A Comparison of Short-Term Load Forecasting Techniques

1st Aastha Kapoor  
School of Electrical Sciences  
Indian Institute of Technology Bhubaneswar  
OD, India, 752050  
aasthakpr21@gmail.com

2nd Ankush Sharma  
Department of Electrical Engineering  
Indian Institute of Technology Kanpur  
UP, India, 208016  
ankushsharma@ieee.org

Abstract—The short term load forecasting plays a crucial role in optimal operation and scheduling of the generation resources in power system. In this work, Auto-Regressive Integrated Moving Average (ARIMA), Multiple Linear Regression (MLR), Recursive Partitioning and Regression Trees (RPART), Conditional Inference Trees (CTREE) with Bootstrap Aggregating (BAGGING), and Random Forest (RF) models have been tested and compared for short term load forecasting. These methods have been tested on a sample electricity load data of a residential area containing data sets for training and testing.

Index Terms—Auto-Regressive Integrated Moving Average, Ensemble, Load Forecasting, Multiple Linear Regression.

I. INTRODUCTION

The load forecasting is an essential part of power system planning and operation. Short Term Load Forecasting (STLF) is of paramount importance in unit commitment, interchange and scheduling of power for power generation and distribution utilities, and Demand Response Management (DRM). To implement various DRM planning strategies, knowledge of the amount of future energy demand at residential level is required [1],[2]. Economically, the precise forecasting of load decreases the operating cost that prompts significant savings for the power companies. However, there are some challenging factors like seasonal effects (daily and weekly cycles, calendar holidays), meteorological conditions, and special events which make load forecasting a complex problem.

Load Forecasting can be achieved for short, medium and long term. Range of short term can be assumed from one hour to seven days, inspite of no explicit explanation of each time horizon [3]. In past few decades, the research have been carried out to develop various techniques of STLF, such as Statistical models, Artificial Neural Network (ANN) based models, fuzzy models and hybrid models [1]. These methods are broadly categorized into Artificial Intelligence (AI) and Statistical approach [4],[5].

ANN has gained utmost attention by both industry and academia [4,6]. ANN has also been implemented in many utility applications such as failure forecast (with varying degree of success) for distribution system [4]. Due to its flexibility and the capability of handling non-linear data, it is highly utilized by several vendors in power industry [4],[6]. Although, as a black-box approach, the ANN models can deliver quite competent forecast but the accuracy of ANN, when combined with other AI-based techniques such as Fuzzy logic, has not been absolutely persuasive [4]. Further, there are some challenges, such as lack of interpretability and overfitting, in applying ANN to electric load modelling and forecasting [4].

Most of the utilities still use similar day approach, one of the earliest techniques, for short term load forecasting due to its simplicity of implementation and adequate results [7]. The idea of STLF is to utilize the historical days as the fundamental reference. Historically, time series approaches such as Auto-Regressive Moving Average (ARMA) and Auto-Regressive Integrated Moving Average (ARIMA) have been used enormously by the utilities for STLF [4]. ARIMA provides short term accuracy for time series models but its interpretability is complicated as compared to regression models [7], [8]. Multiple Linear Regression (MLR) is also one of the extensively used statistical techniques in electrical load forecasting [4], [9], [10]. This approach considers independent dummy variables, such as time of the day and day of the week, for electric load represented as a linear function [4]. In Random Forest (RF), a generic model is considered which is capable of predicting the load demand of one day ahead by a step of half-an-hour (30 minutes). This model imparts precise results giving no importance to any specific day or the season [3]. The accuracy level of RF is as good as optimized ANN and Support Vector Machine (SVM) with no tuning required [3],[6],[11]. Further, poor choice of parameters does not influence the forecasting accuracy severely [3].

This paper compares various STLF techniques, such as ARIMA, MLR, Recursive Partitioning and Regression Trees
(RPART), Conditional Inference Trees (CTREE) and RF. The paper is divided into four sections; Section II gives the overview of STLF techniques, Section III and IV consist of the simulation results and conclusions, respectively.

II. SHORT TERM LOAD FORECASTING TECHNIQUES

A. ARIMA

ARIMA (p,d,q) is a generalized model of Auto-Regressive AR(p) and Moving-Average MA(q) of order p and q respectively. It is applied when time series has some non-stationary behavior. Such a time series can be represented by a generalized autoregressive operator \( \varphi_A \). The corresponding equations [8] are given as:

\[
\phi(A) = \varphi(A)(1 - A)^d,
\]

where, \( \phi(A)X_t = \varphi(A)(1 - A)^dX_t = \Theta(A)e_t \)

where, \( \Theta(A) = \varphi(A)(1 - A)^d \)

and, \( \Delta^dX_t = X_t - X_{t-1} \)

The suitable values of p and q can be promoted by Partial Auto Correlation Functions (PACF) and Auto-Correlation Functions (ACF) and A is the backshift operator.

\[
\mu = E[X_t] = 1/N \sum_{t=1}^{N} X_t \quad (6)
\]

\[
\sigma^2 = E[(X_t - \mu)^2] = 1/N \sum_{t=1}^{N} (X_t - \mu)^2 \quad (7)
\]

\[
\rho_k = \frac{E[(X_t - \mu)(X_{t+k} - \mu)]}{\sigma^2} \quad (8)
\]

Where \( \rho_k \) is the auto-correlation at a lag factor of k, \( N \) is the number of terms, \( \mu \) is the mean and \( \sigma^2 \) is the variance of the series.

The most important process in the time series modelling is to perform a stationary test on a series and transform it into a stationary series, if it shows a non-stationary behaviour. The process is as follows:

- Get load data and convert into time series.
- Visualize, plot time series and stationarize the series.
- Plot ACF and PACF charts and find optimal parameters.
- Build the ARIMA model and make predictions.

B. MLR

The drawback associated with ARIMA approach is the processing of multiple models for whole year which is a time consuming process. The MLR regression method uses independent variables (continuous or categorical) to take multi-seasonality into account, thus requires single model [4]. A few assumptions, to be considered, are as follows.

1) Assumptions of MLR:

- Normal distribution of residuals.
- A linear relationship between the dependent variable and the independent variables.
- Multicollinearity is not present.

To detect multicollinearity, first correlation coefficients are obtained for each pair of independent or predictor variables. If the correlation coefficient, \( r \), is +1 or -1, it is termed as a case of ideal multicollinearity [4], [9].

2) Interactions Among Predictor Variables: The basic interaction between two predictors, inserted into a multiple regression equation, is through the product of the two variables. To express an interaction between predictors \( X_1 \) and \( X_2 \), the following equation is used:

\[
\hat{Y} = a_0 + a_1X_1 + a_2X_2 + a_3X_1X_2. \quad (9)
\]

If the regression coefficient \( a_3 \) is statistically significant, it proves that the relationship between \( Y \) and \( X_1 \) depends on the value of \( X_2 \) (or vice versa).

As the electricity consumption for a week ahead is also predicted, the model should take weekly seasonality into account. Predictor variables are of two kinds : daily \((D_1,...,D_{48})\) and weekly \((W_1,...,W_6)\). During the day, when the consumption level is being measured at a specific time, the daily variable will be given a value 1 otherwise 0. Similarly, week variable will be 1 for consumption at the specific day, otherwise 0. The regression model is represented as [9],[10]:

\[
Y_i = \beta_1 D_{i1} + \beta_2 D_{i2} + \beta_3 D_{i3} + \beta_4 W_{i1} + \beta_5 W_{i2} + \beta_6 W_{i3} + \beta_7 W_{i4} + \beta_8 W_{i5} + \beta_9 W_{i6} + \epsilon_i \quad (10)
\]

where, \( Y_i \) denotes the electricity consumed at the time \( i \), and \( i \) ranges from 1 to \( M \), \( M \) varies from 1 to 48, \( \beta_1...\beta_9 \) are regression coefficients to be estimated, \( D_{i1}...D_{i48} \) and \( W_{i1}...W_{i6} \) are the independent variables, and \( \epsilon_i \) is the random error. Ordinary least squares (OLS) is used to estimate regression coefficients.

\[
Y = \beta X + \epsilon \quad (11)
\]

where, \( \beta \) is a vector of the length \( q \) and \( X \) is a matrix of size \( M \times q \). \( M \) is the length of vector \( Y \). The \( \beta \) estimation is given as follows:

\[
\hat{\beta} = (X^TX)^{-1}X^TY. \quad (12)
\]

The independent variable \( W_{17} \) and the intercept \( \beta_0 \) (as time 0 is not considered) is excluded to remove the collinearity among independent predictor variables. The matrix \( X \) should be a regular and non-singular matrix. This method is considered biased as it is based on the assumption of independence of variables and linearity of data points [4].

C. ENSEMBLE LEARNING FOR LOAD FORECASTING

In Ensemble learning techniques, multiple classifiers are instructed to work out on the same problem [12], [13]. These methods assemble the obtained hypothesis set and use it for further analysis. There are numerous ways to ensemble different models. Some of the frequently used techniques are as follows:
1) Bootstrap Aggregating (BAGGING): Bagging is a meta-algorithm, that takes N number of samples (with replacement) from the initial dataset and trains the model on those samples. The final model is obtained by taking the average of “bootstrapped” models yielding better results by avoiding the issue of over-fitting [12].

2) Decision Tree: A decision tree is a tree like arrangement in which internal(root) node depicts either an attribute or experiment on an attribute. Each branch of a tree comprises outcome of the experiment and each leaf node depicts class label (decision taken after evaluating all attributes).

3) Recursive Partitioning and Regression Trees (RPART): In RPART, the records are being splitted into two parts consecutively, to attain maximum homogeneity inside new parts.

Algorithm for RPART [14] is as follows:
- Choose one of the predictor values \(x_i\).
- Select further a value of \(x_i\), let’s say \(s_i\), which divides the training data into two sets and measure the homogeneity.
- Algorithm tests distinct values of \(x_i\) and \(s_i\) to maximize purity in initial step.
- As “maximum” purity split is obtained, repeat the steps for the second split and so on.

A prevalent way to spot the most informative attribute is to apply entropy based method [12], [14] given below.

\[
H(X) = -\sum_{i=1}^{m} P_i \log_2(p_i), \quad (13)
\]

\[
\text{where, } p_i = P(X = x_i) \quad (14)
\]

Entropy is a measure of amount of uncertainty in the given data. If the level of entropy is high, prediction becomes more difficult. Decision Tree divides the data set such that the entropy of resulting subset is smaller.

\[
\text{InformationGain(I.G)} = H_s - H_{sa} \quad (15)
\]

where \(H_s\) and \(H_{sa}\) are the values of entropy before and after making decision tree.

4) Conditional Inference Trees (CTREE): CTREE also partitions the data recursively by carrying out a uni-variate split on the dependent variable [14]. It adapts statistical hypothesis testing to choose variables rather than choosing them by Information Gain measures as is done in RPART. Chi-square test statistics is used to examine the association among variables and keeps the relevant ones. Therefore, it removes the prejudice ascribed to a substantial number of categories [15].

5) Random Forest Technique: RF is an ensemble learning approach that unites prediction of weak variables (predictors) \(H_i\). Number of trees ‘ntree’ and number of variables to partition ‘mtry’ at each node are the essential parameters which need to be defined for utilizing this technique. The equation is given as [1], [3]:

\[
Y = H(X) = \frac{1}{ntree} \sum_{i=1}^{ntree} H_i(X) \quad (16)
\]

where, \(X\) is a vector having ‘n’ number of features as input.

### III. SIMULATION RESULTS

For simulation study, one year electricity load of a residential area is taken as a sample data. In order to apply ARIMA technique, the complete sample data of one year (1/1/2014 to 31/12/2014) is taken as a training data set. Load data for the first day of next year (1/1/2015) is taken as a test data set. For MLR and Ensemble techniques, a sample data of 17 weeks (May-August) and a training data set of three weeks (May,2-May,22) is taken. Load data of first day of the fourth week (May,23) and upcoming two weeks (May,23-June,5) are considered as test data sets. The programming is done using R, a statistical analysis software and MATLAB.

#### A. Forecasting time series using ARIMA

One year load of a residential area is taken on hourly basis consisting of 24*365 data points (values) from (1/1/2014) to (31/12/2014). Accordingly, a load curve (24 hour schedule) for the next day (1/1/2015) is forecasted. The data is clustered into 24 training sets. Each set is having 365 values in order to reduce complexity and memory required. Load for 24 hour duration on 1-1-2015 is predicted and compared with the real test data as shown in Fig. 1. On comparing the results with test data, it is found that the Mean Absolute Percentage Error (MAPE) for ARIMA model is 13.71%.

![Fig. 1. Load forecasting for 1-1-2015 using ARIMA](image)

#### B. Forecasting time series with MLR

To execute MLR, the dataset of electricity load of consumers from a residential area is used. The time series data of 17 weeks is used to train and test the model. Three weeks (May 2 to May 22) data of electricity consumption is used as a training set. Forecasts are performed for one day and two weeks ahead.

For first multiple linear model, two important parameters in statistics, i.e. p-value for F-test (goodness of fit) and R-squared are tested.

The test results show ‘R-squared’ value as 0.753 and ‘p-value’ for F-test as 0.

In Fig. 2, it can be observed that for weekends, i.e. 9 May (around 300 time samples) and 16 May (around 600 time samples), the fit is not reasonable. The fitted values are plotted against the residuals to observe the characteristics of the model in Fig. 3.
The plot in Fig. 3 shows the occurrence of non-constant residuals as they are not scattered around zero value. As the MLR algorithm needs residuals to be normally distributed, the normal distribution of the residuals is obtained by using a Quantile-Quantile (Q-Q) plot as is shown in Fig. 4.

As it can be observed from Fig. 4 that the scattered points are not as close to the dashed (red in color) line. This shows that the residuals are not normally distributed as stated in MLR assumptions. From Fig. 3 and Fig. 4, it is observed that the model is reasonably not fit for load forecasting. Therefore, some modifications are done which are as follows:

- In above case, the fitted values are moved continuously during the day by the estimated coefficient of weekly variable. However, the response of variables during the day is not recorded. As the behavior for weekends is completely different, therefore, this factor is also included.

- Further, by introducing interactions between day and week variables, the accuracy of the model can be increased. Therefore, each daily variable is multiplied with every week variable.

After testing MLR model with modifications, the new observed ‘R-squared’ value is 0.992, which shows improvement over previous observed ‘R-squared’ value 0.753.

After introducing interactions between variables, the characteristics of the new model are observed in Fig. 5, Fig. 6 and Fig. 7. It is visualized from the above figures that the fitted values are now reasonable and the model is now suitable for load forecasting. Using the above model, the data values for two weeks ahead is predicted. The forecasted plot of fitted values by MLR and real values for two weeks is shown in Fig. 8. Further, forecasted values for one day ahead are observed in Fig. 9. After analyzing the results of Fig. 9, it is found out...
C. ENSEMBLE LEARNING FOR LOAD FORECASTING

1) Forecasting time series using RPART+BAGGING and CTREE+BAGGING:

a) RPART+BAGGING
The training data of three weeks consists of electricity load having double-seasonal fourier terms (daily and weekly). Therefore, electricity consumption is first detrended by Seasonal and Trend decomposition [5], [14] using Loess (STL) decomposition. Trend part is forecasted (modeled) by ARIMA (auto.arima function). Seasonal and remainder part is then, forecasted by regression tree model. Total 100 bootstrapped forecasts by RPART are performed and are stored in the matrix of size 100*48. To sample the training set, sampling ratio is considered in the range of 0.7 to 0.9. The median of forecasts is taken and predictions are visualized accordingly.

In Fig. 10, the dark line (in red color) is the median of forecasts. Forecasts produced by RPART have a rectangular shape, but resultant ensemble forecasts satisfies smooth behaviour as shown in Fig 10. After analyzing the results, the MAPE in RPART+BAGGING is observed as 2.98%.

b) CTREE + BAGGING
In this method, hyperparameters of CTREE which take decisions about splitting a node, are randomized. CTREE does not involve any bias for variable selection which takes place in RPART. It selects variables which possess numerous possible splits. After testing this technique and analyzing the results, as in Fig. 11, the MAPE is 2.94% which is comparatively less than RPART.

2) Forecasting time series using RF: The modelling is done by RF predictors placed for 48 points (value taken in each half an hour) of the day. To predict the electricity load of the specific day ‘D’ and for each continuous half an hour ‘H’, the characteristics of the input variables are the components influencing the electricity load demand. The components include month number, type of the day, load of the previous day ‘D1’ at half an hour ‘H’. The outcome (label) is the predicted load of day ‘D’ at each half an hour ‘H’.

RF also processes number of trees (ntree), number of variables sampled in each split (mtry), including the hyperparameters of previous models (RPART+CTREE) [3]. ‘Importance of variables’ can also be included in the process of prediction [12]. In this model, mtry is considered as q/3 (where q is the number of features). As q is 9 in this case, mtry is computed as 3. The nodesize is 5 and the number of trees is set as 1000. After analysis, on varying the above mentioned hyperparameters, it can be observed that the optimal result with minimum error arrives for mtry as 6 and nodesize as 5.

The forecast is computed and compared with the real values in Fig. 12.

The MAPE for RF technique is observed as 2.86%. The forecasts from all approaches mentioned above are compared
with one another and also with real electricity load in Fig. 13.

Further, the MAPE for all techniques discussed above is compared in Table I.

<table>
<thead>
<tr>
<th>Forecasting Technique</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA</td>
<td>13.71%</td>
</tr>
<tr>
<td>MLR</td>
<td>4.5%</td>
</tr>
<tr>
<td>RPART+BAGGING</td>
<td>2.98%</td>
</tr>
<tr>
<td>CTREE+BAGGING</td>
<td>2.94%</td>
</tr>
<tr>
<td>RF</td>
<td>2.86%</td>
</tr>
</tbody>
</table>

From Table I, it is visualized that BAGGING helps to decrease forecasting error for RPART and CTREE. Further, it can be observed that the error with RF is lesser than the MLR, RPART, CTREE and ARIMA approaches.

IV. CONCLUSION

This paper presents a comparative analysis of five commonly used short-term load forecasting techniques, i.e. Auto-Regressive Integrated Moving Average (ARIMA), Multiple Linear Regression (MLR), Recursive Partitioning Regression Trees with Bootstrap Aggregating (RPART+BAGGING), Conditional Inference Trees with Bootstrap Aggregating (CTREE+BAGGING), and Random Forest (RF). These techniques are tested on a sample electricity load data of a residential area. From the simulation results, it is observed that ARIMA is least accurate. MLR is also tested and found to be better than ARIMA approach. However, the Mean Absolute Percentage Error (MAPE) with MLR is still significant. Therefore, to improve the accuracy, ensemble learning techniques such as RPART with BAGGING, CTREE with BAGGING, and RF are implemented on the same data set. From simulation study, it is observed that CTREE gives better results than RPART. Further, on comparison of MAPE of all techniques mentioned above, it can be concluded that the error associated with RF is least and this approach produces more accurate forecasting results.

REFERENCES