Forecasting Energy Demand For Microgrids Over Multiple Horizons

Fred Otieno[1], Nathan Williams[1], Patrick McSharry[123]

[1]Carnegie Mellon University Africa, Kigali, Rwanda
[2]African Centre of Excellence in Data Science, University of Rwanda, Kigali, Rwanda
[3]Oxford Man Institute of Quantitative Finance, University of Oxford, Oxford, UK

***Abstract*—Access to electricity is one of the key enablers of socio-economic development in Sub-Saharan Africa. Microgrid solutions are currently playing an increasing role in providing access to electricity, especially to rural populations whose electricity is not supplied by the national grid. Microgrid developers need to manage their existing sites and expand to new regions. In order for them to manage this expansion effectively and sustainably, they need to make data-driven decisions. Having access to accurate forecasts of electricity demand at the site level is a key input in designing, managing and up-scaling microgrid solutions. Several forecasting mechanisms are proposed for such microgrid developers. Using daily energy consumption data from seven sites operating in Kenya during 2014-2017, it was established that exponential smoothing offers the best out-of-sample forecasting performance with forecast skill exhibited for horizons up to four months ahead.**

Index Terms--Demand forecasting, Microgrids, Power demand.

I. Introduction

In Sub-Saharan Africa, nearly 600 million people - about 70% of the population - live without electricity [1]. The International Energy Agency foresees that microgrids and other off-grid solutions will have a huge role in providing energy for 70% of all rural populations in developing countries [2]. This has led to the embracing of off-grid and distributed energy approaches. In Kenya, the electrification rate was 36% in 2014 [3]. Access to electricity is crucial for driving economic development in Africa [4]. Therefore, there is a pressing need to increase the number of people that have access to electricity.

Privately owned microgrids are one of many solutions to increasing electricity access in Kenya. Grid extension, public and community microgrids, and stand-alone systems all have a role to play. Private microgrid developers are likely to first target sites and customers viewed as having the highest potential for electricity consumption in order to achieve a commercial return on investment. Without subsidies and other support mechanisms, microgrid rollout and market penetration will be limited to these areas. Nevertheless, overall, private microgrids can provide an important contribution to off-grid electrification efforts for the rural market in Kenya.

The challenge for these offgrid providers is the ability to design systems that will be able to meet the demand of its customers over time. One such offgrid company is PowerGen Renewable Energy.

PowerGen supplies energy solutions to community, home, business and light industrial clients in East Africa. Founded in 2011, PowerGen has set up over 40 microgrids and currently serves thousands of customers across seven countries with clean, renewable energy. It offers system design and engineering, device and technology procuring, implementation and integration, and operations and maintenance services. It is redefining rural electrification in Kenya using these microgrid solutions.

As PowerGen seeks to expand its operations, a challenge they face is customer acquisition and demand assessment. Demand assessment can be conducted through forecasting. This will enable the company to venture into new sites once economic viability has been established. Apart from setting up new sites, this research aims to provide an automated forecasting system for PowerGen. The focus in primarily to facilitate performance monitoring of a site that has a microgrid deployed. This will enable PowerGen to preempt future peaks and troughs in demand and enable the system to be modified accordingly.

An efficient load forecasting model will serve PowerGen with reliability in terms of scheduling, planning and managing their microgrid.

According to McSharry et al. there are various challenges associated with different load forecasting horizons [5]:

- Very short-term load forecasting. This ranges from seconds and minutes to several hours. This is required for controlling the flow.
- Short-term load forecasting. Ranges from hours to weeks. These forecasts are useful in adjusting generation and demand and informing launch of offers to the electricity supply market.
- Medium-term and long-term load forecasting. Ranges from months to years. These forecasts are normally used to plan power generation asset utilities.

The crucial forecasting horizons are weekly, daily and hourly. These horizons have a direct impact on the day to day operation of the power generation company. Being able to foresee an upcoming demand spike can enable the power generation company to adjust accordingly in order to meet the needs of their customers and maintain the reliability of service.

Based on the different forecasting horizons, various forecasting methods have been recommended by Hernandez et al. [6]. Due to the stochastic nature of electricity demand as a function of time the seasonal ARIMA and state space models

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

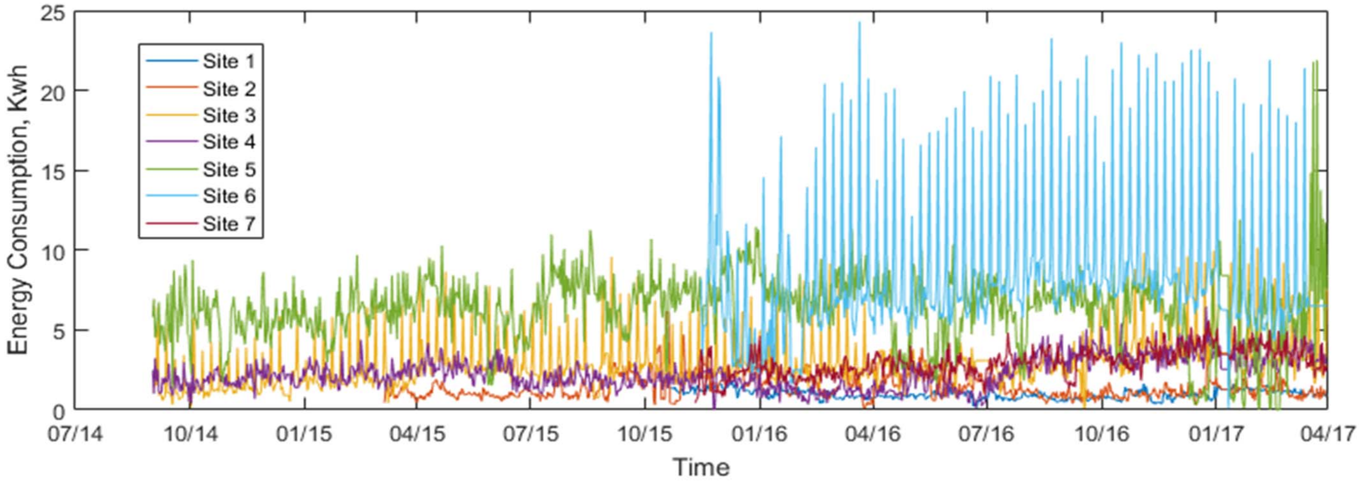

Figure 1. Time series energy consumption in the seven sites

have traditionally been used for modelling. Improvements in technology and advances in the availability of computing power has enabled the adoption of more complex models which are rule-based and use fuzzy logic. These models were designed to deal with the complexity and uncertainty characteristic of electrical load behavior. Some of the suggested models, such as fuzzy logic, enable the incorporation of non-linear behavior of load in a systematic manner when building the models [7].

Complex models, such as neural networks and fuzzy logic, appear more sophisticated and perform well for in-sample evaluation. However, it has been found that in many cases, simpler and more robust methods that are less dependent on domain knowledge can outperform these complex alternatives and have been demonstrated to perform extremely well in out-of-sample forecast evaluations [8]. Occam's razor provides motivation for simple models with few parameters based on the principle of parsimony. Such parsimonious models have practical benefits in terms of avoiding over-fitting whereby in-sample performance improves but out-of-sample performance does not. Further empirical evidence for the advantages of parsimonious models was obtained using an extensive out-of-sample evaluation of models for forecasting intraday electricity demand from ten European countries [9].

Another feature of electricity demand is the existence of intraannual, intraweek and intraday seasonal cycles. Intraweek and intraday cycles are caused by human behavior and repetition of activities at the same time of day and same days of the week. The need to deploy models that take account of intraweek and intraday seasonal cycles was explored in previous forecast comparisons and found that exponential smoothing was most competitive for forecasting demand at a national level [8]-[9].

Intraweek seasonality is relevant for understanding daily demand as described in McSharry et al. [5]. Seasonality with an annual periodicity can be caused by weather patterns and agricultural productivity, which is a key source of income generation for rural communities. Fluctuations in income are often driven by harvests and therefore dependent on favorable weather conditions during the growing seasons.

Building on the previously-conducted work, the aim here is to investigate the intraweek and intraannual seasonality for this microgrid energy consumption data. It is important to note that the electricity consumption of approximately 30 customers at a microgrid site level is clearly very different to aggregate consumption at a national level in a developed country. Several models are constructed and evaluated on the seven microgrid sites in Kenya with the aim of establishing an accurate forecasting approach for microgrid energy consumption.

The paper first provides information about the metered recordings, followed by the descriptive summary of the data, the methods that are compared, the analysis and results obtained, and finally the conclusions.

A. Microgrid Station and Consumer Metering

The typical power generation station is a standardized, modular solar power generation system deployed in 3 kilowatt-peak intervals for roughly every 50 customers. This comprises equipment which converts solar energy into a steady source of AC, 230V electricity. The system may additionally have a diesel generator to cater for situations when consumption exceeds supply, especially during peak seasons.

Each customer's connection includes a smart meter that transmits data to a centralized server. A customer purchases power credit by sending a payment to PowerGen's account using mobile money. Each meter measures consumption while tracking the customer's balance. The smart meters upload this data to servers where they are aggregated and visualized on a web portal. This enables them to manage and monitor the grids remotely.

B. The Energy Consumption Data

The energy consumption data that is primarily of interest in this study is available at both the site level and the customer level. A consumption observation includes a timestamp and the kWh used since the last observation. In our case the consumption data is drawn from seven sites located in four different counties in Kenya. The data resolution is daily energy consumption per customer.

II. Data Preprocessing

Before using the data for analysis, some data cleaning and processing was performed on the data. This involved:

- Converting timestamps from strings to consistently formatted date time values.
- Applying a consistent site naming convention.
- Removing duplicate numerical values.
- Redefining all negative consumption values as invalid information.
- Restructuring the data to have relevant variables. This includes: country, site name, timestamp, access time, access month, and daily consumption.

III. Descriptive Statistics

A variety of statistical quantities were calculated in order to understand the typical amount and variation in energy consumption at each of the seven sites. This descriptive analysis helps to provide some background understanding about how the customers are consuming electricity. Table 1 shows that there are enormous discrepancies between the energy consumption levels across the various sites, with the typical customer in site 5 consuming more than ten times more than a customer in site 1. Meanwhile the site that has been online for the longest time is two-and-a-half years old. The standard deviation reveals the variability in consumption between customers in sites and between sites. For instance, site 6 has a variability of about seventeen times that of site 1.

Figure 1 shows the time series plots of the daily energy consumption in the seven sites and illustrates how diverse the behavior of the sites is in terms of energy consumption. There are no apparent temporal trends to suggest that the customers at any given site are starting to consume increasingly large amounts of electricity. It is also obvious that it will be challenging to find a model that is capable of describing the fluctuations in all the sites being observed. Sites 1 and 3 exhibit extreme fluctuations with certain days contributing to much larger levels of consumption than normal. The intraweek seasonality analysis confirmed that this is due to a particular day of the week in each of these sites.

TABLE I. Descriptive Statistics Calculated Based On Daily Energy Consumption At Each Of The Seven Sites.

Site	1	2	3	4	5	6	7
No. customers	34	28	21	26	32	75	30
Days online	526	756	942	942	942	488	507
Mean	0.03	0.06	0.16	0.15	0.309	0.217	0.11
Standard Deviation	0.01	0.05	0.12	0.09	0.14	0.17	0.04
Median	0.021	0.02	0.08	0.06	0.219	0.05	0.05
Max	0.63	1.51	4.7	3.5	6.2	9.5	1.6
Total	501.7	1073	2637	2122	5909	4113	1489

These statistics have been calculated based on daily energy consumption in Kwh.

Before embarking on an explanation of the models, seasonality tests are performed to better understand what kind of models are likely to perform well. Calculations based on the daily consumption data are used to estimate the average energy consumption on each day of the week and each month of the year in order to detect evidence of intraweek and monthly seasonality across the year respectively. Figure 2 shows the weekly seasonality in the different sites and confirms that there is little evidence of intraweek seasonality in five of the sites. It is noted that there is an isolated day when energy consumption is extremely high for two sites in particular, Monday for site 3 and Friday for site 6. Further analysis confirmed that this behavior is not caused by just one single customer but rather by the majority of the customers in that site.

Agriculture is the second large contributor to Kenya's gross domestic product (GDP) at 35.6% of GDP in 2016 [3]. The agricultural sector is the largest employer in the economy, accounting for 60% of total employment and over 80% of the population. In particular, those living in rural areas mainly derive their livelihoods from agricultural related activities [10]. The investigation of monthly seasonality is motivated by the likely variation in income derived from agriculture and the potential dependence on harvests and weather patterns. This involved consideration of the Kenyan climate using long-term climate averages of temperature and rainfall (1901-2015). The agricultural activities being carried out in the regions where these sites are located are likely driven by the rainy seasons March-May and October-December. Average rainfall and temperature were graphed and compared with normalized monthly energy consumption in order to seek some visual evidence of seasonality in Figure 3. While there is seasonality regarding wet and dry seasons, hot and cold months and a corresponding double peak in the consumption profile, there was insubstantial statistical evidence to justify using this seasonality for the purpose of forecasting energy consumption at the sites. Analysis using correlations between the consumption levels and monthly dummy variables also failed to provide support for using monthly variables in the model.

Having discovered that the intraweek seasonality is most significant and found little evidence for intra-annual seasonality, the process of model building for the purpose of forecasting is now described.

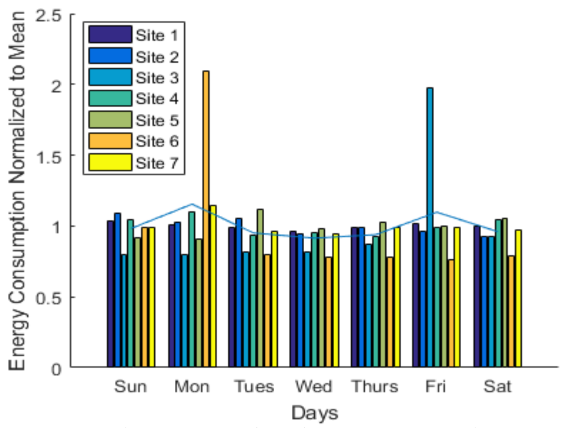

Figure 2. Day of Week Energy Consumption

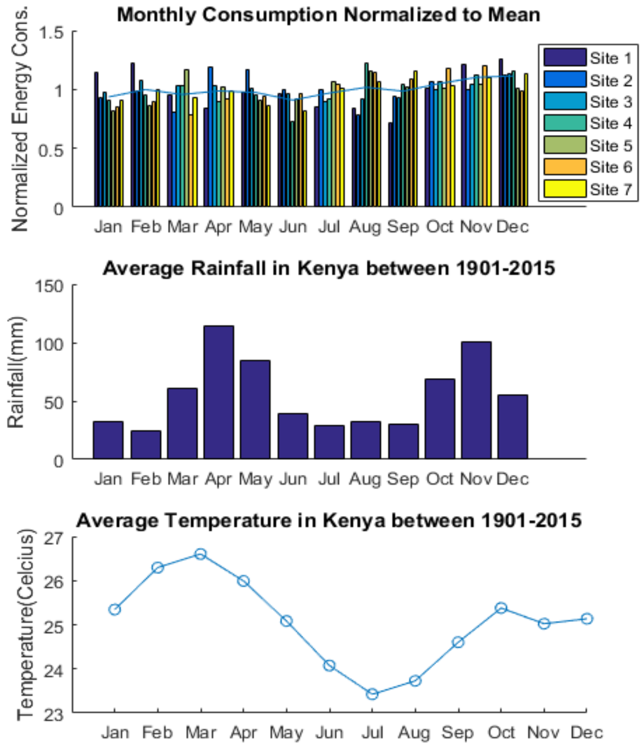

Figure 3. Comparison of monthly climatic seasonality with energy consumption

IV. Methods

This section will introduce the various models that can be used to forecast energy consumption. The list of models includes two simple forecast benchmarks known as the persistence and the unconditional benchmarks. Neither of these forecasts requires the estimation of any parameters, which means they represent the simplest specification of a forecasting model. For a model to be deemed worthy of further consideration for practical implementation, it should be able to outperform both of these benchmarks.

A standard terminology is used to describe each model through mathematical equations. Electricity consumption on day t is denoted by $x(t)$. The forecast consumption at time t with a horizon of k days ahead is given by $\hat{x}(t+k \mid t)$ and, for the purpose of evaluation, this forecast will be verified against the actual observation at time $t+k$, given by $x(t+k)$.

A. Persistence Benchmark

This model takes the last observed energy consumption value from the training set and uses this as the forecast for all subsequent horizons:

$$\hat{x}(t+k \mid t) = x(t). \tag{1}$$

B. Unconditional Benchmark

This model takes the average energy consumption of the training set and uses this as the forecast for all subsequent horizons:

$$\hat{x}(t+k \mid t) = \frac{1}{t}\sum_{i=1}^{t} x(i). \tag{2}$$

C. First Order Autoregressive Model (AR(1))

The autoregressive model forecasts future energy consumption values based on previous values from the time series. When this model is built using the latest preceding energy consumption value in the time series it is known as a first order autoregressive model:

$$\hat{x}(t+1 \mid t) = a + bx(t). \tag{3}$$

Forecasts for horizons with $k > 1$ are obtained by applying the equation recursively:

$$\hat{x}(t+k \mid t) = a + b\hat{x}(t+k-1 \mid t). \tag{4}$$

D. Autoregressive Model with Trend (AR(1)-T)

The previous model only considered the previous energy consumption values when forecasting. But this may not be the only feature that might influence the forecasting results. The energy consumption time series may have a trend. Therefore, there is a possibility of adding a trend variable to the previous model:

$$\hat{x}(t+1 \mid t) = a + bx(t) + ct. \tag{5}$$

E. Autoregressive Model with Seasonality (AR(1)-S)

The previous study of seasonality investigated whether there is any form of seasonality in the microgrid energy consumption. The results of that initial discovery, shown in Figures 2 and 3, suggest that the day of week is important but the month of the year is not when describing energy consumption.

A seasonality feature using a dummy variable for the day of week, $I_n(t)$, is now added to Equation (3) to arrive at a new model specification:

$$\hat{x}(t+1 \mid t) = a + bx(t) + \sum_{n=1}^{7} d_n I_n(t). \tag{6}$$

$$I_n(t) = \begin{matrix} 1 \; if \; t \; is \; n^{th} \; day \; of \; the \; week \\ 0 \; otherwise \end{matrix}$$

F. Autoregressive Model with Seasonality and Trend

This model combines the innovations described in two previous models. After including both seasonality and trend, this model is given by:

$$\hat{x}(t+1 \mid t) = a + bx(t) + ct + \sum_{n=1}^{7} d_n I_n(t) \tag{7}$$

$$I_n(t) = \begin{matrix} 1 \; if \; t \; is \; n^{th} \; day \; of \; the \; week \\ 0 \; otherwise \end{matrix}$$

G. Exponential Smoothing (ES)

This method allows for adaptive changes in the level and incorporates intraweek seasonality as it forecasts energy consumption. A set of equations describes how the exponential smoothing approach iteratively updates the estimates of the level in Equation (9) and seasonality in Equation (10) based on two parameters: α and ω respectively. Finally, these estimates are combined in Equation (10) to provide the m-step ahead forecast for $x(t+m)$.

$$s(t) = \alpha\left(x(t) - y(t-s)\right) + (1-\alpha)s(t-1). \quad (8)$$

$$y(t) = \omega\left(x(t) - s(t)\right) + (1-\omega)y(t-s). \quad (9)$$

$$\hat{x}(t,m) = s(t) + y(t-s+m). \quad (10)$$

V. Analysis and Discussion

The autoregressive model with seasonality and trend was estimated using backward stepwise regression for each of the seven individual sites and the in-sample results were explored. The model that was obtained performed worse than the unconditional benchmark for three out of the seven sites. Furthermore, for three out of the remaining four sites the difference in performance was not significant. Analysis of the stepwise autoregressive model with seasonality and trend showed that the trend variable was inconsistent. Two of the sites models did not select this variable. Of the five sites that selected the trend variable, two were negative and three were positive, although one was very small. This mixed set of results suggests that behavior varies with sites and requires further investigation to identify potential causes. Furthermore, the trend feature is unlikely to improve the out-of-sample forecast performance.

In order to understand the likely forecast performance of these models when deployed online in real-time to facilitate decision-making, it is necessary to undertake an out-of-sample forecast evaluation. The final full year of observations was held back as the testing data set for each of the sites for the evaluation. The root mean squared error (RMSE) is a standard metric for measuring forecast performance. To facilitate averaging the forecast errors across the sites, it is necessary to employ the normalized root mean squared error (NRMSE) which is simply the RMSE divided by the standard deviation of the time series being forecasted. The NRSME therefore takes account of variability in the target data and NRMSE values of less than one are required to indicate forecast skill greater than could be achieved using the unconditional benchmark.

The out-of-sample forecast comparison aims to assess how far into the future one might achieve reliably accurate forecasts based on the proposed models. The results of this multiple step forecast horizon analysis are displayed in Figure 4, which shows the performance of the various models using the NRMSE expressed as a percentage of that achieved using the unconditional benchmark. This forecast horizon analysis

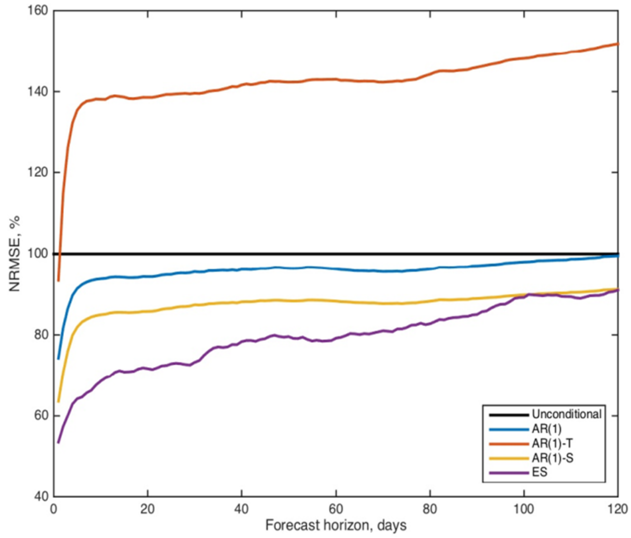

Figure 4. NRMSE of each model as a percentage of that for the unconditional benchmark.

verified that the autoregressive model with trend performs poorly, similar to the in-sample analysis that found the trend feature to be insignificant. The fact that the autoregressive model with trend was outperformed by the unconditional benchmark in this out-of-sample evaluation confirms that including a trend is of no practical use for forecasting.

The unconditional benchmark was outperformed by both the first order autoregressive model and the autoregressive model with seasonality. However, the best performing model was exponential smoothing. The exponential smoothing model showed that reliable predictions could be made up to four months (120 days) ahead with a NRMSE relative to the unconditional benchmark ranging between 55% and 90%.

A portfolio analysis, whereby the individual customers are viewed as contributing to PowerGen's reward and risk, was also conducted. Risk is measured as the volatility of the energy consumption given by the standard deviation, σ, and reward as the average energy consumption, μ. An empirical analysis between the volatility and average gave rise to the following equation:

$$\log\sigma = -0.889 + 0.645\log\mu. \quad (11)$$

Figure 5 shows the average consumption plotted against the standard deviation for each customer and the + symbol gives the values averaged over the customers in each site with specific colors to link to the relevant sites.

A correlation analysis was also undertaken to investigate the similarity between energy consumption between customers in the same site and different sites. It was established that the correlation was 0.0646 and 0.0106 for intrasite and intersite respectively, with the former being higher, as might be expected due to greater similarities for customers within a site. These correlations are both extremely low showing a remarkably high level of independence as measured by linear correlation between customers.

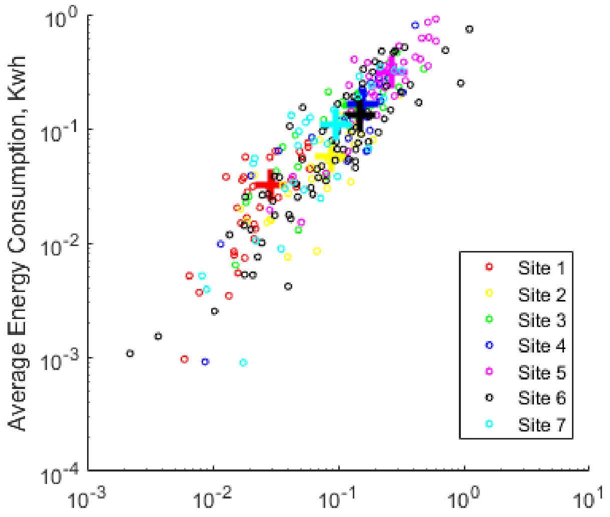

Figure 5. Energy consumption average against standard deviation; the plus signs show the respective site average.

This also suggests that it is likely to be challenging to find a single model to describe energy consumption for all customers since behavior varies dramatically across the population. Fortunately, the analysis suggests that PowerGen could obtain considerable diversification and risk reduction by increasing the number of customers both within existing sites or by adding new sites.

VI. Conclusion

This study explored the characteristics of daily energy consumption collected at seven microgrid sites operated by PowerGen in Kenya. There was evidence of intraweek seasonality, especially at two of the sites, where one day of the week had high levels of consumption. Intraannual seasonality was tested using monthly dummy variables but no consistent patterns were found. A visual comparison, with climate averages for temperature and rainfall, suggested that all three profiles contain two peaks, possibly linked to the rainy seasons. Further analysis based on actual local weather conditions and information about predominant crops is required. As more sites are connected, it might be possible to identify relationships between energy consumption and weather variables. A trend analysis across all the observations at the seven sites found inconsistent results with three sites having positive trends and two sites having negative trends.

Several autoregressive models were built and compared against the unconditional benchmark, persistence and exponential smoothing. Exponential smoothing outperformed all the other models and is recommended as the best model to use for forecasting energy consumption for PowerGen microgrids. The exponential smoothing method was able to compete with the unconditional benchmark, thereby demonstrating considerable forecast skill, for up to four months into the future.

A log-log relationship was identified for forecasting volatility given the average consumption level.

Correlation analysis between customers within a site and between different sites revealed very low correlation between all the customers. This demonstrates that the customers tend to behave differently and hence it will be a challenge to construct models for specific customers. Hence, to forecast consumption at an individual customer level could require several tailor-made models, possibly relying on customer segmentation. The positive outcome of these low correlations is that PowerGen is currently obtaining a diversified portfolio of customers and sites. Increasing the number of customers should smooth out the aggregate consumption thereby reducing the volatility and minimizing risk.

As more energy consumption data is collected from additional microgrid sites, with matching agricultural productivity and weather observations, it may be possible to identify the specific dependence on weather and agriculture.

Acknowledgement

The authors are grateful to PowerGen Renewable Energy for supporting this research and supplying the information and data from their microgrids operation in Kenya.

References

[1] IEA (International Energy Agency), 2013. World Energy Outlook. 〈http://www.worldenergy-outlook.org/publications/weo-2013/〉

[2] IEA (International Energy Agency), 2012. World Energy Outlook. 〈http://www.worldenergy-outlook.org/publications/weo-2012/〉 .

[3] World Bank Indicators. 2016. https://data.worldbank.org.

[4] M. Munasinghe, "Rural electrification in the third world", *Power Engineering Journal*, vol. 4, no. 4, p. 189, 1990.

[5] P. E. McSharry, S. Bouwman, G. Bloemhof, "Probabilistic forecasts of the magnitude and timing of peak electricity demand", *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 1166-1172, May 2005.

[6] L. Hernandez, C. Baladron, J. Aguiar, B. Carro, A. Sanchez-Esguevillas, J. Lloret and J. Massana, "A Survey on Electric Power Demand Forecasting: Future Trends in Smart Grids, Microgrids and Smart Buildings", *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1460-1495, 2014.

[7] F. Ávila, D. Sáez, G. Jiménez-Estévez, L. Reyes, A. Núñez, "Fuzzy demand forecasting in a predictive control strategy for a renewableenergy based microgrid", *Proc. Eur. Control Conf. (ECC)*, pp. 2020-2025, Jul. 2013.

[8] J.W. Taylor, L. de Menezes and P. McSharry, "A comparison of univariate methods for forecasting electricity demand up to a day ahead", *International Journal of Forecasting*, vol. 22, no. 1, pp. 1-16, 2006.

[9] J.W. Taylor, P.E. McSharry. “Short-term load forecasting methods: An evaluation based on European data”. IEEE Transactions on Power Systems 22 (4), 2213-2219.

[10] Kenya Agricultural Research Institute and IFPRI. “Food security report”. Washington, USA. 2012. http://www.foodsecurityportal.org/