Maninder singh al. International Journal of Recent Research Aspects ISSN: 2349-7688, Vol. 1, Issue 1, June 2014, pp. 62-68

Visualization & Analysis of Energy Forecasting Data Using Curve Fitting

Maninder singh[1], Dr. Kamal Sharma[2], Deepak Kumar[3]

[1]Student, M. Tech, ESEAR, Ambala
[2]Professor, Dept. of ECE, E-Max group of Institutions, Ambala
[3]Assistant Professor, Dept. of ECE, E-Max group of Institutions, Ambala

Abstract— **According to current predominant views, the field of energy will undergo significant structural changes in coming periods, making it radically different from what we know today. Thus, many researchers and organizations envision distinct paths of energy development. Managing electrical energy supply is a composite task. In this thesis, it proposes a visualization and analysis of energy forecasting by historical data presented. For this, it uses time series analysis methods like regression analysis and curve fitting mainly. In this work, it takes historical energy data which includes information related to energy at grid and produced energy. It also contains weather report, day type information and temperature. So, it provides complete day information as well as monthly information report. It also provides power estimation at a particular hour in a day. It also checks data with the help of normal distribution. It calculates analysis parameters like mean, S.D and log likelihood values. At last, these values are correlated with each other.The work is done with the help of MATLAB tool.**

Keywords— **data mining, energy forecasting, regression, curve fitting.**

I. Introduction

Energy is vitally important for modern economies. It enables the use of daily appliances (such as computers, medical devices, telecommunication appliances, and transport vehicles) that increase people's quality of life. Most appliances used in daily life are powered by energy and it is generally regarded at least in the developed world to be almost impossible to live without them. As a result, energy is seen as a necessity for social and economic welfare; it is essential to maintain economic activity in modern industrialized nations and social development. Moreover, one of the main reasons for low social and economic progress in developing nations is the limited access to modern energy services given appliances that require electricity (such as computers, televisions and radios) provide access to information that accelerates social progress of societies [1].

Over centuries, humans have changed their lifestyles along with technological progress and innovation. The exceptional economic growth and major improvements in standards of living over the last two decades have mainly come about because of the replacement of manpower with mechanical power through technological progress. Energy consumption and technology have developed through history and modern societies' lifestyles became more energy dependent. These energy dependent lifestyles make energy indispensable for life; societies want uninterrupted light, hot water, warm houses, to travel freely and to power industries. Humans have become accustomed to the benefits that are provided by energy consuming appliances and arguably, it is impossible today to think about life without these appliances.

Energy is becoming scarcer and more expensive, making it an ever more critical factor of production for companies as well as entire countries. Prosperity and growth are increasingly dependent on the efficient use of energy.The prosperity and growth of modern societies depend to a large extent on having sufficient energy available wherever needed – as electric power, fuel, or feedstock. It was long taken for granted that fossil fuels – oil, gas, hard coal, and lignite – make up the majority of the global energy supply. The oil crises in 1973 and 1979 shook the foundations of this assumption for the first time. This period was followed by a surge of innovation in power generation (including renewable energy), as well as a marked rise in energy efficiency in many industrialized countries. However, the share of fossil fuels in the energy supply of western industrialized countries is still over 80 percent.

Although there is a wide diversity among developing economies in terms of socioeconomic conditions (size, economic structure, human resources, and energy endowments, level of urbanization), some common energy system characteristics can be found for most of them. These characteristics include: poor performance of the power sector and traditional energies, transition from traditional to modern energies, and structural deficiencies in the economy, society and in the energy systems which result in "urban-rural divide, inadequate investment decisions and misdirected subsidies". Some authors points out that the existence of large scale inequity and poverty, dominance of traditional life styles and markets in rural areas, transitions of populations from traditional to modern markets, existence of multiple social and economic barriers to capital flow and technology diffusion cause developing countries' energy systems significantly different from that of developed countries.

Energy demand is a derived demand that arises for satisfying some needs which are met through use of appliances. Hence, demand for energy then depends on the demand for energy services and the choice of energy using processes or devices. End-use service demand is affected by the cost of energy but also by other factors such as climatic conditions, affordability (or income of the decision-maker), preference for the end-use service, etc. Similarly, demand for end-use appliances depends on the relative prices of the appliances, relative cost of operation, availability of appliances, etc. The dynamics of energy demand is influenced by the inertia of appliance stocks, which leads to limited flexibility.

The paper is organized as follows. In section II, we discuss related work with energy forecasting data. In Section III, It describes analysis of energy forecasting data and proposed technique. Section IV contains results and performance analysis of system.Finally, conclusion is explained in Section V.

II. RELATED WORK

Authors proposed a new fuzzy logic method for midterm energy forecasting. The proposed method properly transforms the input variables to differences or relative differences, in order to predict energy values not included in the training set and to use a minimal number of patterns. The input variables, the number of the triangular membership functions and their base widths are simultaneously selected by an optimization process. The standard deviation is calculated analytically by mathematical expressions based on the membership functions. Results from an extensive application of the method to the Greek power system and for different categories of customers are compared to those obtained from the application of standard regression methods and artificial neural networks (ANN).

Some proposed that medium-term electric energy demand forecasting is an essential tool for power system planning and operation, mainly in those countries whose power systems operate in a deregulated environment. This paper proposes a novel approach to monthly electric energy demand time series forecasting, in which it is split into two new series: the trend and the fluctuation around it. Then two neural networks are trained to forecast them separately. These predictions are added up to obtain an overall forecasting. Several methods have been tested to find out which of them provides the best performance in the trend extraction. The proposed technique has been applied to the Spanish peninsular monthly electric consumption. The results obtained are better than those reached when only one neural network was used to forecast the original consumption series and also than those obtained with the ARIMA method.

Some proposed that real-time control of wave energy converters requires knowledge of future incident wave elevation in order to approach optimal efficiency of wave energy extraction. We present an approach where the wave elevation is treated as a time series and it is predicted only from its past history. A comparison of a range of forecasting methodologies on real wave observations from two different locations shows how the relatively simple linear autoregressive model, which implicitly models the cyclical behaviour of waves, can offer very accurate predictions of swell waves for up to two wave periods into the future.

Author proposed that handling electrical energy supply is a complex task. The most significant part of electric utility resource planning is forecasting of future load request in regional or national service area. This is usually achieved by building models on relative information, such as climate and previous load demand data. In this paper, a genetic programming method is suggested to predict long term electrical power consumption in the area covered by a utility situated in the southeast of Turkey. The experiential results establish successful load forecast with a low error rate.

III. ANALYSIS OF ENERGY FORECASTING DATA

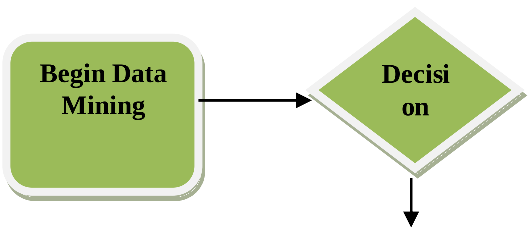
Forecasting methods can be classified as qualitative or quantitative. Qualitative methods generally involve the use of expert judgment to develop forecasts. Such methods are suitable when historical data on adjustable being predicted are either not applicable or unavailable. Quantitative predicting methods can be used when (1) past information about variable being prediction is available, (2) the information can be enumerated, and (3) it is reasonable to undertake that the pattern of the past will continue into the future.

The proposed steps for motion estimation are:

1. Create a database file of energy forecasting data.
2. Interface Excel data with MATLAB Tool.
3. Plot input energy data in 3-D w.r.t system load.
4. Find the mean of energy data to get average day profile.
5. Use curve fitting method to estimate power at a particular hour.
6. Calculate average month profile from energy data.
7. Use regression analysis to find cdf, pdf and normally distribution. Data must be normally distributed.
8. Calculate log likelihood value of data.
9. Estimate performance parameters like mean, standard deviation and its confidence interval of energy data.
10. Separate each day from data and estimate its energy usage.
11. Analysis with the help of graphs.

Process of Proposed System

- The data mining process can be broken down into four distinct phases.
- 1. Decision, whether or not to carry out data mining on a given data set.
- 2. Data preparation, readying the data for analysis.
- 3. Model building, where the work of building a prediction model is carried out.
- 4. Interpretation, which is largely carried out by individuals but which can be greatly assisted using automated means, such as graphical representations of the results. This process is illustrated using a flow chart in Figure 1.



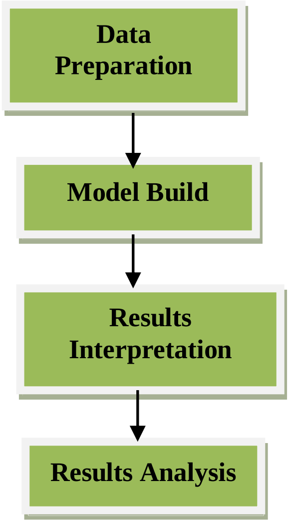


Figure 1: Flow Chart of Proposed System

1. Decision Phase

The first stage in the data mining process is that of deciding whether or not to go ahead with a given analysis. This is one of the most difficult and probably the most crucial of all the stages, as it is here that we decide whether or not we are going to spend our time and other resources investigating a given data set. Although this stage is often given to humans, who are able to ask and answer questions pertaining to the identified problem domain, it is also useful to have automated means to help in the taking of this first step. Unsupervised clustering is one means of determining whether relationships, in the form of concepts, exist in the data. If a clustering finds such concepts, then it can be deduced that a supervised model is likely to perform well, leading us to continue in the data mining process.

2. Data Preparation Phase

Once we have decided to go ahead with our investigation, it is vital that the data be in a format that can be easily interpreted by the model building tool. Data preparation is a vast topic. The scope of this study requires that we understand the importance of data preparation to the process as a whole and the need for such facilities in the tools which we investigate.

Examples of data preparation include:

- The search for outliers,
- The discretization of continuous data
- And normalization.

Certain tools are given almost entirely to this stage of data mining and unsupervised clustering is also widely used in this stage of the process, particularly in the discovery of outliers. The more pre-processing applied to a given data set, the better the results from creating a data mining model of that data set are likely to be. The more pre-processing a tool offers, whether supervised or unsupervised, the better is likely to be the performance of models created by that tool or subsequently applied tools.

The following are a list of summarized data added to the event file:

- Three Excel sheets contain separate data related to forecasting.
- One sheet contains historic data related to grid.
- Second sheet contains data related to energy produced.
- Third sheet contains day type information, HDD value and its temperature.

3. Model Building Phase

Model building is the core of the data mining process. This is where verifiable results are obtained. The scope of this study is limited to supervised learner models. What this essentially means is that the models it creates will have been trained using examples of known cases (from the data set) and then verified using further information from the data which has not yet been presented to the model. This stage is known as the training and testing or validation stage and once completed the model produced can be used to predict future outcomes, instances which have neither been seen by the model nor by individuals. For example, in creating a neural network, a proportion of the instances in a given data set will be used to “train” the network. The neural network is then tested in order to discover what it “knows” using the remaining data and once verified as an effective model is used in classification or prediction of unseen cases.

The structural time series approach enables the formulation of a model that captures the main characteristics of the data in the beginning of the process. After the model has been estimated, the suitability of the model should be assessed by both applying a series of diagnostic tests and checking the consistency of the estimated parameters and hyper-parameters with the economic theory and prior intelligence. Therefore, the estimated parameters, hyper-parameters and the interventions should be consistent with the economic history. In this study, The MATLAB will be the tool used most for the purpose of model creation. It satisfies a number of model selection criteria including:

- Data Coherency: The normality of residuals should be maintained. The residuals should be entirely random white noise disturbance terms that exempt autocorrelation and heteroscedasticity.
- Consistency with theory: The model should present consistent results with the economic theory and economic history.
- Parsimony: The preferred model should be at its possible simplest form.
- Encompassing: The model should present the data better than its rival models.

4. Interpretation Phase

The final stage of the data mining process is that of interpretation. This stage is vital to the process as, “it is the analysis of results provided by the human element that ultimately dictates the success or failure of a data mining project”. As with the decision phase, however, our interpretation of the results can be assisted using automated means. One such means is visualization tools, which illustrate what is known by the model. MATLAB is renowned for its graphical presentation of results and all of the tools Examined. Other automated means of interpretation are mathematical validity measures of the algorithms implemented.

5. Tool Assessment

The first stage of the assessment process is a critical evaluation of the claims of the tool. These claims are evaluated in terms of that stage of the data mining process to which they pertain. Second, the applications of the tool are briefly explored, that is, the stage or stages of the data mining process to which the tool is best suited are highlighted. The final step in the assessment process is to further clarify ideas pertaining to the synthesis of the tool.

6. Data Analysis

The main steps are Sample, Explore, Modify, Model, and Assess. Sample entails choosing a subset of data that is large enough to contain all pertinent information, but small enough to process quickly. This subset is then divided into three subsets—training, validation, and test sets. The training set of data is used to fit the model, the validation set of data is used to prevent over fitting a model, and the test set is used to evaluate how well the model fits the data.

- Explore is the step to gain a better understanding of the data by identifying trends or anomalies in the data either visually or using statistical methods like cluster analysis.
- Modify entails changing the dataset by performing tasks such as creating new variables, eliminating other variables, and eliminating anomalies. Modeling the data is the step where different types of models are chosen for the software to fit to the data automatically.
- Assessing the data is the final step in the iterative process where one checks the validity of the results. This assessment is done by taking a test dataset and applying the model to these data to test if the model predicts the correct result.

B. Statistical Parameters

1. Mean

The Mean, traditionally denoted by x, is the arithmetic average of a set of observations. It can be calculated by the sum of the observations divided by the number of observations and is given by eq.1:

$$x\ (mean) = \frac{1}{n}\sum_{i=1}^{n} xi \qquad (1)$$

2. Variance

The variance is another widely-used measure of dispersion. It estimates the mean squared deviation of x from its mean value x (mean) and is given by eq. 2:

$$var = \frac{1}{n-1}\sum_{i=1}^{n}(xi - x)^2 \qquad (2)$$

3. Correlation

Some numbers can be used to convey the degree of fit to others. Chief among these is the correlation coefficient. The correlation coefficient, often designated by the letter ‘r’, is usually heuristically defined by eq. (3):

$$r = \sqrt{\frac{Explained\ Variation}{Total\ Variation}} \qquad (3)$$

IV. RESULTS

This showcases visualization and analysis (heavy statistics) for forecasting energy usage based on historical data. It has access to hour-by-hour utility usage for the month of January, including information on the day of the week and the Heating Degree Days (defined as 65 minus Average Temperature) of each day. Using this information, we will come up with an algorithm for forecasting future energy usage based on parameters such as day-type, forecasted temperature, and time of day [10].

A. Implementation

“MATLAB has excellent built-in support for many data analysis and visualization routines,” in particular, one of its most useful facilities is that of efficient exploratory data analysis, which is a natural, fit in the context of data mining.MATLAB's portability comes from the fact that all MATLAB users will have the same range of basic functions at their disposal.

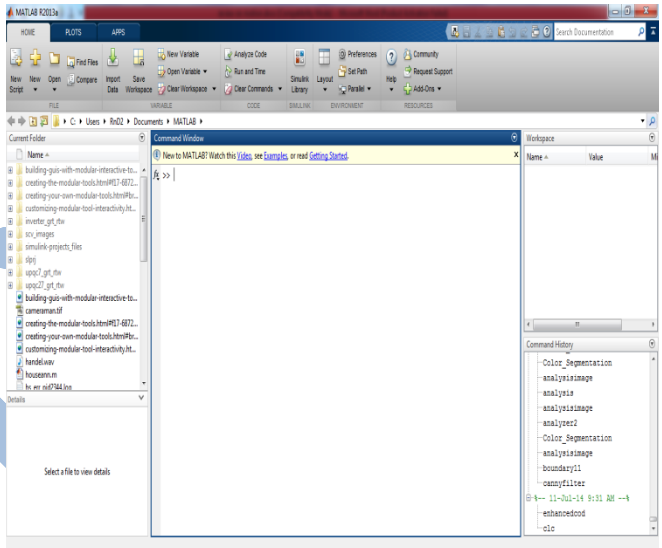


Figure 2: MATLAB Tool

The representation which MATLAB implements, is dealing with all data in the form of matrices. Other advantages of MATLAB include its interactive interface, debugging facilities, object oriented nature and inparticular its high quality graphics and visualisation. MATLAB's add on feature, in the form of toolboxes, makes it possible to extend the existing capabilities of the language with ease.

B. Forecasting Results

This describes visualization and analysis (heavy statistics) for forecasting energy usage based on historical data. It has access to hour-by-hour utility usage for the month of January, including information on the day of the week and the Heating Degree Days of each day. Using this information, it will come up with an algorithm for forecasting future energy usage based on parameters such as day-type, forecasted temperature, and time of day. The figure 3 shows the historic data values.

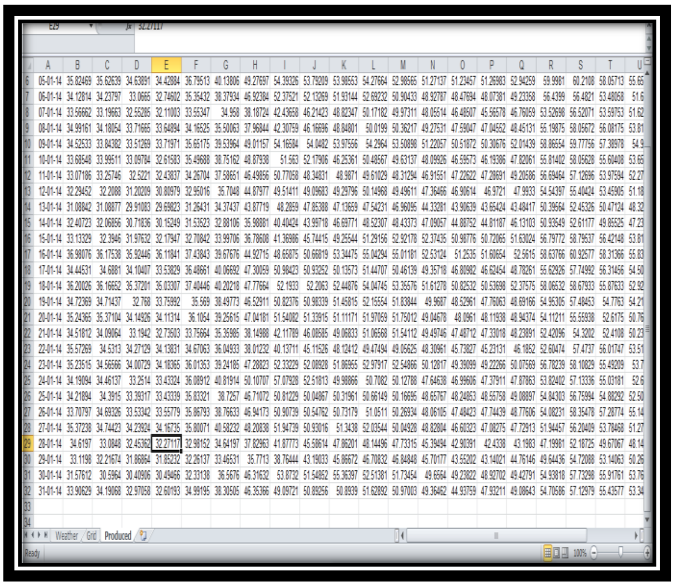


Figure 3: Excel Data

Data must contain energy production values, day type and temperature. Then visualize the data with the help of 3-D plotting as shown in fig 4. Data is plotted w.r.t system load. The axis contains hours, days and system load resp. After this, it has to calculate power at a particular hour as shown in fig 5.

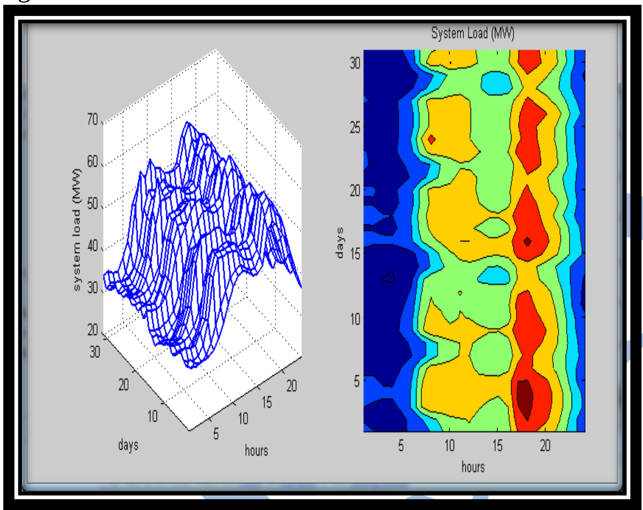


Figure 4: Input Data in 3-D Plot

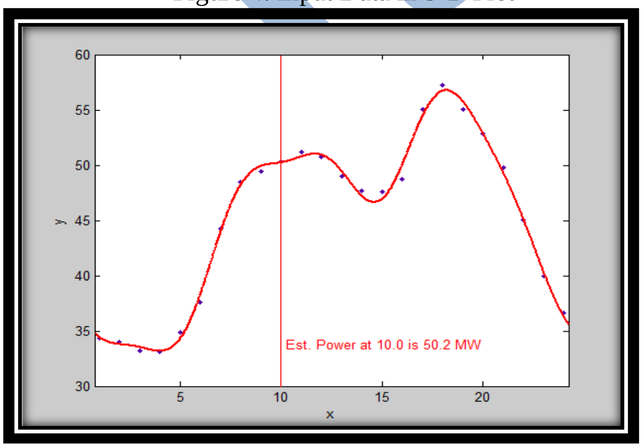


Figure 5: Estimate Power at a Particular Hour

In order to do statistics, it evaluates average month profile. Use curve fitting method for this. This graph is in 3-D view. Axis contains hours, days and system load resp. After this, it is going to determine the distribution type of this data set. Calculate probability distribution function and cumulative distribution function by statistics method. Also find the normal distribution type of data set. Likelihood value is also calculated. The results are shown in fig 6. From the distribution analysis, it will conclude that the data set has a normal distribution. With this information, it can perform more in-depth statistics based on normal distribution. The graphs are plotted between data values and density resp.

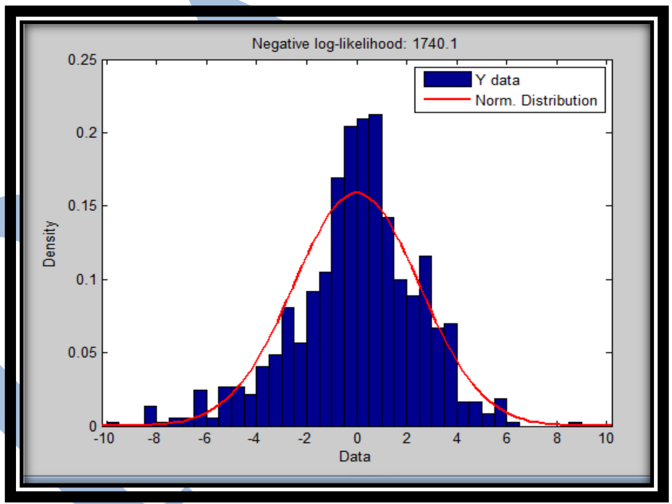


Figure 6: Normally distributed Type (with Log-Likelihood Value)

In fig 7, the daily profile seems to have a very tight confidence interval, suggesting that the general trend throughout the day is similar from day to day.

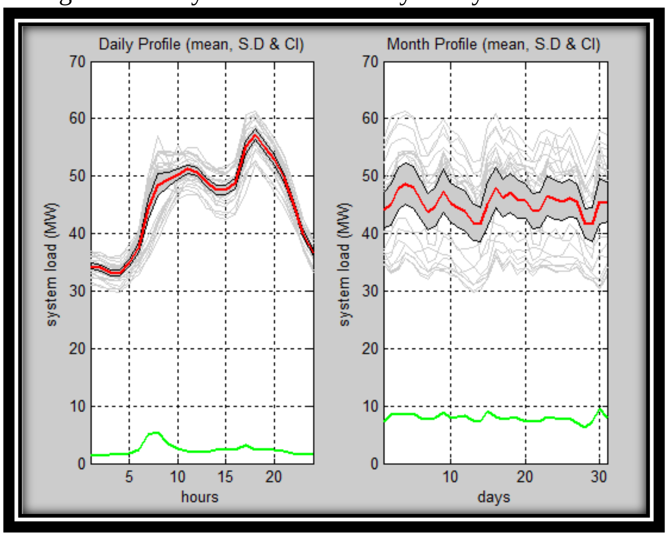


Figure 7: Performance Parameters Profile

By looking at average profiles for each day of the week, it can make some observations on daily trends. It can see that the morning energy spike is not prominent on the weekends. Also, Mondays tend to have more usage throughout the day, and Saturdays have the lowest usage as shown in fig 8.

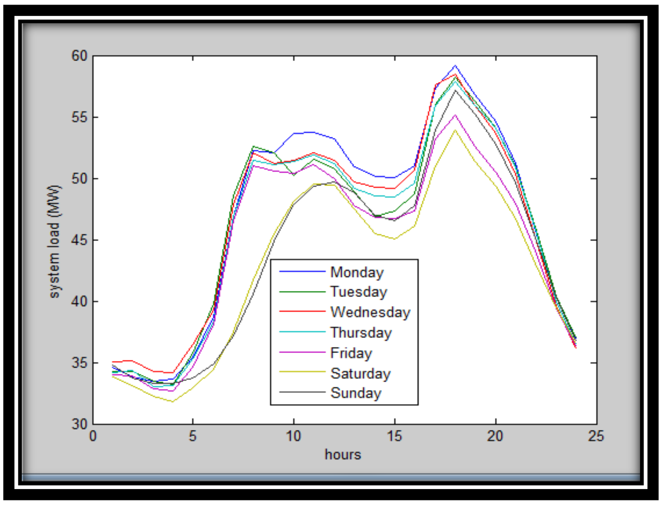


Figure 8: Each Day Profile

Now, Forecast usage based on time of day and day of week. The profile shows energy usage of individual day as well as system load value within confidence interval. The profile is plotted between system load and hours. This profile shows the load value at 3 PM. The profile shows that Monday has more usage than other days and Saturday has minimum usage of energy as shown in fig 9.

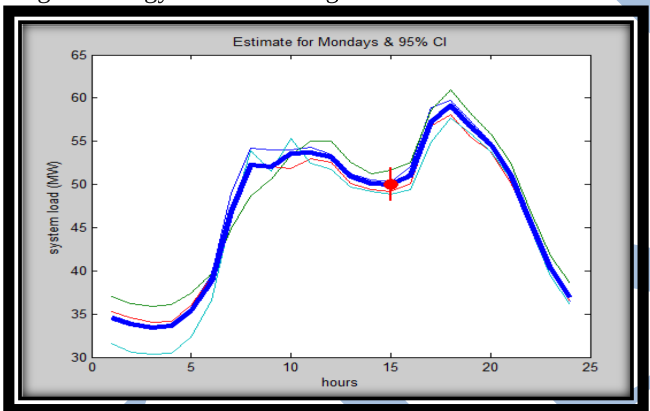


Figure 9: Energy at a Particular Day

The performance analysis is shown below. Table 1 shows the power analysis at a particular hour of a day and Table 2 shows the system load analysis at particular day of profile.

Table 1: Power Analysis at a Particular Hour

Time of Hour	*Power Used*
5 AM	*34.3*
10 AM	*50.2*
3PM	*46.9*
8 PM	*52.8*

Table 2: System Load Analysis at a Particular Day

Days	*System Load (MW)*
Monday	*50.01*
Tuesday	*47.32*
Wednesday	*49.17*
Thursday	*48.43*
Friday	*46.73*
Saturday	*45.03*
Sunday	*46.52*

Now, it can determine this by examining the correlation coefficients between HDD and the various statistical parameters as shown in Table 3.

Table 3: Correlation with Parameters

	Mean	*Median*	*Min*	*Max*	*Range*	*S.D*
Corr:	*0.5720*	*0.4125*	*0.6350*	*0.7606*	*0.4835*	*0.3802*
P-val:	*0.0008*	*0.0211*	*0.0001*	*0.0000*	*0.0059*	*0.0349*

V. CONCLUSION

This paper covers the visualization and analysis of energy forecasting system using historic data. It explored historical energy usage data to develop a forecasting system. Statistics and visualizations revealed that there are usage trends throughout the day, and the trends seem to depend on the day of the week. . It has access to hour-by-hour utility usage for the month of January, including information on the day of the week and the Heating Degree Days of each day. Using this information, it will come up with an algorithm for forecasting future energy usage based on parameters such as day-type, forecasted temperature, and time of day. This knowledge can be used for a gross forecast of the energy usage. In this, single day profile and average month profile is evaluated. It also calculates power of any particular hour observed. In this work, performance parameters like mean, standard deviation and confidence interval is also evaluated. It also fits normally distributed curve with this data. And also used to find cdf and pdf values. It also calculates normalized likelihood values of normalized data. This method is beneficial for large range of data forecasting. It is also helpful to evaluate single day analysis or monthly forecasting. Results shows that Monday has more usage than other days and Saturday has minimum usage of energy. With the help of this, it can evaluate any type of data for forecast. It can say that the this approach can be used for electric utility resource planning and forecasting of the future load demand in the regional or national service area effectively. It also found that there are some correlations between the Heating Degree Days (HDD) and certain statistical parameters of the profile. We may use this information to fine tune our forecast.

REFERENCES

[1] F.M Alvarez, A.Troncoso, J.C. Riquelme, and J.S. Aguilar-Ruiz,"Energy Time Series Forecasting Based on Pattern Sequence Similarity", IEEE Transactions On Knowledge And Data Engineering, VOL. 23, NO. 8, AUGUST 2011.

[2] C.N. Elias and N.D. Hatziargyriou, "An Annual Midterm Energy Forecasting Model Using Fuzzy Logic", IEEE Transactions On Power Systems, VOL. 24, NO. 1, FEBRUARY 2009.

[3] E.G Romera , M.Á. Jaramillo-Morán, and D.C Fernández, "Monthly Electric Energy Demand

Forecasting Based on Trend Extraction", IEEE Transactions On Power Systems, VOL. 21, NO. 4, NOVEMBER 2006.

[4] F.Fusco and J.V. Ringwood, "Short-Term Wave Forecasting for Real-Time Control of Wave Energy Converters", IEEE Transactions On Sustainable Energy, VOL. 1, NO. 2, JULY 2010.

[5] K.Karabuluta, A. Alkanb, A.S. Yilmaz, "Long Term Energy Consumption Forecasting Using Genetic Programming", Mathematical And Computational Applications, Vol. 13, No. 2, pp. 71-80, 2008.

[6] J.Ranjan, "Applications Of Data Mining Techniques In Pharmaceutical Industry", Journal of Theoretical and Applied Information Technology © 2005 – 2007

[7] M. A. Farahat, M. Talaat, "Short-Term Load Forecasting Using Curve Fitting Prediction Optimized by Genetic Algorithms", International Journal of Energy Engineering 2012

[8] R.Behera, B.B.Pati, B.P.Panigrahi, S. Misra, "An Application of Genetic Programming for Power System Planning and Operation" Vol. 03, No. 02, March 2012

[9] A.Botterud, J. Wang, "Wind Power Forecasting and Electricity Market Operations",

[10] N.Padhy, Dr. P.Mishra, and R.Panigrahi, "The Survey of Data Mining Applications And Feature Scope", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.3, June 2012

[11] S.K. Aggarwal, L.M Saini, A.Kumar, "Electricity price forecasting in deregulated markets: A review and evaluation", 14 September 2008

[12] A.Jain, B. Satish, "lustering based Short Term Load Forecasting using Artificial Neural Network", 2009

[13] S.Shakiba, M.Piltan, S.F. Ghaderi, M.S. Amalnik, "Short-term electricity price forecasting in deregulated markets using artificial neural network", Proceedings of the 2011 International Conference on Industrial Engineering and Operations Management Kuala Lumpur, Malaysia, January 22 – 24, 2011

[14] N.Jain, V.Srivastava, "Data Mining Techniques: A Survey Paper", Volume: 02 Issue: 11 | Nov-2013

[15] L.Sharma, M.Chakrawarti, A.Dutta and N. Adhikari, "Neural Network Based Approach For Short-Term Load, Forecasting", International Journal of Science, Environment and Technology, Vol. 1, No 5, 2012