# Forecasting Short-Term Electric Load with a Hybrid of ARIMA Model and LSTM Network

Nevil Pooniwala
*Electronics Department*
*Sardar Patel Institute of Technology*
Mumbai, India
nevilpooniwala@gmail.com

Rajendra Sutar
*Electronics Department*
*Sardar Patel Institute of Technology*
Mumbai, India
rajendra_sutar@spit.ac.in

2021 International Conference on Computer Communication and Informatics (ICCCI) | 978-1-7281-5875-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCI50826.2021.9402461

***Abstract*—Smart Meters in the recent years have led to the generation of large consumer data sets which have enabled more energy forecasting algorithms to be designed. Two such algorithms are discussed in this paper with their minor deficiencies and a hybrid approach is proposed. First algorithm being Autoregressive Integrated Moving Average (ARIMA) model which turns out to be futile in determining nonlinear relationships that are involved. Secondly, Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) can not correctly model seasonal variations in energy consumption. This paper blends Seasonal ARIMA (SARIMAX) with LSTM network by integrating their benefits for a improved electric load forecast. The major contribution is the implementation of the combining algorithm to form the hybrid network. The proposed hybrid implementations provides an almost 13.08% decrease in the mean absolute error when compared with the two algorithms. The slight superior performance of the proposed method in the power load forecasting application is highlighted in the results section.**

***Index Terms*—smart electric meters, load prediction, LSTM, SARIMAX, hybrid approach**

## I. INTRODUCTION

With the rise in smart meters (SM) over the past decade, electric utilities now obtain precise and detailed data of customers. This has led to the generation of various big data sets with previously unattainable information. Such data sets are used in forecasting electric load. Accurate load forecasting promotes the effective functioning of smart grids and power systems, ultimately leading to economic utilization of resources and avoidance of power outages. Moreover, load forecasting is amongst the most important parameters for energy suppliers to determine policies and tariff plans.

For smart grid performance improvement, several algorithms and models have been proposed and implemented till date, many of them being reviewed in [1]. A perfect short-term forecast can sometimes even eliminate unwanted energy generation. An number of techniques have been discussed for the Short-Term Load Forecasting (STLF), including Statistical Methods, Expert Systems, Artificial Intelligence (AI) Techniques, and Hybrid Techniques [2]. Advanced Metering Infrastructure (AMI), a system that includes SMs, data storage units, protocol management systems, and communication channels, plays a vital role in energy grids [3]. Consumer behaviors can be extracted from fine-grained energy data. By the year 2016, the number of SMs introduced in the United Kingdom were 2.9 million [4] and in United States of America were 70 million [5][6]. Most countries are migrating to SMs for most of their power demands.

Fig. 1 shows applications of different energy forecasting time ranges and duration in the smart grid domain.

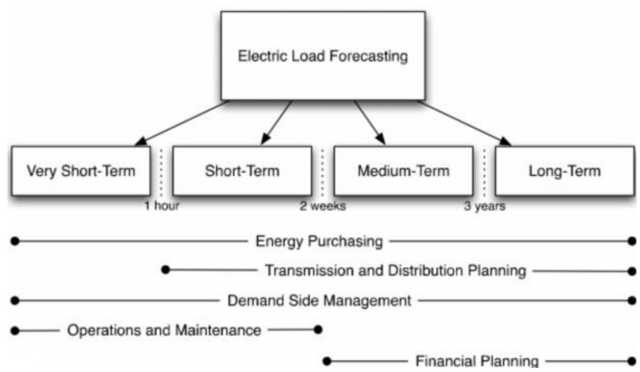


Fig. 1. Applications of load forecasting [7]

To gain economic benefits, energy load forecast should be accurate. Various methods for the same have been studied which include multivariate modeling [8], uni-variate time series prediction [9], artificial neural networks, and various hybrid methods [10], [11], [12], [13]. The design in [14] proved superior in certain aspects when compared with the state-of-the-art model. The authors of that paper used a deep belief network and copula model for short-term load forecasting.

With certain trade-offs, a system performance boost can be achieved by implementing Hybrid models. Although, SARIMAX models are apt for data consisting of linear past and present readings, they tend to be deficient with real-world data. This sometimes becomes a disadvantage especially when real-world data consists of a non-linear relationship. Implementing SARIMAX in such cases may lead to a poor performance. On the other hand, LSTM RNNs are well suited to fit in non-linear relationships. They are even capable to learn through pragmatic data sets. However, when used alone LSTM networks can produce mixed results for linear and non-linear relations. A hybrid combination is proposed in our study. This is helpful as it is sometimes challenging to determine whether the source of energy consumption has a non-linear or linear fundamental process underneath. Furthermore, with a hybrid implementation, the need for selection of a definite model for the system is eliminated. This greatly reduces the influence from factors like sampling rate, thus, minimizing model uncertainty. In essence, a hybrid approach enhances performance by being more robust and raises the possibility of capturing useful underlining patterns.

G. Peter Zhang in [10] took the linear and non-linear relation to describe a hybrid time series forecasting algorithm based upon ARIMA model and artificial neural networks which was further studied by Authors in [11] to implement a similar hybrid model for energy consumption forecasting. Authors in [12] used Zhang's hybrid method to develop a forecasting model by combining the Radial Basis Function (RBF) neural network and Autoregressive model (AR) through the means of binomial smoothing. Authors in [13] compare Zhang's hybrid approach with the one they developed using discrete wavelet transform (DWT) decomposition.

This paper makes use of a back-propagation network (BPN) to form a hybrid model consisting of SARIMAX and LSTM model. In the following section, SARIMAX model, LSTM model, and the hybrid model are reviewed. Section III and section IV describes empirical results and conclusion respectively.

## II. TIME SERIES FORECASTING MODELS

### A. The SARIMAX model

This section first covers the fundamental blocks in the ARIMA model. ARIMA models are one of the most common models used for time series forecasting due to its simplicity. It consists of (p,d,q) where p, d, and q are the order of autoregressive, the differencing level and the order of moving average respectively. The model as in [15] is expressed in (1),

$$X_t = \delta + \emptyset_1 x_{t-1} + \emptyset_2 x_{t-2} + .. + \emptyset_p x_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - .. - \theta_q a_{t-q} \quad (1)$$

where $X_t$ and $a_t$ are the actual and random error values at time period $t$, model parameters are given by $\emptyset_i$ and $\theta_j$, and ($i = 1, 2, 3, \ldots, p$) and ($j = 0, 1, 2, 3, \ldots, q$) with p and q referred to as model orders.

Seasonal ARIMA (SARIMA) is the ARIMA time series which has seasonal variations. A notation of seasonal autoregressive denoted as $P$, seasonal differencing level denoted as $D$ and seasonal moving average denoted as $Q$ forms multiplicative quantities of SARIMA given by $(p, d, q)(P, D, Q)_s$. The subscripted letter denoted by 's' signifies the period length.

To mathematically express the model, it uses a backshift operator (B). Time series equations backward with time by $l$ periods, given in [15], is denoted by $B^l$, such that $B^l y_t = y_{t-l}$.

The stationary transformation, which was formerly being used to present the backshift operator with constant time series statistical parameters like mean and variance, as in [15] is presented in (2):

$$Z_t = \nabla_s^D \nabla^d y_t = (1 - B^s)^D (1 - B)^d y_t \quad (2)$$

where $z$ represents the differencing term of the time series data, $d$ is the degree of non-seasonality used and $D$ is the order of seasonality utilized. Thus, the general equation of the SARIMA model for SARIMA $(p, P, q, Q)$ is as in (3):

$$\emptyset_p(B)\emptyset_p(B^s)z_t = \nabla + \theta_q(B)\theta_Q(B^s)a_t \quad (3)$$

The Box-Jenkins (BJ) theory [16] consists of four different steps - Identification, Estimation, Testing data and Forecasting to predict the time series data, a detailed discussion of BJ theory can be found in [15]. The $X$ in the SARIMAX represents the exogenous regressors. SARIMAX displays the advantage of deciding the correct model to most fit the appropriate time series.

### B. The Long Short-Term Memory model

Long Short-Term Memory (LSTM), a type Recurrent neural network (RNN), has the potential to learn long-term variations and dependencies unlike the traditional artificial neural networks. Although RNNs are suited to makes sense from long-term dependencies, it suffers from some practical difficulties mentioned in [17]. To overcome these disadvantages, LSTM was formulated [18], demonstrating its practical use in tasks which cannot be solved by typical RNNs, e.g. learning to forget [19], and sequence to sequence learning [20]. Authors in [21] have described the use of LSTM network for short-term residential electric load forecasting. LSTM architecture is well suited for in power load forecasting due to the presence of sequential nature in power readings.

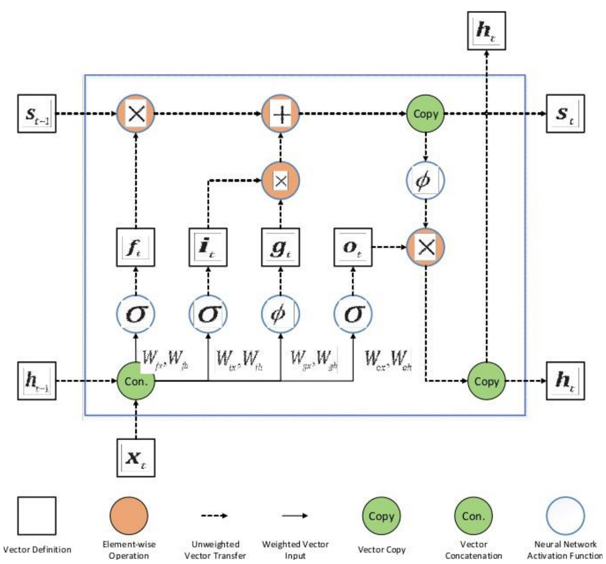


Fig. 2. The LSTM block structure [21]

The LSTM block diagram for a unit time step is shown as in Fig. 2, and the un-rolled architecture joining the information of the next time step is illustrated by Fig. 3. The idea in LSTM is the core memory cell as shown in Fig 2 which maintains most data over time coordinated and handled by the gate units. The block diagram of LSTM shown in Fig. 2 displays data to be maintained in a cell depending on present input and past

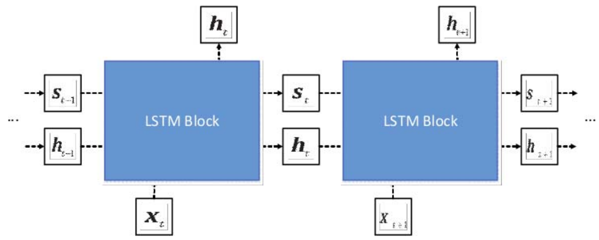


Fig. 3. The sequential architecture of un-rolled LSTM [21]

outputs. Next, it decides on the data to be eliminated and the data to be passed on by braining in gate units, like forget, output and input gate. The input gate controls the overridden of cell state by external data, as in (4).

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \quad (4)$$

where $i_t$ is the vector of the input gate, $b_i$ is the bias vector, $W_i$ and $U_i$ are parameter matrices and $x_t$ is input vector. The output gate whose equation is given in (5) which decides the status in the cell to be affected by other cells.

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \quad (5)$$

State reset is performed by the forget gate of LSTM unit. It is given as in (6),

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (6)$$

Finally, (7) and (8) represent the cell state $c_t$ and output vector $h_t$ obtained from input, forget, and output gate.

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \quad (7)$$

$$h_t = o_t \otimes \sigma_c(c_t) \quad (8)$$

where $\otimes$ represents Hadamard product, and $\sigma_c$ and $\sigma_h$ are hyperbolic tangent functions. LSTM may lead to an over-fitted model which would not be efficient. LSTM networks can learn over a large number of sequences and support multiple timestamps. LSTMs are particularly useful in non-linear function and specifying the category of non-linearity is not needed. LSTM networks can suffer from overfitting due to improper regularisation. LSTM design learns the temporal behavior of electricity consumption which are results of quotidian activities.

### C. The Hybrid Model

The response of SARIMAX and LSTM were merged to retain the advantages and compensate the limitations. This was done by training a network which made predictions determined by the linear and non-linear relationships.

The architecture of back-propagation network (BPN) is shown in Fig. 4 which was trained with six inputs. After modeling SARIMAX and training LSTM network, each of their output were given as inputs to the input layer of BPN for training the neural network based on supervised learning along with the actual energy consumption. In addition to these inputs, the other inputs given to the network were one-hot encoded day of the week, one-hot encoded month, weather and holiday flag. The back-propagation algorithm was used to train the network by updating its weights, more details of the structure can be found in [22]. The hyper-parameters like the number of hidden layer nodes, learning rate and epoch were chosen for optimal performance. The result of the BPN network is $y_t$ which represents the combination of SARIMAX and LSTM, as in (9),

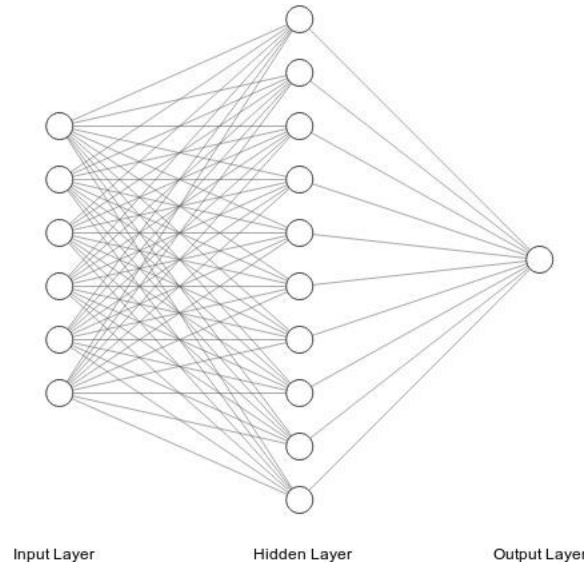


Fig. 4. BPN architecture

$$y_t = f(x_{1t}, x_{2t}, x_{3t}, x_{4t}, x_{5t}, x_{6t}, w_t) + b_t \quad (9)$$

where $f$ indicates the function depending upon the structure of network and their connection weights, $x_{1t}$ is the output of modeled SARIMAX, $x_{2t}$ is the output of LSTM after the network has been trained, $x_{3t}, \ldots, x_{6t}$ are the other inputs like weather cluster, day of week, month and holidays and $w_t$ is the vector weights of parameters. The output $y_t$ which was obtained from the output node was then passed through the activation function to determine the predicted energy consumption. This founded the base of the hybrid approach. After training the hybrid network, $y_t$ generated load predictions with less loss as compared to SARIMAX and LSTM. In the result section, this comparison is discussed in more detail.

An overview of the training flow for the proposed hybrid model can be understood from Fig. 5. In step 1, SARIMAX and LSTM are modeled as explained in section-II A and section-II B respectively. In step 2, a back-propagation neural network architecture is defined. It is then trained, in step 3, with predictions of SARIMAX and LSTM and other inputs like weather cluster, day of week, month and holidays. For example, from the testing data set, as on May 17, 2013, SARIMAX and LSTM energy consumption prediction was 9.281 kW and 10.102 kW respectively. These values along with other inputs were given to the BPN model. The BPN network generated the prediction as 10.076 kW which was closer to the

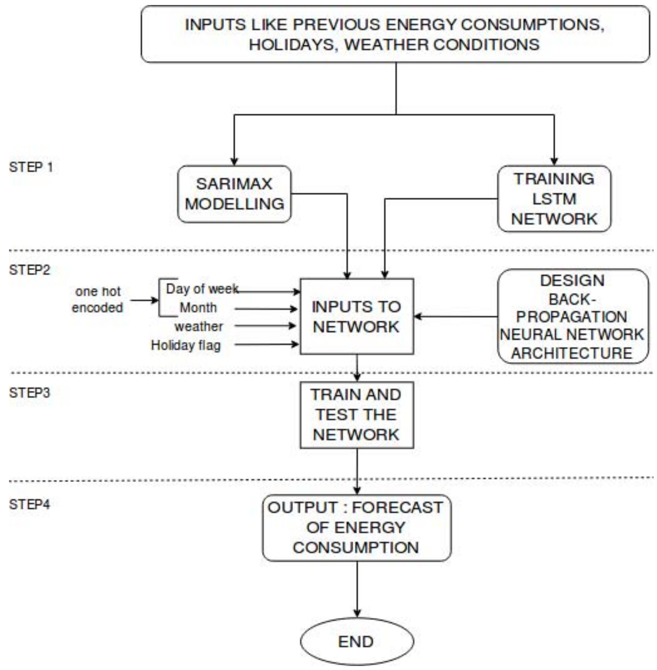


Fig. 5. The proposed hybrid model generation flow

actual value. If May 17, 2013 energy consumption value would have been used as a training data set for BPN, the forecasted values 9.281 and 10.102 obtained from SARIMAX and LSTM respectively, along with other inputs would have been used to train the network as in [21] by updating its weights as per the actual energy consumption to get a prediction closed to the desired output. Algorithm 1 shows the pseudocode for training the proposed hybrid model. After training the model, a simple forward pass would be able to generate load predictions.

**Algorithm 1** Pseudocode of proposed algorithm

**Input:** SARIMAX prediction - $x_{1t}$, LSTM prediction - $x_{2t}$, other inputs - $x_{3t}$ , $x_{4t}$ , $x_{5t}$ and $x_{6t}$ and learning rate $\alpha$
**Output:** Load forecast by hybrid approach - $y_t$

```
Model SARIMAX & train LSTM network
Initialize no. of nodes and hidden layers
BPN network ⇐ ConstructNetworkDesign()
w_t ⇐ InitializeWeights()
for i = 1 to max_iteration do
    y_t ⇐ ForwardPass()
    ComputeError(Actualvalue − y_t)
    BackwardPass()
    Update w_t
end for
```

To sum up, BPN quite accurately identifies the linear and non-linear relationship to accurately predict short term load. It effectively combines SARIMAX and LSTM to achieve comparatively improved results as discussed in the next section.

## III. Results and Discussion

Firstly, we noticed that temperature, humidity, and other weather conditions, sometimes even wind speed, impact the energy consumption. Additionally, we also learnt that holidays too play an import role here. For example, In fig. 6 it can be seen that energy consumption varies inversely with temperature and directly with humidity.

The testing data set for the SARIMAX, the LSTM, and the hybrid model is as shown in Table I. It includes the actual energy consumption values, the predicted values and the residual values of each model. The data sets used for training and validation were also similarly structures. Please note, residual values are the mod of difference between the actual and the predicted readings. It can be calculated as in (10),

$$residual_{mt} = |actual_t - predicted_{mt}| \qquad (10)$$

where $m$ is the model and $t = (1, 2, .., n)$ where n represents the number of samples. The energy consumption values in the table are in kilowatt (kW). Time step represents the number of days in the testing set starting from Friday February 14, 2014 to February 23, 2014.

Table II shows the predicted and actual value of a single household daily electricity consumption for various dates. Case I and case II shows the performance for cold and hot temperature respectively. Thus, the hybrid approach displays better performance even with a change in weather conditions. Case III is an example of load prediction for a holiday which is New Year's day. The linear relationship from the SARIMAX model (case II SARIMAX prediction = 9.281) and the nonlinear relationship from the LSTM model (case II LSTM prediction = 10.102) are combined using a nonlinear function generated by the proposed hybrid model (resulting in case II prediction = 10.076 closer to actual value = 9.825). The hybrid approach dominates as it takes precisely both the linear and nonlinear relationship to generate the load forecast. However, here there is a high risk of overfiting the data.

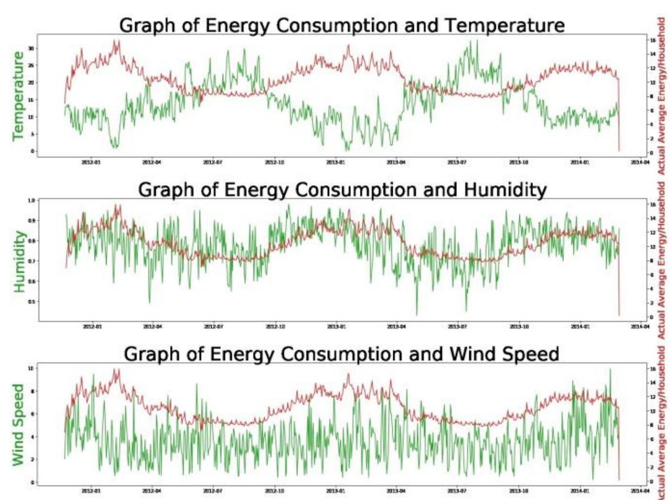


Fig. 6. Relationship of weather conditions with electricity consumption

TABLE I
A SAMPLE OF TEST DATA SET

| Time step | Actual energy | Day of Week* | Month* | Weather cluster | Holiday flag | SARIMAX predicted | SARIMAX residual | LSTM predicted | LSTM residual | Hybrid predicted | Hybrid residual |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 12.024 | 1 | 02 | 1 | 0 | 11.710 | 0.313 | 11.398350 | 0.625 | 12.295 | 0.271 |
| 2 | 11.676 | 2 | 02 | 1 | 0 | 11.542 | 0.133 | 11.599850 | 0.076 | 11.761 | 0.085 |
| 3 | 11.782 | 3 | 02 | 0 | 0 | 11.766 | 0.016 | 11.703016 | 0.079 | 11.859 | 0.076 |
| 4 | 11.164 | 4 | 02 | 0 | 0 | 12.041 | 0.876 | 11.626180 | 0.461 | 11.644 | 0.479 |
| 5 | 10.971 | 5 | 02 | 0 | 0 | 11.658 | 0.687 | 11.489404 | 0.518 | 11.402 | 0.431 |
| 6 | 10.854 | 6 | 02 | 0 | 0 | 12.374 | 1.520 | 11.388611 | 0.534 | 11.741 | 0.887 |
| 7 | 10.754 | 7 | 02 | 1 | 0 | 11.701 | 0.946 | 11.180523 | 0.425 | 11.232 | 0.477 |
| 8 | 10.688 | 1 | 02 | 1 | 0 | 11.498 | 0.810 | 11.128329 | 0.440 | 11.120 | 0.431 |
| 9 | 10.970 | 2 | 02 | 1 | 0 | 11.935 | 0.964 | 11.059685 | 0.088 | 11.131 | 0.160 |
| 10 | 11.673 | 3 | 02 | 1 | 0 | 11.557 | 0.116 | 11.103551 | 0.570 | 11.814 | 0.141 |

* The values given to the model are one-hot encoded

TABLE II
COMPARING ACTUAL VALUE AND PREDICTED RESULTS

| Dates | Power consumption values (in kW) | | | |
|---|---|---|---|---|
| | *Actual Value* | *SARIMAX* | *LSTM* | *Hybrid* |
| 2013-12-19 (Case I: Winter) | 12.413 | 11.754 | 12.918 | 12.147 |
| 2013-05-17 (Case II: Summer) | 9.825 | 9.281 | 10.102 | 10.076 |
| 2014-01-01 (Case III: Holiday) | 10.156 | 9.746 | 10.593 | 10.481 |

TABLE III
PERFORMANCE OF MODEL FOR LOAD FORECASTING

| Performance Metrics | Comparison of models | | |
|---|---|---|---|
| | *SARIMAX* | *LSTM* | *HYBRID* |
| RMSE | 0.772 | 0.448 | 0.418 |
| MAE | 0.507 | 0.390 | 0.339 |
| MAPE | 4.520 | 3.479 | 3.058 |

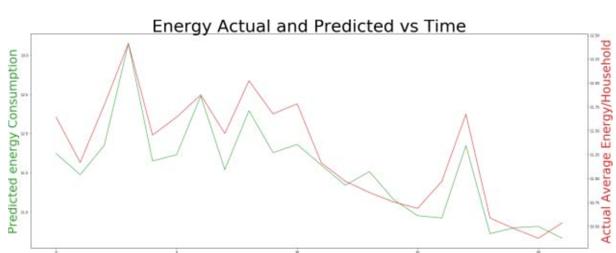


Fig. 7. Actual and predicated energy values with respect to time.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2} \tag{11}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |e_t| \tag{12}$$

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n} \left|\frac{e_t}{y_t}\right| \tag{13}$$

where error $e_t = y_t - \hat{y}_t$, $y_t$ is the actual energy consumption value, $\hat{y}_t$ is the predicted energy consumption, and $t = (1, 2, .., n)$ where n indicates the number of samples.

Various performance metrics are used to judge the accuracy of the forecast. In general, these metrics express the error between prediction and real load observation. Sets of metrics are proposed to measure and evaluate the performance, including the commonly used measures of electric consumption forecasting models like, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) as in [23] given by (11), (12) and (13) respectively. The error values of these performance matrices are shown in the result Table III. These values are derived from the graph plotted in Fig. 7 which shows the actual and predicted load values of the hybrid model with respect to time. It can be observed that the hybrid model has lesser values of RMSE, MAE, and MAPE compared to SARIMAX and LSTM which shows signs of improvement achieved in the load forecasting. Furthermore, it can be seen from Table III that the approximate decrease in error of hybrid model predictions is 6.7%, 13.08% and 12.11% for RMSE, MAE and MAPE respectively.

## IV. CONCLUSION

This paper incorporated the advantages of SARIMAX and LSTM using a simple yet useful neural network technique of back-propagation. As a result, benefits of SARIMAX for linear modeling and LSTM for nonlinear modeling could be achieved in a single system. The main feature is the combined use of traditional time series forecasting and deep learning model for short term meter level demand forecasting. This forecasting often turned out to be a more accurate approach to predict the energy values. The paper described an abstract implementation of the algorithm suggested for better forecast. The fundamental benefit of implementing the algorithm mentioned in this paper is the accuracy that can be achieved as seen in the results.

Thus, combining traditional time series algorithm with a state-of-art deep learning model using a neural network model can provide better results.

Although, in some cases the hybrid model discussed provided better accuracy, it sometimes generates mixed results. The hybrid model may perform worst in very few cases where LSTM performs drastically worse as compared to SARIMAX. Also, the hybrid model even posses the high risk of overfitting the data. Nevertheless, when the performance of models was tested with the statistical performance metrics, the results indicated that a combination performed better. Thus, hybrid model is an effective approach to enhance load prediction accuracy acquired by the two models when utilized independently.

In summary, further research of this algorithm is required by implementing it in different domains like stock price prediction and other time series predictions, to accurately judge the superiority of the hybrid model.

## ACKNOWLEDGMENT

The data set used in this paper was - “SmartMeter Energy Consumption Data in London Households” provided by UK Power Networks [24]. We thank them for providing the data set and making it publicly available. We would also like to acknowledge the financial and the technical support provided by Gulf Automation Technologies Pvt. Ltd. Mumbai, India.

## REFERENCES

[1] Yi Wang, Qixin Chen, Tao Hong and Chongqing Kang, “Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges,” *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3125-3148, May 2019.

[2] A.K. Srivastava, Ajay Shekhar Pandey and Devender Singh, “Short-Term Load Forecasting Methods: A Review,” *International Conference on Emerging Trends in Electrical Electronics & Sustainable Energy Systems (ICETEESES)*, 2016, pp. 130 - 138.

[3] R. R. Mohassel, A. Fung, F. Mohammadi, and K. Raahemifar, “A survey on advanced metering infrastructure,” *International Journal of Electrical Power & Energy Systems*, vol. 63, pp. 473–484, 2014.

[4] Department for Business Energy and Industrial Strategy, *“Smart Meters, Quarterly Report to end December 2016, Great Britain,”* Tech. Rep., 2017.

[5] “Smart meters installed in the united states” [Online]. Available: https://www.eia.gov/tools/faqs/faq.php?id=108&t=3. [Accessed: 28- May- 2020].

[6] “Smart meters deployed in the U.S.A. from 2008 to 2016,” [Online]. Available: https://www.statista.com/statistics/499704/number-of-smart-meters-in-the-united-states/. [Accessed: 28- May- 2020].

[7] Toly Chen and Yu-Cheng Wang, “Long-term load forecasting by a collaborative fuzzy-neural approach,” *International Journal of Electrical Power & Energy Systems*, vol. 43, no. 1, pp. 454-464, Dec. 2012

[8] Yang M, Yu X. “China's rural electricity market – a quantitative analysis,” *Energy*, vol. 29, no. 7, pp. 961-977, June 2004.

[9] Ediger VS, Akar S., “ARIMA forecasting of primary energy demand by fuel in Turkey,” *Energy Policy*, vol. 35, no. 3, pp. 1701-1708, March 2007.

[10] G. Peter Zhang. “Time series forecasting using a hybrid ARIMA and neural network model,” *Neurocomputing*, vol. 50, pp. 159-175, Jan 2003.

[11] Xiping Wang, Ming Meng, “A Hybrid Neural Network and ARIMA Model for Energy Consumption Forecasting,” *Journal of Computers*, vol. 7, no. 5, pp. 1184-1190, May 2012.

[12] Fengxia Zheng, Shouming Zhong, “Time series forecasting using a hybrid RBF neural network and AR model based on binomial smoothing,” *International Scholarly and Scientific Research & Innovation*, vol. 5, no. 3, pp. 419-423, 2011.

[13] Ina Khandelwal, Ratnadip Adhikari, Ghanshyam Verma, “Time Series Forecasting using Hybrid ARIMA and ANN Models based on DWT Decomposition,” *Procedia Computer Science*, vol. 48, pp. 173-179, 2015.

[14] Tinghui Ouyang , Yusen He , Huajin Li , Zhiyu Sun, and Stephen Baek, “ Modeling and Forecasting Short-Term Power Load With Copula Model and Deep Belief Network,” *IEEE Trans. on Emerging Topics in Computational Intelligence*, vol. 3, no. 2, pp. 127-136, April 2019.

[15] Adhistya Erna Permanasari, Indriana Hidayah, Isna Alfi Bustoni, “SARIMA (Seasonal ARIMA) Implementation on Time Series to Forecast The Number of Malaria Incidence,” *International Conference on Information Technology and Electrical Engineering (ICITEE)*, Oct. 2013.

[16] G.E.P. Box, G. Jenkins, *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, CA, 1970.

[17] Bengio, Y., Simard, P., Frasconi, P., “Learning long-term dependencies with gradient descent is difficult,” *IEEE Trans. on Neural Networks*, vol. 5, no. 2, pp. 157–166, March 1994.

[18] Klaus Greff, Rupesh K. Srivastava, Jan Koutn´ık, Bas R. Steunebrink, and J. Schmidhuber, “Lstm: A Search Space Odyssey,” *IEEE Trans. on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222 - 2232, Oct. 2017

[19] F.A. Gers, J. Schmidhuber, F. Cummins, “Learning to forget: continual prediction with LSTM,” *Proc. ICANN’99 Int. Conf. on Articial Neural Networks*, vol. 2, pp. 850-855, 1999.

[20] Ilya Sutskever, Oriol Vinyals, Quoc V. Le, “Sequence to Sequence Learning with Neural Networks,” *Advances in Neural Information Processing Systems*, pp. 3104–3112, 2014.

[21] Weicong Kong, Zhao Yang Dong, Youwei Jia, David J. Hill, Yan Xu, Yuan Zhang, “Short-Term Residential Load Forecasting based on LSTM Recurrent Neural Network,” *IEEE Trans. on Smart Grid*, vol. 10, no. 1, pp. 841 - 851, Jan. 2019.

[22] S. N. Sivanandam, S. N. Deepa, “Supervised Learning Network,” in *Principles of Soft Computing*, Wiley India Edition, New Delhi, 2007, pp. 64-73.

[23] Elham M. Eskandarnia, Sameem Abdul Kareem, Hesham M. Al-Ammal, A Review of Smart Meter Load Forecasting Techniques: Scale and Horizon *International Conference on Information Society and Smart Cities ICS2018*, Cambridge, UK, June 2018.

[24] “Energy Consumption London Data Set” [Online]. Available: https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households [Accessed: 28- May- 2020].