A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for India

Abstract

Selecting a suitable energy demand forecasting method is challenging due to the complex interplay of long-term trends, short-term seasonalities, and uncertainties. This paper compares four time-series models performance to predict total and peak monthly energy demand in India. Indian's Central Energy Authority's (CEA) existing trend-based model is used as a baseline against (i) Seasonal Auto-Regressive Integrated Moving Average (SARIMA), (ii) Long Short Term Memory Recurrent Neural Network (LSTM RNN) and (iii) Facebook (Fb) Prophet models. Using 108 months of training data to predict 24 months of unseen data, the CEA model performs well in predicting monthly total energy demand with low root-mean square error (RMSE 4.23 GWh) and mean absolute percentage error (MAPE, 3.4%), but significantly under predicts monthly peak energy demand (RMSE 13.31 GW, MAPE 7.2%). In contrast, Fb Prophet performs well for monthly total (RMSE 4.23 GWh, MAPE 3.3 %) and peak demand (RMSE 6.51 GW, MAPE 3.01%). SARIMA and LSTM RNN have higher prediction errors than CEA and Fb Prophet. Thus, Fb Prophet is selected to develop future energy forecasts from 2019 to 2024, suggesting that India's annual total and peak energy demand will likely increase at an annual growth rate of 3.9% and 4.5%, respectively.

Keywords: Energy Demand Forecasting, SARIMA, LSTM RNN, Fb Prophet

Highlights

1. Performance assessment of the currently used trend-based model for developing India's monthly total and peak energy demands forecasts.
2. Compare the prediction performance of SARIMA, LSTM RNN and Fb Prophet models to forecast India's total and peak monthly energy demand.
3. Employ the most accurate model to develop future energy demand forecasts for India and its five electrical zones.

List of Abbreviations

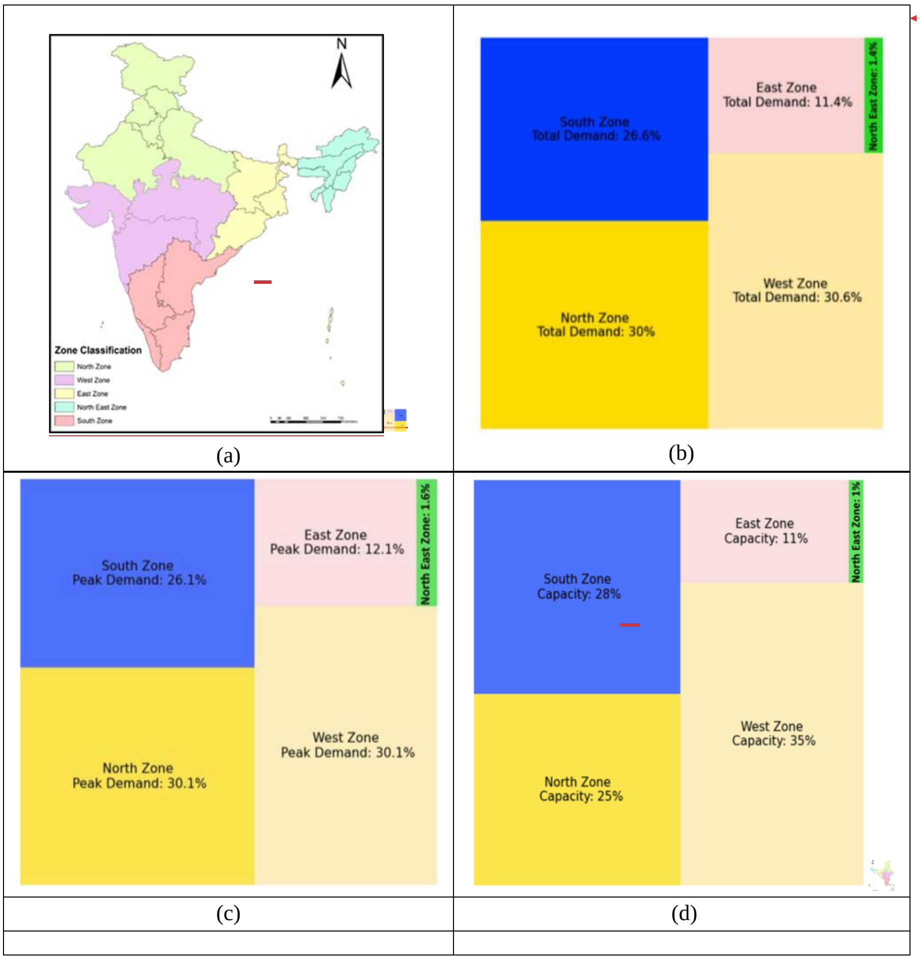
Term	Description
FY	Financial Year starting from 1st April of any given year to 31st March of the next year
CAGR	Cumulative Annual Growth Rate. It can be considered to be an equivalent growth rate which grows the initial investment into its final value assuming that the investment has compounded over this period
RMSE	Root Mean Square Error = $\sqrt{\sum_{t=1}^{n} \frac{\lvert A_t - F_t \rvert^2}{n}}$
MAPE	Mean Absolute Percentage Error MAPE $= \frac{1}{n}\sum_{t=1}^{n} \frac{\lvert A_t - F_t \rvert}{\lvert A_t \rvert}$ where A_t is the actual value, F_t is the forecast value, n is the number of observations
PEC	Per Capita Energy Consumption is the total energy consumption for the year divided by the mid-year national population.
EI	Energy intensity is the quantity of energy required for generating a single unit of gross domestic product at constant prices.

1. Introduction

Energy demand forecasting is defined as the process of employing historical demand data to predict future consumption levels using suitable statistical methods. Energy demand forecasts (EDFs) play a crucial role in planning and policy development for solving critical issues related to expansion of existing infrastructure, scheduling the operation of existing power plants and determining the structure of energy tariffs [1], [2]. Government agencies rely on accurate EDFs to allocate natural resources and power generation facilities across different regions. Resources under-allocation can create energy shortages causing power outages, whereas over-allocation can lead to wastage, both of which are undesirable for a region's socio-economic development. However, consistently producing accurate EDFs is a challenging task as uncertainties linked to various economic activities, climate change, and socio-economic and demographic factors produce considerable variations in daily and monthly energy demands when analysed across multiple years [3].

Depending on the length of the forecasting horizon, EDFs can be classified as long, medium or short term[3], [4]. Long term EDFs (several years ahead) are intended for developmental planning and making important decisions regarding resource allocation and capacity expansion [5]. Medium-term forecasts (several months to a year ahead) help in maintenance planning, fuel supply scheduling and peak demand management. Short-term forecasts (minutes to several hours ahead) are useful for the day-to-day operation of thermal power plants and the scheduling of renewable energy sources such as hydro and gas turbines [6], [7]. For example, accurate next-day forecasts enable producers to calculate the amount of power that needs to be purchased from the national power exchange, which is crucial for preventing supply interruptions, especially during peak demand periods[4], [8].

In India, annual energy demand grew from 775 GWh in 2008-09 to 1,275 GWh in 2018-19, an overall growth of 65% at 5.1% CAGR [9]. Also, Indian government initiatives such as the unification of the national electrical grid and the creation of multiple green energy corridors have doubled the national electrical power generation capacity, from 169 GW in 2008 to 344 GW in 2018 (CAGR 8.6%) [10]. It is noteworthy that a considerable skewness exists in the distribution of India's natural resources and its power generation facilities [4]. Roughly 70% of India's electrical generation is coal based, and almost 70% of the coal is mined from central and eastern regions[11]. Besides, most of India's grid-connected renewable power is produced in the southern states of Karnataka (13.9 GW) and Tamil Nadu (12.7 GW) and the western state of Maharashtra (9.3 GW), mainly due to wind and solar energy [12]. The Indian national grid electrical consists of five zones (Figure 1), with the western zone accounting for the largest share (35%) of total installed electrical capacity followed by south (28%), north (25%), east (11%) and north-east (1%) zones respectively [13].



Formatted Table

Figure 1: (a) The five electrical zones of India and Treemaps showing the split-up of percentage split-up of (b) Electrical Generation Capacity (c) Total Energy Demand (Annual) (d) Peak Energy Demand (Annual) [13]. (The list of states and union territories belonging to the five electrical zones listed in Table A.1 in the appendix)

The west zone accounts for the largest share of the total and peak energy demands annually, followed by north, south, east and north-east zones. It is seen that north, east and north-east zones have capacity deficits of 5%, 1%, and 1%, respectively. These shortfalls are met from the west and south zones that run 5% and 2% surpluses, respectively. Notably, the adverse effects of electrical capacity deficits are more significant for safeguarding power supply over shorter time intervals, such as hours and days. North, West and North East zones have a greater dependence on the national grid for achieving peak demand than total demand as indicated by their ((Peak/Total) >1) ratio.

With the establishment of nationwide power exchange, zonal authorities have become more reliant on accurate EDFs to purchase additional power from the national exchange to balance production shortfalls and schedule the operation and maintenance of generation facilities. Inaccurate short to medium-term EDFs also contribute towards power outages and blackouts due to inefficient scheduling of fuel supply to the power plants [13]. With an anticipated increase of 10 GW in national peak energy demand every year, satisfying instantaneous power requirements shall become more challenging in the future. Nationwide estimates suggest that up to 13 GW (equivalent to 3.8 % annual electrical production) of power outages in India can be eliminated by improving forecasts accuracy, ensuring adequate fuel supply and enhancing coordination between various government and private agencies [10], [12], [13].

1.1 Energy Demand Forecasting in India

In India, the Central Electricity Authority (CEA), a constitutional body under the central ministry of power, is responsible for periodic energy demand forecasting. Each year, CEA publishes a load generation balance report (LGBR) providing zonal and state-wise monthly energy demand data for the previous year and forecasts for the following year. CEA relies on a simple trend-based model to forecast monthly values of total and peak energy demand for the next twelve months. In 2012, Rallapalli and Ghosh reported the CEA model's shortcomings for ignoring the effects of seasonality, non-stationarity and uncertainty present in power data. These issues often lead to a mismatch between the predicted and actual energy demand, increasing financial risk for the generation and distribution companies [4], [14]. For example, if forecasts are lower than actual demand, distribution companies are forced to purchase power from the deregulated power market at a higher price, reducing profit margins [15].

It has been highlighted that a sophisticated time series model such as Seasonal Auto-Regressive Integrated Moving Average (SARIMA) can reduce the estimation errors associated with the CEA model EDFs -[4]. However, with the development and improvement of several new time-series models in the past decade, it is unclear if SARIMA based EDFs can be improved further. To address this issue, we begin by presenting a concise review of existing literature in Section 2, comparing different energy demand forecasting approaches. Examining existing research helped select two widely adopted and a recently developed time series models to predict India's monthly total and peak energy demand. Section 3 describes the data sources and methods. Section 4 contains the results, and Section 5 presents a discussion on the performance and suitability of the three time-series models and CEA's trend-based model for energy demand forecasting applications. Finally, we employ the most accurate model to develop future energy demand forecasts for India and its five electrical zones. Concluding remarks are presented in Section 6.

2. Review of energy demand forecasting models

Researchers have employed a number of energy forecasting techniques depending on the nature, quality and time resolution of available data. We observe that two fundamental approaches are generally used to develop EDFs, i.e. causal and time series models. Causal models are used for developing cause-effect relationships between energy demand and input explanatory variables related to weather, demography and socio-economic factors [16]. In contrast, time series models predict future energy demand values by regressing their previously observed values [4].

Multiple Linear Regression (MLR), Support Vector Regression (SVR) and Artificial Neural Networks (ANN) are the most commonly used causal models for energy demand forecasting. MLR is the easiest to implement and provides estimates for parameter significance and model accuracy [17], [18]. However, as MLR approximates a linear function between the inputs and the output, it does not produce satisfactory results for datasets containing significant amounts of non-linearity and interaction effects [16]. In such situations, adaptive models like SVM and ANN are preferred [15]. ANN contains a collection of nodes that can mimic neurons' working in a biological brain [19].

Several ANN architectures have been tested for energy demand forecasting as they are highly adaptive and efficient in learning complex dependencies between various model inputs and the output [20]–[22]. For instance, Szoplik developed a multi-layer perceptron (MLP) ANN model to predict the natural gas consumption in Szczecin [Poland] considering the weather and temporal (month, day of the week, hour) effects. The authors tested several MLP configurations and identified MLP (22-36-1) containing 22 input neurons, 36 hidden neurons, and one output neuron as the most accurate model configuration [23]. Similarly, Hamzaçebi et al. developed four adaptive ANN models to handle non-linear trends and seasonality effects in Turkey's energy demand data to develop monthly EDFs between 2015-2018 [19]. Intelligent optimization techniques like grid search and nature-inspired heuristics have also been applied for optimal hyperparameter selection to enhance ANN performance [24]–[27]. For instance, Muralitharan et al. compared the performance characteristics of a standalone ANN to a Genetic Algorithm (GA), and Particle Swarm Optimisation (PSO) assisted ANN such that GA and PSO selected optimal hyperparameters. Interestingly, both NN-GA and NN-PSO displayed higher prediction accuracy than conventional ANN. NN-GA faired better for short term load forecasting (hourly, daily), whereas NN-PSO performed better for long term EDS applications (months, years) [28]

The second class of forecasting models, i.e. time series models, are also known as top-down models. They can develop an auto-regressive relationship between the present value and its previous lagged values without relying on other exogenous variables. ARIMA is the most general and widely adopted time-series models to forecast future energy consumption [29]. Proposed by Box and Jenkins, ARIMA deconstructs a given time series to separate its trend, seasonality and error components which are then extrapolated to obtain future values. Owing to the presence of several tuning parameters to account for seasonal and non-seasonal features in the data, ARIMA based models have been used in numerous studies for energy demand forecasting, especially for medium and long-range forecasting horizons [30]–[33]. For example, Ghosh used the multiplicative Holt-Winters multiplicative exponential smoothening (HWMES) and Seasonal ARIMA (SARIMA) model to predict north India's hourly and monthly peak energy demand. SARIMA forecasts were found more accurate than HWMES for both daily and hourly forecasting horizons [8], [14]. Similarly, the multiplicative SARIMA model predicted India's regional monthly peak energy demand with higher accuracy than the CEA's trend-based model [4]. However, ARIMA based models have received some criticism for being 'backwards-looking' and being poor at predicting outputs beyond turning points unless the turning point signifies a long term trend change [34]. For instance, ARIMA forecasts were inferior to SVR and ANN predictions for non-linear short term (0.5 h, 1.0h and 24h) electricity demands in Queensland, Australia [35]. Some researchers have incorporated grey set theory with ARIMA models to resolve uncertainty, missing data, and randomness to achieve higher prediction accuracy [36], [37]. For example, Wang et al. applied single-linear, hybrid-linear and non-linear forecasting techniques based on grey theory to forecast India and China future energy demands. All three models displayed a closer fit, a low error rate and a high fitting precision than the standalone ARIMA model [1].

Another class of neural network architecture known as Long Short Term Memory (LSTM) has gained popularity in energy demand forecasting due to its ability to accurately classify, process and forecast time series data [38]. Unlike standard feed-forward architecture present in ANN, LSTM uses an advanced, recurrent neural network [RNN] architecture where connections between different nodes create a directed graph [DAG[1]] along a temporal sequence [39]. LSTM RNN uses its internal state

[1] DAG: Graph consisting of vertices and edges with edges directed from one vertex to another such that the graph does not contain any closed loops

memory to process variable lengths sequences on inputs to make future predictions. LSTM RNN is modular and highly efficient in handling non-linear complexities and short- and long-term dependencies present in electricity time series data. For example, a recent study found LSTM RNN outperforming ANN, RNN and SVM models to predict city-level daily electricity demand for Chandigarh, India [40]. Similarly, a hybrid model combining a convolutional neural network with LSTM delivered much higher prediction accuracy than MLR, Random Forest, Decision Tree and MLP models for predicting residential hourly electricity consumption [41]. Kim et al. proposed another combined LSTM-CNN EDF framework capable of processing contextual information like temperature, humidity, season etc., apart from historical power time-series data [2]. Similar to ANN, hyperparameter optimisation can also enhance LSTM performance. Recently, Abbasimehr et al. proposed a flexible multi-layer LSTM network using a grid search method to select the best hyperparameters combination. The optimised model predictions were superior to other contemporary ARIMA, ANN and SVM models [42]. In terms of model training and application, LSTM RNN provides more options for parameter tuning than the SARIMA based models. However, LSTM RNN requires a much larger training dataset and a much longer training period than SARIMA based models [43].

A recent development, i.e. Fb Prophet time-series model, promises to solve some of the issues with SARIMA and LSTM RNN models due to its fast computation and high prediction accuracy. Introduced by Taylor and Letham, Fb Prophet is based on an additive framework in which non-linear trends are fitted with yearly, weekly and daily seasonality, plus holiday effects [44]. Fb Prophet exhibits high robustness to missing data, shifts in trends and large outliers, making it an ideal candidate for energy forecasting applications. A recent study compared Fb Prophet's performance with the classical Holt-Winters model to forecast Kuwait power plants' long-term peak energy loads. Fb Prophet model outperformed Holt-Winters predictions over RMSE, MAPE, mean absolute error and R^2 (coefficient of determination) criteria. Further, Fb Prophet displayed better generalizability than Holt-Winters even after introducing random white noise into the dataset[45] . Some studies have also employed hybrid approaches combining multiple prediction models to achieve greater flexibility and higher prediction accuracy than a single model[41], [46]–[51]. For example, Fb Prophet was merged with a Kalman filter to develop an enhanced algorithm for predicting maximum daily power demand in China's power cloud technology company [52]. Similarly, Zhu et al. developed a time series model based on the EMD-Fbprophet-LSTM for short term power consumption prediction. The final EDFs are calculated based on the individual weights of Fbprophet and LSTM model predictions [53]. More in-depth discussions surrounding the strengths and weaknesses of various energy forecasting models can be found by referring to the works of Suganthi and Samuel, Deb et al. and Ghalehkhondabi et al. [16], [29], [54].

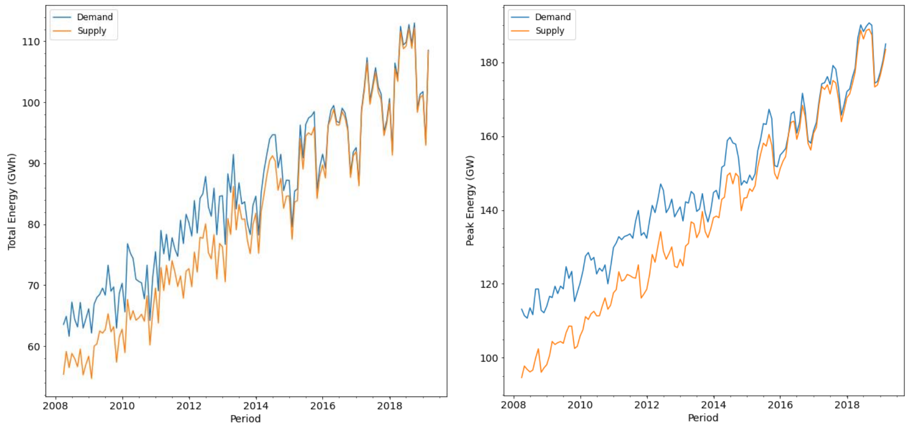
The vast majority of studies aimed at energy demand forecasting applications in India have used HWMES, ARIMA based models such as SARIMA and hybrid Grey-ARIMA models and ANN models [1], [3], [4], [8], [14], [15], [36], [37], [47]. To date, no study has examined the performance of LSTM RNN and Fb Prophet models for regional and national level energy demand forecasting in India. In view of the proven efficacy of SARIMA, LSTM RNN and Fb Prophet models, we perform a comparative assessment to select a suitable method to forecast India's monthly total and peak energy demand. We also compare the prediction performance of these three models with CEA’s trend-based model.

3. Method and Materials

Data for state-level monthly total and peak energy demand for 2008 to 2019 are sourced from eleven previous CEA load generation balance reports to conduct this research. State-level data are added together to produce data at the national level and for the five electrical zones. Referring to Table 1 and Figure 2, India's annual energy demand grew 64% during FY: 2008-09 to FY: 2018-19, whereas energy supply rose by 84%, reducing the demand-supply gap from 11.04% to 0.6%. Likewise, annual peak energy demand grew 61% from FY: 2008-09 to FY: 2018-19, whereas peak energy supply rose by 86%, dwindling the peak demand-supply gap from 13.8% to 0.8%.

Table 1: Year-wise total and peak energy demand for India (FY: 2008-09 to FY: 2018-19)

Financial Year	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016	2016-2017	2017-2018	2018-2019
Total Demand (GWh)	774	831	862	937	998	1003	1069	1115	1143	1214	1275
Total Supply (GWh)	689	747	788	858	911	960	1031	1093	1136	1205	1268
Peak Demand (GW)	1369	1448	1515	1618	1695	1704	1832	1895	1968	2084	2206
Peak Supply (GW)	1180	1270	1374	1457	1534	1626	1749	1844	1949	2058	2188



(a) (b)

Figure 2: Month-wise (a) Total Energy Demand (b) Peak Energy Demand for India during 2008-2019

Statistical properties of total and peak energy demand data from 2008-2019 are summarized in Table A.2 in the appendix. The following section contains a technical description of the three time-series models employed in this study.

3.1 Seasonal Auto-Regressive Integrated Moving Average model

SARIMA or seasonal ARIMA is an extension of the ARIMA model used for modelling univariate time series with a seasonal component. ARIMA model deconstructs any given time series into Auto-Regressive (AR), Integrated (I) and Moving Average (MA) terms [55]. The initial step in building an ARIMA model is to make the data stationary by ensuring its mean, variance and autocorrelation (linear relationship between lagged values) structure does not change over time. It is achieved by performing differencing operation in which each entry of the time series is subtracted from its predecessor. Occasionally, depending on the complexity in data, more than one differencing may be necessary. The order of Integrated term, i.e. 'd', represents the minimum amount of differencing needed to make a time series stationary. The order of the AR term, i.e. 'p', represents the minimum number of predecessors required as inputs to predict the present value. Similarly, 'q' means the order of MA term, denoting the minimum number of lagged forecast errors needed to forecast the current value. Using ARIMA model, value at time t of a time series, y_t can be represented by Equation 1.

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots \theta_q \varepsilon_{t-q} \quad (1)$$

$$\left(1 - \phi_1 B - \cdots - \phi_p B^p\right)(y'_t - \mu) = (1 + \theta_1 B + \cdots + \theta_q B^p)\varepsilon_t \quad (2)$$

where c is the constant term, ϕ and θ are the regression weights for the lagged observations and errors terms, respectively. Using the Backshift operator, Equation 1 is written more elegantly as Equation 2 such that $B^p y_t = y_{t-p}$ and $y'_t = (1 - B)^d y_t$ and μ is the mean of y'_t. In most monthly energy time-series datasets, seasonal variation is a major source of non-stationarity. SARIMA model is recommended for such situations as it can handle trends and seasonal fluctuations present in the time series. SARIMA model is represented as ARIMA (p, d, q) (P, D, Q)s where P denotes the seasonal autoregressive order, D denotes seasonal difference order, Q denotes the seasonal moving average order and s denotes the number of observations in a single year. Seasonal terms (P, D, Q, s) of SARIMA are analogous to the non-seasonal terms in ARIMA but involves backshifts of the seasonal period. For example, ARIMA $(1,1,1)(1,1,1)_4$ model for quarterly seasonality can be represented by Equation 3

$$(1 - \phi_1 B)(1 - \Phi_1 B^4)(1 - B)(1 - B^4) y_t = (1 + \theta_1 B)(1 + \Theta B^4)\varepsilon_t \quad (3)$$

where Φ and Θ are the regression weights for the seasonally lagged observations and errors terms, respectively [56]. Setting up a SARIMA model over any monthly energy time-series datasets involves selecting optimum values for p, d, q, P, D, Q and s parameters. Akaike Information Criterion [AIC] is widely used as a suitable performance metric to compare multiple SARIMA models' prediction performance. AIC compares models by considering accuracy and model complexity as the two main attributes and favours parsimonious models with better fit using fewer features.

3.2 Long Short Term Memory Recurrent Neural Network model

Unlike standard feed-forward neural networks, LSTM is a recurrent neural network architecture containing feedback connections to convey information from several past instances into the present one [55]. LSTM relies on the backpropagation algorithm to find derivatives in the network by moving layer by layer from the final to the first. Quite often RNN architectures get stuck in the vanishing gradient problem where the network cannot propagate useful gradient information from the initial layers of the model to the current one. LSTM units are made up of select units called 'cells' or 'memory blocks' to overcome the vanishing gradient problem [57]. As shown in Figure 3, a single

LSTM unit comprises a cell, an input gate, an output gate and a forget gate. The cell is responsible for maintaining the dependencies between the various elements of an input sequence, and gates are controlled using the sigmoid activation functions. The input gate controls the extent to which a new value flows into a cell, whereas the output gate controls the degree to which a value remains in the cell.

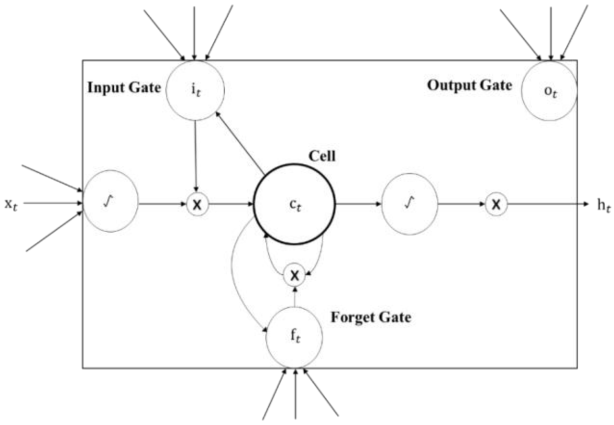


Figure 3: Diagrammatic representation of an LSTM RNN unit

During the model training stage, weights are assigned to the connections moving in and out of the LSTM gates, a few of them being recurrent. These weights are updated continuously over the course of a large number of training cycles (epochs) to improve prediction accuracy. The presence of a forget gate and the additive property of cell state gradients allow LSTM to update connection weights in such a manner reducing chances for a vanishing gradient problem significantly [57]. LSTM RNN prediction performance depends on the value of two key tuning parameters, i.e. the total number of cells or neurons and the total number of training cycles or epochs. Hence, several LSTM models with different neurons and training cycles must be compared based on their prediction accuracy using RMSE or MAPE metrics before selecting the final LSTM RNN architecture.

3.3 Facebook Prophet model

Fb Prophet is an additive time series model based on the Bayesian curve fitting technique. It allows the flexibility to model complicated time series features by fitting trends and multiple seasonalities to incorporate yearly, monthly, weekly and daily patterns along with holiday effects [44]. The three main components of a Fb Prophet model accounting for trend, seasonality and holidays are shown in Equation 4.

$$y(t) = g(t) + s(t) + h(t) + e(t) \tag{4}$$

where $y(t)$ is the output value, $g(t)$ is a trend function which models non-periodic changes in the time series, $s(t)$ represents periodic changes (e.g. weekly, monthly, yearly seasonality), and $h(t)$ represents the effects of holidays. Error term $e(t)$ describes any irregular or random features in the dataset that cannot be explained by this model. The trend part can be represented using a linear piecewise or a or a saturating growth model.

Since both total and peak energy demand datasets do not exhibit saturating growth, a piecewise linear growth model is employed shown in Equation 5.

$$g(t) = (k + a(t)^T \delta)\mathrm{t} + (\mathrm{m} + a(t)^T \gamma) \tag{5}$$

where k represents the growth rate, t represents time steps, δ is the adjustment rate, m is an offset parameter, and γ means trend change points, and is set equal to -$s_j \delta_j$, with a(t) defined in Equation 6 as:

$$\mathrm{a}_j(\mathrm{t}) = \begin{cases} 1, & t \geq s_j\,, \\ 0, & otherwise \end{cases} \tag{6}$$

The trend change points allow the growth model to change trends after a certain number of time steps, increasing modelling flexibility. These change points can be either defined explicitly by the modeller or determined by the model automatically to describe trend altering events in the time series. The second term in $g(t)$, i.e. $s(t)$ represents seasonal component accounting for cyclic changes introduced during a weekly, monthly or annual cycle. Fb Prophet model uses Fourier series to provide a flexible model of periodic effects in Equation 7.

$$s(t) = \sum_{n=1}^{N} (a_n \cos\left(\frac{2\pi \mathrm{nt}}{P}\right) + b_n \sin\left(\frac{2\pi \mathrm{nt}}{P}\right)) \tag{7}$$

Where P = 354.25 for yearly seasonality or P = 7 for weekly seasonality. Fitting seasonality term $s(t)$ requires estimating the 2N parameters $\beta = [a_1\,, b_1 \;\ldots\ldots a_N\, b_N\,]^T$.

Finally, estimation of holiday term $h(t)$ depends on an indicator function for representing whether time t is during a holiday event i, and each holiday is assigned a parameter k_i, responsible for corresponding changes to the forecast. Fb Prophet model also exhibits robustness to problems introduced by outliers and missing data.

As shown in Figure 4, a four-stage research methodology is adopted. In the first stage, national-level monthly total and peak energy demand data are partitioned in an (80-20) ratio to produce a training and testing set. The training set contains initial nine years of data (2008 to 2017), whereas the testing set contains data for the final two years (2017 to 2019). During the model development stage, SARIMA, LSTM RNN and Fb Prophet models are developed inside the Python environment using numpy, pandas, statsmodels, and pmdarima fbProphet and torch libraries [58][59][60][61]. We use the 'pmdarima' library to select the most optimal SARIMA model parameters, i.e. p, d, q, P, D, Q and s for total and peak energy demands as per the 'smallest' AIC scores [62].

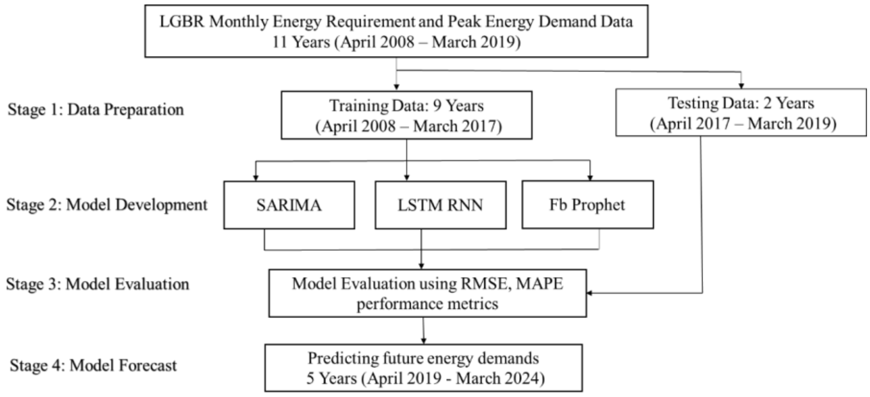


Figure 4: Research methodology adopted in this study

Similarly, to select suitable LSTM RNN models for predicting total and peak energy demand, twelve LSTM models (refer Table 2) are developed by considering (300,500,800,1000) and (200,500,800) values for the number of cells and training cycles, respectively. Out of these twelve models, models with the smallest RMSE and MAPE values are chosen to predict the two target variables. All LSTM RNN models are developed inside python using the pytorch library [62]. Fb Prophet models for both total and peak energy demands are executed in python using the open-source fbprophet library [58].

Table 2: Number of neurons and epochs considered in the twelve LSTM RNN models

Model	M-1	M-2	M-3	M-4	M-5	M-6	M-7	M-8	M-9	M-10	M-11	M-12
Number of Neurons	300	500	800	1000	300	500	800	1000	300	500	800	1000
Number of Epochs	200	200	200	200	500	500	500	500	800	800	800	800

In the model evaluation stage, prediction performance of SARIMA, LSTM, Fb Prophet and CEA trend-based models are analyzed by comparing their test set predictions against observed test set data. The prediction accuracy of these four models is compared using two statistical metrics, namely RMSE and MAPE. In the model forecast stage, the model having the smallest RMSE and MAPE values is selected to forecast the two target variables, i.e. energy requirement and peak demand from April 2019 until May 2024, for India and its five electrical zones.

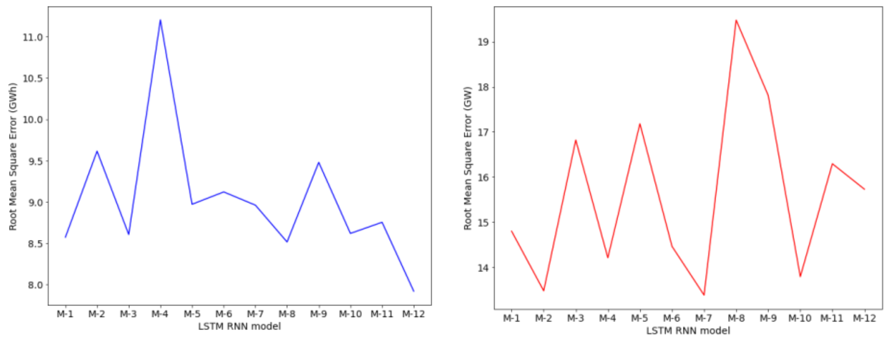
4. Results

Table 3 presents the model coefficients, standard errors, p values and confidence intervals after implementing the SARIMA model. As per AIC criteria, ARIMA (1, 1, 0) $(1,0,1)^{12}$ is the most accurate model for predicting monthly total energy demand as it has the smallest AIC score of 2569. Similarly, ARIMA (2, 1, 2) $(1,0,1)^{12}$ is the most accurate model for predicting monthly peak energy demand as it has the smallest AIC score of 2538. The coefficients of all model parameters, including AR, MA, SAR and SMA, are statistically significant as indicated by their p values (< 0.05).

Table 3: Estimated SARIMA model parameters for monthly total and peak energy demand

Total Energy Demand (MWh)					
Variable	**Coefficient**	**Standard Error**	**z**	**P>|z|**	**95% Confidence Interval**
Intercept	-5.36	48.94	-0.11	0.91	[-101.27, 90.54]
AR(1)	-0.17	0.03	-5.23	0.00	[-0.23, -0.10]
SAR(12)	0.94	0.06	15.42	0.00	[0.82, 1.05]
SMA(12)	-0.82	0.10	-8.27	0.00	[-1.02, -0.63]
Akaike Information Criterion (AIC)	2568.33				
Peak Energy Demand (MW)					
Intercept	604.43	619.27	0.97	0.33	[-609.33, 1818.18]
AR(1)	-1.54	0.03	-95.02	0.00	[-1.57, -1.51]
AR(2)	-0.99	0.01	-73.95	0.00	[-1.03, -0.97]
MA(1)	1.54	0.07	20.77	0.00	[1.39, 1.69]
MA(2)	0.99	0.09	11.15	0.00	[0.82, 1.17]
SAR(12)	0.74	0.16	4.59	0.00	[0.42, 1.05]
SMA(12)	-0.67	0.18	-3.68	0.00	[-1.02, -0.31]
Akaike Information Criterion (AIC)	2538.30	AR: Auto-Regressive SAR: Seasonal Auto-Regressive MA: Moving Average SMA: Seasonal Moving Average			

RMSE values for the twelve LSTM RNN models used for predicting monthly total and peak energy demand are shown in Figure 5. For monthly total energy demand, model M-12 with 1000 neurons and 800 epochs delivers the lowest RMSE of 7.92 GWh. Similarly, model M-7 with 800 neurons and 500 epochs gives the lowest RMSE of 13.38 GW for predicting monthly peak energy demand. It is noteworthy that the best performing LSTM RNN models have a considerably lower prediction accuracy than Fb Prophet models producing 85% and 106% higher RMSE scores for monthly total and peak energy demands, respectively. Therefore, search space for identifying the optimum number of neurons and epochs is not drawn-out further, and M-12 and M-7 LSTM RNN models are selected for predicting monthly total and peak energy demand, respectively.



(a) (b)

Figure 5: Root mean square errors of the twelve LSTM RNN models for predicting (a) monthly total energy demand (b) monthly peak energy demand

4.1 Model Comparison

Test set predictions for monthly total and peak energy demand obtained using SARIMA, LSTM RNN and Fb Prophet models are compared with CEA model forecast and actual observed data between April 2017 - March 2019. Visual comparisons are performed by referring to Figure 6 and Figure 7. RMSE and MAPE values are presented in Table 4 and Table 5.

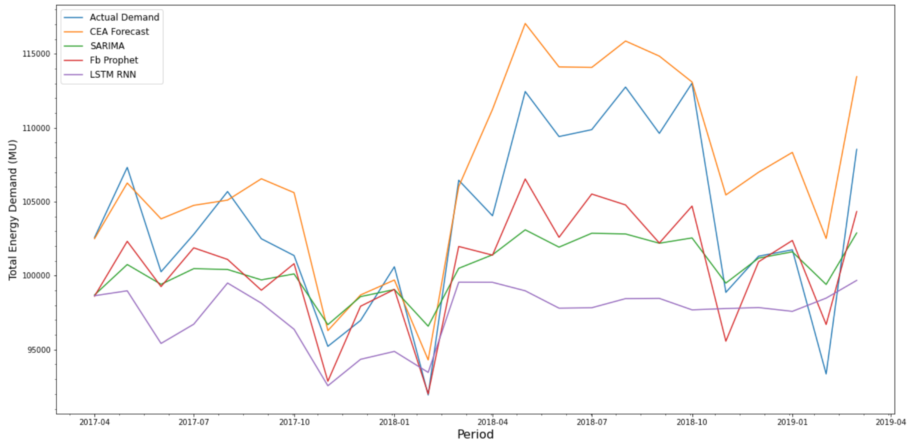


Figure 6: Visual comparison between SARIMA, LSTM RNN, Fb Prophet and CEA models for predicting monthly total energy demand

Table 4: Statistical comparison between SARIMA, LSTM RNN, Fb Prophet and CEA models for predicting monthly total energy demand

Performance Metrics	CEA forecast	SARIMA	Fb Prophet	LSTM RNN
RMSE (GWh)	4.28	5.39	4.28	7.92
MAPE	3.41%	4.12%	3.29%	6.39%

Both CEA and Fb Prophet model predictions deliver the smallest RMSE ~ 4.28 MWh than SARIMA (5.39 GWh) and LSTM RNN (7.92 GWh) models. Further, Fb Prophet model has the smallest MAPE value of 3.29%, followed by CEA (3.41%), SARIMA (4.12%) and LSTM RNN model (6.39%). For the entire duration between March 2017 and April 2019, CEA predictions are higher than the actual demand. Fb Prophet predictions deliver best performance in recreating the surges and valleys seen in actual demand. In comparison, SARIMA and LSTM RNN predictions are less accurate and less efficient in recreating the actual demand's temporal features.

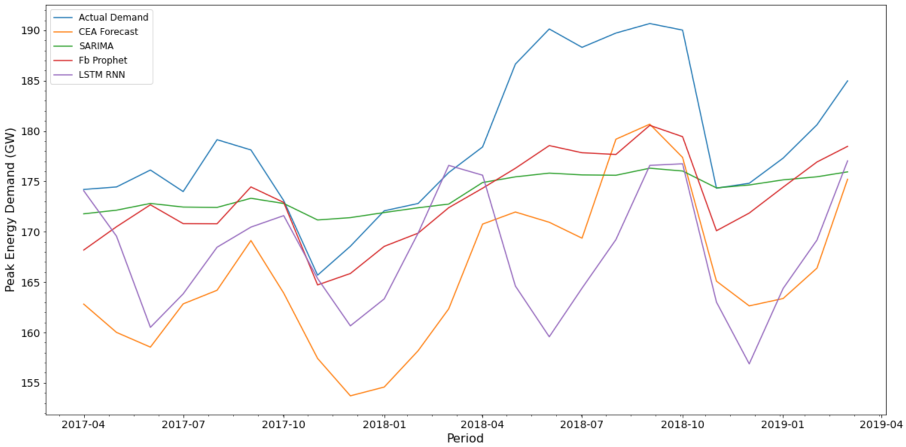


Figure 7: Visual comparison between SARIMA, LSTM RNN, Fb Prophet and CEA models for predicting monthly peak energy demand

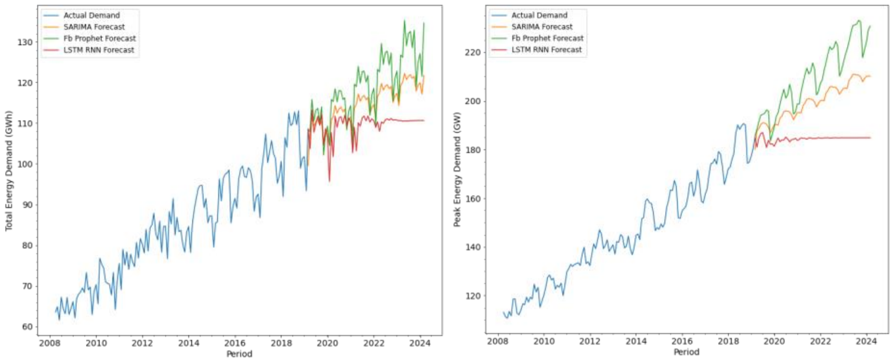
Table 5: Statistical comparison between SARIMA, LSTM RNN, Fb Prophet and CEA models for predicting monthly peak energy demand

Performance Metrics	CEA Forecast	SARIMA	Fb Prophet	LSTM RNN
RMSE (GW)	13.31	7.52	6.51	13.38
MAPE	7.21%	3.04%	3.01%	5.95%

Fb Prophet prediction has the smallest RMSE of 6.51 GW, followed by SARIMA (7.52 GW), CEA forecast (13.31 GW) and LSTM RNN (13.38 GW). Fb Prophet delivers the lowest MAPE of 3.01%, followed by SARIMA (3.04%), LSTM RNN (5.95%) and CEA model (7.21%). All four models under-predict the actual peak energy demand, with the CEA forecast being the farthest from the actual data. Fb Prophet model produces smallest prediction error, and yields the closest match to the actual peak demand data, reconstructing its temporal features. Similar to Rallapalli and Ghosh [4], we find that the multiplicative SARIMA model delivers superior performance than the CEA model for predicting monthly peak energy demand.

4.2 Future forecasts for Total and Peak Energy Demand

Fb Prophet model delivers the highest accuracy for predicting India's monthly total and peak energy demand. Therefore, we employ it to develop future forecasts from April 2019 until May 2024. Estimates from the other two models are also presented alongside in Figure 8 for comparison. Further, the Fb Prophet model is also employed to develop monthly total and peak energy demand forecasts for India's five electrical zones (refer Table A.3 in appendix).



(a) (b)

Figure 8: Monthly forecast (April 2019 - May 2024) for (a) total energy demand and (b) peak energy demand for India using SARIMA, Fb Prophet and LSTM RNN models

Based on the Fb Prophet model forecast (Referring to Table A.3), the national annual total and peak energy demand are expected to grow at CAGR of 3.11% and 3.50%, respectively, during 2019-2024. Further, the total energy demand is expected to grow fastest in the south zone with a CAGR of 3.49%, followed by the east zone at 3.42%, northeast zone at 3.15%, and west zone 3.04% and north zone at 2.63%. In comparison, peak energy demand is expected to grow fastest in the south zone with a CAGR of 4.25%, followed by east zone at 4.02%, west zone at 3.77%, north zone at 3.02% and northeast zone at 2.34%.

5. Discussion

CEA's conventional trend-based model delivers acceptable performance for predicting India's monthly total energy demand with a slight overestimation risk. In comparison, Fb Prophet performs better in capturing the trends and seasonalities present in the actual time series. CEA forecasts can be adopted at a slight risk of overestimation and missing some temporal features present in the time series. However, we recommend adopting Fb prophet total demand predictions when it is necessary to capture the time series' temporal characteristics accurately.

All four models are found to underestimate India's monthly peak energy demand. This issue arises partially due to a sudden increase in peak demand's seasonal component in 2015 (refer Figure 2). The CEA model grossly predicts the actual peak demand having the lowest prediction accuracy and should not be used to develop future peak forecasts. SARIMA predictions have smaller errors but do not capture the seasonal changes efficiently. LSTM RNN peak predictions are inferior to SARIMA and only marginally better than the CEA estimates. Moreover, LSTM RNN models are time-intensive for parameter tuning and model training. Fb Prophet delivers the lowest prediction errors, yielding the closest match with the actual data; therefore, it should be adopted to forecast peak monthly energy demands.

Regarding future forecasts in Figure 8, the Fb Prophet model delivers the best performance in extrapolating trends and seasonalities seen in historical data. In comparison, seasonal effects in SARIMA forecasts are less prominent and gradually becomes weaker towards the later years. LSTM RNN predictions underestimate both target variables and become highly unstable beyond twenty-four time steps. Inspecting future forecasts, we find that the growth rate for peak demand is higher than that for the total demand in north, south, east and west zones. These findings indicate a growing necessity to adopt accurate forecasting models to help stakeholders carry out economic power transactions with the national grid. This study shows that Fb prophet model can train over large datasets to produce highly accurate forecasts within a short period. Due to its additive submodular structure, the Fb Prophet model can efficiently handle anomalies, periodicity and sudden jumps present in time-series data. Plus, it also accounts for demand variations due to seasonal and holiday effects. Our findings agree with Guo et al. [52], who also recommend applying the Fb Prophet model for energy demand forecasting.

6. Conclusion and Policy Implications

Accurate energy demand forecasts are crucial for any country to ensure the efficient operation of its power sector. This paper presented a comparative assessment of SARIMA, LSTM RNN and Fb Prophet models with the currently employed Central Energy Agency (CEA) trend-based model to predict India's total and peak monthly energy demand. All three models were trained using 108 months of historical data from 2008 to 2017 to predict 24 months' data from 2017 to 2019. CEA's trend-based model performs well in predicting monthly total energy demand with low root-mean square error (RMSE 4.28 GWh) and mean absolute percentage error (MAPE, 3.4%), but significantly under predicts monthly peak energy demand (RMSE 13.31 GW, MAPE 7.2%). In contrast, Fb Prophet performs well for monthly total (RMSE 4.28 GWh, MAPE 3.3 %) and peak demand (RMSE 6.51 GW, MAPE 3.01%).

SARIMA predictions are found inferior to Fb Prophet with (RMSE 5.39 GWh, MAPE 4.12%) and (RMSE 7.52 GW, MAPE 3.04%) for monthly total and peak demands, respectively. LSTM RNN forecasts are not found reliable due to large prediction errors, and forecasts tend to become highly unstable after 24 time steps. Fb Prophet forecasts also performed best in recreating the temporal features present in observed data. Thus, the Fb Prophet model should be preferred over others in demand forecasting. We employ it to develop monthly total and peak demand forecasts for India and its five electrical zones from April 2019 until May 2024. Between 2019-2024, India's annual total and peak demand shall increase at a yearly growth rate of 3.9% and 4.5%, respectively.

Both total and peak demands are forecasted to grow fastest in India's south zone, followed by east, west and north zones. Future forecasts also indicate that annual peak demand growth shall be faster than total demand growth in four (north, south, east and west) out of five electrical zones. Findings from this paper shall assist practitioners in the energy sector for medium to long-term planning. Future follow-up studies can investigate the prediction performance of advanced hybrid and ensemble machine learning methods to achieve higher prediction accuracy.

7. Acknowledgements

The authors would like to acknowledge the funding received from the Department of Science and Technology, Government of India DST/TMD/UKBEE/2017/17 project: Zero Peak Energy Demand for India (ZED-I) and Engineering and Physics Research Council EPSRC project: EP/R008612/1.

8. References

[1] Q. Wang, S. Li, and R. Li, "Forecasting energy demand in China and India: Using single-linear, hybrid-linear, and non-linear time series forecast techniques," *Energy*, vol. 161, pp. 821–831, 2018, doi: 10.1016/j.energy.2018.07.168.

[2] M. Kim, W. Choi, Y. Jeon, and L. Liu, “A hybrid neural network model for power demand forecasting,” *Energies*, vol. 12, no. 5, 2019, doi: 10.3390/en12050931.

[3] S. Ghosh and A. Das, “Short-run electricity demand forecasts in Maharashtra,” *Appl. Econ.*, vol. 34, no. 8, pp. 1055–1059, 2002, doi: 10.1080/00036840110064656.

[4] S. R. Rallapalli and S. Ghosh, “Forecasting monthly peak demand of electricity in India-A critique,” *Energy Policy*, vol. 45, pp. 516–520, 2012, doi: 10.1016/j.enpol.2012.02.064.

[5] P. H. Kuo and C. J. Huang, “A high precision artificial neural networks model for short-Term energy load forecasting,” *Energies*, vol. 11, no. 1, pp. 1–13, 2018, doi: 10.3390/en11010213.

[6] S. N. Fallah, R. C. Deo, M. Shojafar, M. Conti, and S. Shamshirband, “Computational intelligence approaches for energy load forecasting in smart energy management grids: State of the art, future challenges, and research directions,” *Energies*, vol. 11, no. 3, 2018, doi: 10.3390/en11030596.

[7] T. Ahmad and H. Chen, “Utility companies strategy for short-term energy demand forecasting using machine learning based models,” *Sustain. Cities Soc.*, vol. 39, no. August 2017, pp. 401–417, 2018, doi: 10.1016/j.scs.2018.03.002.

[8] S. Ghosh, “Univariate forecasting of day-ahead hourly electricity demand in the northern grid of India,” *Int. J. Indian Cult. Bus. Manag.*, vol. 2, no. 6, p. 625, 2009, doi: 10.1504/ijicbm.2009.027180.

[9] Government of India, “Load Generation Load Generation Balance Balance Report 2019-20,” *Cent. Electr. Auth.*, vol. 73, pp. 1–4, 2019, doi: CEA.

[10] M. T. Parray and R. Tongia, “Understanding India’s Power Capacity: Surplus or not, and for how long?,” *Brookings India Center*. https://www.brookings.edu/research/understanding-indias-power-capacity-surplus-or-not-and-for-how-long/ (accessed Mar. 28, 2020).

[11] “Top five states in India with the largest coal reserves,” *Energy, NS*. https://www.nsenergybusiness.com/features/states-india-largest-coal-reserves/ (accessed Oct. 27, 2020).

[12] J. Charles Rajesh Kumar and M. A. Majid, “Renewable energy for sustainable development in India: Current status, future prospects, challenges, employment, and investment opportunities,” *Energy. Sustain. Soc.*, vol. 10, no. 1, pp. 1–36, 2020, doi: 10.1186/s13705-019-0232-1.

[13] N. S. Office and P. Implementation, “Energy Statistics 2020 Energy,” 2020.

[14] S. Ghosh, “Univariate time-series forecasting of monthly peak demand of electricity in northern India,” *Int. J. Indian Cult. Bus. Manag.*, vol. 1, no. 4, p. 466, 2008, doi: 10.1504/ijicbm.2008.018626.

[15] S. S., K. S., and T. C., “India’S Electricity Demand Forecast Using Regression Analysis and Artificial Neural Networks Based on Principal Components,” *ICTACT J. Soft Comput.*, vol. 02, no. 04, pp. 365–370, 2012, doi: 10.21917/ijsc.2012.0057.

[16] I. Ghalehkhondabi, E. Ardjmand, G. R. Weckman, and W. A. Young, “An overview of energy demand forecasting methods published in 2005–2015,” *Energy Syst.*, vol. 8, no. 2, pp. 411–447, 2017, doi: 10.1007/s12667-016-0203-y.

[17] H. M. Al-Hamadi and S. A. Soliman, “Long-term/mid-term electric load forecasting based on short-term correlation and annual growth,” *Electr. Power Syst. Res.*, vol. 74, no. 3, pp. 353–361, 2005, doi: 10.1016/j.epsr.2004.10.015.

[18] D. P. Sharma, P. S. C. Nair, and R. Balasubramanian, “Demand for commercial energy in the state of Kerala , India : an econometric analysis with medium-range projections,” vol. 30, no. 2002, pp. 781–791, 2020.

[19] C. Hamzaçebi, H. A. Es, and R. Çakmak, “Forecasting of Turkey’s monthly electricity demand by seasonal artificial neural network,” *Neural Comput. Appl.*, vol. 31, no. 7, pp. 2217–2231, 2019, doi: 10.1007/s00521-017-3183-5.

[20] A. Azadeh, R. Babazadeh, and S. M. Asadzadeh, “Optimum estimation and forecasting of renewable energy consumption by artificial neural networks,” *Renew. Sustain. Energy Rev.*, vol. 27, pp. 605–612, 2013, doi: 10.1016/j.rser.2013.07.007.

[21] K. B. Sahay, S. Sahu, and P. Singh, “Short-term load forecasting of Toronto Canada by using different ANN algorithms,” *2016 IEEE 6th Int. Conf. Power Syst. ICPS 2016*, pp. 12–17, 2016, doi: 10.1109/ICPES.2016.7584044.

[22] L. Ekonomou, “Greek long-term energy consumption prediction using artificial neural networks,” *Energy*, vol. 35, no. 2, pp. 512–517, 2010, doi: 10.1016/j.energy.2009.10.018.

[23] J. Szoplik, “Forecasting of natural gas consumption with artificial neural networks,” *Energy*, vol. 85, pp. 208–220, 2015, doi: 10.1016/j.energy.2015.03.084.

[24] R. Khalid and N. Javaid, “A survey on hyperparameters optimization algorithms of forecasting models in smart grid,” *Sustainable Cities and Society*, vol. 61, no. May. Elsevier, p. 102275, 2020, doi: 10.1016/j.scs.2020.102275.

[25] M. Kankal and E. Uzlu, “Neural network approach with teaching–learning-based optimization for modeling and forecasting long-term electric energy demand in Turkey,” *Neural Comput. Appl.*, vol. 28, no. s1, pp. 737–747, 2017, doi: 10.1007/s00521-016-2409-2.

[26] A. Anand and L. Suganthi, “Forecasting of Electricity Demand by Hybrid ANN-PSO Models,” *Int. J. Energy Optim. Eng.*, vol. 6, no. 4, pp. 66–83, 2017, doi: 10.4018/ijeoe.2017100105.

[27] P. Jiang, R. Li, N. Liu, and Y. Gao, “A novel composite electricity demand forecasting framework by data processing and optimized support vector machine,” *Appl. Energy*, vol. 260, no. September 2019, p. 114243, 2020, doi: 10.1016/j.apenergy.2019.114243.

[28] K. Muralitharan, R. Sakthivel, and R. Vishnuvarthan, “Neural network based optimization approach for energy demand prediction in smart grid,” *Neurocomputing*, vol. 273, pp. 199–208, 2018, doi: 10.1016/j.neucom.2017.08.017.

[29] C. Deb, F. Zhang, J. Yang, S. E. Lee, and K. W. Shah, “A review on time series forecasting techniques for building energy consumption,” *Renew. Sustain. Energy Rev.*, vol. 74, no. February, pp. 902–924, 2017, doi: 10.1016/j.rser.2017.02.085.

[30] M. S. Al-Musaylh, R. C. Deo, J. F. Adamowski, and Y. Li, “Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia,” *Adv. Eng. Informatics*, vol. 35, no. December 2017, pp. 1–16, 2018, doi: 10.1016/j.aei.2017.11.002.

[31] I. K. Okakwu, E. S. Oluwasogo, A. E. Ibhaze, and A. L. Imoize, “A comparative study of time series analysis for forecasting energy demand in Nigeria,” *Niger. J. Technol.*, vol. 38, no. 2, p. 465, 2019, doi: 10.4314/njt.v38i2.24.

[32] S. A. Ur Rehman, Y. Cai, R. Fazal, G. Das Walasai, and N. H. Mirjat, “An integrated modeling approach for forecasting long-term energy demand in Pakistan,” *Energies*, vol. 10, no. 11, pp. 1–23, 2017, doi: 10.3390/en10111868.

[33] M. Akpinar and N. Yumusak, “Year ahead demand forecast of city natural gas using seasonal time series methods,” *Energies*, vol. 9, no. 9, 2016, doi: 10.3390/en9090727.

[34] M. Personal and R. Archive, “Munich Personal RePEc Archive Forecasting irish inflation using ARIMA models,” no. 11359, pp. 0–8, 2008.

[35] M. S. Al-Musaylh, R. C. Deo, J. F. Adamowski, and Y. Li, “Short-term electricity demand

forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia," *Adv. Eng. Informatics*, vol. 35, no. December 2017, pp. 1–16, 2018, doi: 10.1016/j.aei.2017.11.002.

[36] Y. An, Y. Zhou, and R. Li, "Forecasting India's electricity demand using a range of probabilistic methods," *Energies*, vol. 12, no. 13, 2019, doi: 10.3390/en12132574.

[37] F. Jiang, X. Yang, and S. Li, "Comparison of forecasting India's energy demand using an MGM, ARIMA model, MGM-ARIMA model, and BP neural network model," *Sustain.*, vol. 10, no. 7, 2018, doi: 10.3390/su10072225.

[38] "Long Short-Term Memory (LSTM): Concept | by Eugine Kang | Medium." https://medium.com/@kangeugine/long-short-term-memory-lstm-concept-cb3283934359 (accessed Sep. 20, 2020).

[39] "Recurrent neural network - Wikipedia." https://en.wikipedia.org/wiki/Recurrent_neural_network (accessed Nov. 03, 2020).

[40] J. Bedi and D. Toshniwal, "Deep learning framework to forecast electricity demand," *Appl. Energy*, vol. 238, no. October 2018, pp. 1312–1326, 2019, doi: 10.1016/j.apenergy.2019.01.113.

[41] T. Y. Kim and S. B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy*, vol. 182, pp. 72–81, 2019, doi: 10.1016/j.energy.2019.05.230.

[42] H. Abbasimehr, M. Shabani, and M. Yousefi, "An optimized model using LSTM network for demand forecasting," *Comput. Ind. Eng.*, vol. 143, no. March, p. 106435, 2020, doi: 10.1016/j.cie.2020.106435.

[43] Wikipedia, "Long short-term memory," *https://en.wikipedia.org/wiki/Long_short-term_memory*. https://en.wikipedia.org/wiki/Long_short-term_memory (accessed Mar. 29, 2020).

[44] S. J. Taylor and B. Letham, "Forecasting at Scale," *PeerJ Prepr.*, vol. 35, no. 8, pp. 48–90, 2017, doi: 10.7287/peerj.preprints.3190v2.

[45] U. Prophet, H. W. Models, A. I. Almazrouee, A. M. Almeshal, and A. S. Almutairi, "applied sciences Long-Term Forecasting of Electrical Loads in Kuwait," 2020.

[46] M. Barassi and Y. Zhao, "Combination forecasting of energy demand in the UK," *Energy J.*, vol. 39, pp. 209–237, 2018, doi: 10.5547/01956574.39.SI1.scor.

[47] U. Kumar and V. K. Jain, "Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India," *Energy*, vol. 35, no. 4, pp. 1709–1716, 2010, doi: 10.1016/j.energy.2009.12.021.

[48] M. H. Amjadi, H. Nezamabadi-pour, and M. M. Farsangi, "Estimation of electricity demand of Iran using two heuristic algorithms," *Energy Convers. Manag.*, vol. 51, no. 3, pp. 493–497, 2010, doi: 10.1016/j.enconman.2009.10.013.

[49] W. Li, X. Yang, H. Li, and L. Su, "Hybrid forecasting approach based on GRNN neural network and SVR machine for electricity demand forecasting," *Energies*, vol. 10, no. 1, 2017, doi: 10.3390/en10010044.

[50] I. Shah, H. Iftikhar, S. Ali, and D. Wang, "Short-term electricity demand forecasting using components estimation technique," *Energies*, vol. 12, no. 13, pp. 1–17, 2019, doi: 10.3390/en12132532.

[51] Y. Wang, J. Wang, G. Zhao, and Y. Dong, "Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China," *Energy Policy*, vol. 48, pp. 284–294, 2012, doi: 10.1016/j.enpol.2012.05.026.

[52] C. Guo, Q. Ge, H. Jiang, G. Yao, and Q. Hua, “Maximum Power Demand Prediction Using Fbprophet with Adaptive Kalman Filtering,” *IEEE Access*, vol. 8, pp. 19236–19247, 2020, doi: 10.1109/ACCESS.2020.2968101.

[53] G. Zhu, S. Peng, Y. Lao, Q. Su, and Q. Sun, “Short-Term Electricity Consumption Forecasting Based on the EMD-Fbprophet-LSTM Method,” *Math. Probl. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/6613604.

[54] L. Suganthi and A. A. Samuel, “Energy models for demand forecasting - A review,” *Renew. Sustain. Energy Rev.*, vol. 16, no. 2, pp. 1223–1240, 2012, doi: 10.1016/j.rser.2011.08.014.

[55] C. Dabakoglu, “Time Series Forecasting — ARIMA, LSTM, Prophet with Python,” *https://medium.com/@cdabakoglu/time-series-forecasting-arima-lstm-prophet-with-python-e73a750a9887*, 2019. https://medium.com/@cdabakoglu/time-series-forecasting-arima-lstm-prophet-with-python-e73a750a9887 (accessed Mar. 22, 2020).

[56] “8.9 Seasonal ARIMA models | Forecasting: Principles and Practice.” https://otexts.com/fpp2/seasonal-arima.html (accessed Nov. 02, 2020).

[57] “How LSTM networks solve the problem of vanishing gradients | by Nir Arbel | Data Driven Investor | Medium.” https://medium.com/datadriveninvestor/how-do-lstm-networks-solve-the-problem-of-vanishing-gradients-a6784971a577 (accessed Nov. 01, 2020).

[58] Facebook, “Quick Start | Prophet,” *GitHub*. https://facebook.github.io/prophet/docs/quick_start.html (accessed Mar. 30, 2020).

[59] S. Seabold and J. Perktold, “Statsmodels: Econometric and Statistical Modeling with Python,” *PROC. 9th PYTHON Sci. CONF*, no. December, p. 57, 2010, [Online]. Available: http://statsmodels.sourceforge.net/.

[60] M. Mukherjee and D. Saha, “A Stepwise Approach to Automate the Search for Optimal Parameters in Seasonal ARIMA Models,” no. April, pp. 144–156, 2019.

[61] T. E. Oliphant, “Guide to NumPy,” *Methods*, vol. 1, p. 378, 2010, doi: 10.1016/j.jmoldx.2015.02.001.

[62] “pmdarima · PyPI.” https://pypi.org/project/pmdarima/ (accessed Nov. 02, 2020).

Appendix

Table A.1: List of states and union territories included in the five electrical zones of India (adopted from [9])

Zones	States
Northern	Chandigarh, Delhi, Haryana, Himachal Pradesh, Jammu & Kashmir, Punjab, Rajasthan, Uttar-Pradesh, Uttarakhand
Western	Chhattisgarh, Gujarat, Madhya Pradesh, Maharashtra, Daman & Diu, Dadra & Nagar Haveli, Goa
Southern	Andhra Pradesh, Karnataka, Kerala, Tamil Nadu, Telangana, Puducherry, Lakshadweep
Eastern	Bihar, Jharkhand, Odisha, West Bengal, Sikkim, Andaman & Nicobar
North Eastern	Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Tripura

Table A.2: Statistical summary for the monthly total and peak energy demand data for India (2008 to 2019)

Total Energy Demand (GWh)

Parameter	India	North	South	West	East	North East
Average	85.00	25.76	22.83	26.00	9.34	1.07
Maximum	113.02	38.57	33.61	37.74	13.78	1.67
Minimum	61.61	16.39	15.54	17.84	6.18	0.64
Median	85.10	24.93	23.11	25.39	9.25	1.04
Skewness	0.11	0.47	0.02	0.26	0.38	0.30
Kurtosis	-0.89	-0.34	-0.41	-0.52	-0.60	-0.75
Peak Energy Demand (GW)						
Parameter	India	North	South	West	East	North East
Average	146.47	41.68	35.79	41.55	15.82	2.10
Maximum	190.67	63.17	49.62	56.68	23.14	2.97
Minimum	110.68	29.67	25.77	30.99	11.11	1.46
Median	144.34	40.77	35.90	40.75	15.41	2.07
Skewness	0.21	0.52	0.13	0.52	0.42	0.13
Kurtosis	-0.92	-0.41	-0.57	-0.39	-0.60	-1.08

Table A.3: Sixty months (April 2019 - May 2024) total and peak energy demand forecast for India and its five electrical zones using Fb Prophet model

	Total Energy Demand (GWh)						Peak Energy Demand (GW)					
Period	National Level	North Zone	South Zone	West Zone	East Zone	North East Zone	National Level	North Zone	South Zone	West Zone	East Zone	North East Zone
01-04-2019	107.95	29.49	30.36	34.57	12.48	1.28	188.32	50.32	48.98	54.11	22.09	2.56
01-05-2019	115.81	35.30	29.27	36.23	13.32	1.43	191.46	55.61	46.97	54.02	22.09	2.72
01-06-2019	110.50	36.62	27.51	31.86	13.09	1.47	194.10	60.10	46.30	51.81	22.04	2.66
01-07-2019	113.14	38.59	29.68	30.15	13.16	1.62	194.42	61.39	47.43	49.37	22.48	2.83
01-08-2019	113.65	38.37	29.18	30.79	13.54	1.64	194.86	61.88	46.42	49.70	22.31	2.83
01-09-2019	110.25	35.45	28.91	31.19	13.28	1.62	196.28	59.41	47.58	53.49	22.64	2.81
01-10-2019	114.02	33.16	29.82	36.31	12.84	1.57	195.83	54.33	47.43	57.08	22.71	2.79
01-11-2019	102.13	28.58	26.72	33.97	11.62	1.41	183.54	48.94	46.56	56.03	21.32	2.68
01-12-2019	107.31	31.09	28.66	34.90	11.35	1.42	186.29	51.20	47.38	56.32	20.25	2.63
01-01-2020	109.25	31.75	29.53	34.50	11.95	1.42	189.18	51.92	48.44	56.13	20.81	2.63
01-02-2020	104.47	28.74	30.13	32.39	11.48	1.31	193.60	51.55	50.55	56.88	21.46	2.65
01-03-2020	115.80	30.40	34.02	36.19	13.54	1.37	194.96	48.90	52.03	56.78	22.73	2.64
01-04-2020	115.29	31.19	32.78	36.11	13.69	1.36	198.74	55.35	52.34	57.13	23.32	2.70
01-05-2020	118.42	36.21	31.06	36.38	13.52	1.44	201.96	58.70	50.52	56.70	23.53	2.75
01-06-2020	115.21	38.16	29.08	33.05	13.26	1.58	204.73	63.11	48.84	55.29	23.32	2.86

01-07-2020	117.97	39.99	30.10	32.06	14.00	1.65		201.11	62.84	48.51	50.44	23.46	2.82
01-08-2020	117.93	38.62	31.10	32.40	14.11	1.75		202.51	62.44	50.25	52.81	23.50	2.88
01-09-2020	115.80	37.16	29.65	33.19	13.91	1.67		206.78	61.91	50.40	56.87	24.04	2.91
01-10-2020	116.23	34.43	30.70	36.01	13.69	1.63		203.51	58.18	50.39	58.92	23.79	2.93
01-11-2020	108.29	29.95	29.07	35.63	11.95	1.45		194.62	50.29	49.06	59.49	22.59	2.79
01-12-2020	112.98	32.00	30.10	37.14	11.99	1.49		195.35	53.80	49.87	59.45	21.34	2.69
01-01-2021	114.32	32.68	31.80	35.75	12.58	1.51		198.76	54.43	51.81	59.25	21.90	2.71
01-02-2021	103.59	28.87	29.96	32.67	11.10	1.31		198.71	53.07	52.88	58.55	22.55	2.64
01-03-2021	119.56	31.24	35.71	37.17	13.93	1.42		203.75	51.80	55.05	59.45	23.94	2.71
01-04-2021	118.97	32.02	33.94	37.49	14.08	1.41		207.35	56.51	55.12	59.88	24.56	2.76
01-05-2021	123.95	37.55	32.33	38.28	14.27	1.52		210.68	60.70	53.06	59.53	24.70	2.85
01-06-2021	119.81	39.36	30.29	34.40	14.00	1.63		213.50	65.36	51.60	57.71	24.52	2.90
01-07-2021	122.65	41.31	31.79	33.08	14.54	1.73		211.06	65.63	51.79	53.42	24.78	2.93
01-08-2021	122.79	40.29	32.31	33.53	14.75	1.80		212.17	65.54	52.63	55.19	24.74	2.98
01-09-2021	120.05	38.24	31.19	34.21	14.51	1.74		215.61	64.28	53.13	59.39	25.23	2.99
01-10-2021	121.71	35.54	32.26	38.02	14.20	1.70		213.13	59.84	53.06	62.13	25.07	3.00
01-11-2021	111.90	30.81	29.98	36.92	12.54	1.52		202.41	52.42	51.78	62.13	23.71	2.86
01-12-2021	116.97	33.11	31.39	38.28	12.46	1.54		203.86	55.66	52.64	62.19	22.43	2.78
01-01-2022	118.56	33.81	32.87	37.16	13.09	1.56		207.23	56.34	54.38	61.95	23.02	2.79

01-02-2022	109.42	30.12	31.80	34.27	11.89	1.38		208.78	55.25	55.89	61.74	23.71	2.75
01-03-2022	123.22	32.06	37.40	38.10	14.29	1.46		212.49	54.75	58.08	62.10	25.15	2.77
01-04-2022	122.58	32.83	35.07	38.87	14.45	1.45		215.90	57.58	57.88	62.62	25.79	2.82
01-05-2022	129.56	38.89	33.59	40.21	15.02	1.59		219.36	62.69	55.58	62.36	25.86	2.96
01-06-2022	124.40	40.55	31.50	35.74	14.76	1.68		222.26	67.61	54.38	60.11	25.72	2.95
01-07-2022	127.33	42.64	33.52	34.09	15.07	1.81		221.10	68.46	55.13	56.44	26.11	3.06
01-08-2022	127.65	41.99	33.50	34.66	15.39	1.86		221.87	68.68	54.96	57.54	25.97	3.08
01-09-2022	124.29	39.31	32.75	35.23	15.11	1.82		224.40	66.63	55.87	61.89	26.41	3.07
01-10-2022	127.24	36.64	33.82	40.07	14.69	1.77		222.79	61.46	55.71	65.37	26.36	3.06
01-11-2022	115.45	31.66	30.85	38.18	13.13	1.58		210.12	54.58	54.51	64.75	24.82	2.94
01-12-2022	120.94	34.23	32.68	39.40	12.94	1.60		212.36	57.51	55.42	64.92	23.52	2.87
01-01-2023	122.82	34.95	33.93	38.58	13.60	1.60		215.69	58.24	56.93	64.66	24.15	2.87
01-02-2023	115.37	31.38	33.71	35.90	12.70	1.45		218.96	57.46	58.93	64.96	24.88	2.86
01-03-2023	126.76	32.86	39.09	39.00	14.62	1.51		221.19	57.75	61.10	64.72	26.35	2.83
01-04-2023	126.14	33.63	36.16	40.26	14.81	1.50		224.40	58.55	60.61	65.36	27.03	2.88
01-05-2023	135.25	40.25	34.84	42.18	15.80	1.67		228.00	64.67	58.07	65.20	27.01	3.06
01-06-2023	128.99	41.74	32.72	37.07	15.52	1.72		231.00	69.86	57.18	62.49	26.91	2.99
01-07-2023	132.00	43.97	35.28	35.07	15.59	1.89		231.23	71.33	58.53	59.51	27.44	3.18
01-08-2023	132.53	43.70	34.67	35.79	16.03	1.91		231.60	71.86	57.23	59.86	27.21	3.18

01-09-2023	128.51	40.36	34.33	36.24	15.71	1.89		233.15	68.96	58.61	64.38	27.58	3.16
01-10-2023	132.83	37.74	35.38	42.17	15.18	1.84		232.47	63.03	58.36	68.64	27.65	3.13
01-11-2023	118.93	32.52	31.69	39.43	13.73	1.65		217.75	56.75	57.24	67.34	25.93	3.01
01-12-2023	124.89	35.36	33.96	40.49	13.41	1.65		220.87	59.35	58.20	67.64	24.61	2.95
01-01-2024	127.09	36.10	34.98	40.02	14.11	1.65		224.16	60.15	59.45	67.36	25.27	2.95
01-02-2024	121.46	32.66	35.67	37.55	13.54	1.53		229.27	59.70	61.99	68.22	26.05	2.97
01-03-2024	134.59	34.54	40.25	41.93	15.97	1.59		230.75	56.61	63.74	68.05	27.57	2.96

2