable	Description
Fixed Charge	Fixed	150 KSh
Unit Charge	Variable	1st 0–50 units @ 2.50 KSh/Unit
	Variable	2nd 51–1500 units @ 12.75 KSh/Unit
	Variable	3rd Above 1500 units @ 20.57 KSh/Unit
Fuel Cost Charge	Variable	2.51 KSh/Unit
Forex Fluctuation Adj.	Variable	1 KSh/Unit
Water Resource Management Authority (WARMA)	Variable	0.05 KSh/Unit
Inflation Adj.	Variable	0.23 KSh/Unit
Rural Electrification Program (REP)	Variable	5% of Unit Charge
Energy Regulatory Committee (ERC)	Variable	0.03 KSh/Unit
Value Added Tax (VAT)	Variable	16% of (Unit Charge + Fuel + Forex)

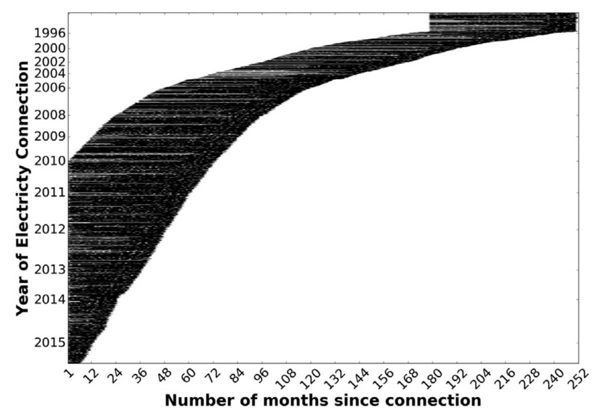
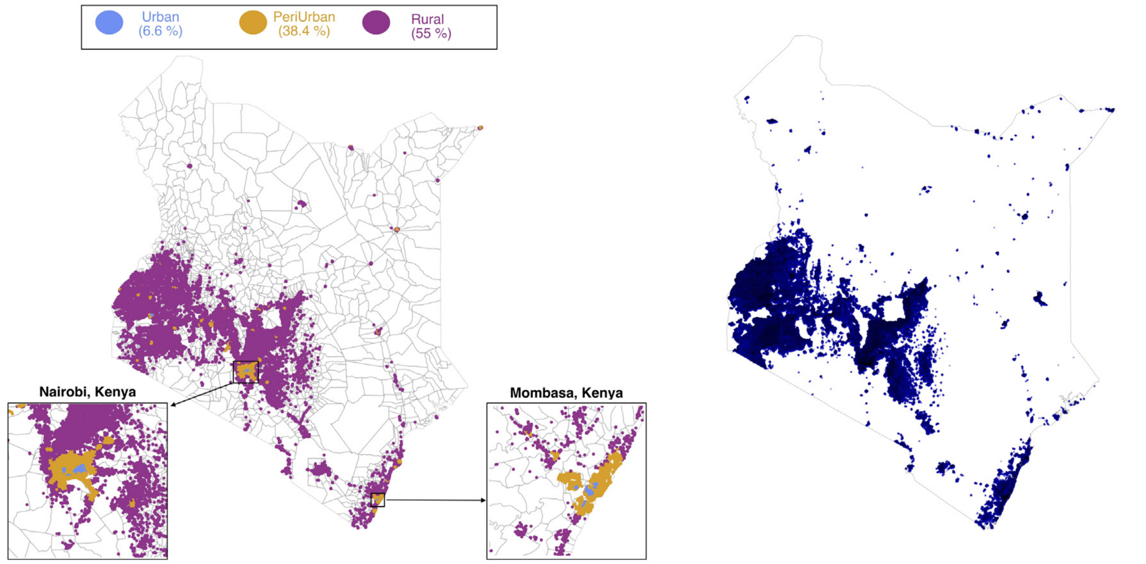


Fig. 2. Year of electricity connection versus number of months since the electricity connection, for each of the 136k customers. Each horizontal line represents a customer over time, with black indicating the presence of data for that given customer in a given month while white represents the absence of data for the given customer in a given month. This figure shows (i.) the data available for each customer and (ii.) the data available over different durations of access.

[1] The implications of analyzing bills only from customers with postpaid meters are discussed in Section 6.2.



(a) Spatial distribution of 136k customers in sample. (b) 2015 Kenya Population – each dot represents 100 people

Fig. 3. This figure compares the locations of electricity customers in our sample with the locations of population in Kenya. The customer locations are also segmented by urbanization level, showing well-defined spatial transitions from urban to peri-urban to rural.

density areas; evidence for this claim is available in Appendix A.

4. Methodology

In order to understand consumption among different groups, this study conducts a combination of spatial and temporal segmentation of the study dataset.

4.1. Spatial segmentation

Most newly-connected and unconnected households are in rural areas. In order to identify these households, we classify the customers in our dataset spatially by urbanization level. While it is common to use an urban-rural classification, unfortunately there is no standard definition of these categories (Christenson et al., 2014).

To address this, we developed a new method for identifying urban and rural areas that makes use of high-resolution data on population density, land use classification, and satellite nighttime light intensity. We provide an abbreviated description of our method here, and describe our method in depth in Appendix A.

Similar to the approach used by Chevez et al., we apply a k-means clustering method, however we apply some constraints to the method (Wagstaff et al., 2001). The constraint k-means algorithm partitions predefined pixels of Kenya into k clusters, such that the euclidean distance between the pixel's features and the cluster centroid are minimized. Eq. (1) shows the objective function of the algorithm, where k represents the number of clusters, k_i the number of pixels in cluster i, x_j a vector of features for pixel j and v_i is the cluster centroid for cluster i.

$$min \sum_{i=1}^{k} \sum_{j=1}^{k_i} (\| x_j - v_i \|^2) \tag{1}$$

Unlike Chevez et al. we apply a non-random initialization to the algorithm in the form of constraints as discussed in Appendix A. Once the clusters are obtained, we use customer GPS locations to assign each customer to a pixel and by consequence a cluster.

From our experience, the clustering algorithm works best with $k = 3$ clusters, which we identify as our urban, peri-urban, and rural areas. Peri-urban represents areas on the urban fringe whose denizens may access urban services and resources.

Fig. 3(a) shows the clustering results from our constrained k-means algorithm. Three customer clusters are shown in blue (urban, 6.6% of customers), yellow (peri-urban, 38.4% of customers) and violet (rural, 55% of customers). Areas classified as urban are mainly the cores of Nairobi and Mombasa, the two largest cities in Kenya, although a few urban areas can be seen in the smaller cities of Kisumu and Nakuru. The peri-urban regions generally envelop the urban locations, although other peri-urban locations border regions classified as rural. For this study, we subsequently add the peri-urban cluster to the urban cluster to form a single urban group; justification for this decision is provided in Appendix A.

4.2. Temporal segmentation

To tease out underline behaviors, the data was decomposed using two methods: by calendar date and by duration since customer electricity connection. For the former, post-paid billing dates are used to aggregate consumption by calendar month. For the latter, the number of months since a customer established their electricity connection is used to group customers. Most of our analysis uses this latter characterization, which aims to provide insight into growth of consumption by the duration of customers' experience with access to electricity. It is important to note that this method conflates customers from different eras into the same group, where bills from customers grouped by the same duration of experience may come from different months or years. We discuss the implications of this approach in Section 6.2.

5. Patterns in consumption

Using customer locations and our clustering method, customers were categorized into rural and urban groups. Table 2 shows the number of customers in each category for the entire study dataset, as well as for those who received an electricity connection before 2009 and after 2009. A majority of customers in our dataset are in rural regions (55%) and most received their electricity connection after 2009 (64.5%). Much of the recent increase in connection is due to efforts by Kenya Power, the Rural Electrification Authority (REA), and the

Table 2
Number of customers in each category (rural, urban). The rural category has more customers, with most added after 2009.

	Rural	Urban	Total
Full Dataset	74,609	60,970	135,579
2009 and After	54,896	32,800	87,696
Pre 2009	19,713	28,170	47,883

Government of Kenya to improve access to electricity in rural areas and slums, especially via the Last Mile Electrification Program (for densification of existing transformers) and the Global Partnership on Output-Based Aid (GPOBA) Program (for formalization of connections in informal settlements) (Kenya, 2016).

5.1. Consumption of a representative residential customer over time

Initially, we characterize the consumption over time of all customers in our study dataset regardless of the time they obtained a grid connection. A representative residential customer is chosen as one whose consumption is the median consumption of all customers in any calendar month. Note that each month this representative customer (here the customer with median consumption in that month) is not necessarily the same customer. Fig. 4 shows electricity consumption of the median customer (and the interquartile range of consumption levels) for each calendar month from 2010 through 2015. The Figure shows a declining trend over time for the median customer's electricity consumption (the solid line in the figure). This in itself is indicative that the utility must service an increasing number of customers whose monthly consumption is reducing.

5.2. Consumption growth over time since connection

The prior section described the consumption of a representative customer as observed by the utility. We wish to now understand whether the consumption of individual customers actually grows over time and if the growth over time varies between rural and urban customers.

We initially examine monthly customer electricity consumption over time as a function of the number of months a customer has had an electricity connection; this draws on the assumption that new electricity customers are similar in their consumption patterns regardless of when they receive their first connection. Fig. 5 shows this organic growth in

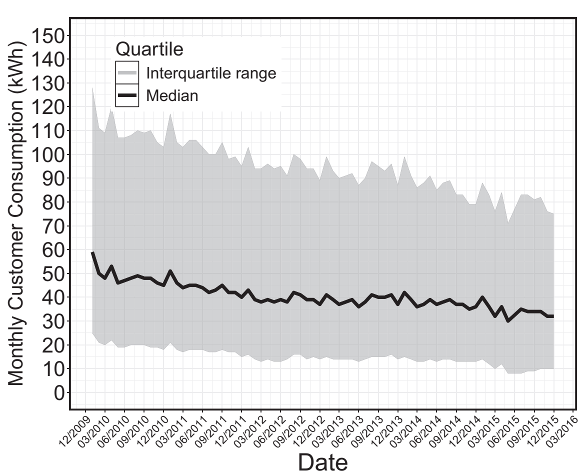


Fig. 4. Monthly customer electricity consumption for 135,579 customers from 2010 to 2015. The solid line represents the monthly median customer's consumption while the grey area represents the interquartile range of the study dataset. From the utility's perspective, there is an increasing number of lower-consuming customers.

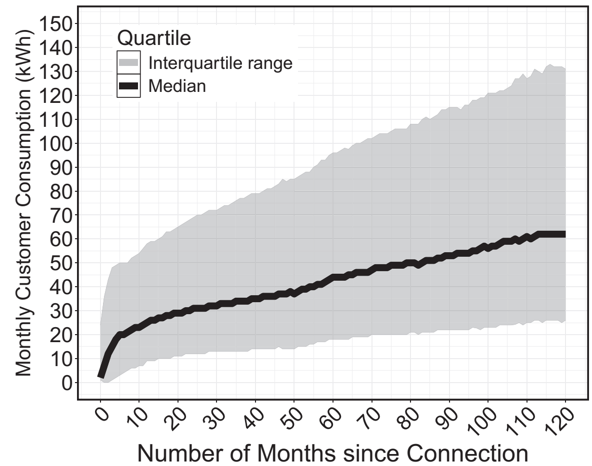


Fig. 5. Monthly customer electricity consumption for 135,579 customers by duration of customer's electricity connection, for the first ten years of access. The solid line represents the monthly median customer consumption while the grey area represents the interquartile range. Electricity consumption for the whole dataset initially increases sharply followed by continual, though decreasing, growth.

consumption amongst residential customers in our study dataset. In this figure, the solid line indicates the monthly median electricity consumption, and the grey area shows the interquartile range. From this figure, it is apparent that monthly electricity consumption for the whole study dataset continually increases upon access.

Not all customer groups will experience the same organic growth pattern. We use the previously-defined customer categories (urban and rural) to further segment the consumption data. Fig. 6 shows electricity consumption for urban and rural customers. Solid lines represent monthly median customer consumption while dashed lines represent the interquartile range.

Across all quartiles, rural customers consumed less electricity during their first decade of access than urban customers. This distinction is most pronounced with high-consuming rural consumers, who use significantly less electricity than their high-consuming counterparts in urban areas. Nonetheless, each group shows the same characteristic pattern of fast initial growth followed by persistent though slowing

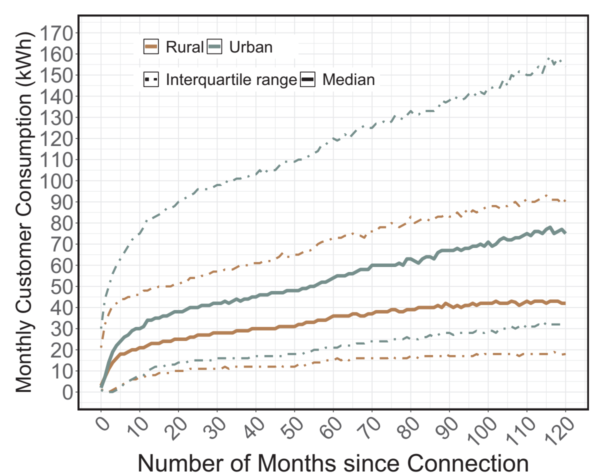
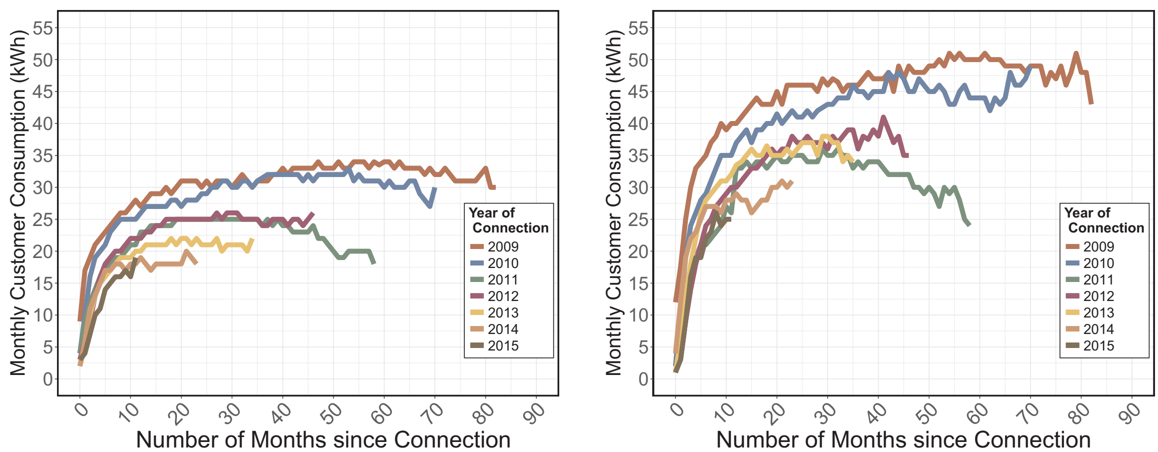


Fig. 6. Median monthly customer electricity consumption during the first decade of access, by urbanization level. The distribution for rural customers is shown in red and the distribution for urban customers is shown in green. Solid lines are median monthly customer consumption while dashed lines show the interquartile range. Rural customers consistently consume less than urban customers. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) Median monthly customer electricity consumption for rural customers

(b) Median monthly customer electricity consumption for urban customers

Fig. 7. Monthly median customer consumptions, separated by the year customers received a connection. The year the median customer received a connection matters, as more recently-connected customers consume less electricity and peak sooner than customers connected at earlier times.

growth thereafter.

5.3. Does the year of connection matter?

So far we have shown that customers grow their consumption upon receiving access, irrespective of their urbanization level. This perspective hides the possibility that customers connected to the grid earlier in calendar time – possibly those who were urban and started out with the means to afford a connection – might have different consumption levels from those who were connected more recently through a wave of subsidized rural electrification. Here we examine the effect of different waves of connection by grouping customers into the year they received an electricity connection.

Fig. 7 (a) and (b) shows median customer electricity consumption for rural and urban customers, respectively. In order to ensure that we can compare consumption of customers with the same age of electricity connection, we consider only customers who received an electricity connection in 2009 or later. Looking at the figure, it is apparent that the year of connection is an important consideration for both the rural and urban cohorts, as earlier connected customers (2009, 2010) tend to peak and level off. Further, it is evident that more recently-connected customers peak sooner and at lower consumption levels than those customers with earlier connections. This pattern is fairly consistent, showing that the most recently-connected customers simply do not consume as much electricity as earlier customers even after their consumption growth has abated. *In fact, the median customer whose connection began in 2009 consumes almost twice the electricity of the median 2014 or 2015 customer.*

Although consumption patterns are similar across urban and rural cohorts, it is clear from Fig. 7 that median urban customers consume more electricity than median rural customers. To further explore how much more electricity median urban customers consume, we computed the ratios of consumption for each year of connection.

Fig. 8 shows these ratios of consumption for median urban to median rural customers, separated by the year customers received an electricity connection. From the figure we see that beyond the stabilization period of 6–12 months the median urban customer consumes 50% more electricity than the median rural customer. This ratio provides a concise way to understand electricity consumption at varying levels of urbanization.

5.4. Sample size considerations

Each step of segmentation reduces the sample size of customer bills

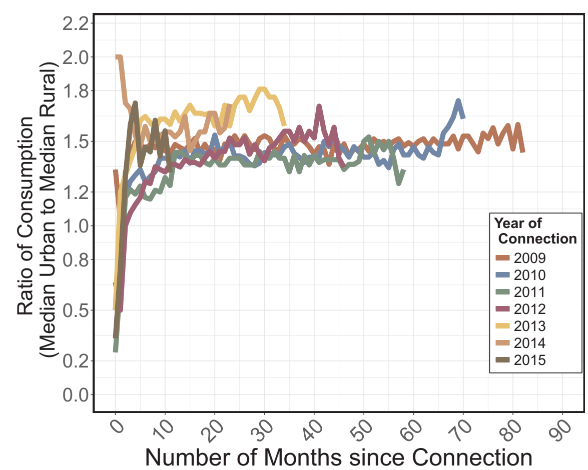


Fig. 8. Ratio of Monthly Consumption for median urban to median rural customers. Median urban customers consume 50% more electricity than their median rural counterparts.

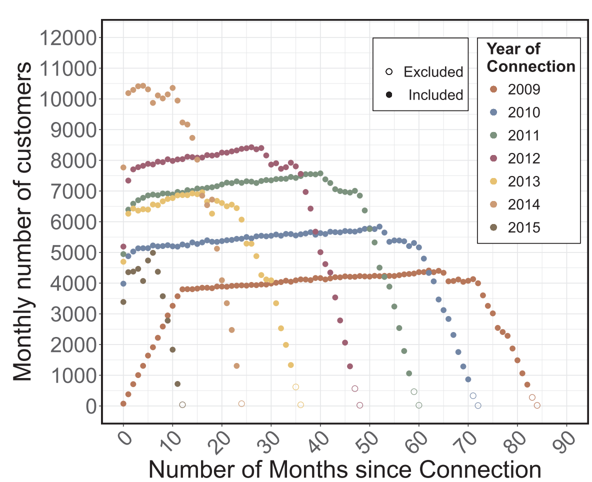


Fig. 9. Monthly number of customers in the rural category, separated by electricity installation dates. The large number of customers, numbering in the thousands of bills, allows for confidence in the significance of our finding.

available in the segment. To ensure that our conclusions are durable, we examine the sample sizes of customer bills for these segments. Fig. 9 shows the monthly customer sample size for each year of connection. To remove points with perhaps too few samples, we filtered out months for which the sample size was less than 10% of the median sample size for a given year of connection. Since each line in Fig. 9 is comprised of distributions numbering in the thousands of bills, we have confidence in the significance of our finding.

We apply the same sample size filtering approach to customers in the urban segment.

5.5. Whose consumption is reducing?

To orient our observations towards the implications of increasing electrification, we look specifically at the rural consumers, who will comprise much of the further potential growth in the electricity customer base in Kenya. We must realize that not all rural customers have the same patterns in consumption; Fig. 7 (a) shows a drop in consumption in the later months of access. This pattern stands out for rural customers in 2009, 2010, and 2011 especially, whose consumptions reduce anywhere between 12% and 28%. While these drops appear to be synchronized in calendar dates, their appearance only among customers who started their connections in particular years along with the lack of any known macroeconomic change over the period raises questions about what caused the drop. A drop in the median could be the result of either an equally-distributed “broad” reduction or a deeper reduction focused on a particular group of customers. To investigate this question, we selected the rural customers from 2009 and 2011 and looked at their consumption in two different time periods: all of 2013, when both groups have reached their steady-state peak in consumption, and the last five months of 2015, when the drop in consumption occurs. Fig. 10(a) and (b) are migration charts that show the percentage of customers that change their consumption bin from 2013 to 2015. Bin boundaries measured by monthly consumption in kWh were chosen to be consistent for the 2013 and 2015 groups. The 2013 customer sample sizes of each group (n) are shown at the bottom of the chart.

We can see that for both groups more customers reduced consumption than increased it and that reductions in consumption are more concentrated in the lower portion of the distribution. We also see that a larger percent of the 2011 customers dropped to the lowest consumption bin in 2015 than the earlier 2009 customers. Table 3 compares the percentage of customers (2009 and 2011) who were in the lowest consumption bin in 2013 and 2015. For 2009 customers, 29.3% of all customers were in the lowest consumption bin ($\leq 20\ kWh$) during the 2015 period compared to 24% during the 2013 period; for 2011 customers, this number is more pronounced, at 43.9% of all customers during the 2015 period compared to 34.5% during the 2013 period. Thus, for the 2011 customers, the reduction in consumption is relatively more concentrated in the lower end of the distribution. Although there is some migration to higher consumption bins, customers at the lower end of the distribution are far more likely to reduce their consumption and sometimes stop consuming entirely.

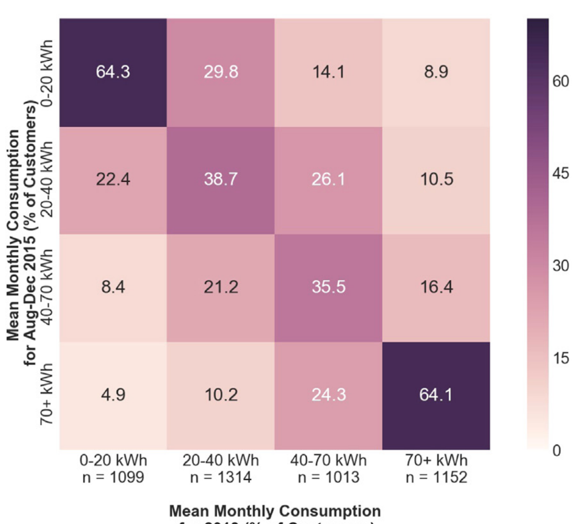
Table 3
Comparing the proportion of customers in the lowest consumption bin for two groups of customers: those starting in 2009 and those starting in 2011. For customers who received an electricity connection in 2011, more customers started in the lowest bin and a larger proportion moved there by 2015.

Customer Start Year	% in [0,20] kWh in 2013	% in [0,20] kWh in 2015
2009	24.0	29.3
2011	34.5	43.9

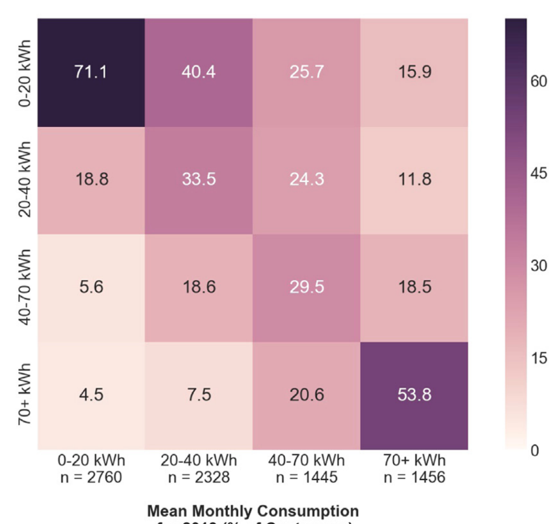
While some customers may actively elect to reduce their consumption by purchasing more efficient lighting and appliances, others may be deprived from enjoying the economic and quality-of-life benefits of electricity consumption due to high electricity costs, poor reliability, lack of access to financing for equipment purchases, damaged equipment, or a combination of factors. We note that only a small proportion of customers in our sample went to zero consumption, which might imply a disconnection or other billing issue. Understanding the motivations for reductions in consumption among these lower-consuming customers, perhaps via surveys and other measurements, is a critical next step for improving customers' experiences and outcomes with electricity access as well as building more sustainable and durable electricity-providing institutions.

6. Policy implications and discussion

Examination of grid-connected Kenya Power customers shows that the monthly median electricity consumption of the recently-connected customers is lower than that of grid-connected customers from several



(a) Consumption changes among 2009 rural customers



(b) Consumption changes among 2011 rural customers

Fig. 10. Migration within the electricity consumption distribution for (a) rural customers with start dates during 2009 and (b) rural customers with start dates during 2011. Horizontal axis shows breakdown of customers by mean monthly consumption for the year 2013 and vertical axis shows breakdown of customers by mean monthly consumption for the last five months of 2015.

years ago, comparing at the same point in time after connection. For example, a median customer in an urban area who received a connection in 2009 consumed 43 kWh per month after 18 months while a median customer in a rural area who received a connection in 2014 consumed 18 kWh per month after 18 months. This result shows that electricity planning based on earlier consumption estimates may be misleading. In this section we consider implications of our results, some limitations, and the sensitivity of our analyses to important methodological choices.

6.1. Implications for electricity planning

Countries with low GDP *per capita* must make critical decisions on how to allocate precious public-sector resources amongst competing priorities, especially when it comes to spending on infrastructure. For example, if Kenya tried to connect 1 million households annually to the grid, the investment in distribution infrastructure alone would exceed 4% of the annual government budget. We are assuming here that investments in generation and transmission can come from private sources. It is equally difficult to recover the investment cost from cross-subsidies applied to industrial customers. Recovering an investment of $1 billion USD from the 3575 presently-connected industrial consumers with an average consumption of 95,000 kWh per month would require an additional tariff of $0.25 USD/kWh levied on industrial customers; this is clearly an unreasonable expectation. Hence a least-cost investment approach suited to anticipated electricity demand is crucial for low-income countries. The results of this study can potentially help Kenya Power to reduce the cost of providing electricity to households. We propose three cost reduction approaches based on our findings: (i) Solar Home Systems (SHS) for low consuming customers; (ii) Reforming technical standards to connect more low-consuming customers within the existing connection radius; and (iii) Extending the existing connection radius.

Median consumption levels below 20 kWh/month for a residential customer may provide a crucial tipping point when compared to planning based on historical estimates of consumption – typically closer to 50 kWh/month. For example, a 20 kWh/month consumption level could possibly be met by an off-grid system that would deliver 500 Wh/day or a 150 Watt peak SHS costing $500 if such a shift did not limit a customer's anticipated consumption growth. If a 20 kWh/month consumption level were met with a grid connection, the connection cost would be 2–3 times higher. On the other hand, for a 50 kWh/month consumption level, the investment cost of an off-grid system is likely to be higher than that of a grid connection. This simplistic example illustrates how the results of this study impact electrification planning in a resource-constrained economy. The real planning scenario is likely to be much more nuanced and might depend on specifics of sub-populations that are being addressed.

Kenya's connection policy states that the utility charges customers who wish to connect a flat fee if those customers reside within 600 m of any transformer on the grid. This fee is 34,980 KSh (≈$340 USD), or 15,000 KSh (≈$145 USD) under the subsidized Last Mile Connectivity Program (LMCP). Customers outside of this radius who wish to connect may do so at the full cost of the connection, on average $1200 or more as the distance grows. The reasoning behind this 600 m policy is a combination of engineering and cost constraints; the voltage drop experienced as well as the cost of poles and conductors needed both increase with a longer distance from the transformer. Knowledge of anticipated demand can shape appropriate engineering requirements of the grid. For example, one could easily and safely reduce the service standard, sized for a peak 3 *kW* load to perhaps 1 *kW* for lower-consuming customers. This would in turn lower cost of transformers, conductors, and cables as more customers can be added onto the same transformer. Less stringent yet still sufficient technical standards enable the utility to densify existing transformers at the current connection radius, lowering the per customer transformer cost, as more low consuming customers can be accommodated on the same transformer.

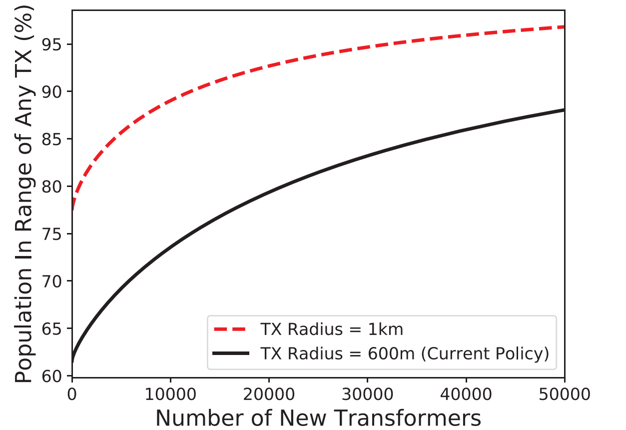


Fig. 11. The proportion of Kenya's population within range of any of Kenya Power's transformers under two different connection fee policies: (1) customers within 1 km of any existing or new transformer can connect for a flat fee (red line) and (2) the existing policy, where customers within 600 m of any existing or new transformer can connect for a flat fee (black line). Note that Kenya Power presently has a total of roughly 58k transformers, and those transformers are within range of 62% of the country's population. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Alternatively, extending the connection radius with the same wire standards would potentially also allow a low-voltage wire to reach customers located further away from the transformer. In Fig. 11, we show the implications for Kenya Power if the connection radius were increased. For this analysis, we use a greedy algorithm that places new transformers in the locations that maximize the population covered.

At present, 62% of Kenya's population lives within 600 m of Kenya Power's roughly 58,000 transformers, and the Government of Kenya has a stated goal of providing access to electricity to 100% of the population by 2020. According to the figure, maintaining the same connection policy and attempting to reach 85% of Kenya's population with the grid would require an additional 35,000 transformers. However, newer transformers are in rural areas where customers are further apart, but voltage drops are lessened due to lower consumption per customer. Thus, relaxing the 600 m constraint no longer poses as much of an engineering challenge and would enable the grid to reach more customers with existing or fewer additional transformers. If the policy were changed to allow any customer within 1 km of any transformer to connect for a flat fee, it would take fewer than 5000 additional transformers to reach the same 85% of the population.

While the cost of connections is still a heavy burden achieving those connections by extending existing low-voltage infrastructure, as opposed to deploying new transformers, may present a lower-cost option. Further, this strategy would align well with the LMCP, which aims to densify existing underused transformers using a budget of roughly $450 million USD. It is important to note that existing plans for the three phases of LMCP (an investment of roughly $450 million USD) include only 1400 additional transformers, challenging the Government of Kenya's stated goals of reaching 70% electrification by the end of 2017 and universal electrification by 2020.

Without a significant change of direction on alternative means of electrification, massive reductions in connection costs, or unexpectedly high growth in electricity consumption, the utility model faces severe challenges in meeting the dual mandate of universal electrification and investor profitability. Sustained low consumption levels will hinder the financial viability of utilities whose goal is to increase electricity access. It may be possible to boost consumption and by consequence financial viability via targeted programs such as appliance financing and tariff subsidies. These can create more growth in electricity consumption, support higher quality-of-life and have potential income benefits for customers while supporting the dual mandate of electricity providers.

Although our discussions have focused on Kenya, we believe that Kenya Power's experience can highlight broader lessons that are relevant for utilities in other developing countries.

- Customer consumption may not grow at a constant percentage over time.
- Performing better customers analytics, prior to deciding how to connect these customers can result in fewer underutilized grid connections, allowing more customers to be reached at a cheaper cost.
- The assumption that everyone must be connected in the same manner has both benefits and costs, and it is important to quantify the costs to design evidence-based policy.

6.2. Additional considerations

Urban/Rural Sensitivity: Urbanization levels were defined using a combination of datasets. However, we recognize that there are a range of classification methods for determining urbanization level, and that our results are sensitive to the method we used. Additionally, not all rural regions are similar – localized economic effects will not be captured by this approach, but we attempt to deal with this by primarily considering medians as well as interquartile ranges, so as to not be affected by extremes in the distribution. Further, definitions of urbanizations are hardly static as captured in our clustering analysis. These definitions change with time and are influenced by changing socioeconomic factors and migration. Thus our definition of urbanization levels only capture one snapshot, which is at the start of the analysis period (2010). Future work on this topic is to examine how consumption evolves in areas that experience slower or faster changes in urbanization levels.

Other Temporal Effects: Analyzing customer growth on a calendar basis conflates the effects of a *growing* customer base with those of an *evolving* customer base, a typical situation for grids in sub-Saharan Africa. In an effort to disaggregate these two, we spend the majority of our analysis analyzing customers via the lens of time since electricity connection. While transforming the temporal axis from *calendar dates* to *time since electricity connections* reveals relevant information for electricity access, there are also adverse effects to consider. This approach obscures the effects of cyclic and seasonal changes, macroeconomic shocks, and, as we show in this work, differences among newer and older customers. While we acknowledge that these exogenous events occurred during our study period, we believe that a six-year duration to our study should allow examination of larger trends in growth of consumption among these customers.

Tariff and Meter: We use kilowatt-hours as the measurement of consumption over time, with limited consideration of the various tariff structures in place for these customers. Some of these tariff components changed during the course of the study period; for example, in mid-2014, the fixed tariff increased from 120 *KSh* per month to 150 *KSh* per month. Some of the variable tariff components also had small changes during the study period, and others, such as the Foreign Exchange (Forex) and Fuel Cost Charges changed on a monthly basis to reflect market conditions. While many of these changes were seemingly negligible, more in-depth analysis is needed to estimate the scale of these effects on longitudinal consumption.

In particular, our sample consists of customers only on *postpaid* electricity meters. Initially, we do not have any clear evidence that differentiates these customers from customers with prepaid electricity meters. However, since customers with postpaid meters tended to receive their connections earlier, as a class they are likely more wealthy than their prepaid counterparts, potentially depressing the consumption values reported throughout this paper. We take it as future work to understand the implications of examining only customers with postpaid meters, and seek to compare the consumption patterns among those customers with postpaid and prepaid electricity meters.

Equity: Different electricity delivery technologies within the same community challenge notions of equity in electricity connections and may pose political barriers. Quantifying the costs of equity of connections, though not necessarily equity of service, are worthy of further study, though beyond the scope of this work.

7. Conclusion

Developing economies are undergoing rapid growth in the number of customers with electricity access. This work analyzes the dynamics of electricity consumption among newly-connected customers in Kenya over a recent six-year period. While reaching the entire population with some form of electricity access is a goal of all countries, it is vital to consider the challenges therein. If, as our results show, the expected consumption plateau is lower for newer customers, then the lowest-cost technology for initially providing electricity access to some customers, at least until the demand grows significantly, may not be grid power.

Appendix A. Classification

We applied a constrained *k*-means clustering method (Wagstaff et al., 2001) to identify three clusters (urban, peri-urban, and rural). We initially used two clusters, representing urban and rural areas, but discovered that the numerical uniqueness of the urban cores of Nairobi and Mombasa – with high population density and intense nighttime lighting – set those areas apart into their own cluster. The peri-urban surroundings of the cities and the rural areas were quantitatively more similar and therefore grouped into the same cluster. This does not agree with conventional definitions of urban areas. By identifying three clusters, the algorithm is able to separate these “peri-urban” areas from the rural areas, arriving at a much more justifiable classification. We also note that electricity consumption levels across all quartiles were largely similar between the urban and peri-urban categories, so we felt comfortable pairing these two clusters into a single category representing urban consumption.

The constrained k-means clustering method works by exploiting accepted characteristics about urban areas, which is used to apply initial constraints on the clustering algorithm. For identifying these initial constraints, we leverage three methods for determining urban and rural locations from the literature. We use the spatial regions of overlap of these three methods to bootstrap our algorithm, effectively identifying consensus-urban regions. The methods include:

1. The Global Rural-Urban Mapping Project (GRUMP) (Socioeconomic Data and Applications Center, SEDAC, 2010), which combines census and satellite data to produce various datasets, including urban masks used in this analysis;
2. LADA Land Use Systems of the World data which provides 40 land-use classes for the world including urban areas (Land Degradation Assessment in Drylands, 2010); and
3. The UN population estimate (The United Nations Population Divisions World Urbanization Prospects, 2010), which uses a national urbanization level (23.6% in the case of Kenya) to determine a threshold of population density at which to separate urban areas and rural areas.

The constraints (consensus regions) describe which items in the dataset must be or cannot be “linked” (appear in the same cluster). These areas

Table 4
Comparison of our clustering method with other definitions of urbanization, in classifying the total population of Kenya in 2010.

Urbanization methods	Urban (%)	Rural (%)	Peri-Urban (%)
Our Method (2 clusters)	15	85	NA
Our Method (3 clusters)	5.4	81.7	12.9
GRUMP	22.6	77.4	NA
Land Use Systems	11.9	88.1	NA
UN Population Estimate	23	77	NA

Table 5
Comparison of our method with other definitions of urbanization, in classifying the 136k customers in our sample, by urbanization level.

Urbanization methods	Urban (%)	Rural (%)	Peri-Urban (%)
Our Method (2 clusters)	32.9	67.1	NA
Our Method (3 clusters)	6.6	55	38.4
GRUMP	53	47	NA
Land Use Systems	22	78	NA
UN Population Estimate	46	54	NA

provide the initial conditions of the clustering algorithm, effectively bootstrapping the cluster definitions with areas that must appear in the same cluster. With this guidance for initial cluster relationships, the algorithm can then proceed to assign the remaining areas to any of the three clusters. To determine cluster membership, the algorithm uses features obtained from three 2010 data sources, all of which are available publicly and in a raster format at a maximum common resolution of $1\ km \times 1\ km$:

- Population Density via WorldPop (AfriPop, 2010);
- Nighttime Lights via the DMSP-OLS satellite imagery dataset (NOAA's, 2010); and
- LADA Land Use Systems of the World data which provides 40 land-use classes for the world including urban areas (Land Degradation Assessment in Drylands, 2010).

Various methods for urban-rural classification in the literature employ one or two of these datasets, but we were unable to find any methods that used all three data sources. Prior to applying the clustering algorithm, the features are normalized by their mean and standard deviation. The algorithm is able to classify each $1\ km \times 1\ km$ grid cell of Kenya as urban, peri-urban, or rural. Based on this classification, customers in our sample can be assigned to an urbanization level using the GPS locations of their electric meters.

Table 4 compares our method under 2 and 3 clusters to the other urbanization methods, in classifying the total population of Kenya. We show that our method under 3 clusters better allows us to extract the most rural population of Kenya, compared to when we only apply 2 clusters. Although our method performs similar to existing methods when defining urbanization levels, our method offers a robust clustering approach because it leverages regions which existing definitions all agree to be urban, and uses these regions to initialize the clustering thereby providing a more trustworthy definition of urbanization.

In Table 5 we also show the performance of our method in classifying our study dataset of about 136k customers. The peri-urban customers defined in our 3 cluster approach tend to be carved from mostly the urban customers in a 2 cluster approach- although there are some from the rural cluster. This result aligns with our decision to group urban and peri-urban customer consumption while understanding the behavior of the most rural customers.

References

AfriPop, 2010. Kenya 100m Population.
Amarala, Silvana, et al., 2005. Estimating population and energy consumption in Brazilian Amazonia using DMSP night-time satellite data. Comput. Environ. Urban Syst. 29, 179–195.
Chévez, Pedro, et al., 2017. Application of the k-means clustering method for the detection and analysis of areas of homogeneous residential electricity consumption at the Great La Plata region, Buenos Aires, Argentina. Sustain. Cities Soc. 32, 115–129.
Christenson, Elizabeth, et al., 2014. Examining the in uence of urban definition when assessing relative safety of drinking-water in Nigeria. Sci. Total Environ. 490, 301–312.
Elvidge, C., et al., 1997. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. Int. J. Remote Sens. 18, 1373–1379.
Energy Regulatory Commission, 2014. Gazette Notice no. 280: The Advocates Act, The Complaints Commission, 92nd Quarterly Report.
Esmaeilimoakher, Parisa, et al., 2016. Identifying the determinants of residential electricity consumption for social housing in Perth, Western Australia. Energy Build. 133, 403–413.
Jones, Rory V., Fuertes, Alba, Lomas, Kevin J., 2015. The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings. Renew. Sustain. Energy Rev. 43, 901–917.
Kenya National Bureau of Statistics, 2009. The 2009 Kenya population and housing census.
Kenya Power, 2016. Kenya power annual report and financial statements for the year ended 30 June 2016.
Land Degradation Assessment in Drylands, 2010. LADA land use systems of the world.
Louie, Henry, Dauenhauer, Peter, 2016. Effects of load estimation error on small-scale off-grid photovoltaic system design, cost and reliability. Energy Sustain. Dev. 34, 30–43.
Mensah, Justice, Marbuah, George, Amoah, Anthony, 2016. Energy demand in Ghana: a disaggregated analysis. Renew. Sustain. Energy Rev. 53, 924–935.
NOAA's National Geophysical Data Center & US Air Force Weather Agency, 2010. Version 4 DMSPOLS Nighttime Lights Time Series.
Socioeconomic Data and Applications Center (SEDAC), 2010. Global rural-urban mapping project (GRUMP), v1.
Sun, Chuanwang, et al., 2014. Household pathway selection of energy consumption during urbanization process in China. Energy Convers. Manag. 84, 295–304.
The United Nations Population Divisions World Urbanization Prospects, 2010. Urban Population (% of total).
Villareal, Maria José Charfuelan, Moreira, Jo ao Manoel Losada, 2016. Household consumption of electricity in Brazil between 1985 and 2013. Energy Policy 96, 251–259.
Wagstaff, Kiri et al., 2001. Constrained K-means clustering with background knowledge. In: Proceedings of the Eighteenth International Conference on Machine Learning, pp. 577–584.
Wolfram, Catherine, Shelef, Orie, Gertler, Paul, 2012. How will energy demand develop in the developing world? J. Econ. Perspect.
Xie, Yanhua, Weng, Qihao, 2016. Detecting urban-scale dynamics of electricity consumption at Chinese cities using time-series DMSP-OLS (Defense Meteorological Satellite Program-Operational Linescan System) nighttime light imageries. Energy 100, 177–189.
Yan, Huijie, 2015. Provincial energy intensity in China: the role of urbanization. Energy Policy 86, 635–650.