Demand Forecasting Models for Academic Campus Buildings

Johann Bierman
Dept. of Electrical and Electronic Engineering
Stellenbosch University
Stellenbosch, South Africa
jbierman@sun.ac.za

Bernard Bekker
Dept. of Electrical and Electronic Engineering
Stellenbosch University
Stellenbosch, South Africa
bbekker@sun.ac.za

***Abstract* — The use of demand forecasting, benchmarking and response software is proposed as a solution for the Facilities Management division at Stellenbosch University to reduce the maximum demand cost component of campus facilities. This paper aims to address the demand forecasting segment by investigating three commonly used forecasting methods, namely Least squares estimation (LSE), Gaussian process regression, and Artificial neural networks. Forecast models implementing these methods are designed, trained on historical demand data and weather data from 2013 to 2015, and their response for 2016 is analyzed. The LSE forecast model yields the most accurate results for the following input parameters: forecasted temperature and ultraviolet A-rays (UVA), time of the day, and the corresponding demand measurements of one and two weeks earlier.**

Keywords — demand forecasting, academic campus buildings, least squares estimation, gaussian process regression, artificial neural networks

I. Introduction

South Africa's electricity public utility, Eskom, first introduced the time of use (TOU) tariff structure in the 1990s to make the cost of electricity more cost reflective and to encourage customers to shift load out of the peak season and peak times [1]. This tariff structure also includes billing customers for their monthly peak demand that was utilised. As a result, energy consumers have complex electrical tariff structures that consist of two aspects; electrical power demand (from here on refer to as demand) and TOU related charges.

Stellenbosch University (SU) has various sustainability initiatives to reduce its energy consumption [2]. The energy usage at SU's campuses are measured by smart electricity meters. Electrical data is used for billing purposes and to perform energy audits in order to identify potential savings of energy-intensive campus buildings.

The Facilities Management (FM) division at SU maintains most campus facilities and services. After investigating FM's current demand management approaches and tools, it was concluded that there is a need for a monitoring system that acts as a warning system which sends out alerts when a building's measured demand exceeds a predefined threshold. Therefore, the use of demand forecasting, benchmarking and response software, with a focus on specifically reducing the peak demand of campus buildings, is proposed to provide a solution for FM's need.

This paper aims to address the demand forecasting segment required for the proposed solution. Thus, it strives to identify the most suitable demand forecast model for academic campus buildings. The paper is structured as follows: Section II presents an overview of the system components required for the proposed solution. Section III discusses the selection of input parameters for demand forecasting by analyzing the correlation between these parameters. Section IV investigates three forecasting methods, and respectively designs forecast models. Section V compares the response of these models for various input structures, and extends the best performing model to be self-learning. Section VI concludes the study.

II. System Components

The proposed solution, as discussed in Section I, would allow FM to monitor, forecast, be alerted and respond to various measured demand profiles in order to optimise the energy usage and manage the demand of campus facilities. Thus, a solution that consists of the following components are required:

- **User Interface**: Dashboard software that effectively conveys information to the user regarding the current and forecasted demand profile, alerts and allows modification of systems settings.
- **Communication Network**: Allows communication between system components.
- **Measured Data**: Measured demand data.
- **Exogenous Data**: Input of external variables such as hyperparameters and weather forecasts.
- **Forecast Model with a self-learning feedback loop**: Implementation of a forecasting model that forecasts the demand profile and dynamically evolves.
- **Benchmarking**: Logical monitoring system that compares the measured demand profile with the forecasted demand profile.
- **Demand Response**: A system that is able to modify the demand profile by managing various loads.

A conceptual diagram of such a system, showing the relationships between components, is displayed in 0.

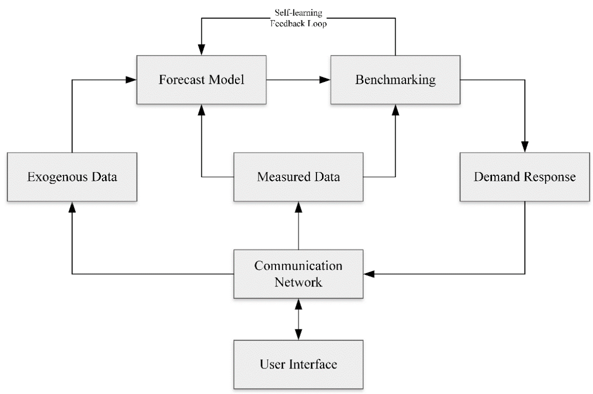

Fig. 1. System diagram for demand forecasting, benchmarking and response software

The demand forecasting segment incorporates four components: Exogenous Data, Measured Data, Forecast Model, and Benchmarking. As shown in Fig. 1, the forecast model component is interconnected with the other three. These components are discussed in rest of the paper with the aim to construct a demand forecast model suitable for academic campus buildings.

Note that MATLAB was used for the processing of data, to develop the forecast models and perform detailed analyses. MATLAB was chosen because of its available toolboxes, fast matrix operations, support services, and well-documented function descriptions [3].

III. Input Parameters

This section discusses various input parameters appropriate for demand forecasting, generates suitable hyperparameters, and analyses the correlation between these parameters.

The maximum demand (MD) means the highest average demand measured in kVA at the per point of delivery (POD) over a 30-minute integrating period during a billing month [4]. This excludes the demand registered during off-peak times [1]. Network access charges (NACs) are based on the notified maximum demand (NMD). The NMD is the maximum capacity in kVA, as measured over a 30-minute integrating period, at the POD which the customer contract to be made available during all time periods [5]. If the MD exceeds the NMD, the consumer will have to pay an additional amount that is incorporated in the NAC. Customers are allowed to reduce their NMD only once per 12-month cycle [6].

As the forecast models set out to predict only the demand, all input parameters need to be measured or processed as 30-minute resolution data. Three categories of available input parameters are defined: demand data and weather data, and hyperparameters.

A. Demand Data

An energy audit [7] on the top 44 largest energy consumption buildings of 2017 at SU, identified the top two excessive buildings as: Arts and Social Sciences (ASS) and JS Gericke Library (SU Library). These two buildings' demand data are used in this study to compare the forecast models. However, this paper refers mainly to ASS as SU Library yields similar results.

Due to the limited availability of data, it was decided that the demand data of 2016 will act as live measurements and that the weather data of 2016 will act as forecasted weather data. This will allow caparison of the designed forecast models across all months of the selected year. Thus, these models are trained on historical data from 2013 to 2015 and evaluated for 2016.

B. Weather Data

The weather data used in this study was obtained from Southern African Universities Radiometric Network's (SAURAN) website [8]. Weather data from the SUN weather station located at SU was downloaded. The data consists of solar irradiance, ultraviolet (UV), air temperature, barometric pressure (BP), relative humidity (RH), wind speed (WS), and wind direction measurements.

C. Hyperparameters

It may be beneficial to include hyperparameters as input to forecast models. The following time and occupancy based hyperparameters were created for this study:

- **Time**: Time of the day, Day of the month, Month of the year, Year, Weekday, and weekend.
- **Occupancy**: Public holidays, Class periods, Exam periods, University recess periods, and University office periods.

The time of day hyperparameter was at first a linear relationship of 1:48 which represented every half-hour of a single day, but this linear relationship is not ideal when using a linear forecasting method such as Least squares estimation (LSE). An innovative method to optimise this hyperparameter is to represent time as a sinus function where zero is at the start of the day, peak at noon and zero at the end of the day. Day of the month, Month of the year and Year hyperparameters represents the current day, month and year of each entry. The weekday hyperparameter is equal to one when the current day is a weekday and zero otherwise. The weekend parameter is the complement of weekday.

The public holiday hyperparameter is equal to one if the current day is a public holiday and zero otherwise. If the current day falls within a class, university recess or university office period the corresponding hyperparameter is equal to one, otherwise zero. Exams at SU stretches over two periods; first exam opportunities period and second exam opportunities period. Therefore, the exam hyperparameter is set to be equal to two during the first exam opportunity period, and equal to one during the second exam opportunity period (which represents a lower occupancy), and zero otherwise.

D. Correlation

The autocorrelation function (ACF) plot of ASS for 2015's demand data lagged in weeks for four weeks, is displayed in Fig. 2. It is evident that there are daily and weekly seasonality in the demand profile. It was found that the ACF plot for 2013 till 2015's demand data lagged in years, indicated no annual seasonality. However, the demand profile of SU Library indicates annual seasonality, but the correlation gradually weakens as the number of lags is increased. Note that the ACF only captures linear relationship between lagged values [9].

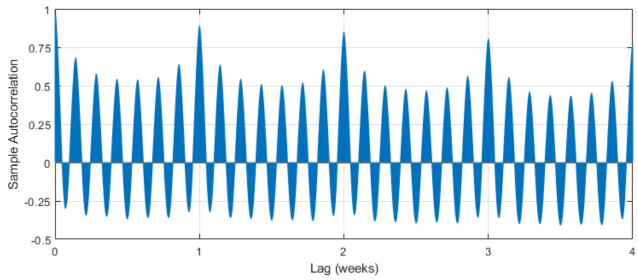

Fig. 2. ASS sample ACF plot for 2015

The correlation coefficients of the demand profile and available weather variables were analysed. BP, RH and WS is omitted because of the weak correlation that exists between these variables and the demand profile. DNI, ultraviolet A-rays (UVA) and ultraviolet B-rays (UVB) measurements are similar in shape and relative change in magnitude. Thus, only one of the three variables should be considered to reduce duplication of input data. UVB is omitted because ultraviolet radiation reaching the earth surface are mostly UVA. The correlation coefficients between the demand of versus DNI and UVA are equal to 0.3884 and 0.4112 respectively. Therefore, UVA is chosen and DNI omitted.

The correlation of both building demand profiles versus the selected weather variables (UVA, temperature and time) for 2013 to 2015 is shown in Fig. 3. It is observed that the demand does not significantly co-relate with temperature.

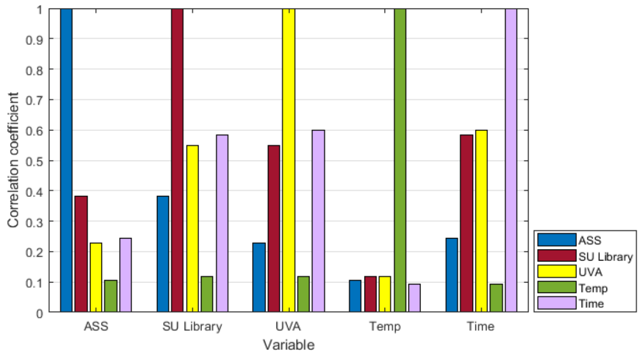

Fig. 3. Correlation between demand and weather variables for 2013 to 2015

Scatter diagrams and heat maps can be useful to visually identifying correlation and patterns between data variables. Scatter plots for the demand profile of ASS versus time, for three days and one month, are shown in Fig. 4. Normalised values, ranging [0 1], was used in the scatter plots. It is noticed that demand profiles have a similar shape to a Gaussian distribution, usually peaks at the same time (midday), and the regression line has a positive gradient which indicates a slight positive correlation between time.

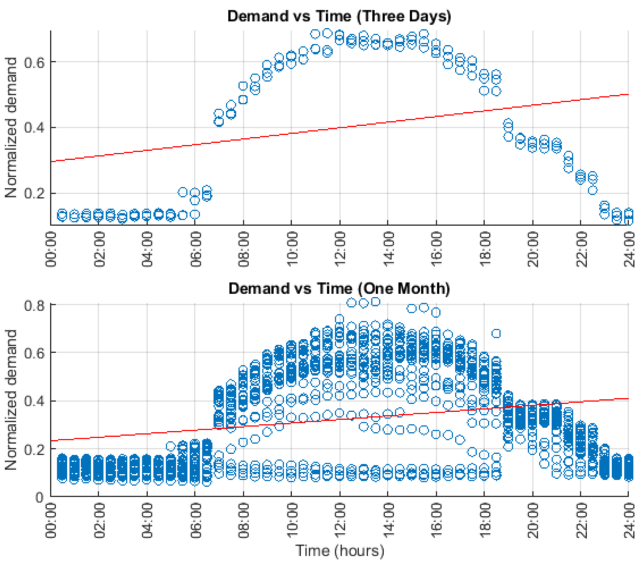

Fig. 4. ASS scatter plots for demand versus time, regression line indcated in red

The heat map of ASS for each week for a month in 2015 versus the day of the week is displayed in Fig. 5. It is evident that there exists daily and weekly seasonality. The demand is relatively low until around 7 am, then it peaks shortly after and remains relatively high until the end of the day. Also, the demand profile during weekdays is much higher opposed the weekend. On Thursday the 24th of September 2015 the demand profile is abnormal from the existing pattern. This was due to that day being a public holiday, therefore explaining the anomaly.

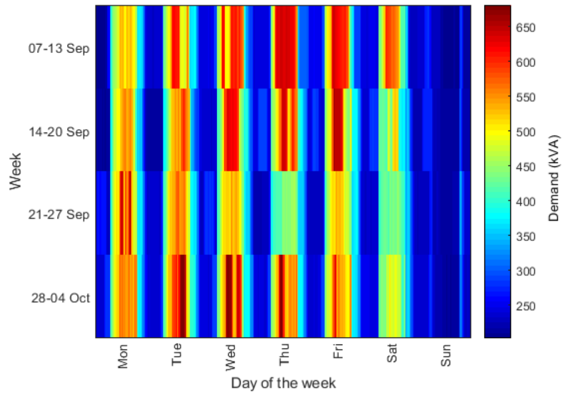

Fig. 5. ASS sample heat map for 2015

IV. Forecast Models

This section presents an overview of three commonly used forecasting methods, namely LSE, Gaussian process regression (GPR), and Artificial neural networks (ANNs). LSE is based on the least error squares criterion that has been used for load forecasting over the past 50 years [10]. GPR and ANN are popular but effective machine learning methods used for load forecasting [11]. This section also designs different forecast models, incorporating these methods. Section V compares the response of these models for various input structures.

A. Least Squares Estimation

Regression analysis study is often used to generate regression models for the purposes of forecasting future values of a response variable given regressor variables [12]. The static estimation problem as described by Soliman and Al-Kandari follows [10].

Given the linear equation

$$\underline{y} = X\underline{\theta} + \underline{e} \tag{1}$$

where $\underline{y}$ is an m×1 vector of known observations, $\underline{\theta}$ is an n×1 vector of the regression parameters to be estimated, X is an m×n matrix describing the mathematical relationship between the observations and the regression parameters, and $\underline{e}$ is an m×1 error vector. The state estimation problem is to estimate values for $\underline{\theta}$ such that $\underline{e}$ is minimised.

By minimising the least error squares cost function, the best regression parameter estimate can be calculated as

$$\hat{\underline{\theta}} = [X^T X]^{-1} X^T \underline{y} = [X]^{+} \underline{y} \tag{2}$$

where $\hat{\underline{\theta}}$ is the optimal least error squares estimate of θ.

This method of parameter estimation (from here on refer to as training) a regression model by the least error squares is referred to as LSE. More complex methods include weighted linear, constrained, recursive and nonlinear estimation. LSE is biased with outliers in the regressor variables as it does not ignore outliers during training. LSE estimates best when the regressor variables have a Gaussian error distribution [10].

The LSE forecast model is formulated as in (3). The model's regression parameters, $\hat{\underline{\theta}}$, are trained using (2).

$$\hat{y}(t) = \hat{\theta}_0 + \sum_{i=1}^{n} \hat{\underline{\theta}}_i X_i(t) \tag{3}$$

where $\hat{y}(t)$ is the forecast demand at time t, $X_i(t), \ldots, X_n(t)$ are the input variables at time t, and $\underline{\hat{\theta}_i}$ is the estimated regression parameters.

B. Gaussian Process Regression

A Gaussian process is a collection of random variables which have a multivariate Gaussian distribution [13]. GPR is a supervised machine learning method that incorporates probabilistic programming to estimate a set of training inputs as a known distribution by a defined mean and variance [14].

A Gaussian process can be described as

$$f(y) \sim GP\left(E(y), k(y,y')\right) \tag{4}$$

where f(y) is a real process, *GP* represents a Gaussian distribution with a mean function E(y) and covariance function $k(y,y')$.

A zero-mean function is generally chosen [13]. The covariance function is also known as a kernel. A popular kernel choice is the squared exponential:

$$k(y, y') = \sigma_f^2 \times \exp\left[\frac{-(y-y')^2}{2l^2}\right] \tag{5}$$

where σ_f^2 is the maximum covariance [15].

Kernels are used to capture characteristic patterns in data. A typical application for using the squared exponential kernel is to estimate cyclic variations. A constant kernel is typically used for noisy data and a linear kernel for one-dimensional linear data structures. Kernels can also be multi-dimensional and combined to optimise estimation. [14]

The Regression Learner application as part of MATLAB's Statistics and Machine Learning Toolbox is able to create and train GPR models. There are many specifications that can be altered to optimise the GPR models for the purposes of forecasting. The application has four pre-set GPR models and allows the user to specify one of the following kernels: Rational Quadratic GPR, Squared Exponential GPR, Matern 5/2 GPR and Exponential GPR.

The Simple input structure, defined in Section V, was used to test which model would estimate 2013 to 2015's data best in terms of the root mean square error (RMSE). The Exponential GPR model estimated the input parameters best. It was found that the model performs better when data validation is disabled. Not validating the data might not ensure protection of overfitting which can be detrimental to the application forecasting. However, this was not the case for the set of input parameters. Therefore, Exponential GPR with no validation is selected as the GPR forecast model.

C. Artificial Neural Networks

The human brain consists of billions of neurons, non-linearly connected by synapses, which are capable of solving complex problems. ANN is a machine learning method that attempts to emulate how the brain processes information. An artificial neuron can be mathematically described as

$$\varphi\left(\underline{w}^T\underline{x} + b\right) \tag{6}$$

where φ is the activation function, $\underline{x}$ is an vector of input variables, $\underline{w}$ is an weight vector, and b is the bias value [9].

Artificial neurons, also referred to as a node, is connected via synapses that has a synaptic weight to emulate the intensity of passing signals between biological neurons. Each node's output is transformed by an activation function. A visual representation of an artificial neuron with three inputs is shown in Fig. 6. ANNs consist of many nodes that are connected in layers.

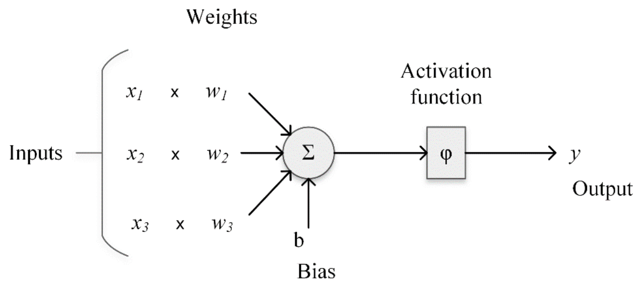

Fig. 6. Artificial neuron, adapted from Bartoš [9]

There are many strategies for choosing an ANN's architecture (activation function, number of layers and layer structure). A feedforward network is a simple network that consists of layers in series. ANNs are generally used as a "black-box" type model to capture various linear and non-linear relationships [12]. ANNs can be used for both supervised and unsupervised learning. During training, the synaptic weights are estimated. Generally, gradient descent methods are used to estimate these weights. The backward propagation of errors algorithm is a popular example of a gradient descent method. The use of binary hyperparameters to indicate the presence or absence of an event is frequently used. Thus, ANNs are typically over-parametrised and prone to overfitting [12].

The ANN forecast model can be described as follows:

- Network structure – Feedforward network
- Hidden layers – 12
- Training function – Bayesian regularisation backpropagation
- Training data – Training, testing and validation data are chosen at random at a distribution of 70%, 15% and 15% respectively.

The number of hidden layers was determined by training 30 different feedforward neural networks, with hidden layers ranging from 1 to 30, for each meter on the Simple input structure from 2013 to 2015. The demand profile for 2016 was forecasted (using its corresponding input data) and evaluated according to accuracy. The ANN model performed best with eight and 29 hidden layers for ASS and SU Library, respectively. Through inspection, an ANN model with 12 hidden layers serves as an adequate middle ground.

V. Forecasting Comparison

The response of the three forecast models, as designed in Section IV, for various input parameters are compared in this section. Firstly, the input structures used to compare these models are defined. Thereafter, accuracy measures and comparison results are discussed, and the best performing forecast model is extended to be self-learning.

The input structures used to compare the forecast models are listed below.

- **Simple**: Temperature, UVA, time of the day and the corresponding demand values from one and two weeks earlier.
- **Extended**: Temperature, UVA, time of the day, month of the year, year, weekday, public holidays and the corresponding demand values from one and two weeks earlier.
- **Complete**: Temperature, UVA, the corresponding demand values from one and two weeks earlier, and all the hyperparameters created in Section III (time of day, day of the month, month of the year, year, weekday, weekend, class periods, exam periods, public holidays, university recess periods and university office periods).

A. Comparison Results

In the context of forecasting, accuracy is difficult to quantify as the specific purpose for forecasting may differ [9]. For the purposes of peak shaving, accuracy of spikes or peak loads are of most importance. For the purposes of reducing energy usage, accuracy of the demand profile over the whole day may be of most importance.

Most accuracy measures are based on residuals and generally extends the percentage error or mean squared error measure. For this study, the accuracy of each forecast model is measured by two different accuracy measures, (7) and (8): Mean percentage error (MPE) and Scaled root mean squared error (SRMSE).

$$\text{MPE} = \text{E}\left(\frac{|\text{y-}\hat{\text{y}}|}{\text{y}} \times 100\%\right) \tag{7}$$

$$\text{SRMSE} = \frac{\sqrt{\text{E}((\text{y-}\hat{\text{y}})^2)}}{\text{E(y)}} \tag{8}$$

For the defined input structures, the forecast accuracies for the respective forecast models are tabulated in TABLE I.

TABLE I. FORCAST MODELS ACCURACY

Method	Input Structure	Accuracy Measure	
		MPE	*SRMSE*
LSE	Simple	8.804	0.1269
	Extended	10.79	0.1399
	Complete	649.1	6.683
GPR	Simple	9.572	0.1323
	Extended	15.27	0.1891
	Complete	13.05	0.1622
ANN	Simple	9.232	0.1292
	Extended	12.53	0.1525
	Complete	10.62	0.1343

The Simple input structure delivered the best results for all the forecast models. Fig. 7. illustrates the accuracy of each forecast model for the Simple input structure.

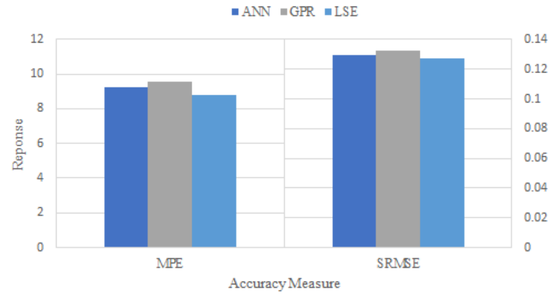

Fig. 7. Forecast model accuracy for the Simple input structure

From the results, the LSE model delivered the best results. However, the linearity of the LSE model does not react well to time and occupancy based hyperparameters, excluding the time of the day, as included in the Complete input structure. It is observed, for both GPR and ANN, that the forecast accuracy increased from only including some time and occupancy based hyperparameters to including all hyperparameters.

B. Self-learning Extension

To extend the LSE forecast model with a self-learning feedback loop, as described in Section II, an iterative approach is implemented. The LSE model is trained on data from 2013 to 2015, thereafter as time progresses (day by day) the previous day's data is appended to the training data, and the model is trained again. By incorporating the iterative feedback loop, the model is considered self-learning.

For ASS, the extended LSE model trained, using the Simple input structure, and forecasted the demand for 2016 in only 2.509 seconds. The response of the best and worst forecast day is displayed in Fig. 8 and Fig. 9. The best and worst day, best and worst month, and year average accuracies are shown in Fig. 10.

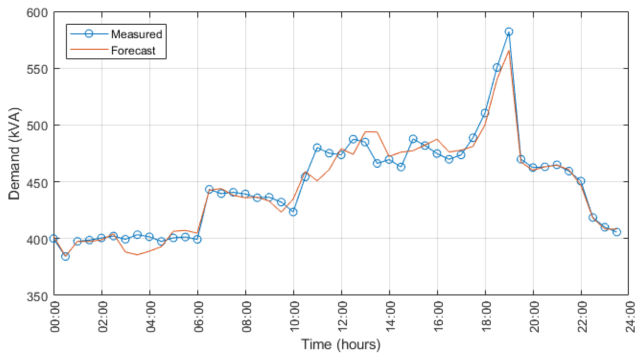

Fig. 8. ASS best day forecast: 12 June 2016

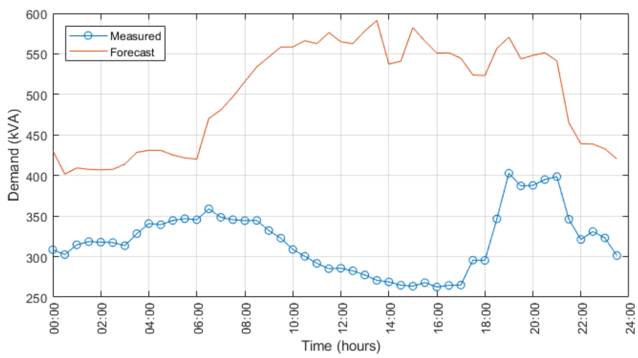

Fig. 9. ASS worst day forecast: 23 August 2016

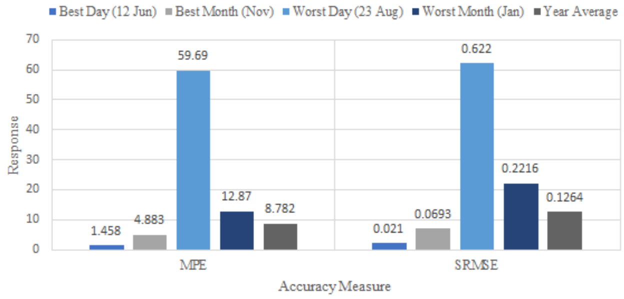

Fig. 10. Best and worst forecast accuracies for the extended LSE model

VI. Conclusion

From the investigated forecasting methods, suitable forecast models, for the purposes of demand forecasting of academic campus buildings, were constructed and evaluated. The LSE forecast model yielded the most accurate results when using the Simple input structure. Forecast accuracy of the LSE model was significantly worsened when additional time and occupancy based hyperparameters were used. The extended LSE model delivered more accurate results compared to the non-iterative approach. However, the improvement is not significant compared to changes in the input structure. Therefore, optimising the input structure may yield more accurate results.

Although the forecast models were designed for academic campus buildings, these models can be adapted for commercial buildings smart grids or utility-scale power systems. Future work could focus on the design and optimising of more complex forecast models from the forecasting methods investigated in this paper, to explore and compare other forecasting methods, or to explore innovative ways of combining various forecast models.

References

[1] D. Ramsbottom, "Time-of-use tariffs in the Eskom Western region," Energize, vol. 1, no. 28, pp. 29–33, 2009.

[2] Stellenbosch University, "SU Sustainability: Energy on campus," 2018. [Online]. Available: www0.sun.ac.za/sustainability/pages/services/energy/on-campus.php. [Accessed: 05-Oct-2018].

[3] MathWorks®, "MATLAB vs. Python: Top Reasons to Choose MATLAB," 2018. [Online]. Available: www.mathworks.com/products/matlab/matlab-vs-python.html. [Accessed: 30-Aug-2018].

[4] Eskom, "Eskom Schedule of Standard Fees 2018/19." 2018.

[5] Eskom, "Notified maximum demand rules," vol. Rev 2, no. February. 2009.

[6] Ethekwini municipality, "Notified Maximum Demand Rules," vol. Rev 3, no. June. 2014.

[7] F. Duvenhage, "Stellenbosch University Energy Audit.", Internal Report, Unpublished, CRSES, 2017.

[8] SAURAN, "Southern African Universities Radiometric Network." [Online]. Available: https://sauran.ac.za/. [Accessed: 30-Nov-2019].

[9] S. Bartoš, "Prediction of energy load profiles," Charles University, 2017.

[10] S. A. Soliman and A. M. Al-Kandari, Electrical Load Forecasting: Modeling and Model Construction. Elsevier, 2010.

[11] W. Xie, C. Wette, and B. Yan, "Improving University Energy Efficiency," Harvard University, 2014. [Online]. Available: http://cs109-energy.github.io. [Accessed: 07-Aug-2018].

[12] D. C. Montgomery, C. L. Jennings, and M. Kulahci, Introduction to Time Series Analysis and Forecasting. Wiley, 2008.

[13] C. E. Rasmussen and C. K. I. Williams, Gaussian Processes for Machine Learning. MIT Press, 2006.

[14] J. Maritz, F. Lubbe, and L. Lagrange, "A Practical Guide to Gaussian Process Regression for Energy Measurement and Verification within the Bayesian Framework," Energies, vol. 11, no. 4, 2018.

[15] M. Ebden, "Gaussian Processes for Regression: A Quick Introduction," 2008. [Online]. Available: https://arxiv.org/abs/1505.02965v2.