Towards sustainable developing cities: A simplified forecasting model for sizing grid-tied PV using monthly electricity bills

S. Gerber[a], A.J. Rix[a], M.J. Booysen[a,*]

[a]*Department of E&E Engineering, Stellenbosch University, Stellenbosch, South Africa*

Abstract

The recent reductions in PV costs and the convenient concurrence of insolation and schools' energy usage have resulted in increased interest in augmenting supply with solar PV to save on energy costs and unburden the grid in developing countries. However, optimal sizing of PV solutions requires a detailed analysis and hourly simulation to match demand, council tariff structures (incl. import vs. export rates) and geography. This complexity exposes already resource-constrained authorities to abuse by overzealous and unwitting suppliers, and currently impedes large-scale deployment of solar PV in a sunshiny cities depraved of its requisite energy. We present a novel approach to forecast schools' hourly demand using only monthly utility bills and a trained forecasting model. We also propose an iterative solar sizing technique to assess the economic viability of using PV at schools to support government and city council officials. The results show that the method is able to forecast annual energy usage and monthly demand to within 5% and 6% respectively, while accurately determining the potential return on investment.

Keywords: PV; Developing cities; Forecasting; Optimisation; Schools; Load profiles; Cost saving; Sustainable development.

1. Introduction

Driven by a surge in economic growth, the combined share of total electricity production within developing countries surpassed that of developed Organisation for Economic Co-operation and Development (OECD) countries, and in 2016 the share of global electricity produced by non-OECD countries reached 56% (International Energy Agency, 2018). Increased public awareness and political pressure have resulted in a universal push to move away from carbon emitting energy supplies (Dietz et al., 2015; Kocak et al., 2019). As a result, renewable energy is the fastest growing source of electricity production, accounting for 26% of the worlds' energy supply in 2018 (International Energy Agency, 2018). However, in many developing countries and cities, electrification necessary for economic development would not be possible without coal-powered plants (World Energy Council, 2016; Azizalrahman and Hasyimi, 2019), leading to an increase in global carbon emissions from 2017 to 2018 (Jackson et al., 2018). This is an unfortunate consequence of developing nations being unable to afford cleaner alternatives without sacrificing spending in other, more essential sectors (Future of Life Institute, 2016; Kocak et al., 2019). In sunny South Africa, Eskom is the largest electricity utility company, generating approximately 95% of the countrys electricity demand (Eskom, 2017). Coal accounts for nearly 80% of the country's electricity needs (Department of Energy, 2018) and produces 85% of the country's total carbon dioxide emissions, resulting in South Africa being one of the largest CO_2 emitting countries in the world (World Energy Council, 2016).

Due to a poor development strategy, energy spending remained restricted from 1998 until 2004 as national spending was focused on generating economic growth (Eskom, 1999, 2008). However, rising energy demand promoted by the steady economic growth led to a diminished energy supply, and by 2007 Eskom was unable to reliably provide electricity, leading to a national energy crisis and rolling "load shedding". To help fund generational capacity the average cost of electricity across all sectors increased by 357% from 2007 to 2018

*Corresponding author
Email address: mjbooysen@sun.ac.za (M.J. Booysen)

(Eskom, 2019a). This placed a large financial strain on the country and its citizens, especially in poorer communities (Baker and Sovacool, 2017; Smit et al., 2019).

The educational sector is by no means unaffected by the crisis and soaring costs and generally accommodate large number of people during the daytime and place a substantial burden on the grid and harbour substantial inefficiencies (Sanchez-Torija et al., 2017; Samuels and Booysen, 2019). A study (Van der Berg et al., 2011) reported on how the low quality of education provided in schools within disadvantaged communities can lead to exclusion and marginalisation, limiting the future prospects of learners from a young age (Van der Berg et al., 2011). In 2006, 49% of education spending reached the poorest 40% of households (Van der Berg, 2009; Van der Berg et al., 2011), and has slowly increased to 54% in 2017 (McLaren, 2017). However, after personnel spending and conditional grants, only 10% to 20% of the budget remains for non-personnel expenses, which includes textbooks, laboratory equipment, stationary, school-maintenance and utility costs for which schools are currently billed on a commercial or industrial tariff (McLaren, 2017; Eskom, 2019b). Myende (2015) and Myende and Hlalele (2018) reiterated these statements in two studies focused on strength- and asset based approaches to improving academic performance in rural schools, stating that rural education improvement in South Africa has been hindered by the marginalisation of people living in poorer communities. This has led to lasting problems within these communities that inhibit the possibility of education providing a way out of poverty (Van der Berg et al., 2011). Reducing the financial strain placed on schools is therefore imperative. Ripunda and Booysen (2019), and Visser et al. (2019) have investigated ways to reduce water consumption within schools using smart meters and behavioural change with favourable results. However, further intervention is possible. Relief could come by employing alternative energy generation methods such as renewable energy sources, providing a pathway to sustainable development.

In January 2018 the International Renewable Energy Agency (IRENA) released figures indicating that renewable energy has emerged as a sufficient method for providing new power needs and found that the levelised cost of energy (LCOE) from solar photovoltaics (PV) decreased by 70% between 2010 and 2016 (International, Renewable Energy Agency, 2019). However, many governmental institutions within South Africa still lack the necessary resources required to determine the cost vs. benefits of renewable energy solutions, particularly due to the complexity involved in determining an accurate cost versus benefit forecast, as it will vary on a case by case basis (Sedibe, 2011; Myende and Hlalele, 2018).

Schools provide the potential benefit of having a load-profile suited for solar intervention without energy storage, with electricity usage coinciding well with solar generation curves, especially considering air conditioners that typically operate when it is sunny and warm. Furthermore, maximising the effectiveness of a solar system will allow for the greatest economical benefit (Osuri et al., 2015). This requires a deep understanding of a building's energy usage throughout hours, days, and varying seasons to correctly size a PV system (Voss et al., 2010), which can usually be achieved by installing expensive smart meters to measure hourly energy usage over a long representative period.

1.1. Energy forcasting

The ability to accurately forecast energy usage has become a widely discussed topic during recent years as the interest in energy saving within today's energy-conscious society continues to grow. The effects of reliably determining a building's energy usage are far reaching as arrangements can be made in advance to mitigate demand during peak times, resulting in large monetary savings (Raza et al., 2016). The same can be argued for optimally sized systems. If the system is too small the probability of the power supply being insufficient increases, however, if the system is too large it results in unnecessary capital expenditure. Between 2006 and 2019 there were several investigations regarding load forecasting through time series analysis (Zala and Abhyankar, 2014; Heylman et al., 2015; Yilmaz et al., 2019; Zhu et al., 2019), and research done regarding optimal energy system design (Yang et al., 2009). These studies and their relevant results are discussed below.

Zhu et al. (2019) developed a data-driven process capable of using smart metering data to create daily load profiles for various building types, and accounts for weather and holiday periods. The constructed profiles built using regression algorithms are then compared to the observed usages to detect energy consumption anomalies. The process requires smart-meters to be installed at all times, and will only be able to provide reliable detection results for a specific building once a large number of data points have been measured. Further research into data-driven predictive modelling was performed by Bourdeau et al. (2019) and Kuster et al. (2017). The papers evaluated various algorithms including regression, clustering, k-means, classification and neural networks, and concluded that a universal protocol capable of evaluating each algorithm for a specific set of input parameters is necessary to determine the optimal technique.

Heylman et al. (2015) investigated energy usage trends from over 200 buildings on the University of Virginia's campus to determine the most effective techniques for forecasting building energy usage. The study also examines the clustering of buildings based on their energy usage trends rather than the building's functional use. It was determined that one year's hourly usage data was insufficient to accurately predict long term usage patterns. Short term seasonal auto-regressive integrated moving average (SARIMA) models used for 2-day forecasts accurately determined the overall usage trend, but failed to capture demand fluctuations during the day as the hourly data points were insufficient to produce minute-by-minute forecasts. Clustering was done by grouping buildings based on energy usage with mixed results as some buildings displayed seasonal patterns while others did not, and it was concluded that additional information was required to more precisely group the campus buildings.

A recent study by Yilmaz et al. (2019) presented a methodology for creating representative electricity demand profiles for households in Switzerland. By examining the energy usage profiles of 656 multi-family apartments with a temporal resolution of 15 minutes, the study determined that five defining features including mean daily usage, morning-, midday-, evening- and night usages were enough to adequately define the households usage patterns. Households were then clustered into three distinct groups allowing for improved load forecasting and enhanced demand response targeting. The paper considers electricity use patterns within individual households, and states that daily profiles for a given household significantly differ from the averaged profile, and concludes that more features defining the underlying traits of households need to be explored. These feature requirements were investigated by Zhang et al. (2018) and introduced three engineering methods including visualisation, selection and extraction. The three methods were tested using measured data from 1000 homes, evaluating 132 features to create lists containing the ten most important features. The study concluded that the three methods discussed resulted in three different feature lists. However, certain features appeared in all rankings, and by increasing the detail with which these elements are trained using machine learning models, the accuracy of the predictions improved.

Yang et al. (2009) developed an optimal design method for a hybrid solar-wind system needed to power a continuous-load telecommunication station. The method is based on a genetic algorithm with two goal parameters including minimising the loss of power supply probability and the annualised cost of the system. The method provided favourable results with a high system utilisation rate while the battery's state of charge remained above 50% for most of the 1-year field test. An advantage of the method described includes taking into account system design characteristics including PV modules specifications, slope angle and the installation area which affect the resulting energy production and system installation costs.

Zala and Abhyankar (2014) examines the ability to design a time-of-use tariff for three major sectors (residential, commercial and industrial) by decomposing the measured load known to the utility company using representative load profiles and the elasticity of different types of loads. The study developed a k-means clustering method to find the usage patterns of consumers without the need for expensive smart-meters. However, the method was required to subdivide each sector into smaller sub-sectors as usage patterns still varied within each tariff type. These variations were apparent but still accurately identified the individual major sector usages, and could possibly be applied to cluster individual rooms and buildings within a larger load measurement.

1.2. Contribution

Although various load forecasting- and system optimisation techniques have been discussed (Yang et al., 2009; Zala and Abhyankar, 2014; Heylman et al., 2015; Yilmaz et al., 2019; Zhu et al., 2019), none of the solutions present a method for forecasting individual building (or property) load through easily accessible input parameters, essential for the South African educational context, in which smart-meters are not commonly used, access to smart meter data (where used) is difficult to obtain, and the requisite skills for detailed simulations and available resources (time and money) at government level are limited. Furthermore, none of the surveyed forecasting approaches present a method for optimally sizing renewable energy systems from a forecast load profile. A load forecast will also give insight into possible demand charges, crucial for institutions billed on a demand based-tariff structure.

In this paper, we assess the development of a load forecasting method that uses only their monthly utility bills and building specifications, including school type, available roof area and roof tilt. In conjunction, we also

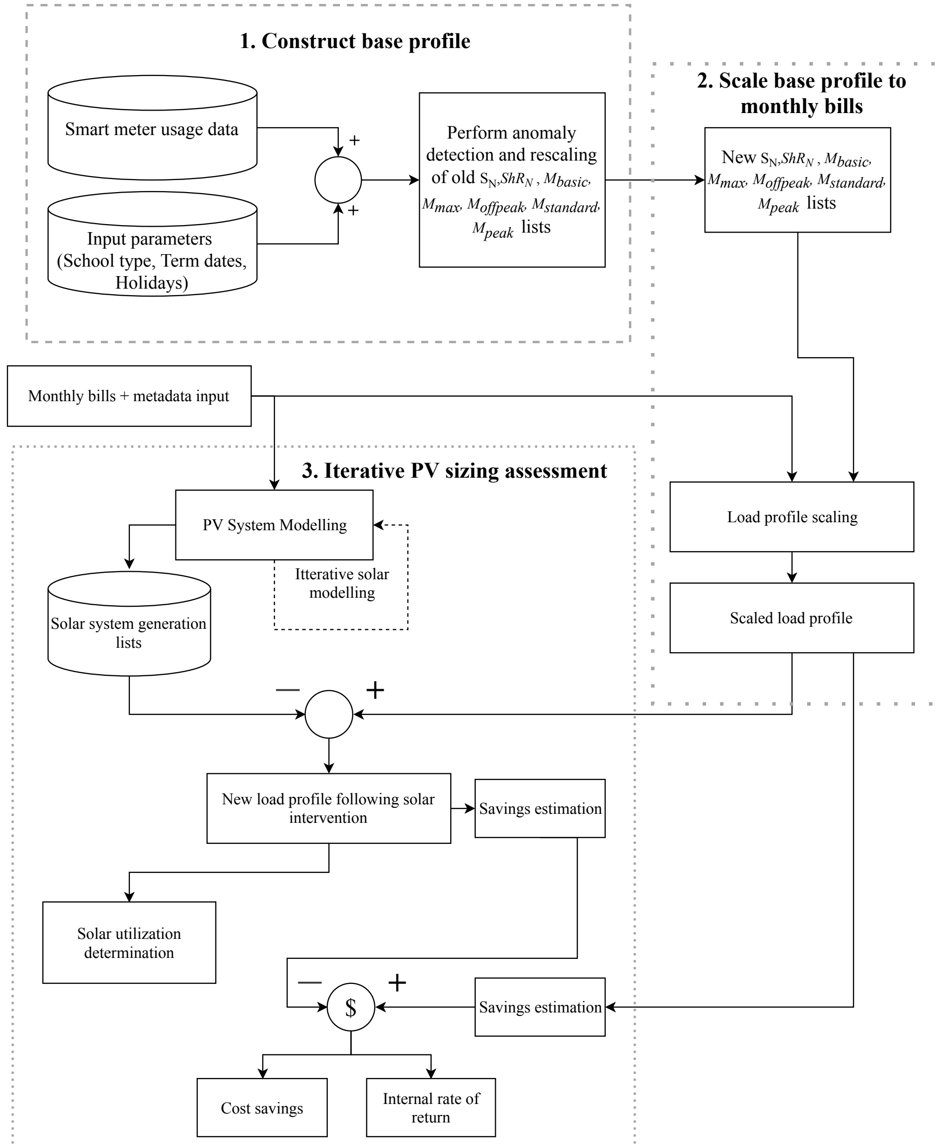


Figure 1: Simplified diagram of the proposed method.

propose and evaluate an iterative evaluation model for a grid-tied PV installation, to support governmental bodies to size and evaluate the economic viability of a solar investment for a specific school. The method removes the need for investing in smart meters, overcomes the dependence on external contractors or suppliers, and negates a technical understanding and time-intensive analysis of solar systems.

2. Proposed forecasting and optimisation technique

We propose the following three-staged forecasting approach, illustrated in Figure 1, to overcome the challenges in sizing a grid-tied PV solution for a school without the benefit of high resolution smart meter data:

1. Construct base profile: We use smart meter data from a limited subset of schools in combination with other input parameters to perform a time-series decomposition, which is used to determine for each type of school a normalised base profile, which can be used to forecast hourly usage for the year.

2. Scale base profile to monthly bills: The base type-specific profile is scaled to a representative hourly usage using the school's monthly bills, which capture the cumulative totals and maximum peak loads.

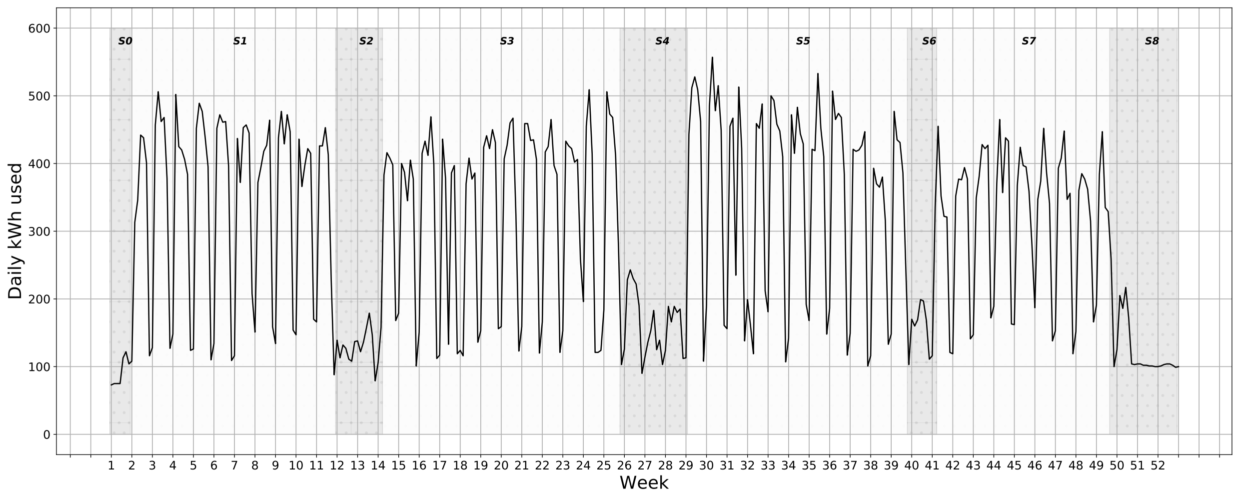


Figure 2: A typical school's daily energy usage profile over a year with the seasonal periods highlighted.

3. Iterative PV sizing assessment: We use an iterative process based on the representative profile, geography, solar prediction and solar system model to perform a cost benefit analysis over various PV sizes. The stages are presented in detail below.

2.1. Construct base profile

Most forecasting models require large datasets to ensure a degree of certainty when predicting future outcomes through commonly used regressive methods. Unfortunately this is not always possible as smart-meters necessary for data capturing are expensive and not common in residential or commercial settings in South Africa. The proposed forecasting method solves this by minimising the computational complexity and required input parameters of the forecast. It does this by converting hourly-measured smart meter data into easily usable scalar lists. The input parameters needed to decompose the time series are: the measured energy usage data, the school type (primary or secondary), school term dates and public holidays.

Firstly, the days are divided into nine seasonal periods, presented by the date-ranges of the four school terms and five vacation periods throughout the year. The vacation periods succeed each school term, with the fifth period occurring between New Year and the first day of school. Distinctive usage patterns can clearly be seen on a typical daily usage profile over a year, presented in Figure 2. Energy usages during individual public holidays are not used when constructing the base profile.

Next, the hourly usages for each of the seasonal periods are separated for each day of the week. If a weekday occurs more than once during a seasonal period the usages are averaged for that specific day and hour. New lists now contain the measured hourly usages for each weekday of each seasonal period, containing 24 hourly entries for each of the 7 weekdays.

These usages are normalised to contain the energy usage relative to the entire day's, resulting in a normalised profile, U_S, describing the usage for each hour of each day of the week, for each of the nine seasonal periods, S.

$$U_S = \begin{bmatrix} \frac{E_{0-0}}{\left(\sum_{23}^{i=0} E_{0-i}\right)} & \cdots & \frac{E_{0-23}}{\left(\sum_{23}^{i=0} E_{0-i}\right)} \\ \vdots & \ddots & \vdots \\ \frac{E_{6-0}}{\left(\sum_{23}^{i=0} E_{6-i}\right)} & \cdots & \frac{E_{6-23}}{\left(\sum_{23}^{i=0} E_{6-i}\right)} \end{bmatrix} \tag{1}$$

The process can easily be applied to large datasets. If new usage data is added the same steps outlined previously are followed and averaged with the existing hourly consumption ratios, in effect, re-scaling the seasonal usage ratios. Samples with an error of 1 standard deviation (34% above the mean) is classified as an anomaly and discarded.

The constructed usage profile consisting of normalised hourly usage values ranging from 0 to 1 can then scaled using the aggreagted usages listed on the monthly utility bills.

2.2. Scaling of base profile to bills

Schools in South Africa are billed according to one of three tariff structures (Eskom, 2019b, 2017). The first is a basic commercial tariff whereby the school is billed at a flat rate for energy (kWh) used. The second is a demand-based tariff, by which the school is billed a lower flat rate for the energy used, but billed additionally for the peak demand during the month (kW), which is sampled every 30 minutes. The third tariff structure is based on time-of-use (TOU), for which schools are billed for usage according to the the applicable time of day, day of the week, and season.

To forecast the usage of a school, its monthly bills are used to scale the hourly base profile of the same school type to forecast the hourly usages. The method uses the nine U_S lists to first construct an unscaled version of the specific year's forecast for individual days, using the term dates and public holiday input parameters – for public holidays, the preceding weekend's usages are averaged. As such, a standard year would represent all U_S lists summing to 365 entries.

Using matrix multiplication the new scaled values can be calculated to determine a year's load profile with estimated peak- and basic usages using only the relevant monthly bill components and the school's term dates as input parameters. For a basic commercial tariff scaling is done by estimating the energy usage for all months using M_{basic} to scale the given monthly usages to the remaining months. The same process is followed for a demand based tariff with the addition of M_{max} scaling maximum monthly demand. For a time-of-use tariff the off peak, standard and peak times are scaled using the $M_{offpeak}$, $M_{standard}$ and M_{peak} matrices respectively, resulting in a scaled hourly energy usage forecast as the method output.

However, the scaling method does not depend on all twelve bills being available – it requires at least one monthly bill, and scales the value to each of the remaining months using the months for which bills are available. The factors by which the months are scaled are calculated from the relational proportion of each month's energy to every other month's energy usage obtained from the measurement data. This results in 12x12 matrices containing scalars between 0 and 1 for basic energy usage, maximum monthly demand and the three time of use periods (M_{basic}, M_{max}, $M_{offpeak}$, $M_{standard}$ and M_{peak}).

The generated load profile is then combined with solar generation estimates to determine an optimal solar system size.

2.3. Iterative PV sizing assessment

The goal of the solar optimisation method is to maximise the profitability and utilisation of the system using an iterative technique. The scaled load profile obtained from the forecasting model contains hourly energy usage values with estimated demand peaks. A validated solar model is used in addition to location specific weather data to obtain a generation forecast for each iteration of solar evaluation, returning the estimated energy generated for each hour. Taking the hourly usage forecast and subtracting the hourly generated solar now presents a new hourly load profile. Solar system size optimisation is done by estimating the profitability of the investment using the internal rate of return (IRR) metric shown below:

$$NPV = \sum_{t=1}^{T} \frac{C_t}{(1+r)^t} - C_0 = 0 \tag{2}$$

where C_t is the net cash inflow during the period t, C_0 is the initial investment cost, r is the rate of return and T is the number of time periods representing the total years of the system's lifetime.

The IRR is calculated as the rate of return r that equates to a net present value (NPV) of zero. The higher the IRR, the more attractive the investment opportunity becomes. In addition, the system utilisation is determined by the percentage of the total solar energy used. A larger percentage is desired, resulting in less energy being exported and sold to the utility company at an unfavourable rate (Eskom, 2019b).

Combining the methods discussed in this section allows schools to determine the economic viability of a solar investment using only their monthly utility bills.

3. Experimental setup

This section describes the experimental setup, detailing the load forecast and PV optimisation process steps during simulation. We also explain how the performance of the simulation results are evaluated.

Table 1: Input parameters used for experimental setup

Parameters	Description
Time series decomposition	
Measured usage data	8760 hours
School type	Primary school or High school
School term dates	Start and end dates for each of the 4 terms
Public holidays	A list of dates
Load profile construction	
School type	Primary school or High school
Tariff structure	Basic, demand-based or time-of-use
Monthly energy usage s	kWh used, max monthly kVA or TOU usages (minimum one month)
School term dates	Start and end dates for each of the 4 terms
Public holidays	A list of dates
Solar PV simulation	
School location	Latitude and longitude
Estimated available roof area	m^2
Roof tilt	degrees
Azimuth	degrees
Custom module specifications* (Required for 6-paramater model)	Max power voltage, Max power current, Open circuit voltage, Short circuit current, Number of Cells per module, NOCT, Module Area

*Note:** These are optional simulation input parameters.

3.1. Input data and data cleaning

The data used for the simulations was obtained from municipal smart meters installed within each school and used to determine their monthly energy bill. In order to generate reliable forecasts the smart meter data was cleaned to ensure that no missing values were present. Of the 131 520 entries roughly 3% of values were missing from the raw data. Missing data can be a result of equipment failure, power outages or connectivity issues. Data cleaning was performed by taking into account the seasonality of the school's energy usage patterns. The data cleaning was done as follows:

If no data was present at a specific entry the entry exactly 7 days prior would be selected granted that it was during the same season and of the same day-type (normal day or holiday). If this was not possible the value 7 days later would be selected. Lastly, if a value could still not be retrieved the previous or next day's hour would be selected if it was of the same weekday-type (weekday or weekend day). This process was then repeated for all empty data entries.

3.2. Simulation Setup

The simulation uses measured energy usage data of 5 high income (quintile 5) schools in South Africa. The temporal resolution of the data is 60 minutes and is available from January 2016 until December 2018. The schools were selected due to having measured energy data available for multiple seasons.

3.2.1. Load profile determination

The school data was divided amongst the two types of schools, three primary schools with learners from Grade 1 to Grade 7, and two high schools with learners from Grade 8 to Grade 12. This was done as enough data was available, and due to noticeable changes regarding their energy usage characteristics with high schools presenting larger demand surges as a result of more classrooms, hostels and after school curricular activities.

The measured data for both schools is used to create scalar lists as presented in Figure 1. Each year is individually added to calculate new non-scaled lists for the two types of schools. The yearly trend from the usage data was not taken into account due to newly constructed buildings on the premises only affecting a single school, and carrying over the trend from a single building would be nonsensical. Once more data from a larger number of schools is available the trend data can be considered.

The input parameters for the load profile determination is the school type, the tariff structure and monthly energy usages, and the term dates and public holidays for the specific year. All input parameters used during the experimental setup is listed in Table 1.

Table 2: Input data from tested schools

School	School type	Tariff structure	Number of learners	Average yearly energy usage* (MWh)	Est. available roof area (m^2)	Roof tilt (°)	Azimuth Azimuth (°)
A	Primary	TOU	671	150.4	1500	24	340
B	Primary	Demand	910	165.6	2100	28	350
C	Primary	TOU	820	112.6	1800	24	340
D	High	TOU	550	182.6	1300	25	350
E	High	TOU	718	334.3	2000	35	340

*Note:** Average usage per year from 2016 to 2018

3.2.2. Solar PV simulation

The solar PV generation modelling is performed using SAM (System Advisor Model) open-source libraries that employ a photo-voltaic performance model developed by NREL (2018), and calculates the AC electrical output for each hour over a one year period. The input parameters required for the simulation is the school location, allocated budget, estimated available roof area, roof tilt and azimuth angles. A standard 320W poly-crystalline module is used. The simulation does allow for customised modules to be specified from manufacturers specifications by using an improved 6-parameter sub-model implementation by Dobos (2012). SAM's inverter model is an implementation of a model described by L King et al. (2007), which uses manufacturer specifications with empirically derived coefficients to simulate the AC output power for a specific DC input. During the simulation the inverter best suited for the solar generation capacity is selected from a California Energy Commission (CEC) database (NREL, 2018). Weather data is automatically obtained from the European Commission's photo-voltaic geographical information system database for the specified location (PVGIS, 2018). The method iterates through installed capacity steps of 2 kWp, storing one year's hourly generation estimates for each iteration within a database until the maximum system size determined by the investment budget or available roof area is reached.

3.3. Evaluation and metrics

By obtaining the new load profile after solar intervention, the performance of the system can be evaluated. The complete system is evaluated by determining the performance of the forecasting method for school energy use, and determining the financial viability of a solar system installation for the building. The effectiveness of the forecasting technique is determined by training the non-scaled load profile for each school type, and then testing it using an independent school as to avoid over-fitting. In the case of the primary schools: School A and B were used to train the profile, and school C's monthly usage data obtained from utility bills was used to test it. The viability of the PV system is determined by output metrics including the internal rate of return and the utilisation of the system's generating capacity.

4. Results and discussion

This section details the results obtained using the simulation setup and parameters discussed. Measured- and parameter data from the five schools were used as explained in Section 3 and can be found in Table 2.

School's A,B and C are primary schools, and school's D and E high schools. The yearly load forecast for School A is presented in Figure 3, and was obtained by training the non-scaled load profile using decomposed time series data from 2016 until 2018 of the remaining primary school's B and C. The monthly utility usages for school A from 2017 were then used to scale the obtained load-profile and compared to the measured usage of school A from 2018. Scaling the profile using 12 months of utility bills resulted in an overall energy usage accuracy forecast of 97%. However, the model does allow for usages to be scaled from only 1 month's usage. Using 3 months of data resulted in an average yearly consumption forecast accuracy of 88% across all 5 schools. A single month's consumption data resulted in a forecast accuracy of 83%.

Figure 4 presents the average maximum monthly demand forecast for primary- and high schools versus the measured maximum monthly demand during 2018 using 12 month's demand data. Overall the results are favourable with the demand forecast accuracy averaging 97% for primary schools and 93% for high schools. From the figure however it is clear that some months are more accurate than others, and the maximum forecast error recorded was 27% and 46% for primary- and high schools. Reducing the maximum demand

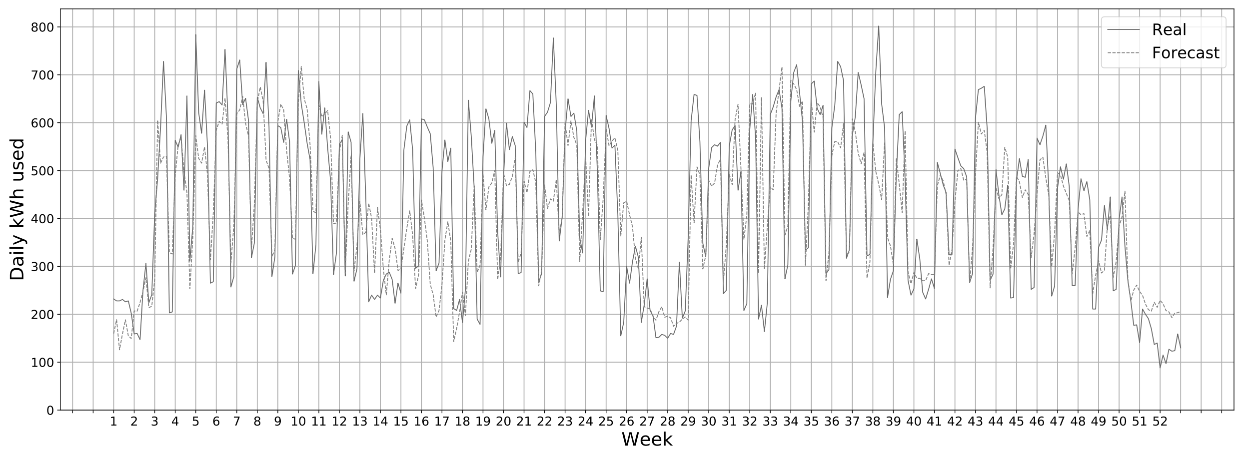


Figure 3: Energy usage forecast versus measured usage for school A in 2018.

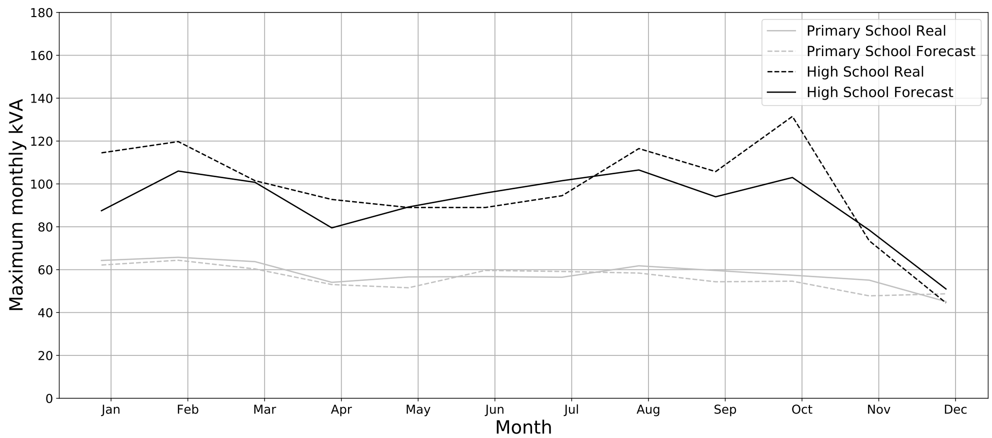


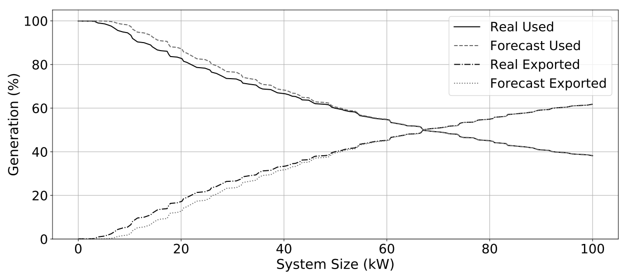
Figure 4: Maximum monthly demand forecast versus measured maximum demand for 2018.

input length to 3 months and 1 month lowered the average demand forecast accuracy of school A to 90% and 86% respectively.

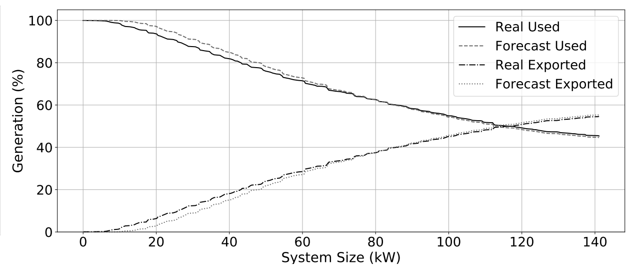
All usage and demand forecasting results are presented in Table 3. It can be seen that the primary schools fared better than the high schools. This is due to more training data being available for the primary schools, as the process of averaging scaling data results in a "smoothing" effect, reducing the impact of usage surges only seen at a specific school. But, since high school data is only available for two schools their usage profiles are directly compared as a single training data source is analysed against a single testing data source. This results in large discrepancies within peak demand and daily load forecasts as some usage trends within a school only exist at that specific school.

Effectively sizing a solar system without the use of smart-meters is a difficult task and requires many assumptions to be made. Considering the case for schools. The assumption is made that most of the energy usage occurs during the day. Therefore, that the school's demand profile is similar to that of a solar system's generation profile. With this in mind, an approach to this task would be to use the available monthly energy usages and the estimated kWh per installed kW of solar generational capacity for each month. The optimal size can then be determined by sizing a PV system such that the grid-consumption for the smallest possible solar array during any one month is zero kWh. This will allow for considerable savings for all other months and provide the optimal return on investment for this sizing strategy. Using this simple sizing method for the available schools resulted in an averaged internal rate of return of 12.4% for primary school solar systems and 13.3% for solar systems installed on high school premises. This is considered a baseline result and the forecasting model discussed in this paper will seek to improve upon this result.

After the scaled load profile is determined a series of solar systems with varying generation capacities are simulated. The hourly generation estimates are then subtracted from the hourly load forecast to calculate the energy- and demand savings, the system utilisation and if the results justify an investment using an IRR metric. Figures 5a and 5b present the averaged percentage of solar generation capacity used versus the generated energy exported to the utility for primary- and high schools respectively. Both figures show similarly shaped graphs, indicating similar inter-day consumption profiles for the forecast and measured loads. However, the forecast consumption profile is slightly better suited for solar intervention due to a larger percentage of generation capacity being used.



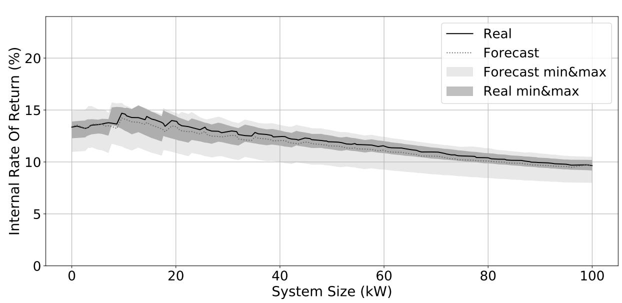
(a) Primary school solar generation results



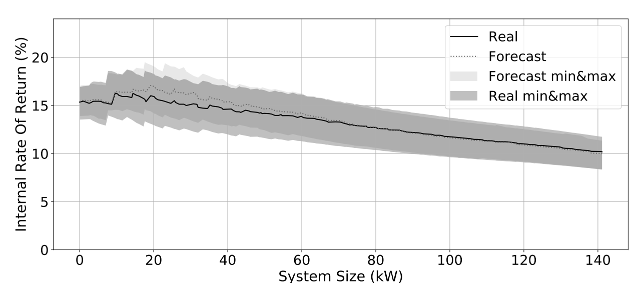
(b) High school solar generation results

Figure 5: Solar generation usage for varied system sizes.

Figures 6a and 6b present the averaged internal rate of return for primary- and high schools over a 25 year system lifespan. The IRR gives an indication of the profitability of an investment, with a larger percentage equating to a greater return. As explained in Section 2, it is seen that the calculation takes net cash flow into account. The initial investment cost was estimated using recent module, inverter, balance of system (BOS) and labour costs (International, Renewable Energy Agency, 2019). The jagged graphs on the figures are due to differences in system costs for various generation capacities ranging from R16 ($1.13) per Watt installed to R12.88 ($0.93) per Watt installed. The first year cash flow does not consider the benefits obtained through tax incentives, even though South African law permits a full deduction of the system's value as depreciation during the first year of operation for commercial installations (Dippenaar, 2018). Further assumptions made include electricity prices rising by 7% per year, considered modest within the South African context, while export rates increase by 5% per year (Eskom, 2019a). A yearly maintenance fee of R1000 ($70.40) was deducted and an inverter replacement was made every 10 years. The shaded areas present the minimum and maximum IRR graph paths for the real- and forecast simulation results, indicating a greater discrepancy within the primary school results as the minimum forecast IRR was 1.0% to 1.5% lower than the real value. Differences in the forecasted and real IRR rate are due to variations in their daily load profile. This, in addition to dissimilar monthly demand estimations led to mismatched yearly cash-flows. From the figures it is clear however, that the results far exceed the rate of return determined using the previously discussed simple sizing method, considered a baseline, and that the discussed model and accompanying methods provide an improved approach to sizing solar systems only using monthly measurement data.



(a) Primary school internal rate of return.



(b) High school internal rate of return.

Figure 6: Internal rate of return for varied system sizes.

Table 3: Forecast simulation results

School	Maximum Forecast error		Average Forecast accuracy	
	Annual energy (%)	Peak demand (%)	Annual energy (%)	Peak demand (%)
A	23.6	18.7	97.2	96.9
B	15.0	26.8	98.6	99.6
C	36.1	15.5	90.7	93.4
D	26.7	36.9	95.4	97.3
E	29.9	46.6	96.9	88.1

5. Conclusion

This paper addresses the problem that current techno-economic sizing of PV solutions for schools in South Africa rely on hourly smart meter measurements, costly equipment and scarce skills. The paper presented a novel mechanism by which the laborious and expensive task of sizing a PV solution can be performed by using one or more monthly bills and some commonly known metadata.

The proposed approach creates a generic energy consumption profile for a building from measured energy usage data of a subset of schools in developing countries and cities. This method is expanded to scale the load profile using only usage data and seasonal dates to produce a load forecast. A PV optimisation technique is implemented using the forecast to determine the potential profitability of the solar system through metrics including the internal rate of return and the utilisation of the system's generating capacity.

This proposed technique was applied to five schools using their hourly energy usages obtained using smart meters. In each assessment, the schools used for training did not overlap the school being assessed, ensuring not overfitting and reliable assessments.

With the provided data the method was capable of forecasting the yearly energy consumption of the schools to within an averaged accuracy of 5% while estimating the maximum monthly demand to 6% of the measured usage. Furthermore, through an iterative solar sizing technique it was possible to accurately estimate the potential profitability of a solar system tailored to the school's budget, presenting itself as a valuable tool for reducing the financial burden many schools face, and the assessment challenges the city councils and government departments face.

Funding

This research was funded by MTN South Africa.

6. References

Azizalrahman, H., Hasyimi, V., 2019. A model for urban sector drivers of carbon emissions. Sustainable Cities and Society 44, 46 – 55. doi:10.1016/j.scs.2018.09.035.

Baker, L., Sovacool, B.K., 2017. The political economy of technological capabilities and global production networks in south africa's wind and solar photovoltaic (pv) industries. Political Geography 60, 1 – 12. doi:10.1016/j.polgeo.2017.03.003.

Bourdeau, M., Zhai, X.Q., Nefzaoui, E., Guo, X., Chatellier, P., 2019. Modeling and forecasting building energy consumption: A review of data-driven techniques. Sustainable Cities and Society 48, 101533. doi:10.1016/j.scs.2019.101533.

Dietz, T., Frank, K.A., Kelly, J., Kelly, R., Whitley, C.T., 2015. Political influences on greenhouse gas emissions from us states. Proceedings of the National Academy of Sciences 112, 8254–8259. doi:10.1073/pnas.1417806112.

Dippenaar, M., 2018. The role of tax incentives in encouraging energy efficiency in the largest listed South African businesses. South African Journal of Economic and Management Sciences 21, 1 – 12. doi:10.4102/sajems.v21i1.1723.

Dobos, A.P., 2012. An Improved Coefficient Calculator for the California Energy Commission 6 Parameter Photovoltaic Module Model. Journal of Solar Energy Engineering 134. doi:10.1115/1.4005759.

Department of Energy, S.A., 2018. Coal resources overview. http://www.energy.gov.za/files/coal_frame.html. Accessed on 2019-07-01.

Eskom, 1999. Eskom 1990 -1999 - african renaissance - powering transformation. http://www.Eskom.co.za/sites/heritage/Pages/1990.aspx. Accessed on 2019-07-01.

Eskom, 2008. Eskom 2000 - 2008 - our recent past -shift performance and grow sustainably. http://www.Eskom.co.za/sites/heritage/Pages/2000.aspx. Accessed on 2019-07-01.

Eskom, 2017. Eskom yearly integrated report for the year 2017. http://www.Eskom.co.za/IR2017/Documents/Eskom_integrated_report_2017.pdf. Accessed on 2019-07-01.

Eskom, 2019a. Tariff history. http://www.Eskom.co.za/CustomerCare/TariffsAndCharges/Pages/Tariff_History.aspx. Accessed on 2019-07-04.

Eskom, 2019b. Tariffs and charges booklet 2019/2020. http://www.Eskom.co.za/CustomerCare/TariffsAndCharges/Pages/Tariffs_And_Charges.aspx. Accessed on 2019-07-02.

Heylman, C., Kim, Y.G., Wang, J., 2015. Forecasting energy trends and peak usage at the University of Virginia, in: 2015 Systems and Information Engineering Design Symposium, pp. 362–368. doi:10.1109/SIEDS.2015.7117006.

International Energy Agency, F., 2018. Electricity statistics: Detailed, comprehensive annual data on electricity and heat. https://www.iea.org/statistics/electricity/. Accessed on 2019-06-24.

International, Renewable Energy Agency, 2019. Renewable Power Generation Costs in 2018. https://www.irena.org/publications/2019/May/Renewable-power-generation-costs-in-2018. Accessed on 2019-06-26.

Jackson, R.B., Quéré, C.L., Andrew, R.M., Canadell, J.G., Korsbakken, J.I., Liu, Z., Peters, G.P., Zheng, B., 2018. Global energy growth is outpacing decarbonization. Environmental Research Letters 13, 120401. doi:10.1088/1748-9326/aaf303.

Kocak, E., Ulucak, R., Dedeoglu, M., Ulucak, Z.S., 2019. Is there a trade-off between sustainable society targets in Sub-Saharan Africa? Sustainable Cities and Society 51, 101705. doi:https://doi.org/10.1016/j.scs.2019.101705.

Kuster, C., Rezgui, Y., Mourshed, M., 2017. Electrical load forecasting models: A critical systematic review. Sustainable Cities and Society 35, 257 – 270. doi:https://doi.org/10.1016/j.scs.2017.08.009.

L King, D., Gonzalez, S., M Galbraith, G., E Boyson, W., 2007. Performance model for grid-connected photovoltaic inverters. Sandia Natl. Lab. 38.

Future of Life Institute, U.S., 2016. Developing countries can't afford climate change. https://futureoflife.org/2016/08/05/developing-countries-cant-afford-climate-change/?cn-reloaded=1. Accessed on 2019-06-23.

McLaren, D., 2017. Funding basic education. http://section27.org.za/wp-content/uploads/2017/02/Chapter-2.pdf. Accessed on 2019-07-04.

Myende, P., 2015. Tapping into the asset-based approach to improve academic performance in rural schools. Journal of Human Ecology 50, 31–42. doi:10.1080/09709274.2015.11906857.

Myende, P., Hlalele, D., 2018. Framing sustainable rural learning ecologies: A case for strength-based approaches. Africa Education Review 15, 1–17. doi:10.1080/18146627.2016.1224598.

NREL, 2018. Solar advisor model technical reference. https://www.nrel.gov/docs/fy15osti/64102.pdf. Accessed on 2019-06-22.

Osuri, S.O., Ngoma, J.L., Chowdhury, S.P., 2015. Cost-effective renewable energy technologies in universities of south africa: A case study, in: International Conference on Renewable Power Generation (RPG 2015), pp. 1–5.

PVGIS, E.C., 2018. Photovoltaic geographical information system. http://re.jrc.ec.europa.eu/pvg_tools/en/tools.html#TMY. Accessed on 2019-06-24.

Raza, M.Q., Nadarajah, M., Quoc Hung, D., Baharudin, Z., 2016. An intelligent hybrid short term load forecast model for seasonal prediction of smart power grid. Sustainable Cities and Society doi:10.1016/j.scs.2016.12.006.

Ripunda, C., Booysen, M.J., 2019. Understanding and affecting school water behaviour using technological interventions. URL: engrxiv.org/c7ujh, doi:10.31224/osf.io/c7ujh.

Samuels, J.A., Booysen, M.J., 2019. Chalk, talk, and energy efficiency: Saving electricity at south african schools through staff training and smart meter data visualisation. Energy Research & Social Science 56, 101212. doi:10.1016/j.erss.2019.05.022.

Sanchez-Torija, J., Larrumbide, E., Bedoya, C., 2017. The incorporation of the study into water consumption in energy audits in schools. Revista de la construccin 16, 361–373.

Sedibe, M., 2011. Inequality of access to resources in previously disadvantaged south african high schools. Journal of Social Sciences 28, 129–135. doi:10.1080/09718923.2011.11892937.

Smit, S., Musango, J.K., Brent, A.C., 2019. Understanding electricity legitimacy dynamics in an urban informal settlement in south africa: A community based system dynamics approach. Energy for Sustainable Development 49, 39 – 52. doi:https://doi.org/10.1016/j.esd.2019.01.004.

Van der Berg, S., 2009. Fiscal incidence of social spending in South Africa, 2006. Working Papers 10/2009. Stellenbosch University, Department of Economics. URL: https://ideas.repec.org/p/sza/wpaper/wpapers82.html.

Van der Berg, S., Burger, C., Burger, R., de Vos, M., du Rand, G., Gustafsson, M., Moses, E., Shepherd, D., Spaull, N., Taylor, S., van Broekhuizen, H., 2011. Low quality education as a poverty trap. Working Papers 25/2011. Stellenbosch University, Department of Economics. URL: https://ideas.repec.org/p/sza/wpaper/wpapers255.html.

Visser, M., Booysen, M.J., Berger, K., Brühl, J., 2019. Saving water at cape town schools by using smart metering and behavioural change. doi:10.31224/osf.io/7kzwv.

Voss, K., Sartori, I., Napolitano, A., Geier, S., Gonalves, H., Hall, M., Heiselberg, P., Widn, J., Candanedo, J., Musall, E., Karlsson, B., Torcellini, P., 2010. Load matching and grid interaction of net zero energy buildings, in: Concepts and strategies for zero emission buildings, pp. 101–108. doi:10.18086/eurosun.2010.06.24.

World Energy Council, U.K., 2016. World energy resources: 2016. https://www.worldenergy.org/wp-content/uploads/2016/10/World-Energy-Resources-Full-report-2016.10.03.pdf. Accessed on 2019-06-25.

Yang, H., Wei, Z., Chengzhi, L., 2009. Optimal design and techno-economic analysis of a hybrid solarwind power generation system. Applied Energy 86, 163 – 169. doi:https://doi.org/10.1016/j.apenergy.2008.03.008. iGEC III.

Yilmaz, S., Chambers, J., Patel, M., 2019. Comparison of clustering approaches for domestic electricity load profile characterisation - implications for demand side management. Energy 180, 665 – 677. doi:https://doi.org/10.1016/j.energy.2019.05.124.

Zala, H.N., Abhyankar, A.R., 2014. A novel approach to design time of use tariff using load profiling and decomposition, in: 2014 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), pp. 1–6. doi:10.1109/PEDES.2014.7042027.

Zhang, C., Cao, L., Romagnoli, A., 2018. On the feature engineering of building energy data mining. Sustainable Cities and Society 39, 508 – 518. doi:https://doi.org/10.1016/j.scs.2018.02.016.

Zhu, J., Shen, Y., Song, Z., Zhou, D., Zhang, Z., Kusiak, A., 2019. Data-driven building load profiling and energy management. Sustainable Cities and Society 49, 101587. doi:10.1016/j.scs.2019.101587.