Parametric Time-series Modelling of London Smart Meter Data for Short-term Demand Forecasting

Arijit Mondal
Department of Computer Science, Faculty of Environment, Science and Economy, University of Exeter, North Park Road, Exeter, Devon EX4 4QE, United Kingdom. Email: arijit.mondal94@outlook.com

Saptarshi Das, *Member, IEEE*
Centre for Environmental Mathematics, Faculty of Environment, Science and Economy, University of Exeter, Penryn Campus, Cornwall TR10 9FE, United Kingdom. Email: saptarshi.das@ieee.org, s.das3@exeter.ac.uk

***Abstract*—Electricity being one of the most important components behind economic growth in 21st century, accurate electricity demand forecast became essential. Now with the deployment of smart meters that are capable of providing half-hour energy usage data comes new opportunities for short-term demand forecasting. In this research two statistical timeseries models known as the seasonal auto-regressive integrated moving average (SARIMA) and with exogenous inputs (SARIMAX) are employed to study half-hourly energy demand forecast and daily peak forecast capability over a week at half-hourly interval. The models are tuned and tested on a half-hourly aggregate level data and individual meters data extracted from London smart-meter dataset. The models are also cross validated over different seasons to evaluate model robustness over different training data size and forecasting under different temperature conditions. The SARIMA model performed better at consistently forecasting daily-demand peaks, while the SARIMAX was overall more accurate as compared to the SARIMA at more computational cost. This is because of the exogenous temperature variable used in SARIMAX which explains some of the demand profile volatility due to temperature changes. This also resulted in a better fit for the SARIMAX model. The models tested in this paper can accurately forecast energy-demand at half-hour intervals and daily-peaks for a week-ahead forecast at a regional demand profile over different seasonal condition.**

Keywords—smart meter data, time series models, forecasting

I. Introduction

In 21st century electricity is deemed as one of the primary forces behind economic growth and an absolute essential in our daily life. Over the last decade with rapid economic growth demand for electrical power has increased significantly. While UK is expected to see additional growth in demand as a result of electrification of portion of heating needs, multi-vector interaction between energy, transport and waste sector [1]. To address this fast-growing energy demand in UK distribution network operators (DNOs) are migrating to become distribution system operators (DSOs). As DSOs have greater visibility into energy generation and distribution assets [2]. This enables DSOs to make the resources economical while reducing carbon footprint by contracting services from distributed energy resources and renewables developers using low carbon generation technologies. However, with access to a more distributed grid, generation assets and due to variability in outputs from renewable developers, grid balancing has become a major concern for UK based DSOs. If a distributed grid is not balanced properly, it can lead to collapse of the grid with massive rolling blackouts. The solution to this problem is to avoid traditional and expensive redundancy-based energy system model and designing a smart grid. Here, DSOs have access to load forecasts through a demand-side management system (DSM) and can actively optimise the grid by increasing generation by switching on more wind turbines or requesting users to shift/reduce the demand.

Although this was almost impossible in the past, it is possible now with implementation of Advanced metering infrastructure (AMI). Since 2011, over 19.5 million smart meters have been installed in the UK. The high-resolution data generated by these smart meters can be used for energy demand forecasting [3]. Thus, the models need to be robust enough so that it can learn different demand features in different seasons from training data and produce accurate forecast and demand-peaks for optimisation. A large number of studies are carried out historically and in recent years on load forecasting which has practical and quantifiable impact on efficient power distribution and in-turn on the economy [4], [5]. Electricity demand/load-forecasting is a time-dependent problem and it is affected by factors such as temperature, appliances, occupancy and various other factors, which makes it a complex, multi-variate time-series forecasting problem.

A. Aims and Objectives

The aim of this paper is to compare two statistical time-series forecast models with a focus on short-term load forecasting (a week ahead) at half-hour interval. Models selected in this research are known as SARIMA and SARIMAX. The specific research objectives are:

- Cleaning, analysing and interpolating London smart meter data and associated weather data to identify and extract significant and meaningful features.
- Extensive hyper-parameter tuning and multi-step cross validation to ensure optimal performance of the models without overfitting.
- Evaluating effects of weather variables (temperature) on model forecasting accuracy and performance.

B. Related Previous Work

Now a days, due to very high electricity consumption, in recent years load forecasting and management systems are heavily being researched. This also provides practical and quantifiable impact on efficient power distribution and in-turn on the economy. The load forecasting methods can be divided into two main categories: data-driven or artificial intelligence (AI) model based and engineering/physical model based techniques. Engineering solutions involve very complex mathematical models, large number of system parameters to represent physical components. Not to mention that high-level of domain expertise is necessary to perform such complex model based forecast evaluation. Whereas, the data-driven models do not need such complex and detailed parameters and expertise. Instead AI models can learn from real-time historical smart meter data and also with season effects [6], [7]. Electricity demand/load-forecasting is a time-dependent problem and it is affected by factors such as temperature, appliances, occupancy and various other factors. These factors

979-8-3503-1997-2/23/$31.00 ©2023 IEEE

make it a complex, multi-variate time-series forecasting problem. Due to this complexity, various exogenous factors, strong seasonality combined with missing data and uncertainty, long-term forecasting becomes unreliable [8]. As a result, in recent years most of the research have been concentrated on short-term (day-week ahead) forecasting, which are more effective and usable [9]. AI models for smart meter can further be divided into three categories: traditional statistical models, machine learning (ML) models and deep learning (DL) models. Short-term can be hourly, daily or even weekly ahead forecasting.

For energy usage forecasting several statistical models have been used. Some of the most commonly used statistical models for short-term load forecasting are regression analysis [10], state-space model [11], Gaussian mixture model (GMM) [12], spatial-temporal model [13], high order Markov chain model [14], Kalman filter model [15]. Some research [16] also demonstrated the effect of time-resolution on forecast accuracy using linear regression models. Researchers have also used multiple regression combined with genetic algorithm to model short-term electricity demand [17]. Backward elimination based regression models which uses climatic variables to improve model accuracy has also been found effective for energy-demand forecast in for a building [18]. GMM, conditional demand model, autoregressive moving average (ARMA) models have also been the baseline for energy demand forecasting and is widely used [19]. However, some statistical models may have limitations or assumptions such as the data needs to be stationary, or follow normal distribution etc. which may not be optimal for the data processing and hyper-parameter tuning of those models.

Thus, researchers have also used machine learning models for demand/load forecasting. ML models like support vector regression (SVR) along with meta-heuristic optimization also showed great potential in demand forecasting while not suffering from the same limitations as traditional statistical models [20]. Chen *et al.* [21] used temperature variable to improve performance of the SVR model. Some researchers used hybrid wavelet SVM and neural networks to convert the time domain data [22]. In recent days, artificial neural networks (ANNs) are widely being tested for electricity-consumption forecasting. ANN models such as multi-layer perceptrons (MLPs), recurrent neural networks (RNNs), long-short-term-memory (LSTM) models for load forecasting have shown promising results [23]. However, one of the primary limitations of ANNs are, they are all black-box models and are not easily interpretable. Neural networks also need larger training data than traditional machine learning methods and can also be computationally expensive. In this research model interpretation has been given importance and therefore parametric statistical time-series models have been employed. Both SARIMA and SARIMAX are built on the ARMA model. However, it does not have the same limitations as ARMA model since both the models can capture non-stationary data with seasonal components.

II. SMART METER DATA AND METHODS

The primary objectives of this paper is to test the generalisability and robustness of the two statistical time-series models – SARIMA vs. SARIMAX. Therefore, the models are trained on highest-resolution, half-hourly load consumption values in KWh at an aggregate and individual level. The aggregate and individual data are extracted from the London smart meter dataset [24]. After pre-processing the data, half-hourly aggregated data was used for tuning and validation of the model over different seasons. Then the models were tested on the aggregate data. The next sections discuss the pre-processing of dataset, two models, and performance metrics used for model evaluation.

A. London Smart Meter Data and Exploratory Analysis

The dataset used in this study to evaluate the statistical time-series models is refactorized version of original dataset Smart Meter Energy Consumption Data [24]. We use the smart meter energy consumption data available in the dataset ranging from November 2011 to February 2014 from 5,567 London households which were part of low carbon London project led by UK Power Networks. The dataset includes unique Smart Meter identifier and their respective power consumption in kilowatt hour (KWh) in different resolution, half-hourly, hourly and daily divided into 110 blocks of houses. It also includes hourly and daily weather data from London. The blocks are grouped by 18 ACORN groups. ACORN is a segmentation tool which categorises the UK's population into demographic types based on socio-economic information about the household [25], [26]. According to the London Smart meter dataset documentation, A, B, C group belongs to affluent achievers' group. These are middle-aged to old people who are financially successful and are of the 'baby-boomer' generation. Here, group A lives in large plush mansions where group B lives in large metropolitan apartments and group C lives in large, detached houses. These groups likely to have high energy-demand which is analysed in detail in this section. Historical energy demand data from 2013-01-01, 00:00:00 to 2014-02-28, 23:30:00 have been used due to relatively low missing data and bad timestamps compared previous years.

The aggregate level historical half-hourly energy-consumption data was calculated and used to train the models. The data transformation method is given as:

$$\acute{y}_t = \frac{1}{n}\sum_{i=1}^{n}(y_t)_n \tag{1}$$

where, n is the number of smart metres, t is the time lag.

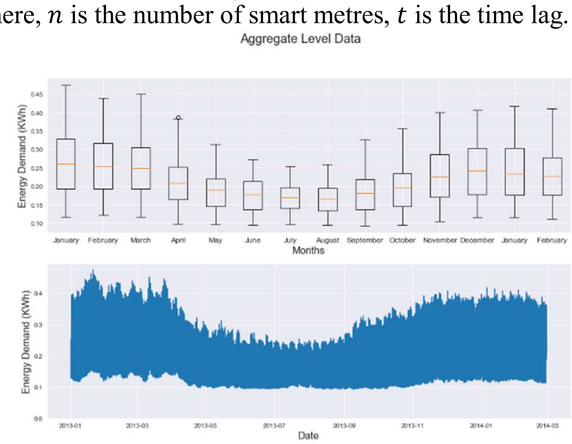


Figure 1. Aggregate level smart meter data over one year (2013) with boxplots showing distributions for each month.

The aggregate level data is used to tune, and cross validate the models. This dataset is also used to evaluate seasonal temperature effects on forecast by splitting the training sets (3, 6, 12 months) over Spring, Summer and Winter. In pre-processing stage first, the meter IDs with lowest missing data are selected. Then the meter IDs with over unusually high (20,000 kWh) or unusually low (2,000 kWh) consumption are removed. Then the houses with zero mean and zero standard deviation are filtered out. And lastly meter IDs with less than 3 kWh consumption for more than one month are filtered out as well. To test the effects of weather data on forecasting accuracy, hourly temperature data from London smart-meter dataset is used in SARIMAX model as an exogenous input. However, the temperature data, provided in the dataset is hourly observations of temperature in London. This is reindexed at half-hour interval and linearly interpolated to be used in this research. For data extraction, cleaning and interpolation of the aggregate data the Pandas library was used [27] as shown in Figure 1.

B. Time-series Model Description

In this paper two statistical models are used for short-term demand forecasting viz. seasonal autoregressive moving average (SARIMA) and seasonal autoregressive moving average with exogenous variable (SARIMAX) [28]. The exogenous variable used in SARIMAX model is the temperature data as shown in Figure 2. Both models belong to the traditional statistical timeseries model category. The SARIMA $(p, d, q)(P, D, Q, s)$ is based on Box-Jenkins ARMA (p, q) model. However, unlike the ARMA (p, q) model, SARIMA $(p, d, q)(P, D, Q, s)$ is not restricted to stationary data. It can model non-stationary data due to its additional Δ^d differencing term. This model also takes into account seasonality by applying ARIMA (p, d, q) to lags that are integer multiples of seasonality. On contrary, the SARIMAX $(p, d, q)(P, D, Q, s)$ model takes additional exogenous variables x_t with timestamp t being same for both x_t and the timeseries y_t.

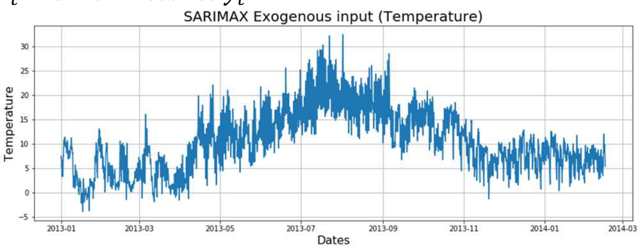


Figure 2. Temperature data for the SARIMAX model exogenous input.

C. Performance Metrics

Evaluation of the model performance is carried out by calculating forecasting error, which is the difference between actual and forecast values. Out of all the different error metrics a combination of scale dependent/independent error, percentage error and goodness of fit metrics are selected. Performance metrics used in this study are Akaike Information Criteria (AIC), coefficient of determination (R^2), root mean squared error (RMSE). RMSE is a scale-dependent metrics as a result the errors are on the same scale as the data [29]. Hence, it can be used to compare model performance over the same time-series. If f_t is the forecast and y_t is the ground-truth at time t over n number of test observations, the RMSE and MAE can be represented as:

$$\text{RMSE}\ (t) = \sqrt{\frac{\sum_{t=1}^{n}(y_t - f_t)^2}{n}} \tag{2}$$

Two additional performance metrics used in this study are AIC and R^2 to assess relative fit quality of the statistical models used in this study. Akaike Information Criteria (AIC) is an extension of maximum likelihood in multi-model situation. The AIC evaluates both risk of underfitting and overfitting of a model. If k is the number of estimated parameters in the time-series and L is the maximum value of the likelihood function, AIC can be expressed as:

$$\text{AIC} = 2k - 2\ln(L) \tag{3}$$

Coefficient of variation or R^2 represents the proportion of variance of y_{pred} explained by the independent variables y_{true}. Thus, R^2 can be used as an indication for how well the model is going to predict unseen variables through proportion of explained variance. Hence, R^2 is used as a goodness-of-fit indicator. R^2 can be calculated as:

$$R^2 = \frac{\sum_{i=1}^{n}(y_{true} - y_{pred})^2}{\sum_{i=1}^{n}(y_{true} - \bar{y}_{true})^2} \tag{4}$$

III. Simulations and Results

We first develop and tune both the chosen statistical time-series models to ensure best performance. Then we evaluate the model performance using the aggregate smart meter data. The models are implemented primarily using three Python libraries, Statsmodel, Scikit-learn and pmdarima [30], [31]. Model building, visualisation and forecasting is conducted on MacBook Pro with an Intel Core i7 2.9 GHz and 16 GB memory. While hyper-parameter tuning, and cross-validation were run on Windows 10 PC with an AMD Ryzen 7 3.6 GHz and 64 GB memory parallelised over 12 CPU cores. Integrated development environments used for this work are Jupyter notebook, PyCharm, Spyder using Python version 3.7. Tuning, forecasting results with and without exogenous temperature variable are discussed next.

A. Model Selection and Hyper-parameter Tuning

For both SARIMA and SARIMAX models, there are seven parameters $(p, d, q)(P, D, Q, s)$, divided into two categories as model order (p, d, q) and seasonal order (P, D, Q, s). The model order d is estimated through Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test and Augmented Dickey-Fuller (ADF) test as:

$$\text{KPSS} = \left(T^{-2} \sum_{t=1}^{T} \hat{S}_t^2\right) / \hat{\lambda}^2 \tag{5}$$

where, $\hat{S}_t = \sum_{j=1}^{t} \hat{u}_j,$ $\hat{u}_t$ is residual, $\hat{\lambda}^2$ is estimation of variance of u_t using $\hat{u}_t$.

$$\text{ADF}(\Delta_{y_t}) = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta_{y_{t-1}} + \delta_2 \Delta_{y_{t-2}} + \cdots \tag{6}$$

where, y_t is the data we are performing regression on.

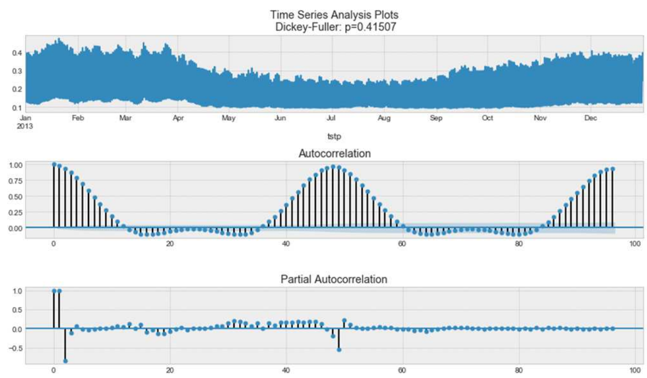


Figure 3. ACF/PACF plots for parameter estimation (time-lags on *x*-axis).

The KPSS test is designed to complement unit-root test such as ADF to determine the d parameter. The null hypothesis for KPSS test is the data is stationary while the null hypothesis for ADF test is that the data is non-stationary. To reject the null-hypothesis the *p*-value should be less than 0.05. The seasonal D parameter is estimated using Osborn-Chui-Smith-Birchenhall (OCSB) test which is similar to the ADF test. The null hypothesis of OCSB test is that timeseries contains a seasonal unit root.

The model order $(p,\ q)$ and seasonal order $(P,\ Q)$ is initially estimated through auto-correlation function (ACF) and partial auto-correlation function (PACF) plot using Box-Jenkins methodology as shown in Figure 3. From the PACF plot in Figure 4 for the differenced data AR(p) can be determined at $p = 2$, while the correlation around lag 47 can be interpreted as a seasonal corelation with seasonal $P = 1$. Thus, all combinations for ARMA (p, q) ranging from 0 to 3 and seasonal order (P, Q) ranging between 0 to 2 are tested to find the best parameters for both SARIMA and SARIMAX model. The best models above with highest R^2 value are then selected for cross validation to identify the best SARIMA and SARIMAX model as shown in Table I.

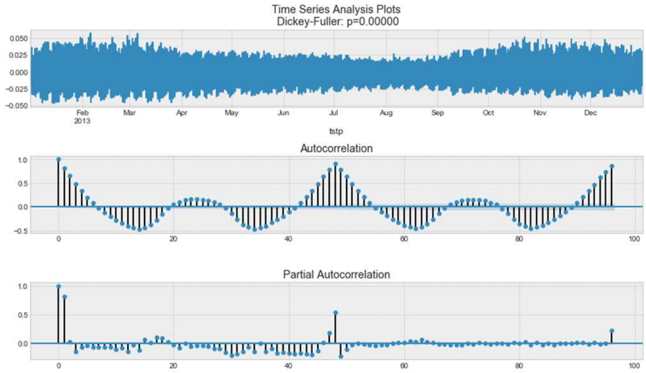


Figure 4. ACF/PACF plots with 1st order differencing for parameter estimation (time-lags on the *x* axis).

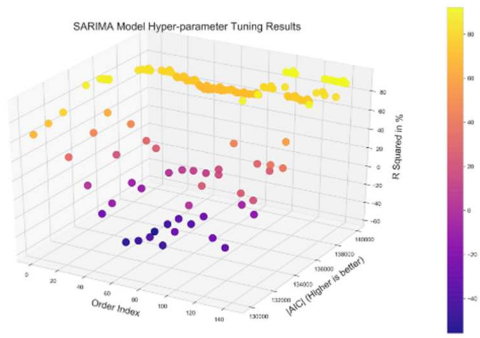


Figure 5. Hyper-parameter tuning of SARIMA model.

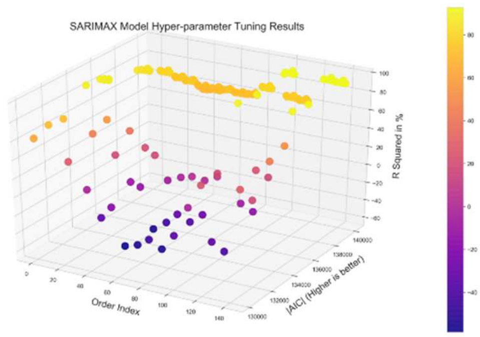
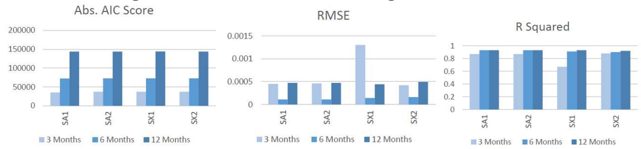


Figure 6. Hyper-parameter tuning of SARIMAX model with exogenous temperature variable.

B. Cross-Validation Results

Size of training set can also be very influential for predictive accuracy. If a model is trained on inadequate training data, the model might be too simple and never learn complex seasonal features. While large training sets with redundant features might result in very complex models with high computation time. Thus, after parameter tuning a cross-validation was performed to understand model performance and accuracy depending on the length of training sets and forecasting season. However, for time-series forecasting the usual k-fold cross-validation could not be applied. This is due to the temporal correlation between time-lags and presence of non-stationarity and seasonality. Thus, for this research a modified rolling-window cross validation method was used. Where training window is varied between three, six, twelve months of half-hourly aggregate energy consumption data for a-week-ahead forecast. The cross-validation results are shown in Figure 7 and Table II using AIC, RMSE and R^2.



(a) (b) (c)

Figure 7. Model cross-validation for 3, 6, 12 months. (a) Abs. AIC Score, (b) RMSE, (c) R^2. SA1: SARIMAX (2, 1, 3) (2, 1, 2), SA2: (2, 1, 3) (1, 1, 1), SX1=SARIMAX (2, 1, 3) (1, 1, 1), SX2: SARIMAX (3, 1, 2) (1, 1, 1).

TABLE I. BEST PARAMETERS (P, D, Q) OF THE THE TWO MODELS

Model	Order Index	Model Order $(\mathbf{p, d, q})$	Seasonal Order $(\mathbf{P, D, Q, S})$	Abs. AIC	R^2
SARIMA	107	$(2, 1, 3)$	$(2, 1, 2, 48)$	139623.84	0.92
	103	$(2, 1, 3)$	$(1, 1, 1, 48)$	139938.46	0.92
SARIMAX	103	$(2, 1, 3)$	$(1, 1, 1, 48)$	140090.92	0.93
	130	$(3, 1, 2)$	$(1, 1, 1, 48)$	140127.10	0.93

From Figure 7, the AIC score improves in proportion with training set size and R^2 also follows a similar pattern with marginal improvement between the training sets. This is relatively more evident on SARIMAX model as compared to the SARIMA model. In terms of daily peak demand forecast, both models trained on three, six months data performed better than the models trained on twelve months training data. Similar trends can be observed in overall accuracy (RMSE) as well. This trend can be explained by the effect of temperature on forecast error. With lower temperature demand profiles volatility increases which results in more forecasting error. Models trained on three, six months are forecasting over spring and summer time when temperature is relatively higher compared to models trained on twelve months data which is forecasting over the winter. Also, models trained on limited data might learn bias or variance so selecting models on overall accuracy is not a good strategy. This is why R^2 and AIC are primarily used for model selection in this research where R^2 shows the variance of ground truth explained by the forecast and AIC ensures the model does not overfit the data.

TABLE II. CROSS-VALIDATION RESULTS FOR THE TWO MODELS

Model	Order Index	Model Order $(\mathbf{p, d, q})$	Seasonal Order $(\mathbf{P, D, Q})$	Abs. AIC	R^2	RMSE
SARIMA	107	$(2, 1, 3)$	$(2, 1, 2, 48)$	84010.17	0.91	0.00035
	103	$(2, 1, 3)$	$(1, 1, 1, 48)$	84375.40	0.91	0.00035
SARIMAX	103	$(2, 1, 3)$	$(1, 1, 1, 48)$	84453.32	0.84	0.00062
	130	$(3, 1, 2)$	$(1, 1, 1, 48)$	84422.38	0.90	0.00034

C. Model Performance Evaluation

In this section we have selected the best SARIMA and SARIMAX models based on the procedure discussed in the previous section. The models are selected mainly based on the R^2 and AIC score. Where R^2 is the same, the AIC score is used as secondary selection criteria. The best models are described in Table III.

TABLE III. TUNED MODEL PROFILES, TRAINING SIZE AND TIME

Model	Order (p, d, q)	Seasonal Order (P, D, Q, S)	Training Data Size (in Months)	Training Time (minutes over single core CPU)
SARIMA	$(2, 1, 3)$	$(1, 1, 1, 48)$	12	68
SARIMAX	$(3, 1, 2)$	$(1, 1, 1, 48)$	12	92

The best models are then trained on twelve months aggregate level data and tested on one week ahead out of sample forecast at half-hour resolution to measure the accuracy. The generalizability of the models tested on three individual meters selected from three different socio-economic classifications ACORN-A, B, C.

1) SARIMA Fit Statistics and Forecast Results

Best parameters estimated through parameter tuning and cross validation are order (p, d, q): $(2,\ 1,\ 3)$ and seasonal order (P, D, Q): $(1,\ 1,\ 1)$ and seasonal S: 48. The week ahead out of sample forecast result are shown in Figure 8. A zoomed version in the predicted part is shown in Figure 9.

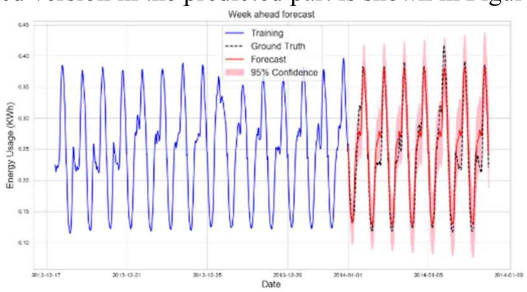


Figure 8. SARIMA week-ahead forecast with training and testing data.

It is evident that the model is performing well, as the forecast very closely follows the ground truth and the departures are mostly within 95% confidence. The model does tend to underestimate the daily peak. However, the margin of error is very low over a week ahead prediction. The way SARIMA works, it fits separate ARIMA components to lags that are integer multiples of seasonality and differenced data. Now there is some evidence, fitting a higher order ARIMA might cause common factor problem and risk of overfitting. To ensure that the model is not overfitting, and it is extracting adequate temporal features from the dataset the model has been evaluated using a set of residual ($e_t = y_t - \hat{y}_t$) diagnostic plots as shown in Figure 10.

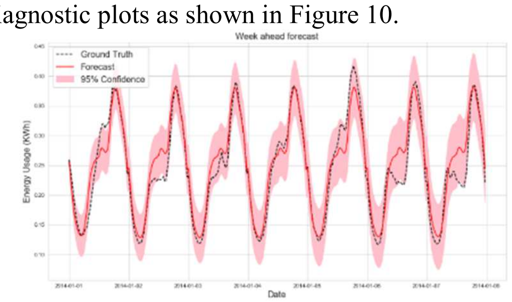


Figure 9. SARIMA week-ahead forecast for test data (zoomed Figure 8).

The correlogram plot shows no significant correlation between the residual and the standardized residuals closely following close to the delta function. Normal Q-Q plot and histograms also show the residual is closely following a normal distribution with mean close to 0. This indicates that the model is performing well at extracting features from the data without overfitting and without learning significant bias. In addition, the ground truth vs forecast regression also fits a straight line. This demonstrates the model fit quality is good.

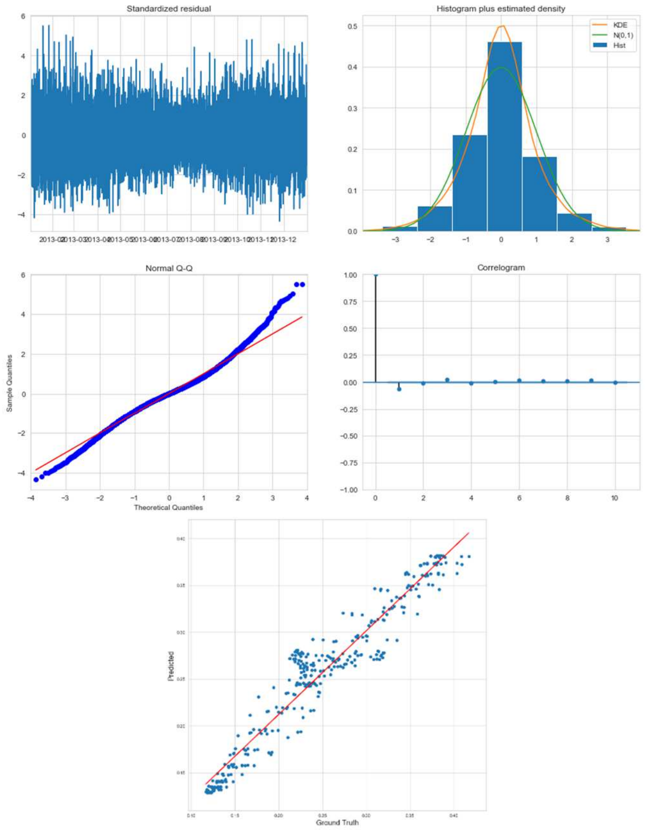


Figure 10. SARIMA residual diagnostics. (a) standardised residual, (b) histogram of residuals with estimated density, (c) normalised Q-Q plot (d) correlogram of residuals, (e) predicted vs ground truth regression.

2) SARIMAX with Exogenous Temperature Data Forecast and Fit Statistics

Through cross-validation and hyperparameter tuning, the best model SARIMAX model was selected with order (p, d, q): $(3,\ 1,\ 2)$ and seasonal order (P, D, Q): $(1,\ 1,\ 1)$ and the seasonality S: 48. The week ahead out-of-sample forecast a half-hour interval are shown in Figure 11 for both training and testing data with a zoomed version in Figure 12.

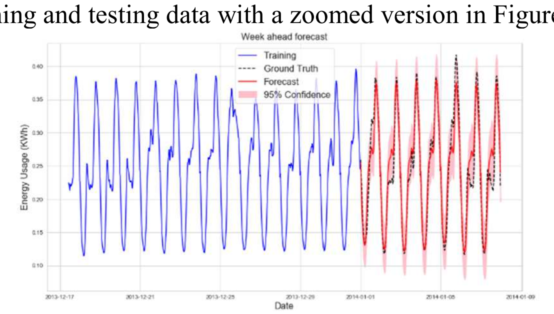


Figure 11. SARIMAX week-ahead forecast with training and testing data.

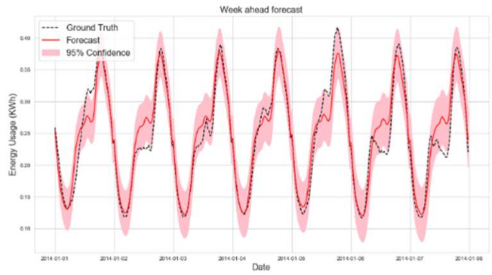


Figure 12. SARIMAX week-ahead forecast for test data (zoomed Figure 11).

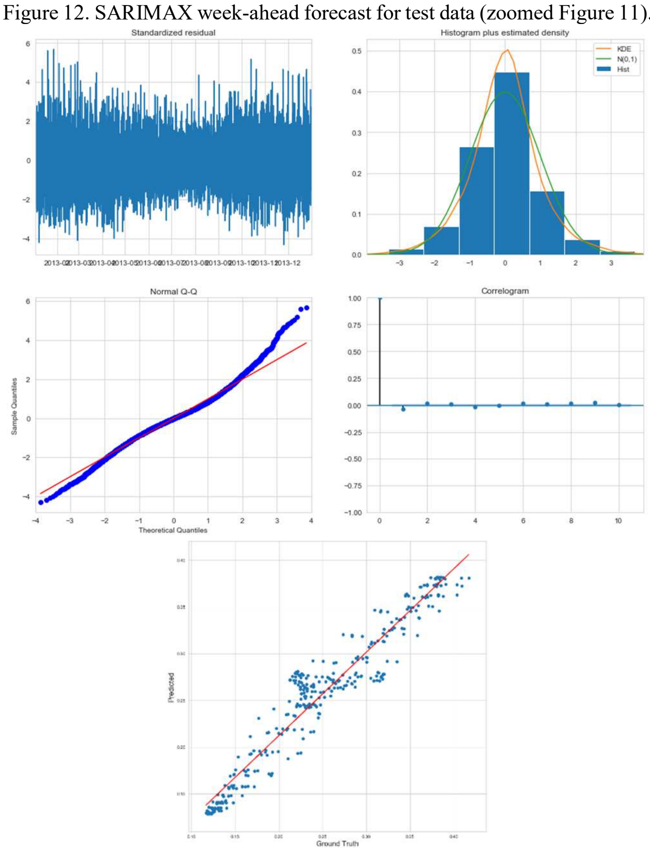


Figure 13. SARIMAX residual diagnostics plots. (a) standardised residual, (b) histogram of residuals with estimated density, (c) normalised Q-Q plot, (d) correlogram of residuals, (e) predicted vs ground truth regression.

Similar to the previous model, SARIMAX is also underestimating the daily peak over a week ahead forecast. Except for some variations between 11am and 4pm, overall the model is closely following daily demand profiles, mostly within 95% confidence intervals. Residual diagnosis reveals that SARIMAX does not show sign of bias and overfitting. The correlogram shows the model is effective at extracting temporal features from the smart meter data. The histogram shows the residual closely following a Gaussian distribution which means the model is optimally learning features from the training dataset. In addition, the ground truth vs forecast regression also fits a straight line which demonstrates the model fit quality is good.

D. Discussions

The best models were selected through hyper-parameter tuning and cross validation resulted in accurate half-hourly forecast over global consumption profile. The tuned SARIMA and SARIMAX models has forecast error of RMSE 0.00053, 0.00049 KWh respectively as shown in Table IV along with other criteria and associated model parameters. While in terms-of week-ahead daily demand-peak forecast the models do tend to underestimate. DSOs can integrate these forecasting techniques for week-ahead prediction at half hour interval and daily demand-peak forecast at a regional level. Because the global mean profile used in this research resembles regional demand profiles used in other research. As evident from the diagnostic plots, both statistical models are able to effectively learn the temporal features from the global profile. Although the dataset includes 5,567 houses, increasing the house number does not necessarily mean model accuracy will increase. To boost model performance, the additional data must have new features that the model can learn.

TABLE IV. SARIMA AND SARIMAX ERRORS ON THE TEST SET

Model	Order Index	Model Order $(\mathbf{p},\mathbf{d},\mathbf{q})$	Seasonal Order $(\mathbf{P},\mathbf{D},\mathbf{Q})$	Abs. AIC	R^2	RMSE
SARIMA	103	$(2, 1, 3)$	$(1, 1, 1, 48)$	139938.46	0.92	0.00054
SARIMAX	130	$(3, 1, 2)$	$(1, 1, 1, 48)$	140127.1	0.93	0.00049

According to recent research, lower temperature can cause more load variations which makes forecasting difficult during those weeks. This is evident in the reported RMSE where the models reported relatively low error during UK summer month July, as compared to the spring and winter months (April and December) which are relatively colder months. Although these variations could not be explained in SARIMA, the SARIMAX can translate some of these variations due to exogenous temperature input. This is reflected through higher R^2 of 0.93 achieved by SARIMAX model as compared to 0.92 for the SARIMA model. However, this comes at additional computational cost of 24 minutes more processing time for a single core CPU. However, the models tuned on global profile did not perform well in forecasting individual meters selected from three different socio-economic groups ACORN-A, B, C. The individual datasets show very high variations between the each other's load profile and also shows high-level of volatility of the demand data. As a result, the models were not sensitive enough to learn those features in individual meters.

IV. CONCLUSIONS

This parametric time-series modelling study investigates the comparative performance of two statistical timeseries models called SARIMA and SARIMAX for short-term (week-ahead) load forecasting at half-hourly time resolution. These models are trained on twelve months of historical London smart-meter power consumption data. Generalisability of the model is tested on three different levels, on global mean, different seasons and individual house demand profiles and tested on respective test sets. The significant findings from this research are summarised as:

- When optimised both SARIMA/SARIMAX models perform well and can provide accurate forecasting for aggregate demand at a regional level.
- Adding more data does not mean more accurate model performance, unless those additional data comes with new features that the model can learn.
- Cross validation over different seasons also reveal that, low temperature makes the prediction difficult.

- SARIMAX model can partially overcome this issue with additional exogenous temperature variables that explained some of the variations. This resulted in lower overall forecast error during cross validation.
- However, the SARIMA model performed relatively accurately as compared to the SARIMAX model in predicting daily demand peak.
- While the SARIMA model is simpler and more accurate at daily peak demand forecasting, the SARIMAX is relatively more accurate overall, but at additional computational cost.
- Although the models are generalisable over different seasonal forecast, it is not generalisable over different demand profile from different smart meters.

Future research on short-term load forecasting can be improved by adapting these time series models to the individual demand profile clusters. The demand profiles can be identified using clustering of individual meters as the ACORN group is not a good classifier for different demand profiles. To make the predictors more accurate over all the houses in a profile cluster, more generalisable models such as the gradient boosting regression tree (GBRT) or LSTM can be used in future research. Similar season/day-based forecasting can also be used to improve predictive accuracy over different seasons.

ACKNOWLEDGMENT

SD was partially supported by the ERDF Deep Digital Cornwall project number: 05R18P02820.

REFERENCES

[1] P. J. Baruah, N. Eyre, M. Qadrdan, M. Chaudry, S. Blainey, J. W. Hall, N. Jenkins, and M. Tran, "Energy system impacts from heat and transport electrification," *Proceedings of the Institution of Civil Engineers-Energy*, vol. 167, no. 3, pp. 139–151, 2014.

[2] S. Ruester, S. Schwenen, C. Batlle, and I. Pérez-Arriaga, "From distribution networks to smart distribution systems: Rethinking the regulation of European electricity DSOs," *Utilities Policy*, vol. 31, pp. 229–237, 2014.

[3] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3125–3148, 2018.

[4] C. Deb, F. Zhang, J. Yang, S. E. Lee, and K. W. Shah, "A review on time series forecasting techniques for building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 74, pp. 902–924, 2017.

[5] H. Cai, S. Shen, Q. Lin, X. Li, and H. Xiao, "Predicting the energy consumption of residential buildings for regional electricity supply-side and demand-side management," *IEEE Access*, vol. 7, pp. 30386–30397, 2019.

[6] B. Yildiz, J. I. Bilbao, J. Dore, and A. B. Sproul, "Recent advances in the analysis of residential electricity consumption and applications of smart meter data," *Applied Energy*, vol. 208, pp. 402–427, 2017.

[7] M. N. Fekri, H. Patel, K. Grolinger, and V. Sharma, "Deep learning for load forecasting with smart meter data: Online Adaptive Recurrent Neural Network," *Applied Energy*, vol. 282, p. 116177, 2021.

[8] A. Mehdipour Pirbazari, M. Farmanbar, A. Chakravorty, and C. Rong, "Short-term load forecasting using smart meter data: A generalization analysis," *Processes*, vol. 8, no. 4, p. 484, 2020.

[9] Z. Mohamed and P. Bodger, "Forecasting electricity consumption in New Zealand using economic and demographic variables," *Energy*, vol. 30, no. 10, pp. 1833–1843, 2005.

[10] F. Wang, Y. Yu, X. Wang, H. Ren, M. Shafie-Khah, and J. P. Catalão, "Residential electricity consumption level impact factor analysis based on wrapper feature selection and multinomial logistic regression," *Energies*, vol. 11, no. 5, p. 1180, 2018.

[11] K. Nagbe, J. Cugliari, and J. Jacques, "Short-term electricity demand forecasting using a functional state space model," *Energies*, vol. 11, no. 5, p. 1120, 2018.

[12] R. Debnath, R. Bardhan, A. Misra, T. Hong, V. Rozite, and M. H. Ramage, "Lockdown impacts on residential electricity demand in India: A data-driven and non-intrusive load monitoring study using Gaussian mixture models," *Energy Policy*, vol. 164, p. 112886, 2022.

[13] J. de Assis Cabral, L. F. L. Legey, and M. V. de Freitas Cabral, "Electricity consumption forecasting in Brazil: A spatial econometrics approach," *Energy*, vol. 126, pp. 124–131, 2017.

[14] W. Zhao, J. Wang, and H. Lu, "Combining forecasts of electricity consumption in China with time-varying weights updated by a high-order Markov chain model," *Omega*, vol. 45, pp. 80–91, 2014.

[15] H. Al-Hamadi and S. Soliman, "Short-term electric load forecasting based on Kalman filtering algorithm with moving window weather and load model," *Electric Power Systems Research*, vol. 68, no. 1, pp. 47–59, 2004.

[16] N. Fumo and M. R. Biswas, "Regression analysis for prediction of residential energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 47, pp. 332–343, 2015.

[17] S. Yu, K. Wang, and Y.-M. Wei, "A hybrid self-adaptive Particle Swarm Optimization-Genetic Algorithm-Radial Basis Function model for annual electricity demand prediction," *Energy Conversion and Management*, vol. 91, pp. 176–185, 2015.

[18] D. H. Vu, K. M. Muttaqi, and A. P. Agalgaonkar, "A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables," *Applied Energy*, vol. 140, pp. 385–394, 2015.

[19] K. Li and T. Zhang, "Forecasting electricity consumption using an improved grey prediction model," *Information*, vol. 9, no. 8, p. 204, 2018.

[20] A. Kavousi-Fard, H. Samet, and F. Marzbani, "A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting," *Expert Systems with Applications*, vol. 41, no. 13, pp. 6047–6056, 2014.

[21] Y. Chen, P. Xu, Y. Chu, W. Li, Y. Wu, L. Ni, Y. Bao, and K. Wang, "Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings," *Applied Energy*, vol. 195, pp. 659–670, 2017.

[22] C. Xia, M. Zhang, and J. Cao, "A hybrid application of soft computing methods with wavelet SVM and neural network to electric power load forecasting," *Journal of Electrical Systems and Information Technology*, vol. 5, no. 3, pp. 681–696, 2018.

[23] S. S. Subbiah and J. Chinnappan, "Deep learning based short term load forecasting with hybrid feature selection," *Electric Power Systems Research*, vol. 210, p. 108065, 2022.

[24] "Smart meters in London." [Online]. Available: https://www.kaggle.com/datasets/jeanmidev/smart-meters-in-london

[25] Q. Raillard-Cazanove and E. Barbour, "Analysis of Smart Meter Electricity Consumption Data for PV Storage in the UK," *Energies*, vol. 15, no. 10, p. 3732, 2022.

[26] S. Singh, A. Yassine, and R. Benlamri, "Consumer segmentation: Improving energy demand management through households socio-analytics," in *2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, 2019, pp. 1038–1045.

[27] W. McKinney, "pandas: a foundational Python library for data analysis and statistics," *Python for High Performance and Scientific Computing*, vol. 14, no. 9, pp. 1–9, 2011.

[28] J. Lee and Y. Cho, "National-scale electricity peak load forecasting: Traditional, machine learning, or hybrid model?," *Energy*, vol. 239, p. 122366, 2022.

[29] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006.

[30] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with python," in *Proceedings of the 9th Python in Science Conference*, 2010, vol. 57, no. 61, pp. 10–25080.

[31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and others, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.