

Article

Electricity Consumption Forecast Model Using Household Income: Case Study in Tanzania

Kyohei Shibano * and Gento Mogi

Department of Technology Management for Innovation, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan; mogi@tmi.t.u-tokyo.ac.jp

* Correspondence: shibano@tmi.t.u-tokyo.ac.jp

Received: 20 March 2020; Accepted: 12 May 2020; Published: 15 May 2020



Abstract: When considering the electrification of a particular region in developing country, the electricity consumption in that region must be estimated. In sub-Saharan Africa, which is one of the areas with the lowest electrification rates in the world, the villages of minority groups are scattered over a vast area of land, so electrification using distributed generators is being actively studied. Specifically, constructing a microgrid or introducing a solar system to each household is being considered. In this case, the electricity consumption of each area needs to be estimated, then a system with enough capacity could be introduced. In this study, we propose a household income electricity consumption model to estimate the electricity consumption of a specific area. We first estimate the electricity consumption of each household based on income and the electricity consumption of a specific area can be derived by adding up them in that area. Through a case study in Tanzania, electricity consumption derived using this model was compared with electricity consumption published by TANESCO, and the validity of the model was verified. We forecasted the electricity consumption in each region using the household income electricity consumption model, and the average forecast accuracy was 74%. The accuracy was 87% when the electricity consumption in Tanzania mainland was forecasted by adding the predicted values.

Keywords: electricity consumption forecast; household electricity consumption; electrification; Tanzania

1. Introduction

1.1. Overview

In 2018, 860 million people, or approximately 11% of the world's population, do not have access to electricity. In comparison, 46% of Africa's population does not have access to electricity and, in particular, 55% of the people in sub-Saharan Africa do not have access to electricity. Sub-Saharan Africa is considered as one of the least electrified areas in the world [1]. In this study, we focused on sub-Saharan Africa and considered how to proceed with electrification. Research into electrification methods is therefore important, including how much electric energy is necessary in the electrified area. Equipment must be introduced in accordance with the electricity demand of the area.

In sub-Saharan Africa, where the population is spread over a large area, the introduction of distributed power sources into each region has often been considered. In such cases, the electricity consumption in a specific area instead of the whole country must be considered. In this study, we constructed a model (household income electricity consumption model) to estimate the electricity consumption according to household income. Electricity consumption in a specific area can be estimated by multiplying the number of households.

1.2. Literature Review

Many studies have estimated the electricity demand or consumption of a country, also for the purpose of electrification in developing countries. Several methods can be used to estimate electricity demand or consumption, and Bazilian et al. [2] categorized the methods into six types: Trend method, End-use method, Agent-based models, Time series method, Econometric method, and Neural network techniques. For the end-use method (or engineering based method), the following studies have been conducted. da Silva et al. [3] estimated the electricity consumption of the industrial sub-sector of pulp and paper by 2035 in Brazil based on the hierarchical structure using FORECAST software. Debnath et al. [4] estimated household energy consumption, but not limited to electricity, in Bangladesh based on the number of typical appliances owned and the operating hours. Regarding the econometric method, Amarawickrama and Hunt [5] used six econometric models to forecast peak demand and electricity consumption in Sri Lanka by 2025. In models using neural networks, Azadeh et al. [6] used an artificial neural network (ANN) to forecast annual electricity consumption of energy intensive manufacturing industries in Iran. Oğcu et al. [7] estimated monthly electricity demand in Turkey using ANN and support vector regression, and verified the prediction accuracy. Inglesi [8] calculated electricity consumption in South Africa by 2030 using the Engle–Granger methodology. Morales-Acevedo [9] estimated electricity consumption of 2050 in Mexico by using a model with the evolution of population, gross domestic product per capita, and energy intensity. Rosnes and Vennemo [10] expressed the electricity consumption of sub-Saharan Africa by the sum of market demand, suppressed demand, and new connections. McQueen et al. [11] assumed that each household's electricity demand follows the gamma distribution, and estimated the maximum electricity demand on distribution networks using Monte Carlo simulation. Regarding the relationship between income and electricity consumption discussed in this study, McNeil and Letschert [12] and Daioglou et al. [13] constructed a model that describes the relationship between household income and appliance ownership rate using the Gompertz curve. In order to determine the capacity of minigrids for electrification, there are methods of estimating the electricity consumption by survey, but it is pointed out that the accuracy is low.

Blodgett et al. [14] compared survey-predicted electrical energy use to actual measured consumption of customers of eight minigrids in rural Kenya. They showed that the predictions of the general survey approach were poor, which had a mean absolute error of 426 Wh/day per customer on a mean consumption of 113 Wh/day per customer and also an alternative data-driven proxy village approach, which uses average customer consumption from each minigrid to predict consumption at other minigrids, was more accurate and reduced the mean absolute error to 75 Wh/day per customer. Hartvigsson and Ahlgren [15] compared load profiles and performance metrics based on interviews and on measurements relating to a rural minigrid in Tanzania. This study showed distinct differences between load profiles based on interviews and measured data. Moreover, suggestions for how the interview process can be improved were presented.

The purpose of this study is to predict the electricity consumption necessary for electrification in a specific area. Many studies estimated electricity consumption for a whole country and these methods are not suitable for forecasting electricity consumption in a specific area of a country. The method of forecasting electricity consumption based on the historical data of electricity consumption cannot be applied to new areas that are not electrified yet or where no historical data exist. One study, in which the electricity consumption of the household was estimated from the number of owned home appliances [4], has been used as a reference. For the relationship between income and the number of home appliances owned, we referred to the models in [12,13]. The electricity consumption of each appliance was assumed to follow the gamma distribution; the demand at each home was also assumed to follow the gamma distribution in [11]. In order to estimate electricity consumption using the proposed model, it is sufficient to conduct a survey of income to the target area, and a survey of income and number of appliances owned for already electrified areas similar to the target area.

2. Modelling Electricity Consumption per Household by Income

We developed a new model for estimating annual electricity consumption of a certain area, named the household income electricity consumption model. By summing the electricity consumption of some group of households, we estimated the electricity consumption of the group. Figure 1 depicts a flowchart of the model.

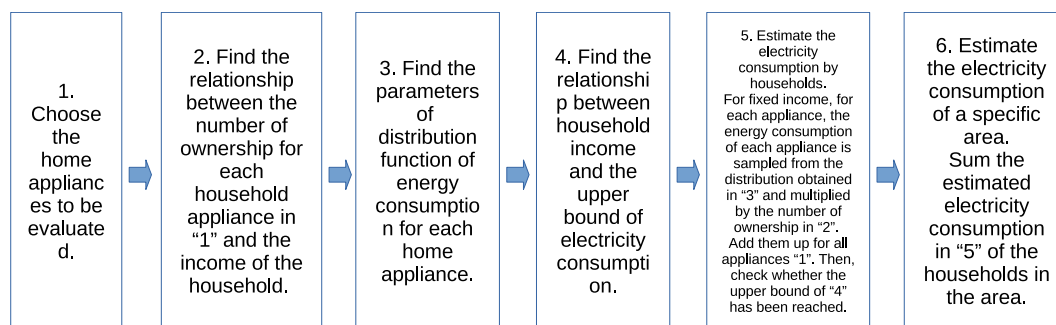


Figure 1. Flowchart of the model.

Our model uses statistical data to consider the change in the number of appliances owned as income increases and derives electricity consumption as a probability distribution. To construct the household income electricity consumption model, we required data on household income, food expenditure, and the number of electric appliances owned in a target area. First, we derived the relationship between household income and the number of owned household appliances. We assumed that the annual energy consumption of each household's appliances follows the gamma distribution, as the frequency of use, the amount of electricity consumed, and preferences do not change with income. To calculate annual electricity consumption, we considered the rated output of home appliances and the amount of time they are used in a year; we assumed, in this model, that neither will change due to income. Let E be the set of electric appliances to be evaluated in the model. The elements of E are refrigerator, washing machine, TV, and so on. Although it is possible to formulate a model that can accurately estimate electricity consumption if E consists of as many products as possible, there is a limit to the actual data that can be acquired. However, it is sufficient if E covers products that are generally owned by high-income groups in the target area and have a significant influence on household electricity consumption. Appliances with high electricity consumption that are not owned by many households in the high-income group do not need to be included in E , and conversely, appliances that are owned by many households in the high-income group with quite low electricity consumption are also not included in E . Therefore, when constructing the model, E must consider whether it contains sufficient appliances.

2.1. Appliance Ownership Model

We modeled what kind of appliances households tend to have according to income in the target area. In [12,13], the relationship between household income (or expenditure) and the ownership rate of electric appliances was examined. The rate was approximated using the Gompertz curve. In this study, we used the Gompertz curve based on these results. For household income I and arbitrary $a \in E$, the number of owned appliances per household $S(a, I)$ is given by:

$$S(a, I) = S_{max}(a) \times (1 - e^{-\alpha I})^\beta \quad (1)$$

where α and β are coefficients that determine the shape of the Gompertz curve. $S(a, I)$ increases monotonically and converges to $S_{max}(a)$ as I increases. $S_{max}(a)$ is the number of appliances in a household when it has sufficient income. In this study, $S_{max}(a)$ is the average number of a owned by

the group of the top several percentages of income, for example, 1% of the incomes in the data given. We perform a regression using Equation (1) to obtain α and β .

2.2. Electricity Consumption per Household Estimation Model

We assumed that the annual electricity consumption of households (kWh/year/hh) for each $a \in E$ follows a gamma distribution. Among continuous probability distributions, we examined the appropriate one to express the annual energy consumption. Even for the same appliance, differences exist in the usage time for each household and in the rated output of different products. Therefore, the annual electricity consumption varies by households, but it is at least zero, which is the case of no use. We assumed that households with a large number of hours of annual use and a large rated output of the products owned therein will have a considerably large annual electricity consumption. However, this is limited to a few households, and the greater the electricity consumption, the fewer such households. Many households are expected to consume a similar amount of electricity each year. Based on the above discussion, two non-negative and long-tail continuous probability distributions—the gamma distribution and the log-normal distribution—were considered. However, the log-normal distribution tends to have a relatively thicker base so is more likely to produce unusually large values of annual energy consumption.

To simplify the model, we assumed that the electricity consumption of each home appliance is independent. A gamma distribution $Gamma(\kappa, \theta)$ is a type of continuous probability distribution for which the expected value of the variable in question is greater than 0. It is characterized by two parameters—a shape parameter $\kappa > 0$ and a scale parameter $\theta > 0$ —and the probability density function is characterized by

$$f(x) = x^{\kappa-1} \frac{e^{-x/\theta}}{\Gamma(\kappa)\theta^\kappa} \quad \forall x > 0 \quad (2)$$

where $\Gamma(\kappa)$ is the gamma function. κ and θ are coefficients that determine the shape of the probability density function. A distribution function is expressed by

$$F(x) = \int_0^x f(u)du = \frac{\gamma(\kappa, x/\theta)}{\Gamma(\kappa)}, \quad (3)$$

where $\gamma(\kappa, x/\theta)$ is an incomplete gamma function. In a gamma distribution,

$$\begin{aligned} \mu &= \kappa\theta \\ \sigma &= \sqrt{\kappa}\theta \end{aligned}$$

and mean μ and SD σ derive κ and θ .

Once κ_a, θ_a is found for each a , the distribution $Gamma(\kappa_a, \theta_a)$ of electricity consumption is determined. As we obtain the number of ownership for each a and income I from the appliance ownership model (Section 2.1) $S(a, I)$, the household electricity consumption $D(I)$ (kWh/year/hh) with income I is

$$D(I) = \sum_{a \in E} S(a, I) \times X_a \quad (\text{kWh/year/hh}), \quad (4)$$

where sampling $X_a \sim Gamma(\kappa_a, \theta_a)$ for $a \in E$. Though it is possible for a relationship to exist between income and electricity consumption for some appliances, we assumed there is no relationship between them for each appliance to simplify the model.

However, the electric electricity consumption above is insufficient as a sample. We must set an upper bound on electricity consumption defined by owned appliances and by payable money for the electricity bill. The former is simple and the upper bound is 24×365 times the rated output of the owned all home appliances. For the latter, the upper bound is the amount of household income

minus food expenditure. If an amount exceeding the upper bound has been sampled, this is solved by resampling from uniform distribution (In resampling, we assume uniform distribution from zero to the upper bound. When sampling from the same gamma distribution, it has a possibility not to obtain a value below the upper bound even after repeated so many trials.). To use electricity, households have to pay for the electricity bill (In case households purchase SHS (Solar Home System) at home, households also have to pay for the cost of SHS. In this paper, we call the necessary cost to use electricity “electricity bill”.) in reality. The amount of electricity consumption derived only from the number of appliances owned tends to be too large, and the upper bound derived from electricity bill expenditure should be considered as set. If we plan to invest equipment for electrification based on electricity consumption derived without the upper bound, then there will be an equipment overinvestment. The purpose of setting the upper bound is to avoid overinvestment by setting the upper bound in consideration of electricity bill expenditure. The upper bound obtained here is better to be small in the set of all upper bounds. On the other hand, if the upper bound value obtained here is lower out of the set of upper bounds and we plan to invest based on this, the amount of electricity will be insufficient. Therefore, the upper bound should be set higher.

3. Case Study on Mainland Tanzania

We estimated the electricity consumption in each region using actual income data from mainland Tanzania, and verified the accuracy compared with the actual electricity consumption data. The United Republic of Tanzania is located in East Africa, and is one of the least developed countries in the world [16] with a GDP of approximately USD 58.001 billion (2018) [17]. The population of Tanzania is 44.93 million, and the population density is 51 persons/km² (2012) [18]. The electrification rate is 37% (2018) [1]. The electricity in mainland Tanzania is supplied by TANESCO (Tanzania Electric Supply Company Limited, Dar es Salaam, Tanzania). Tanzania was a suitable country for case study because of the low rate of access to electrification and low economic development. The United Republic of Tanzania consists of mainland Tanzania and Zanzibar. We conducted a simulation case study only on the mainland. A model was constructed based on household income data obtained from work in [19], and the electricity consumption predicted from this model was compared with actual electricity consumption data described in [20] to verify the validity of the model.

3.1. Income Data

We examined the changes in electricity consumption corresponding to changes in household income based on theory in Section 2. In this study, we calculated household income using the data from [19]. The data included socioeconomic surveys such as agricultural production, non-agricultural income, and consumption. The data included at least 5000 households, categorized based on the following sources of income; agricultural business, small business, wages, and other income, considering the periodic variation in income for each category. We corrected these variations for each year, and used the sum of each household's income for each category of household income.

3.2. Upper Bound of Electricity Consumption

We derived the relationship between household income and food expenditure so that the total of gross income minus food expenditure would be the upper bound of the electricity bill expenditure. Here, we identified the ratio of food expenditure to household income. We excluded non-electrified households to build models considering only electrified households. Data on food expenditure in the past week were obtained from work in [19], and these data were used to calculate the amount of food expenditure for one year. We then used this result to obtain the ratio of food expenditure to household income.

We regressed the log curve results in the relationship between household income and food expenditure as follows; let x be the annual household income and y be the ratio of food expenditure to total household expenditure, then

$$y = -0.12 \log(x) + 1.40 \quad (5)$$

where R^2 is 0.31. Figure 2 shows data plots and the obtained curve.

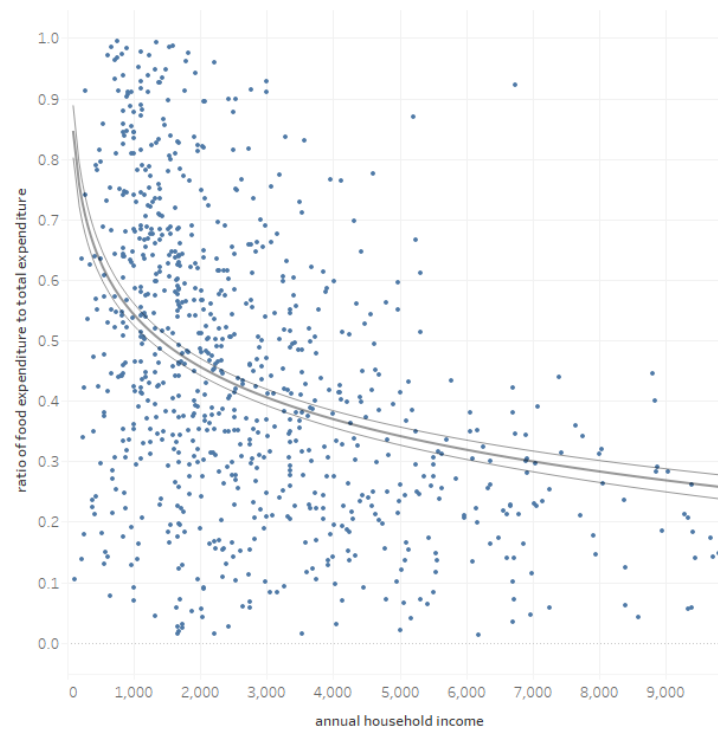


Figure 2. Relationship between annual household income and ratio of food expenditure to total household expenditure.

We used the TANESCO T1 electricity unit cost of 306 TSH/kWh (USD 0.14/kWh) to convert electricity bill expenditure into electricity consumption. As described in Section 2.2, the upper bound should be set to an appropriate or a larger value, and the following were assumed in this case study.

- (1) Allocate all income except food expenditure to pay for the electricity bill.
- (2) For the equation for calculating food expenditure, $R^2 = 0.31$, which is quite small. Where income is low, many actual values deviate above the value obtained from the equation and, conversely, actual values after middle income often fall below.
- (3) It is converted to electricity consumption using the electricity cost of TANESCO.

In (1), in reality, it is hard to think that households pay such a large electricity bill. This does not contradict the setting that derives the upper bound value to a larger value. Regarding (2), in the low-income group, which is a large number of people in Tanzania, many actual households that pay more for food than derived from the equation. Therefore, adopting the value derived by the equation means that food expenditure is lower than in reality, namely, household can pay more for electricity bill. This leads to setting the upper bound higher. For (3), TANESCO's electricity bills are inexpensive considering electrification of un-electrified areas. In other words, assuming the same electricity bill expenditure, the electricity consumption will be calculated higher, which leads to setting the upper bound to a larger value. For all three assumptions, the upper bounds are set to increase electricity

consumption, and when planning electrification based on the consumption calculated derived from the upper bound, electricity shortages will not occur.

3.3. Relationship between Household Income and Number of Household Appliances

The difference in the number of household appliances owned according to household income is discussed in Section 2.1. It is important to consider E when constructing this model. In this simulation, there are 14 items of E , of which 11 are from the results of the questionnaire in [19] asking each household what they own, and 3 are generally deemed important for estimating electricity consumption: washing machines, water heaters, and lights [21]. Here, $E = \{ \text{a complete music system, computer, dish antenna decoder, fan/air conditioner, iron (charcoal or electric), radio and radio cassette, refrigerator or freezer, sewing machine, telephone (mobile), television, DVD player, washing machine, water heater, and lights} \}$.

Based on the data in [19], we only analyzed the households that are electrified, which included 1091 households. Household income and the ownership of 11 items out of the elements a of E were arranged in income order, and we assumed the average ownership of the top 10 households based on income as $S_{max}(a)$, top 1% of the data. Then, by calculating the average value of household income and ownership per 100 households in order of income, we performed a regression on the Gompertz curve against them and estimated α and β , respectively, and obtained a result. There were no data for washing machines, water heaters, and lights in [19]. Therefore, for the value of $S_{max}(a)$, we adopted the high-income value described in [21]. For appliances without data in [19], we adopted α and β of appliances that seem to be similar in ownership change and for which data existed. Therefore, for α and β , the value of a washing machine was assumed to be the same as the value of an iron (charcoal or electric), the value of a water heater was assumed to be the same as the value of a refrigerator or freezer, and the value of lights was assumed to be the same as the value of a telephone (mobile). Table 1 shows estimated parameters. R^2 is the coefficient of determination, and for most a , the data are represented by the Gompertz curve model. Figure 3 shows the relationship between household income and the number of owned household appliances.

Table 1. Estimated parameters for each appliance in the electric appliance retention model.

a	R^2	$S_{max}(a)$	α	β
Complete music system	0.81	0.3	2.88	0.00031
Computer	0.99	1.5	4.16	0.00021
Dish antenna decoder	0.83	1	1.91	0.00019
Fan/air conditioner	0.83	1.4	1.93	0.00027
Iron (charcoal or electric)	0.85	1.3	1.43	0.00048
Radio and radio cassette	0.26	1.2	0.70	0.00029
Refrigerator or freezer	0.94	1.2	2.26	0.00027
Sewing machine	0.52	0.2	5.18	0.019
Telephone (mobile)	0.77	5.5	1.47	0.00022
Television	0.75	1.6	1.31	0.00033
Video DVD	0.81	1.3	1.34	0.00040
Washing machine		1	1.43	0.00048
Water heater		1	2.26	0.00027
Light		13	1.47	0.00022

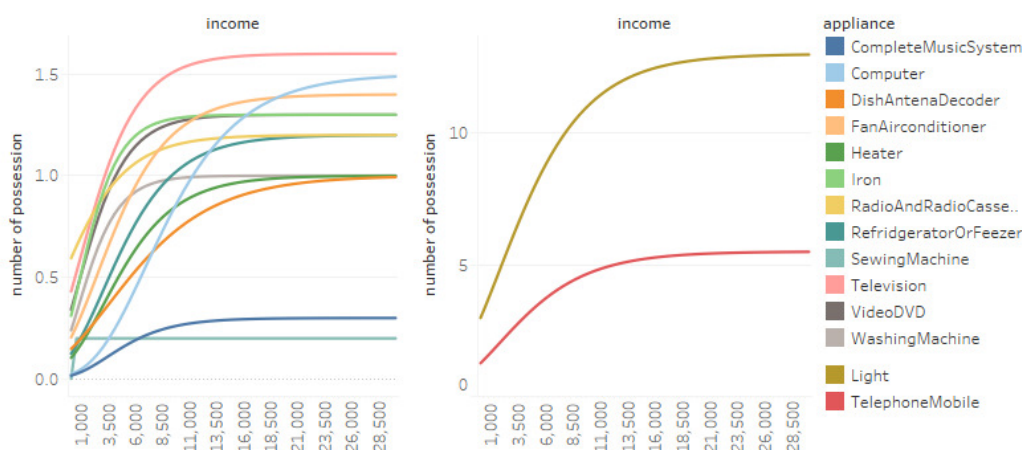


Figure 3. Annual household income and number of owned appliances per household.

3.4. Relationship between Household Income and Electricity Consumption

We then prepared for the distribution of electricity consumption (gamma distribution) for each electric appliance. To achieve that, we created the standard rated output of each household electrical appliance and the rated output of a heavy user's product, regardless of income and specific hours of the day. In this model, although the number of ownership of household electric appliances changes according to income, the type of product and its use is independent of income and depends on the users, which we assumed are stochastically distributed. There is a tendency to hold high-grade products according to income; however, these are not necessarily products with large outputs and their frequency of use is low. The consumption for electricity is expressed by the output of each product and usage time. Therefore, we prepared the rated output for two patterns, standard consumption and heavy user consumption, and the average usage time per day for each electric appliance (Table 2). We prepared this table referring to the information in [21], products sold in [22], and, if there were no data, we made reasonable assumptions. In this case study, we compare with the electricity consumption of TANESCO in 2010. Therefore, we assumed values reflecting trends as of early 2010s. We derived the parameters of electricity annual consumption in each gamma distribution using the standard "Consumed electricity (kWh/year)" in this table as the mean μ , and the difference between standard "Consumed electricity [kWh/year]" and heavy users as 2σ .

Prediction of Household Electricity Consumption Using the Model

With regards to the model, we constructed a graph and confirmed how the household annual electricity consumption changes with income. We sampled 5000 electricity consumptions for each income level in increments of USD 500, ranging from USD 250 to USD 14,750. Then, that sampling values were plotted and drawn to box plot charts.

For each income level and appliance $a \in E$, the number of possessions was derived from the Gompertz curve and Table 1 as discussed in Section 2.1. Then, the electricity consumption of a was derived from the gamma distribution and Table 2, as discussed in Section 2.2. By summing the multiplication of the number of possessions and the amount of electricity consumption for all a , we obtained the sampling of the household electricity consumption of that income. If it exceeded the upper bound, it was resampled from the uniform distribution. The simulation was implemented in Java, and Apache Commons Math [23] was used as a library for random number generation. A plot of the relationship between annual household income in USD and annual electricity consumption in kWh/year is shown in Figure 4. If households' annual income is USD 10,000, then the median of annual consumption is 3000 kWh, after which it increases gradually.

Table 2. Standard and heavy user electricity consumption.

	Standard				Heavy User			
	Rated Output (W)	Used Hours (h/day)	Used Hours (h/year)	Consumed Electricity (kWh/year)	Rated Output (W)	Used hours (h/day)	Used Hours (h/year)	Consumed Electricity (kWh/year)
Complete music system	30	1	365	10.95	150	8	2920	438
Computer	50	4.32	1576.8	78.84	100	4.32	1576.8	157.68
Dish antenna decoder	1	24	8760	8.76	1.5	24	8760	13.14
Fan/Air conditioner	574	2.94	1071.79	615.20	1148	5.87	2143.57	2460.82
Iron (charcoal or electric)	908	0.51	186.15	169.02	1362	1	365	497.13
Radio and radio cassette	10	1	365	3.65	100	8	2920	292
Refrigerator or freezer	200	8.11	2960.15	592.03	1200	8.11	2960.15	3552.18
Sewing machine	30	0.071	26.07	0.78	70	0.071	26.07	1.83
Telephone (mobile)	2.5	1	365	0.91	5	4	1460	7.3
Television	55	7.03	2565.95	141.13	82.5	7.03	2565.95	211.69
Video DVD	20	1	365	7.3	50	8	2920	146
Washing machine	490	1.74	635.1	311.20	735	1.74	635.1	466.80
Water heater	1535	0.071	25.84	39.66	2302.5	0.071	25.84	59.50
Light	60	7.58	2766.7	166	90	16	5840	525.6

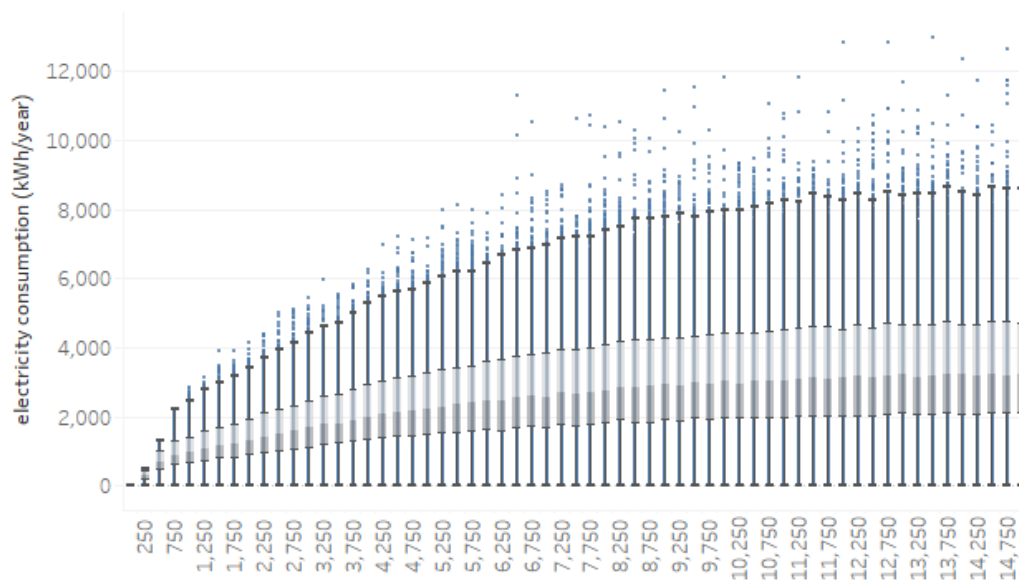


Figure 4. Relationship between annual household income and annual electricity consumption.

3.5. Model Evaluation

We compared the electricity consumption derived with the value of electricity consumption reported by TANESCO [20] to verify the validity of the model (Here, as the electricity consumption in households is forecasted, we use the sum of D1 and T1 as TANESCO's electricity consumption.). According to the electricity consumption in Tanzania mainland in 2010 for each region published by TANESCO, we compared it with the value calculated from the model based on the number of subscribers and verified the prediction accuracy of the model. We used the data in [19], where household income differed on each region. In [19], a survey of households throughout Tanzania was conducted, which included items related to regions and income. From this, only the income of households receiving electricity supply from TANESCO was extracted and used as income data for each region. These data have a small sample size, so, for example, using the average household income in that region would be unreliable. Therefore, we assumed that each region has their own income distribution function, and the parameters of the distribution function were obtained using the maximum likelihood estimation method. The income distribution function uses the generalized beta of the second kind (GB2), which was explained in [24,25]. We randomly sampled from the distribution function using the parameters obtained for each region to determine the income of a household, and then sampled the electricity consumption at that income from the model. The samplings were performed for the number of customers in 2010, TANESCO, and the sum of the electricity consumption was used as the predicted electricity consumption in that region. The probability density function of GB2 is expressed as follows. The parameters $a > 0, b > 0, p > 0, q > 0$ were estimated, where $B(p, q)$ is a beta function.

$$f(x) = \frac{a(x/b)^{ap-1}}{bB(p, q)(1 + (x/b)^a)^{p+q}}$$

For the estimation, we used the GB2 package of R [26]. Table 3 lists the GB2 parameters obtained.

Table 3. The estimated parameters of the income distribution function GB2 for each region.

Region	<i>a</i>	<i>b</i>	<i>p</i>	<i>q</i>
Arusha	6.9	2008.72	0.16	0.14
Dar es Salaam	2.41	2035.67	0.65	0.53
Dodoma	6.88	604.16	3.18	0.13
Iringa	1.54	3199.26	0.67	1.03
Kagera	2.08	6026.8	0.42	2.58
Kigoma	0.39	804.31	8.57	7.39
Kilimanjaro	3.31	900.05	0.31	0.2
Lindi	0.95	4015.67	0.71	1.16
Manyara	4.35	191.31	5.81	0.18
Mara	1.06	6094.13	1.3	3.56
Morogoro	6.32	565.1	0.26	0.1
Mbeya	6.09	1187.47	0.11	0.1
Mtwara	0.39	810.47	11.18	7.86
Mwnza	3.78	1572.5	0.4	0.26
Pwani	0.98	5787.98	2.28	7.29
Rukwa	6.38	1462.84	0.19	0.47
Ruvuma	0.47	240.81	10.99	3.94
Singida	0.78	5681.56	3.5	6.54
Shinyanga	0.85	234.75	5.12	1.27
Tabora	1.16	829.08	1.41	0.94
Tanga	3.15	4265.33	0.26	4.42

Table 4 lists electricity consumption, the number of customers of TANESCO, estimated electricity consumption, and errors used in the evaluation in each region. Considering the sum of each region's electricity consumption as the overall consumption in Tanzania mainland, the error between the actual measurement value and the predicted value was about -0.13 , confirming that the prediction has sufficiently high accuracy. The average absolute error of each region was 0.26. The average households in each region was about 40,000, indicating that the prediction accuracy for a region of this size was 0.74. However, we observed a large variation in the accuracy of each region. It has been pointed out that the estimation of electricity consumption based on surveys was not very accurate [14,15]. Blodgett et al. [14] estimated electricity consumption for each of the 8 minigrids in Kenya by the general survey approach. As the error tends to decrease as the number of customers increases, the accuracy of total electricity consumption of eight minigrids was compared with our results. The estimated value was 72,359 Wh/day and the measured value was 17,365 Wh/day, where the number of customers was 154 of all eight minigrids, and the error was 317%. The electricity consumption estimated in this study is for each region in Tanzania, and the number of customers varies from several thousand to several hundred thousand. Although it is difficult to make a simple comparison due to the vastly different conditions, the mean prediction accuracy of 74% shown in this study is nevertheless considered to be sufficient.

To apply this method, for example, suppose there is a business operator in a developing country that builds a microgrid in each village. This can be used to estimate the total capacity of solar power generation equipment needed by such an operator considering the installation of microgrids in thousands of villages.

Table 4. Actual electricity data in 2010 from [20], number of subscribers, electricity consumption predicted by model, and the errors.

Region	D1+T1 Consumption [kWh] [20]	Number of Customers [hh] [20]	Predicted Electricity Consumption [kWh]	Relative Error
Arusha	152,296,238	60,936	135,039,661	−0.11
Dar es Salaam	971,897,812	285,139	642,595,899	−0.34
Dodoma	71,452,940	31,207	63,121,380	−0.12
Iringa	40,356,479	34,592	75,127,329	0.86
Kagera	30,830,559	17,646	32,589,882	0.06
Kigoma	8,641,101	9291	16,784,770	0.94
Kilimanjaro	100,271,313	70,987	144,100,095	0.44
Lindi	12,825,731	6575	14,958,498	0.17
Manyara	21,909,654	7443	12,170,727	−0.44
Mara	30,758,998	16,222	33,165,221	0.08
Morogoro	93,716,113	43,866	82,165,218	−0.12
Mbeya	93,822,410	52,219	104,293,895	0.11
Mtwara	23,752,924	10,406	22,865,148	−0.04
Mwnza	108,620,452	49,313	108,074,980	−0.01
Pwani	41,142,762	29,919	52,312,584	0.27
Rukwa	16,042,885	8566	12,011,898	−0.25
Ruvuma	17,635,828	12,774	29,964,785	0.70
Singida	25,342,782	12,618	27,582,409	0.09
Shinyanga	47,541,884	22,088	45,610,877	−0.04
Tabora	33,986,978	19,404	37,952,329	0.12
Tanga	82,259,509	43,803	62,825,612	−0.24
all	2,025,105,352	845,014	1,755,313,196	−0.13

4. Conclusions and Future Work

In this study, we developed a household income electricity consumption model to estimate residential electricity consumption in a specific area based on household income. The residential electricity consumption of a specific area can be derived by adding up the electricity consumption of each household based on income in that area.

As a result of verification using the electricity consumption in Tanzania in 2010, the accuracy when predicting electricity consumption in the entire Tanzania mainland was 87% and the average accuracy of forecasting electricity consumption in each region was 74%.

Further research is required to improve the accuracy of prediction. The number of persons per household could differ in the number of products owned, rated output, and usage time. As there were no data for verification, we did not model the number of persons per household in this paper. However, detailed analysis and verification of the number of persons per household will help to improve the accuracy. The connection to the electric power company may be shared by multiple households [20], so it may be useful to analyze the number of people per connected customer as an object of discussion.

As our model can be adapted to other countries beyond Tanzania, we hope that our findings will help lead to suitable electrification in many developing countries.

Author Contributions: K.S. designed and performed research and wrote the article; G.M. supervised the research. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. IEA. SDG7: Data and Projections, Access to Electricity. Available online: <https://www.iea.org/reports/sdg7-data-and-projections/access-to-electricity> (accessed on 15 May 2020).
2. Bazilian, M.; Nussbaumer, P.; Rogner, H.H.; Brew-Hammond, A.; Foster, V.; Pachauri, S.; Williams, E.; Howells, M.; Niyongabo, P.; Musaba, L.; et al. Energy access scenarios to 2030 for the power sector in sub-Saharan Africa. *Util. Policy* **2012**, *20*, 1–16. [[CrossRef](#)]
3. Da Silva, F.L.; Souza, R.C.; Oliveira, F.L.C.; Lourenço, P.M.; Fagundes, W.D.C. Forecast of long-term electricity consumption of the industrial sub-sector of pulp and paper in Brazil using a bottom-up approach. *Procedia Comput. Sci.* **2015**, *55*, 514–522. [[CrossRef](#)]
4. Debnath, K.B.; Mourshed, M.; Chew, S.P.K. Modelling and forecasting energy demand in rural households of Bangladesh. *Energy Procedia* **2015**, *75*, 2731–2737. [[CrossRef](#)]
5. Amarawickrama, H.A.; Hunt, L.C. Electricity demand for Sri Lanka: A time series analysis. *Energy* **2008**, *33*, 724–739. [[CrossRef](#)]
6. Azadeh, A.; Ghaderi, S.; Sohrabkhani, S. Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. *Energy Conv. Manag.* **2008**, *49*, 2272–2278. [[CrossRef](#)]
7. Oğcu, G.; Demirel, O.F.; Zaim, S. Forecasting electricity consumption with neural networks and support vector regression. *Procedia Soc. Behav. Sci.* **2012**, *58*, 1576–1585. [[CrossRef](#)]
8. Inglesi, R. Aggregate electricity demand in South Africa: Conditional forecasts to 2030. *Appl. Energy* **2010**, *87*, 197–204. [[CrossRef](#)]
9. Morales-Acevedo, A. Forecasting future energy demand: Electrical energy in Mexico as an example case. *Energy Procedia* **2014**, *57*, 782–790. [[CrossRef](#)]
10. Rosnes, O.; Vennemo, H. The cost of providing electricity to Africa. *Energy Econ.* **2012**, *34*, 1318–1328. [[CrossRef](#)]
11. McQueen, D.H.; Hyland, P.R.; Watson, S.J. Monte Carlo simulation of residential electricity demand for forecasting maximum demand on distribution networks. *IEEE Trans. Power Syst.* **2004**, *19*, 1685–1689. [[CrossRef](#)]
12. McNeil, M.A.; Letschert, V.E. *Forecasting Electricity Demand in Developing Countries: A Study of Household Income and Appliance Ownership*; European Council for an Energy Efficient Economy-2005 Summer Study, LBNL-58283; European Council: Mandelieu, France, 2005; Volume 85, p. 90.
13. Daioglou, V.; Van Ruijven, B.J.; Van Vuuren, D.P. Model projections for household energy use in developing countries. *Energy* **2012**, *37*, 601–615. [[CrossRef](#)]
14. Blodgett, C.; Dauenhauer, P.; Louie, H.; Kickham, L. Accuracy of energy-use surveys in predicting rural minigrid user consumption. *Energy Sustain. Dev.* **2017**, *41*, 88–105. [[CrossRef](#)]
15. Hartvigsson, E.; Ahlgren, E.O. Comparison of load profiles in a minigrid: Assessment of performance metrics using measured and interview-based data. *Energy Sustain. Dev.* **2018**, *43*, 186–195. [[CrossRef](#)]
16. UN. The Least Developed Countries Report 2019. United Nations. 2019. Available online: <https://unctad.org/en/pages/PublicationWebflyer.aspx?publicationid=2571> (accessed on 15 May 2020).
17. WorldBank. The World Bank Home Page. Available online: <http://data.worldbank.org/country/tanzania> (accessed on 14 May 2020).
18. NBS. Statistical Abstract 2013. National Bureau of Statistics Ministry of Finance. 2014. Available online: <https://www.nbs.go.tz/index.php/en/tanzania-statistical-abstract> (accessed on 15 May 2020).
19. NPS. Tanzania National Panel Survey Year 3: 2012–2013. The World Bank. 2013. Available online: <https://microdata.worldbank.org/index.php/catalog/2252> (accessed on 15 May 2020).
20. Tanesco. Power System Master Plan 2012 Update. Tanzania Electric Supply Company Limited. 2013. Available online: <http://www.tanesco.co.tz/index.php/investments/investment-report/22-power-system-master-plan-2012-update> (accessed on 14 May 2020).
21. Kassim, M.F. Research on Estimation Methods for Household Energy Demand in Developing Countries—The Case of Tanzania. Ph.D. Thesis, Department of Architecture, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan, 2013.
22. Alibaba. Alibaba.com. Available online: <https://www.alibaba.com/> (accessed on 14 May 2020).

23. CommonsMath. Commons Math: The Apache Commons Mathematics Library. Available online: <http://commons.apache.org/proper/commons-math/> (accessed on 14 May 2020).
24. McDonald, J.B.; Ransom, M. The Generalized Beta Distribution as a Model for the Distribution of Income: Estimation of Related Measures of Inequality. In *Modeling Income Distributions and Lorenz Curves*; Chotikapanich, D., Ed.; Springer: New York, NY, USA, 2008; pp. 147–166. [[CrossRef](#)]
25. Chotikapanich, D.; Griffiths, W.E.; Hajargasht, G.; Karunaratne, W.; Rao, D. Using the GB2 income distribution. *Econometrics* **2018**, *6*, 21. [[CrossRef](#)]
26. GB2, P. GB2: Generalized Beta Distribution of the Second Kind: Properties, Likelihood, Estimation. Available online: <https://cran.r-project.org/web/packages/GB2/index.html> (accessed on 4 May 2020).



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).