

# Mapping Induced Residential Demand for Electricity in Kenya

Douglas Fabini<sup>\*†</sup>, Diego Ponce de León Baridó<sup>\*‡</sup>, Akomeno Omu<sup>\*</sup>, and Jay Taneja<sup>\*</sup>

<sup>\*</sup>IBM Research, Africa  
Nairobi, Kenya  
{akomeno\_omu,taneja}  
@ke.ibm.com

<sup>†</sup>Materials Research Laboratory  
University of California, Santa Barbara  
Santa Barbara, CA, U.S.A.  
fabini@mrl.ucsb.edu

<sup>‡</sup>Energy and Resources Group  
University of California, Berkeley  
Berkeley, CA, U.S.A.  
diegoleon@berkeley.edu

## ABSTRACT

Despite substantial gains in the past few decades, 550 million people in sub-Saharan Africa still lack access to electricity. Rural areas present the largest electrification challenge, with access levels below 12% in most countries. Public rural electrification efforts, where undertaken, have generally effected slow and limited change. Further, to motivate the substantial investment required for traditional large-scale generation and transmission projects, strong demand for electricity services is required, and this demand is not easily demonstrated in rural African settings in which little data and substantial uncertainty exist. In this paper, we develop a predictive model for mapping induced residential demand for electricity – the hypothetical demand that would exist if access to electricity services were made available. We apply this model on a fine geographic basis to Kenya to demonstrate the applicability of the approach to informing public or private electrification efforts. Together with spatially explicit cost models for generation, transmission, and distribution, these induced demand predictions can be used to evaluate various technology options, business models, and tariff structures, or to guide public sector electrification program development.

## 1. INTRODUCTION

Though 1.7 billion people gained access to electricity between 1990 and 2010, the global rate of electricity access increased only modestly from 76% to 83% over the same period [12]. Furthermore, access remains uneven – roughly 550 million people lack access to electricity in sub-Saharan Africa, and the continent’s population is growing rapidly. Access levels are particularly low in rural areas, as shown in Figure 1 [11]. In sub-Saharan Africa, a multitude of economic, social, and political factors have inhibited widespread electrification for sparsely settled rural populations by traditional, public-sector efforts. Where conventional strategies have struggled, steep reductions in the costs of distributed

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM DEV-5 (2014), December 5–6, 2014, San Jose, CA, USA.  
Copyright 2014 ACM 978-1-4503-2936-1/14/12 ...\$15.00.  
<http://dx.doi.org/10.1145/2674377.2674390>.

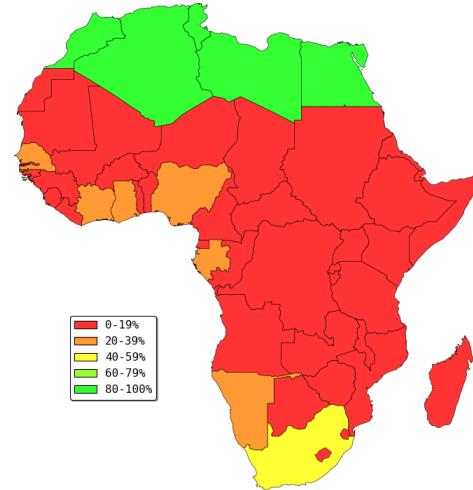


Figure 1: Rural Electrification Rates in Africa.

electricity generation (DG) present an alternative pathway to expanding access – it may now be possible for private rural electrification business models based on DG to emerge and scale.

Advocates of centralized and distributed power systems are engaged in a vigorous debate over the economic, developmental, environmental, and ethical implications of each approach for expanding electricity access to the un-electrified rural poor [8] [23]. Large, interconnected power systems capture significant economies of scale and resource efficiencies and are quite flexible with regards to future demand growth. On the other hand, they require large upfront investments in generation and transmission projects that can be hard to justify in the absence of robust, accurate demand projections. Small, distributed power systems that serve individual households, villages, or towns are of increasing interest due to sustained technology and cost improvements in solar photovoltaics and other distributed generation, solid-state power conversion, and metering and management systems. If implemented properly, these nascent technologies may present a cost-effective, low-carbon approach to expanding electricity access that bypasses some of the financing, execution, and corruption challenges that can plague large energy infrastructure projects. However, poorly designed and executed build-outs of distributed power systems run the risk of locking rural populations into small quantities of high cost,

low reliability electricity with little room for demand growth. In order to develop sound analyses and inform rural electrification stakeholder decision-making among various centralized and distributed approaches, accurate estimates of electricity costs and demand for electricity services are needed.

In this work, we focus on quantifying *induced residential demand for electricity* in off-grid areas. Induced demand represents the potential additional electricity consumption if reliable electricity services were made available. Quantifying this potential for demand is critical in evaluating the feasibility of a particular electrification program or business in a particular location; if there is enough demand to generate sufficient revenue to cover operational costs, an electrification business can grow to serve more and more customers or the program can be expanded to more areas. We develop a model that employs geo-spatial survey data and nearest-neighbors techniques to predict induced residential electricity demand and demonstrate its application in Kenya.

This paper is organized as follows. Section 2 briefly examines related work on estimation of electricity supply costs and residential demand. To better contextualize this work, Section 3 broadly summarizes the set of rural electrification technology options that are currently pursued and gives examples of recent implementations in Kenya. Section 4 explains the induced residential demand methodology and identifies the data sources we employ to demonstrate the methodology in Kenya. Results are presented and discussed in Section 5.

## 2. RELATED WORK

Existing work on evaluating rural electrification technology options can broadly be divided into supply-side models, which estimate the costs of providing electricity services, and demand-side models, which seek to determine the demand for these services.

### 2.1 Supply-side Models

Supply-side models use resource characteristics, equipment and fuel costs, and spatial data to estimate the costs of electrifying new areas by various centralized or decentralized means. One study with a broad geographic scope employs simple models for the production costs of diesel and solar photovoltaic electricity that are dependent on solar resource characteristics, national diesel prices, and costs of fuel distribution [25]. Others employ techniques from combinatorial optimization to estimate optimal network design of electricity transmission and distribution systems in addition to modeling generation costs [20] [18] [31]. Across most of these studies, demand for electricity is given limited treatment, though it is equally important to the overall cost-efficacy of expanded electrification, as the range of feasible electrification technology options and business models is heavily dependent on the quantity and price of electricity sold.

### 2.2 Demand-side Models

Most approaches to off-grid electricity demand estimation, as well as studies that evaluate users' ability and willingness to pay for electricity services, have traditionally used social science methods such as surveys [16] [22], field and longitudinal studies [6] [26], and stated preferences (contingent valuation, ability and willingness to pay) [4] [28]. These approaches are extremely valuable as they usually provide detailed knowledge about a consumer and the intricacies of

daily life in a region, village, or town. They can be used for evaluating the preference and decision making process that goes into buying and using different energy services (fuelwood or gas for cooking, kerosene or solar lamps for lighting, for example), and perhaps later be used for design of optimal tariff structures and demand-side management schemes. However, although extremely insightful about a particular place, they are time- and resource-intensive, and results are not usually generalizable.

End-use methods can allow the researcher to incorporate different scenarios (behavioral dynamics, energy efficient devices, and income and energy transitions, for example), and data sources (census and appliance ownership data, technology characteristics, and usage patterns, among others) to make assertions about electricity consumption in different sectors and areas of village life [7]. These approaches can facilitate generalization across larger spatial footprints.

## 3. RURAL ELECTRIFICATION TECHNOLOGY OPTIONS

To further contextualize this work and to share the particulars of rural electrification approaches currently employed in Kenya, we examine a variety of technology options and example implementations of each. These approaches can be coarsely condensed into three categories: 1) centralized grid extension, 2) solar home systems, and 3) micro- and mini-grids. The suitability of each electrification concept depends on local geographic factors (topography, renewable resources, etc.), electricity demand, and ability to pay for electricity services. In the absence of any of these technology options, rural consumers sometimes resort to automotive batteries and small commercial charging services to meet their electricity needs. These modes of minimal access are not consistent with sustained human and economic development and are not considered here.

### 3.1 Centralized Grid Extension

The extension of national- or regional-scale electric power systems ("centralized grids") to rural areas has traditionally been the main strategy for rural electrification. However, grid extension becomes significantly less cost-effective for sparsely settled areas with low demand intensity. Due to these fundamental factors and to the typically high cost of materials in sub-Saharan Africa, grid extension into remote rural communities is often economically prohibitive, as budgets for electrification are constrained and utilities are unable to recoup the full costs through connection fees and revenue from electricity sales. The result is that grid extension becomes a negative profit endeavor, giving little incentive for utility companies to undertake such programs in the absence of government mandates or subsidies. Another major challenge with this approach is the lackluster reliability of many African power systems. According to the World Bank Enterprise Survey, typical commercial consumers with grid electricity in Kenya experience an average of 6.3 power outages per month, with the average outage lasting 5.6 hours [30]. Outage rates and durations vary significantly across sub-Saharan Africa but are non-trivial in most nations. Low reliability significantly diminishes the value of electricity services, particularly for commercial or industrial uses and cold storage, and mitigating this problem requires costly investments in backup genera-

tion systems. On the positive side, by capturing economies of scale and efficiencies associated with large generation facilities and large interconnection footprints, the grid offers the lowest marginal costs for electricity and the greatest potential for demand growth. As of this writing, the national utility, Kenya Power and Lighting Company (KPLC) charges a fixed monthly cost of KES 120 (US\$1.36) and consumption tariffs of KES 2.50/kWh (US\$0.03/kWh) for the first 50 kWh, KES 13.68/kWh (US\$0.16/kWh) for consumption between 50 and 1500 kWh, and KES 21.57/kWh (US\$0.25/kWh) for all consumption above 1500 kWh for residential customers [21].<sup>1</sup> However, there is an additional upfront connection fee of KES 75,000 (US\$852), which raises the barrier to entry to beyond the means of the typical potential rural customer. In recent years, the Kenyan Rural Electrification Authority has focused its efforts on electrifying health clinics, public secondary schools, and market centers and subsequently offering subsidized connections to nearby homes and businesses.<sup>2</sup> Despite significant progress in electrifying these public facilities, financing residential and commercial connections remains a barrier [17].

### 3.2 Solar Home Systems

In Kenya, solar home systems are standalone solar energy kits that typically consist of a 5-100 W solar panel, a charge controller, a lead-acid battery, and a suite of DC appliances like LED lights, phone charging connections, radios, and televisions. The solar panel and battery are sized to provide reliable electricity to power the typical usage of the associated appliances. However, while reliability is high, the total electricity generation is small and essentially fixed over the life of the equipment, limiting demand growth with time. Additionally, these systems typically only support DC appliances, which are less widely available and are typically more expensive than their AC counterparts. M-KOPA is one of the leading solar home system companies in Kenya and Uganda, with 100,000 units sold in less than two years of commercial operation [3]. Its primary product is a solar energy kit, consisting of a 5W solar panel, a charge controller, a sealed lead-acid battery, three LED lights, a hub for mobile phone charging, and a radio. M-KOPA utilizes an innovative pricing model, whereby customers pay a KES 2999 down payment (US\$34) and KES 50 per day (US\$0.57), paid via a mobile money platform, for one year before owning the kit outright [2]. This results in electricity that is more than sixty times the cost of centralized grid electricity for every unit of energy consumed (assuming a two-year battery lifetime). However, the low down payment cost in comparison to the high grid connection cost makes this option more accessible to potential rural consumers.

### 3.3 Microgrids and Minigrids

Microgrids and minigrids are small electric power systems that typically comprise one or more generation sources – such as solar panels, small wind turbines, or diesel gensets – a battery bank, a distribution network, and associated metering and management hardware [29]. They are arbitrarily

<sup>1</sup>In addition, there are substantial variable surcharges including fuel price corrections that vary significantly with oil prices and hydrology.

<sup>2</sup>As of this writing, there is ongoing debate about the level of subsidy, and the proportion of the KES 75,000 connection fee that rural customers would pay may change rapidly.

distinguished by their size, with microgrids here referring to systems of 1-100 kW and minigrids referring to systems larger than 100 kW. If designed and operated properly, these systems can provide high power quality and reliability – however, this requires substantial data on variability of load and renewable resources or conservative overdesign of generation and storage components. In minigrids and larger microgrids, the larger load footprint and power ratings of generation and storage may allow accommodation of sharply transient productive loads like welding and grain milling. In Kenya, both small village-scale and large town-scale microgrid projects are being implemented by a variety of private and public entities. access:energy is a Kenya-based microgrid design and operation company that aims to provide affordable electricity to rural populations in East Africa [1]. It has installed 5 systems to serve remote communities in Western and Central Kenya. These solar PV and/or wind turbine systems are typically 1-10 kW in generation capacity and serve 10-100 people in a village. The distribution network of these microgrids is currently limited to a 100m radius due to distribution line losses. Much like the centralized grid option, access:energy consumers are charged a connection fee and then a consumption tariff. However, while the electricity is around KES 400/kWh (US\$4.55/kWh), the connection fee is in the low thousands of KES [1]. Therefore, while rural microgrid customers pay roughly twenty times more for every unit of electricity than their centralized grid counterparts, they are more easily able to afford the connection fee that allows them “first access.” A number of large towns in Kenya that are situated far from the national grid are served by larger-scale (130-3400 kW) minigrids built and operated by the utility, KPLC. These minigrids tend to be fully diesel, though recently some have been hybridized with small amounts of solar and wind [10].

## 4. STUDY METHODOLOGY

We develop estimates of induced residential demand for electricity using datasets describing socioeconomic, demographic, and household infrastructure characteristics present in Kenya. Though Kenya in 2008-2009 is the initial empirical vehicle for demonstrating our approach, we have developed the methodology in a way that is not specific to any particular country or point in time. In this section, we document our data sources as well as the methodology employed to arrive at estimates of current and induced residential electricity demand.

### 4.1 Data Sources

The data sources used in this study are briefly described in Table 1. Two principal sources of socioeconomic data were utilized for this analysis. *Exploring Kenya’s Inequality: Pulling Apart or Pooling Together?* [24], a joint publication of the Society for International Development and the Kenya National Bureau of Statistics (herein, “SID-KNBS”), includes information on household-level demographics, employment, education, and poverty indicators for each ward in Kenya (wards are the smallest administrative unit, with each covering roughly 30,000 people, and numbering 1455 in Kenya). These data were derived from the 2009 Kenyan census and the 2005-2006 Kenya Integrated Household and Budget Survey. Details on the disaggregation of the data from administrative units preceding a constitutional restructuring in 2010 to the current boundaries and the sampling

errors associated with these small areas are included in the report. Table 1 summarizes the types of features which are incorporated in our analysis. For brevity, a full description of the 100+ features in the dataset is omitted.

A subset of data from the 2008-2009 Kenya Demographic and Health Survey (herein, “DHS”) [15] is also utilized. Specifically, detailed household-level information is available from ~9000 households across ~400 sample sites nationally on electricity access and ownership of a number of electricity-related assets. These data are summarized in Table 1. The 2008-2009 survey represents the most recent information available, though a forthcoming 2014 survey can be used to update results.

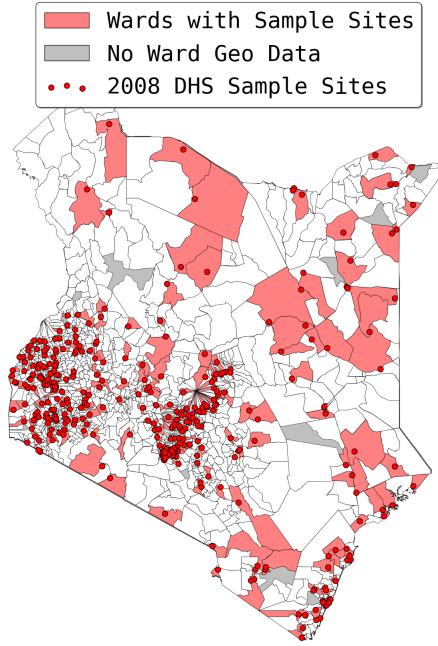
In addition to these socioeconomic and electricity-related data, geographic data describing the ward boundaries of Kenya were utilized for visualization purposes and to derive secondary data such as population densities. These geographic data were compiled by a private consultant [19] from sources available on the website of the Independent Electoral and Boundaries Commission of Kenya.

## 4.2 Data Transformation

The electricity access and asset ownership data from DHS were transformed into a format equivalent to that of the SID-KNBS socioeconomic data. That is, for each variable of interest, the transformed data is a proportion of households in each ward possessing a certain characteristic. This was accomplished by first aggregating the original binary variables (e.g. possession/non-possession of a television, access/non-access to electricity) for each household into a proportion of households at a given sample site (~400 across Kenya) that possess the relevant asset or have the relevant access. Next, the DHS sample sites were matched to their corresponding wards via their geographic coordinates. Where multiple sample sites exist within a single ward, the value for the ward was considered to be the mean of the associated sample site values for lack of a more rigorously justifiable method. A map of Kenya’s wards and the locations of the 2008 DHS sample sites (with corresponding wards highlighted) is given in Figure 2. After necessary data cleaning, 1401 wards of the original 1455 remain. This ~4% attrition is due to missing socioeconomic or geospatial data or to irreconcilable differences in placenames between the datasets.

## 4.3 Analytical Methods

This work seeks to estimate, at a fine geographic resolution, the demand for residential electricity services (and implicitly, the ability to pay for such services) under two scenarios: (1) current levels of electricity access; and (2) expanded access to electricity services in localities which do not currently have access. The approach employed here is to develop estimates of electricity-consuming appliance ownership as a proxy for an economically sustainable level of residential electricity demand under each of these scenarios. We will refer to these proxies as *current ownership* and *total ownership*, respectively, for the current access and expanded access scenarios outlined above. *Induced ownership*, the additional appliance ownership one would expect when electricity is made available, is simply the difference between total ownership and current ownership (analogously, we will refer to *current demand*, *total demand*, and *induced demand* for electricity). To develop these proxy estimates, we employ k-nearest neighbors regression to predict detailed ownership



**Figure 2: Kenyan Wards and 2008 DHS Sample Sites.**

information across a finer and wider geographic basis than the DHS using socioeconomic similarity from the SID-KNBS data.

Implicit in this approach is the assumption that localities that share socioeconomic characteristics will also have similar demand for electricity services and similar ability to pay for them. Additionally, this assumes that electricity prices are uniform across the locations where ownership observations are available (in this case, the DHS sample sites). This price uniformity will likely not be the case in a world with a diverse range of potential technologies and business models for rural electrification, and Section 5.5 in this work will address the ways in which our results may be interpreted for different electricity price regimes.

We choose to employ k-nearest neighbors regression for its simplicity and intuitive interpretation. Other supervised approaches, such as multivariate linear regression, were explored. The underlying structure of the data was found to be highly non-linear, and in the absence of a domain-informed rationale for more complex pre-supposed relationships between the dozens of socioeconomic characteristics and appliance ownership levels, this technique was abandoned. Similarly, k-means and hierarchical clustering techniques were explored to determine whether the wards form natural groupings by their socioeconomic characteristics. While the resulting clusters do build intuition about the socioeconomics of Kenya’s wards, the distributions of the data are rather continuous (rather than tightly clustered) so the validity of any hard-assignment clustering techniques is dubious. Lastly, while principal components analysis was explored, the small datasets obviated the need to compress the data for computational reasons, and domain knowledge-driven feature reduction is preferred for interpretability reasons.

Data Source	Feature Types Used	Year	Reference
SID-KNBS	Water Source, Human Waste Disposal, Lighting Fuel, Cooking Fuel, Floor Material, Wall Material, Roof Material, Demographics, Employment	2009	[24]
DHS	Appliance Ownership & Electricity Access	2009	[15]
IEBC	Ward Boundaries	2012	[19]

Table 1: Data Sources

#### 4.3.1 Defining Socioeconomic Similarity

In defining a similarity metric for k-nearest neighbors regression, care must be taken to include characteristics that impact the quantities of interest but not to allow regional differences that are exclusively a function of geography to skew the analysis. As an example, one might expect that economic status is an important predictor of asset ownership and energy appetite and that home building materials could be an indicator of this status. However, households in one region may build with grasses or reeds where those materials are widely available, while households of similar socioeconomic status in another region may use mud and dung because of different soils and a prevalence of cattle.

To address this challenge, features in the socioeconomic dataset are aggregated into similar classes (e.g. natural building materials, improved sanitation) in an attempt to account for the heterogeneity of poverty without obscuring important differentiators that may impact energy behaviors. Wherever possible, these aggregation choices follow accepted standard definitions in the development community [15] [24] [27].

Additionally, a number of extensive properties that are dependent on the absolute size or population of a ward are transformed into intensive properties (via normalization by population, ward area, etc.) to facilitate comparison across wards of somewhat arbitrary boundaries. Lastly, redundant features (those that are repeated or not linearly independent from the others) are removed from the dataset so that they do not contribute disproportionately to the determination of socioeconomic similarity. These choices of feature aggregation are summarized in Table 2 while choices of feature reduction and normalization are omitted for brevity.

Next, each of the features are translated by their mean value and scaled by their standard deviation so that the data take the form of z-scores. This is a common choice of feature standardization that facilitates comparison based on the underlying structure of the data rather than the absolute breadths of the feature distributions. Lastly, each of the feature classes is scaled by the number of features in the feature class (e.g. three features in the “roof material” class: natural, rudimentary, and finished). This is a design choice, and reflects the fact that no more rigorous method for determining the importance of the various data in predicting appliance ownership is known. Some efforts were undertaken to establish the predictive power of the socioeconomic data in this regard via principal components analysis and multivariate regression, but further work is needed to establish a conclusive answer.

These choices of feature aggregation and reduction result in a 42-dimensional feature space for the socioeconomic dataset (each ward  $i$  is described by an observation vector  $x_i \in \mathbb{R}^{42}$ ). We define the socioeconomic distance, or dissimilarity, between two wards to be the L-1 norm (Manhattan

length) of the difference between the two vectors that describe the ward characteristics.

#### 4.3.2 k-Nearest Neighbors Regression

With this distance metric in hand, k-nearest neighbors regression can be performed. The training set comprises the socioeconomic data for wards for which labels exist (asset ownership and electricity access data from the DHS) and the labels themselves. We refer to these wards as  $w_{obs}$  and the associated socioeconomic data and labels as  $X_{obs}$  and  $y_{obs}$  respectively (the notation ‘obs’ refers to observations). The test set comprises the socioeconomic data for wards where no labels are known. We refer to these wards as  $w_{test}$  and the associated socioeconomic data as  $X_{test}$ . For each ward in the test set, the label is estimated to be the average of the labels of the  $k_{nn}$  nearest neighboring wards in the training set, where ‘nearest’ refers to those with the least socioeconomic distance from the ward at hand.

The value of  $k_{nn}$  in the nearest neighbors regression algorithm is chosen via k-folds cross validation. In this non-exhaustive technique, the original training set is randomly partitioned into  $k_f$  subsets. One at a time, each of the  $k_f$  subsets are withheld from the training set, and the regression is performed using the withheld subset as the test set. The root mean square prediction error in accurately predicting the label values for each of these subsets is averaged and recorded. This is repeated across a range of potential  $k_{nn}$  values, and  $k_{nn}$  is chosen to be the value that minimizes this error metric.

The number of subset combinations needed for exhaustive cross-validation is intractably large, but the random partitioning in the non-exhaustive k-folds cross validation can produce inconsistent results. As a compromise between speed and accuracy, we perform this validation across a range of  $k_f$  values and a number of random seedings for the partitioning process and average the results to arrive at a consistent choice for  $k_{nn}$ . This approach is illustrated graphically in Figure 3. The error metric described above is plotted against potential values for  $k_{nn}$ . This relationship is plotted for a range of  $k_f$  values and the entire process is conducted for 20 different random seedings (hence the multiple lines for each  $k_f$  value). The average error metric across all of these  $k_f$  values and all of the random seedings is also shown. In the case presented in Figure 3, we choose  $k_{nn} = 20$  to minimize this average error.

#### 4.3.3 Current & Induced Ownership Estimation

Once the data have been imported and transformed, and the features in the data have been aggregated, normalized, selected, and scaled as discussed above, k-NN regression can be directly used to ascertain an estimate for current ownership of the relevant electricity-consuming appliances. Specifically, the current ownership levels are predicted for the test set socioeconomic vectors  $X_{test}$  using the training set vectors

Feature Class	Aggregated Feature	Original Features
CookingFuel	Transitional	Paraffin, Charcoal
CookingFuel	Advanced	LPG, Biogas, Electricity
FloorMaterial	Finished	Tiles, Cement
LightingFuel	Kerosene or Paraffin	PressureLamp, Lantern, TinLamp, GasLamp
RoofMaterial	Natural	Makuti, Grass, Mud or Dung
RoofMaterial	Rudimentary	Corrugated Iron Sheets, Tin
RoofMaterial	Finished	Tiles, Concrete, Asbestos Sheets
WallMaterial	Rudimentary	Mud and Wood, Mud and Cement, Wood Only, Corrugated Iron Sheets, Tin
WallMaterial	Finished	Stone, Brick or Block
WasteDisposal	Unimproved Sanitation	Pit Latrine Uncovered, Bucket, Bush
WasteDisposal	Improved Sanitation	Main Sewer, Septic Tank, Cess Pool, VIP Latrine, Pit Latrine
WaterSource	Surface Water	Pond, Dam, Lake, Stream or River
WaterSource	Unimproved Sources	Unprotected Spring, Unprotected Well, Water Vendor
WaterSource	Improved Sources	Protected Spring, Protected Well, Borehole, Piped, Rainwater Collection, Jabia

Table 2: Feature Aggregation

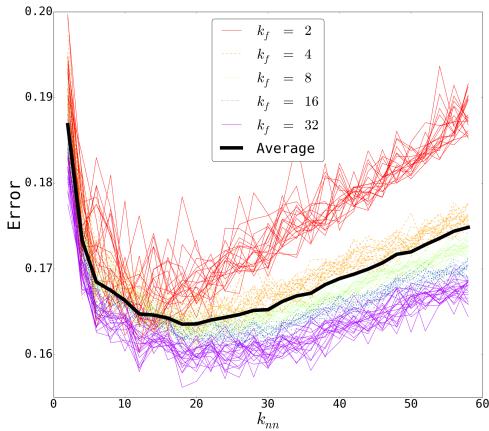


Figure 3: Example of k-folds cross validation for prediction of current television ownership (choose  $k_{nn} = 20$ )

and labels  $X_{obs}$  and  $y_{obs}$ . However, to produce an estimate for induced ownership, a more involved process is required.

First, electricity access is predicted across  $w_{test}$  via the k-NN algorithm using  $w_{obs}$  as the training set. More specifically, the estimated electricity access  $\hat{y}_{test,e}$  is predicted for the socioeconomic vectors in  $X_{test}$  using the vectors and labels  $X_{obs}$  and  $y_{obs,e}$ . Next, combining the observed and predicted electricity access data, the wards are divided into those ‘with’ and ‘without’ electricity access as defined by a threshold proportion of households with electricity access. This threshold is chosen to be 10% in this analysis, which reflects a balance between domain knowledge considerations and limited availability of data. This threshold is reasonable despite its ostensibly low value because of the low penetration rates of actual connections that often exist in rural areas to which KPLC’s distribution infrastructure extends. A higher choice of threshold value would also lead to an unacceptably small set of wards that are deemed to have electricity access for the next step in the induced ownership estimation.

Next, the original training set,  $w_{obs}$ , is partitioned into wards without electricity access,  $w_{obs,no}$  and wards with

	With Elec.	No Elec.	Total
Observation Set	121	211	332
Test Set	439	630	1069

Table 3: Data Partitioning Sample Sizes (Number of Wards)

electricity access,  $w_{obs,yes}$ . Similarly, the test set is partitioned into  $w_{test,no}$  and  $w_{test,yes}$ . The socioeconomic data and labels are partitioned according to the same nomenclature. The sizes of each subset of the data are given in Table 3. The total ownership for wards in  $w_{test,no}$  is predicted using the training set  $X_{obs,yes}$ ,  $y_{obs,yes}$ . Similarly, the total ownership for wards in  $w_{obs,no}$  is predicted using the same. As described in Section 4.3, the induced ownership is simply the difference between total ownership and current ownership.

The basic rationale for this approach is a slight extension of the core assumption of our methods: one would expect that a ward without electricity would have similar adoption of electrical appliances (need or desire for the services they provide and ability to pay for them) as a ward with electricity if the two closely share socioeconomic characteristics.

This approach treats electricity access as an exogenous factor. If an un-electrified village were to become electrified, its socioeconomic characteristics and the ownership of electricity-consuming appliances would not change overnight. However, there would likely be some appliance adoption transient that depends on the costs (in time, money, and inconvenience) of the incumbent energy sources, the costs of electricity, public awareness, the psychology of behavior change, and other factors. The estimates made here pertain to a steady-state level of ownership, once this transient has decayed. One might suppose that in some communities, the DHS may have been conducted shortly after electrification and thus during this transient period. This phenomenon is likely rare if it exists at all, and it is ignored here.

#### 4.4 Residential Energy Usage Model

For the purpose of illustrating the end-to-end use of our approach, we employ a rudimentary model for translating electricity-consuming appliance ownership levels into resi-

dential electricity demand. Knowledge of this demand figure is essential in evaluating the viability of various technology options and business models for rural electrification, whether evaluating potential tariff structures or specifying and costing equipment.

Based on evidence from rural Kenyan households in the literature [13], we assume a daily energy use of  $210\text{ Wh/day}$  for a 14-inch color TV,  $960\text{ Wh/day}$  for a small refrigerator,  $6\text{ Wh/day}$  for a radio, and  $6\text{ Wh/day}$  for a mobile phone. These appliance consumption levels are translated into an average household daily electricity demand for each ward using predicted and observed appliance ownership levels.

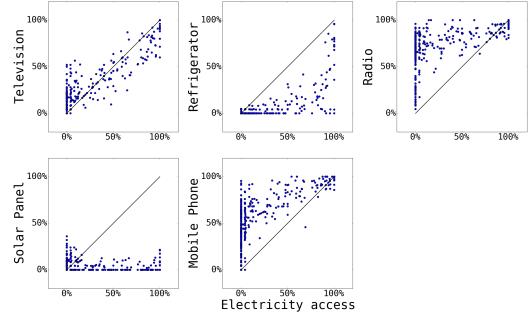
## 5. RESULTS & DISCUSSION

### 5.1 Observed Trends in Ownership & Access

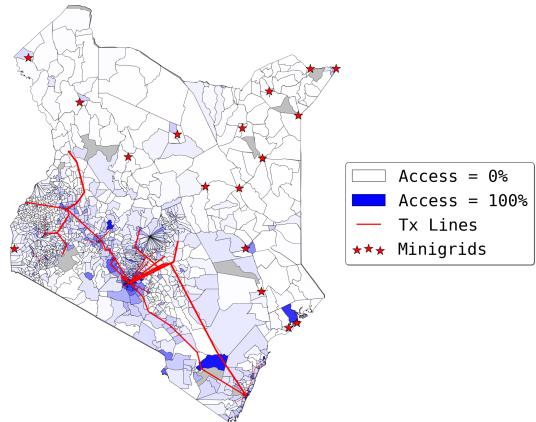
The relationship between reported appliance ownership (television, refrigerator, radio, mobile phone, and solar panel) and electricity access from the DHS is presented in Figure 4. The raw observations are transformed as described in Section 4.2 into a proportion of households in the ward with a given appliance or with electricity access. It should be noted that the DHS appears to define electricity access as connection to an external power system (in this case, to the KPLC distribution system), and thus, household possession of solar panels does not constitute electricity access in this dataset.

These data reflect important trends about the ‘appliance ladder,’ which describes the order and manner in which electrified households acquire electricity-consuming appliances [9]. Radio and mobile phone ownership is often in excess of electricity access, which suggests that people use batteries and charging services (for instance, from a shop in a nearby electrified town) to power these devices. Levels of refrigerator ownership are significantly below electricity access levels (except for a handful of affluent urban wards), which reflects the high capital cost of the appliance, the relatively high electricity consumption, and the high reliability of electricity service needed to make cold storage practical. Comparing the television ownership with the other appliances, one observes that TVs are more common than refrigerators, but less common than radios and mobile phones. This reflects the moderate capital cost and energy consumption and the less stringent reliability requirements as compared to cold storage. Additionally, ownership levels exceeding electricity access suggest that televisions are sometimes powered by batteries or perhaps larger solar/battery systems, and it suggests a strong desire for television viewing (given the high cost or inconvenience of these approaches to powering TVs).

The insights from the solar panel ownership data are less clear, as the technology and costs of solar photovoltaic systems have changed significantly in the years since the survey was conducted. However, recalling the narrow DHS definition of electricity access, one observes that the highest solar panel ownership occurs in wards with little to no electricity access. The presence of non-zero ownership in wards with some electricity access suggests either fuel-stacking (where households utilize multiple energy sources to enhance reliability or to navigate fluctuating prices, as for kerosene or charcoal) or it suggests a social effect in electricity usage: households that may not be able to afford a connection to the utility nevertheless become aware of electricity and its



**Figure 4: Observed Appliance Ownership Versus Electricity Access**



**Figure 5: Electricity Access**

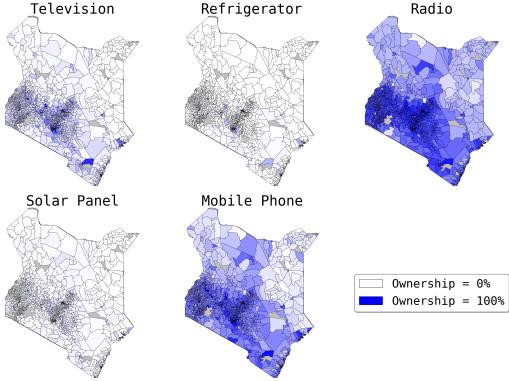
uses and acquire solar panels for low energy applications (LED lighting, radio/mobile phone charging).

### 5.2 Electricity Access

Predicted and observed electricity access are given in Figure 5. For wards with DHS data, the observed access is shown, while for wards without, the predicted access is shown. The Kenyan electricity transmission network down to  $33\text{kV}$  [5] and the locations of most minigrids [10] [14] are overlaid. Sources indicate that most minigrids predate the 2008 DHS, though the exact commissioning dates are not available. Wards with significant predicted electricity access are mostly in proximity to the transmission network (the  $11\text{kV}$  distribution network is not shown, but its extent outwards from the transmission backbone is limited by loss or voltage drop considerations) or to known minigrids. A few border towns appear to be connected to neighboring power systems in Ethiopia and Tanzania (Moyale, Taveta, Oloitokitok).

### 5.3 Current Ownership

The observed and predicted current ownership levels are plotted for each ward in Figure 6. For wards with DHS data, the observed ownership is shown, while for wards without, the predicted ownership is shown. Mobile phone and radio ownership is widespread, though noticeably higher in the western and central regions, parts of the southern Rift Valley, and the cities and large towns of the coast and northern regions. Refrigerator ownership is low overall and concen-



**Figure 6: Current Ownership**

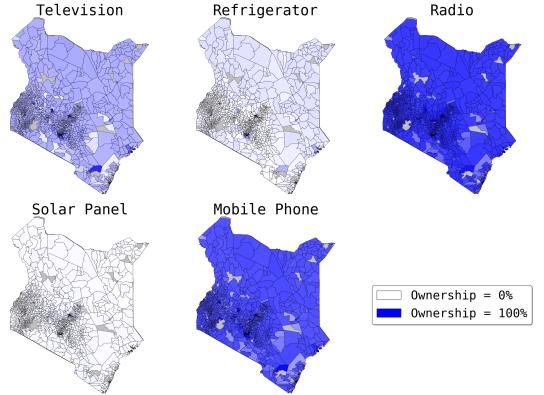
trated in urban areas. Television ownership is moderate and follows a similar regional distribution as mobile phones and radios. All of these appliances show higher concentrations in the large towns and small cities of the north and coast than the surrounding rural areas. Most of these locations, which include Wajir, Garissa, Marsabit, Lodwar, Mandera, and others, are sites of minigrids that provide traditional electricity services despite their distance from Kenya’s national grid [10] [14]. Solar panel ownership is limited, and concentrated in rural parts of western and central Kenya.

A number of large-area wards that are fairly sparsely populated appear to be outliers due to high ownership and access levels compared to their similarly sparse neighboring wards. Closer investigation reveals that these wards have towns of significant size, and the population is in fact relatively urban and concentrated in one part of the ward. These locations include Nyahururu, which sits at a major node in the electricity transmission system, Voi, a major town on the road, rail, and electricity transmission corridor between Nairobi and Mombasa, and parts of Lamu, a tourist destination and site of a KenGen-owned minigrid.

#### 5.4 Total Ownership

Total ownership is shown in Figure 7. For wards with electricity access levels already above the threshold value, current ownership is equivalent to total ownership and is shown here. For all other wards, the model-predicted value of total ownership is shown. Despite differences in absolute ownership level, appliances share a high degree of homogeneity in predictions across wards. As one would expect significant socioeconomic differences to result in a broader distribution of predicted ownership levels, these results call into question the validity of the analysis, a topic we explore below.

To explore the validity of induced ownership predictions, we examine the relative socioeconomic distance to the nearest neighbors with and without electricity access. One would expect that for a ward without electricity (the wards of concern here,  $w_{test,no}$  and  $w_{obs,no}$ ) the distance to the nearest wards of any kind should be less than that to the nearest wards with electricity. However, unless the latter distance is vastly greater than the former, the estimation of total ownership is likely reasonable. Conversely, a much greater distance to electrified wards would indicate that the prediction based on electrified wards is so inaccurate relative



**Figure 7: Total (Current + Induced) Ownership**

to prediction based on all wards that the results are not of practical value.

Figure 8 presents one possible validation metric. On the lefthand side, each dot represents one ward. The value on the x-axis is the average ‘fractional distance’ to a given ward’s  $k$ -nearest neighbors of any kind, while the value on the y-axis is the average ‘fractional distance’ to the ward’s  $k$ -nearest neighbors with electricity. Fractional distance  $d_f$  is defined as  $d_f(d) = (d - d_{min}) / (d_{max} - d_{min})$  where  $d_{max}$  is the distance from the ward to its furthest neighbor and  $d_{min}$  is the distance to its closest neighbor. Here we observe that as expected, distance to nearest neighbors with electricity is always further than distance to nearest neighbors of any kind for this set of wards that does not have electricity. The ratio of the latter distance to the former, which should be an indication of the validity of the induced demand approach for this dataset, ranges from roughly 1.2:1 to 4.6:1. The right-hand side of Figure 8 presents the ratio of these distances for each relevant ward in geographic form. We observe that wards in the former central province and along the infrastructure-rich Nairobi-Mombasa corridor often have the lowest values because of the abundance of nearby electrified wards with which they have significant socioeconomic similarity (this similarity is not explicitly demonstrated here for brevity, but visualization of the various socioeconomic parameters and clustering analysis confirms this assertion).

Further investigation that is omitted here for brevity indicates that a paucity of appliance ownership data for electrified wards in certain regions is the root cause of the homogeneity of induced demand predictions (even aggregating features to control for heterogeneity of poverty, the socioeconomic characteristics are quite regional, reflecting significant geographic patterns in incomes and access to improved water and sanitation). For rural areas of much of the coast and northern regions and for rural areas far from the cities in the western and Rift Valley regions, there are so few similar wards with electricity access that the prediction from  $k$ -nearest regression is overwhelmed by electrified wards that are in fact rather dissimilar. This is because the values of  $k_{nn}$  determined via cross validation are quite high relative to the total number of electrified wards to learn from (these values are given in Table 4). The current simple cross validation approach presents an objective and repeatable approach to choosing the tunable parameter  $k_{nn}$ , but a more

Predict	Current	Induced	Induced
From	All Obs.	Obs. with Elec.	Obs. with Elec.
For	All Test	Test no Elec.	Obs. no Elec.
Television	20	22	18
Refrigerator	14	12	14
Radio	18	24	26
Solar Panel	34	12	12
Mobile Phone	18	22	28
Electricity	12	n/a	n/a

Table 4:  $k_{nn}$  values determined by cross-validation

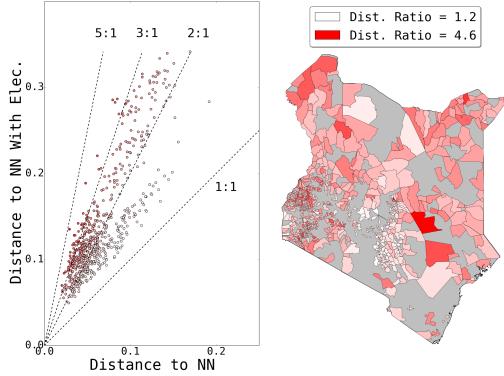


Figure 8: Validation of Total Demand Estimation Methodology

nuanced approach could potentially yield better results from the same limited input data.

It should be noted that while this does not appear to be a conceptual limitation of our induced ownership estimation method, it is a practical limitation. Data for other countries with low rural electrification may be similarly lacking. In the case of Kenya specifically, model performance should improve with use of the forthcoming 2014 DHS because there has been significant progress in rural electrification in the intervening years.

Despite the challenges discussed above, we present the ownership predictions in another form in Figure 9 to illustrate the interpretation of results produced via this method. For each ward without electricity, total ownership is plotted against current ownership (observed or predicted, depending on the ward). The high, tight distributions for mobile phones and radios in total ownership suggest that regardless of the current, unelectrified ownership level, these appliances would be adopted broadly and that their adoption is nearly independent of socioeconomic factors. The conclusions regarding television ownership are nearly the same except for the lower average ownership level. The modest total ownership in the refrigerator data (a bit wider in absolute terms, much wider in relative terms) and its concentration on the low current ownership end reflects the near non-existence of residential cold storage in unelectrified wards, and limited ownership in similar electrified areas (likely rural and relatively poor). Lastly, the most marked trend in the solar panel results is lower ownership levels in the electrified case – this derives from the definition of electricity access as a grid connection in the DHS.

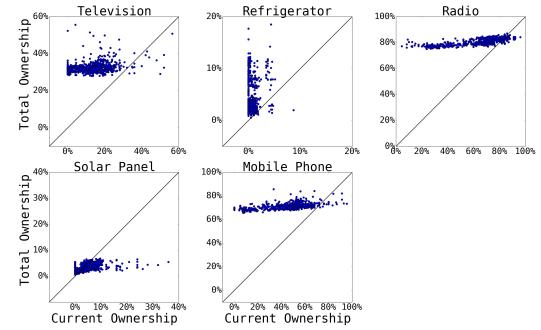


Figure 9: Current & Total Ownership

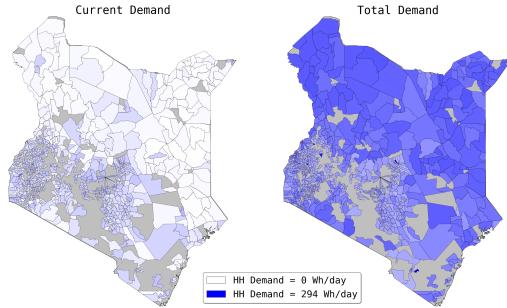


Figure 10: Household Demand Estimation (Currently Un-Electrified Wards)

## 5.5 Current & Total Residential Demand

Translating the current and total ownership levels into estimates of residential electricity demand yields the results in Figure 10. Due to the significant difference in ownership levels between urban and rural wards (and corresponding difference in demand estimates), currently electrified wards are omitted to maximize the range of demand levels that can be expressed on a linear color scale. As noted above, the estimates of induced demand for certain regions are likely of limited accuracy. Nevertheless, this demonstrates the end-to-end methodology of our data-driven approach to electricity demand prediction.

The implicit assumption in this supervised learning process that electricity prices are uniform (as discussed in Section 4.3) presents another challenge. When evaluating potential technology options and business models for sustainable and scalable rural electrification, prices are likely to markedly vary for different strategies. One possible way to address this challenge is to transform the ownership level estimates into a total household budget for the services these appliances provide. By coupling this budget with data on the ownership costs of these various appliances and with domain-knowledge about the appliance ladder, these ownership estimates (derived under one price assumption) can be transformed into ownership estimates under arbitrary electricity rate scenarios.

## 6. CONCLUSION

This work represents a step towards understanding the potential for novel technologies and strategies to enable rural electrification. We have used an approach that draws

from socioeconomic, demographic, geospatial, and domain-relevant data to build a model of induced residential demand for electricity in Kenya. This model helps to address an important gap: understanding future demand for electricity is essential for evaluating the wide range of technologies and business models in this space. Continuing in this direction, we recognize that there is much more to the problem of understanding future electric demand. We aim to use a similar approach to understand the potential for growth in electricity demand for commercial purposes by analyzing specific industries and business types that emerge in these communities as electricity becomes available. Further, we aim to build a tool for various public and private entities to employ our model to make business, funding, and policy decisions. With refinement, we believe that this type of approach may be relevant in other domains as well, such as water and waste management, and in other countries beyond Kenya.

## 7. REFERENCES

- [1] access:energy. <http://accessenergy.org/>.
- [2] M-KOPA Solar. <http://www.m-kopa.com/>.
- [3] M-kopa solar powers up 100,000th home in east africa. <http://www.m-kopa.com/media/latest-news>.
- [4] S. Abdullah and P. Wilner. Willingness to pay for renewable energy: Evidence from contingent valuation survey in kenya. *Renewable and Sustainable Energy Reviews*, 15, 2011.
- [5] Africa Infrastructure Knowledge Program. Kenya Electricity Transmission Data. <http://www.infrastructureafrica.org/library/doc/734/kenya-electricity-transmission-data>.
- [6] P. Alstone, K. Radecsky, A. Jacobson, and E. Mills. Field study methods and results from a market trial of night market vendors in rural kenya. *Light and Engineering*, 22(2):23–37, 2014.
- [7] R. Banerjee, R. Pandey, and A. Bhure. Electricity Demand Estimation for Village Electrification, 2000.
- [8] Breakthrough Institute. Our High Energy Planet: A Climate Pragmatism Project, 2014.
- [9] Catherine Wolfram and Orie Shelef and Paul Gertler. How Will Energy Demand Develop in the Developing World?, 2012.
- [10] Eng. Henry Gichungi. Mini Grid PV Business Opportunities in Kenya. <http://www.giz.de/fachexpertise/downloads/2013-en-gichungi-pep-workshop-pv-kenia.pdf>.
- [11] International Energy Agency. World Energy Outlook 2012, 2012.
- [12] International Energy Agency. Global Tracking Framework, 2013.
- [13] A. Jacobson. The Market for Micro-Power: Social Uses for Solar Power in Kenya. [http://users.humboldt.edu/arne/AJacobson\\_TegemeoWP9\\_v2.pdf](http://users.humboldt.edu/arne/AJacobson_TegemeoWP9_v2.pdf).
- [14] KenGen. Off Grid Stations. <http://www.kengen.co.ke/index.php?page=business&subpage=offgrid>.
- [15] Kenya National Bureau of Statistics (KNBS) and ICF Macro. Kenya demographic and health survey 2008-09.
- [16] C. Kirubi, A. Jacobson, D. Kammen, et al. Community Based Electric Micro-grids Can Contribute to Rural Development: Evidence from Kenya. *World Development*, 37(7), 2008.
- [17] K. Lee, E. Brewer, C. Christiano, F. Meyo, E. Miguel, M. Podolsky, J. Rosa, and C. Wolfram. Barriers to Electrification for “Under Grid” Households in Rural Kenya. Working Paper 20327, National Bureau of Economic Research, July 2014.
- [18] T. Levin and V. Thomas. Least-cost network evaluation of centralized and decentralized contributions to global electrification. *Energy Policy*, 2012.
- [19] M. Maron. Kenya Election Open Data. <https://github.com/mikelmanon/kenya-election-data/blob/master/output/ward.results.geojson>.
- [20] L. Parshall, D. Pillai, S. Mohan, A. Sanoh, and V. Modi. National electricity planning in settings with low pre-existing grid coverage: Development of a spatial model and case study of Kenya. *Energy Policy*, 37(6), 2009.
- [21] Regulus. Electricity Cost in Kenya. <https://regulusweb.com/kenyan-electricity-costs>.
- [22] Schnitzer, D. and Shinde Lounsbury, D. and Carvallo, J.P and others. Microgrids for Rural Electrification: A Critical Review of Best Practices Based on Seven Case Studies., 2014.
- [23] Sierra Club. Clean Energy Services for All: Financing Universal Electrification, 2014.
- [24] Society for International Development. Exploring Kenya’s Inequality: Pulling Apart or Pooling Together?, 2013.
- [25] S. Szabo, K. Bodis, T. Huld, and M. Moner-Girona. Energy solutions in rural Africa: mapping electrification costs of distributed solar and diesel generation versus grid extension. *Environmental Research Letters*, 2011.
- [26] C. Thom. Use of grid electricity by rural households in south africa. *Energy for Sustainable Development*, IV(4), 2011.
- [27] WHO/UNICEF. Improved and unimproved water and sanitation facilities. <http://www.wssinfo.org/definitions-methods/watsan-categories/>.
- [28] Wilson, M. and Jones, J.B. and Audinet, P. A New Slant on Slopes: Measuring the Benefits of Increased Electricity Access in Developing Countries., 2010.
- [29] Woodrow Wilson School of Public and International Affairs. Rural Energy Alternatives in India: Opportunities in Financing and Community Engagement for Renewable Energy Microgrid Projects. <https://wws.princeton.edu/sites/default/files/content/591f%20Rural%20Energy%20Alternatives%20in%20India.pdf>.
- [30] World Bank. Data Bank. <http://data.worldbank.org/>.
- [31] A. Zvoleff, A. Kocaman, W. Huh, and V. Modi. The impact of geography on energy infrastructure costs. *Energy Policy*, 2009.