ANALYTICAL STUDY OF CORRELATION BETWEEN DEMAND AND RENEWABLE ENERGY FORECASTING USING DATA MINING/ANALYTICS

Disha Harwalkar[1A], S P Gahana[1B], Theertha Pilakkal[1C], Siddharth Ganesh Sreeganesan[1D]
Students, Dept. of E&EE, Acharya Institute of Technology, Bengaluru-560107
Shreyas Karnick[2]
Assistant Professor, Dept. of E&EE, Acharya Institute of Technology, Bengaluru-560107.

Abstract - The demand for electricity is increasing day by day owing to the growing population and urbanization. Therefore, in order to meet the demand of the consumers and also to ensure no mismatch of supply/demand, development of efficient data forecasting methods is necessary. Load data, solar data, wind data of Germany was used for the purpose of time series forecasting. The load forecasting was done for renewable energy sources- solar, wind and for the conventional load. Linear Regression and LSTM based RNN are the two methods of forecasting used in the paper. The criteria for selecting the best load forecasting method amongst the two was based on the error obtained.

Correlation is a statistical technique that can show whether and how strongly paired variables are related. In this paper in order to find the correlation between the load data and solar data, solar data and wind data and load data and wind data Pearson's and Spearman's correlation coefficient algorithms were used. The Pearson's and Spearman's rank correlation coefficient clearly shows how much energy generated from solar and wind helps in meeting the day to day load requirement.

The platform used by us for coding python is SPYDER. For the cross verification of the theoretical and practical values obtained, mathematical calculations were done.

***Index Terms*: Load forecasting, Correlation, Solar and Wind generation forecasting**

1. INTRODUCTION

Electricity is correlated with the economic growth of our country. It provides a working flow for the industries. Meeting the increasing electricity demand is the key issue. Electricity can be produced by conventional sources of energy- coal petroleum and nuclear energy; non-conventional sources of energy- solar, wind, biomass etc. The energy generated from solar and wind is highly variant in nature. It depends on various factors. The modernization of power systems has brought a revolution in the electricity generation (which usually consists of a mix of renewable and non-renewable energy sources) and distribution sectors in recent years.

Accurate power demand forecasting is very important in the performance of the grid. The vast data collected on daily generation of electricity through various means it can help in forecasting the day ahead generation. Electric load forecasting plays a vital role in overall operation and planning of power systems. Accurate electric load forecasting helps to run the power system efficiently and effectively. Areas such as cause of power interruptions, coordination between supply and demand, operating costs, maintenance and infrastructure development can benefit from electric load forecasting.

Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate together. It can be useful in data analysis and modeling to better understand the relationships between variables. The statistical relationship between two variables is referred to as their correlation. Every dataset uses variables and observations. In this dataset studying how the various data are related to each other will help us draw useful results. With reference to electrical factors, it is important to know the contribution of the renewable energy resources in meeting the required demand which can be achieved using various correlation methods.

2. LITERATURE SURVEY

Rajesh Kumar et.al,[1] has discussed various ways of electricity generation by renewable resources and the challenges faced for grid integration. In this paper, various issues regarding integration of large scale renewable which are variant and intermittent are addressed. Various suggestions have been proposed.

Subrata.M et.al.[2] inspects the progress of renewable energy integration with smart grid. India being a huge demography there is vast variety in parameters, hence production remains unpredictable. Various Regulatory mechanisms in India, some already in vogue and others gradually coming to be in force, encourages to promote Distributed Generation and Renewable and protects the concerned Green Energy Sources to meet the electricity demand.

Use of persistence method for forecasting of electricity generation using renewable sources is studied in [3]. In this paper, persistence technique (today equals tomorrow) was chosen as the forecast algorithm for two reasons: it is simple to implement and unlike most other forecast algorithms, it relies neither on weather forecast data nor on in-built toolboxes in software for implementation.

It is very important for grid operators to know how energy RES will produce in future which is studied in [4]. In this paper, the importance of accuracy of energy forecasting is discussed as it provides flexibility in energy management for future power systems. It proposes using data driven techniques for future forecasting.

Importance of big data analysis for efficient working of Energy Management System (EMS) is studied in [5]. The paper is about analyzing big data, which helps in achieving a smart energy management system. Several prediction forecasting methods are used like, univariate forecasting (time series) and multivariate forecasting.

Electricity Price and Load Forecasting using Enhanced Convolution Neural Network and Enhance Support Vector Regression in Smart Grids is studied in [6]. In this paper, Deep Learning (DL) and data mining techniques are used for electricity load and price forecasting. The performance metrics MSE, RMSE, MAE, and MAPE are used to evaluate the Performance of the proposed models.

Electricity Power Load Forecast via Long Short-Term Memory Recurrent Neural Networks is studied in [7]. This paper studies electricity forecasting for large scaled grids. Load data from Estonia country has been studied and analyzed for forecasting using Long short term memory recurrent network (LSTM RNN). LSTM gives better accuracy and performance of forecasting, thus helps efficient planning in large scale like national electrical grids.

Yeongik Son et.al. [8] Discusses developing accurate electrical load forecasting methods for ensuring energy efficiency through use of neural networks and ARIMA models for load forecasting in Jeju island and the error in forecasting is calculated.

Ms. Seema Kore et.al.[9] about data mining techniques, such as data cleaning, data smoothing to get the data required for prediction. The Artificial Neural Network (ANN) plays a great role in forecasting the electricity consumption.

Tanwalai Panapongpakorn et.al [10] discusses various models for time series forecasting. This paper employs mainly four models namely (ARIMA), SARIMA as time series models and Recurrent Neural Network(RNN), Artificial Neural Networks(ANN) as machine learning models is used for forecasting. This paper studies one-step ahead prediction (30 minutes).

Drawbacks of short term load forecast can be overcome by using LSTM based RNN which is discussed in this paper [11]. Various time series models are used if linear characteristics is assumed for load forecasting, in order to achieve more accuracy using Long Short Term Memory based Recurrent Neural Networks is proposed in this paper.

Aleksandra Dedinec et.al. [12], analyzes the correlation between the consumption of electricity in the Republic of Macedonia and eight other variables. The correlation can be used as an indicative measure of the relationship between the variables. From this paper, it can be concluded that the correlation between the consumption in the current (or next) hour is mostly correlated (above 0.9) with the historical data for the same hour the previous day, combination of the previous week and the consumption in the previous day.

Yi Li, Vassilios G. Agelidis et.al. in this paper [13] presents analysis of wind and solar data for the same geographical location and correlates them.

Dipti Srinivasan and Sujana Gundam et.al.in [14]have emphasized about several models that are successfully developed to predict electrical load demand. The models were then used to predict day ahead average and day ahead hourly load for Singapore. A rough estimation has been made for the solar power capacity in Singapore, and even with an experimental panel with not very high efficiency, a maximum of 1377 MW of power can be saved an hour by using solar power generating sources

3. SIGNIFICANCE OF FORECASTING

The modernization of electric grids toward a smart grid are carried out to improve reliability, facilitation of integration of renewable energies, and improvement in the power consumption management. Due to continuous depletion of primary fuel resources and global concern about the environmental pollution, the development of smart grids based on renewable energy resources has gained huge strategic significance to resolve the today's energy crisis.

With the advancement of technology and more number of units, particularly with wind power plants, being feasible at a place grid connectivity has improved. At the same time gradual evolution of intelligent or smart grid employing extensively communication and information technology is paving the way for connection to the grid even for remotely located plants.

Electric load forecasting plays a vital role in overall operation and planning of power systems. Accurate electric load forecasting helps to run the power system efficiently and effectively. Areas such as cause of power interruptions, coordination between supply and demand, operating costs, maintenance and infrastructure development can benefit from electric load forecasting. Load forecasting is an effective method to predict the power/energy needed to meet the demand and supply equilibrium. Load forecasting can be short-term (a few hours), medium-term (a few weeks up to a year) or long-term (over a year).

4. CORRELATION

In statistics, correlation or dependence is any statistical relationship, whether causal or not, between two random variables or vicariate data. In the broadest sense correlation is any statistical association, though it commonly refers to the degree to which a pair of variables are linearly related. Correlations are useful because they can indicate a predictive relationship that can be exploited in practice. For example, an electrical utility may produce less power on a mild day based on the correlation between electricity demand and weather. In this example, there is a causal relationship, because extreme weather causes people to use more electricity for heating or cooling.

However, in general, the presence of a correlation is not sufficient to infer the presence of a causal relationship. The correlation coefficient is a statistical measure of the strength of the relationship between the relative movements of two variables. The values range between -1.0 and 1.0. A calculated number greater than 1.0 or less than -1.0 means that there was an error in the correlation measurement. A correlation of -1.0 shows a perfect negative correlation, while a correlation of1.0 shows a perfect positive correlation. A correlation of 0.0 shows no linear relationship between the movement of the two variables. There are several types of correlation coefficients, but the one that is most common is the **Pearson correlation.**

The Pearson correlation coefficient denoted by ***r*** is used to measure the strength of a linear association between two variables, where the value ***r* = 1** means a perfect positive correlation and the value ***r* = -1** means a perfect negative correlation.

A value of exactly 1.0 means there is a perfect positive relationship between the two variables. For a positive increase in one variable, there is also a positive increase in the second variable. A value of -1.0 means there is a perfect negative relationship between the two variables. This shows that the variables move in opposite directions i.e, for a positive increase in one variable, there is a decrease in the second variable. If the correlation between two variables is 0, there is no linear relationship between them. The closer r is to +1 or -1, the more closely the two variables are related.

$$r_p = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2 \Sigma(y-\bar{y})^2}}$$

Spearman's rank correlation coefficient or Spearman's **ρ,** named after Charles Spearman and often denoted by the Greek letter rho, is a nonparametric measure of rank correlation (statistical dependence between the rankings of two variables). It assesses how well the relationship between two variables can be described using a monotonic function. The Spearman correlation between two variables is equal to the Pearson correlation between the rank values of those two variables.

$$r_s = 1 - \frac{6\Sigma D^2}{n^3-n}$$

Pearson's correlation assesses linear relationships; Spearman's correlation assesses monotonic relationships (whether linear or not). If there are no repeated data values, a perfect Spearman correlation of +1 or −1 occurs when each of the variables is a perfect monotone function of the other. Intuitively, the Spearman correlation between two variables will be high when observations have a similar (or identical for a correlation of 1) rank (i.e. relative position label of the observations within the variable: 1st, 2nd, 3rd, etc.) between the two variables, and low when observations have a dissimilar (or fully opposed for a correlation of −1) rank between the two variables. Spearman's coefficient is appropriate for both continuous and discrete ordinal variables.

5. ALGORITHM FOR FINDING PEARSON SPEARMAN RANK CORRELATION FOR ELECTRICAL DATA IN PYTHON

Data required

The data required for finding the correlation between load data, solar and generation data using Pearson correlation or Spearman rank correlation is predicted value of each load data, solar and wind generation data. Correlation is done in Python. The predicted value is obtained from the output of LSTM based RNN.

Algorithm

Step 1: START

Step2: The forecasted value from LSTM based RNN output is imported and stored in a data frame in python.

Step3: Appropriate code: correlation= dframe.corr (method='Pearson') Based on whether Pearson or Spearman correlation has to be calculated the method is changed.

Step 4: Correlation is obtained for load data &solar data, load data & wind data, solar data & wind data.

Step 5: Stop

6. HEAT MAPS

Heat maps are unique two dimensional maps in which data is represented or displayed in the form of colors. These heat maps display complex data in a simplified easier way. Therefore, from such simplified displays one can easily understand the trends in the data distributions. The data collected is represented in a such a way that warmer colors indicate where the users have spent most of their time and cooler colors represent lack of interest or reduced levels of attention of the users. These maps can simultaneously display spatial as well as numeric information. These maps can also be used to find the contribution of renewable sources to meet electrical demand.

Heat maps divide the graph into rectangular (or hexagonal) bins and use colors to show how many observations fall in each bin. If you have a large number of data values, then ordinary scatter plots, fit plots, residual plots, etc. become difficult to interpret. Heat maps distinguish more clearly between the denser and less dense parts of the data. While the patterns in heat maps are clear, because the colors are used to display the frequency of observations in each cell of the graph making it easier to understand. In this paper heat maps were used to represent the contribution solar and wind in meeting the load demand.

7. RESULTS & DISCUSSION

Results

Correlation between the forecasted data of load, solar and wind was obtained by using an appropriate algorithm in python. Heat map is obtained and the relation was studied. Correlation was also calculated theoretically for a sample size of 60 Pearson and Spearman coefficient correlation was used in this paper.

- **Pearson Correlation**

Remarks	Correlation Coefficient (Practical)	Correlation coefficient (Theoretical for 60 samples)
Load & Solar	0.29	0.08147
Load & Wind	0.046	0.6228
Solar & Wind	-0.17	-0.8768

Table 1. Results of Pearson correlation calculated theoretically and practically

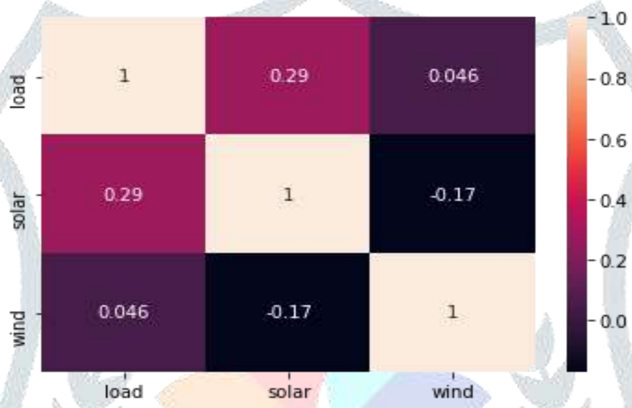

Fig 1 Heat map of Correlation for forecasted values using Pearson's Coefficients

- **Spearman's rank correlation**

Remarks	Correlation Coefficient (Practical)	Correlation coefficient (Theoretical for 60 samples)
Load & Solar	0.47	0.34330
Load & Wind	0..014	0.52103
Solar & Wind	-0.19	-0.00665

Table 2 Results of Spearman correlation calculated theoretically and practically

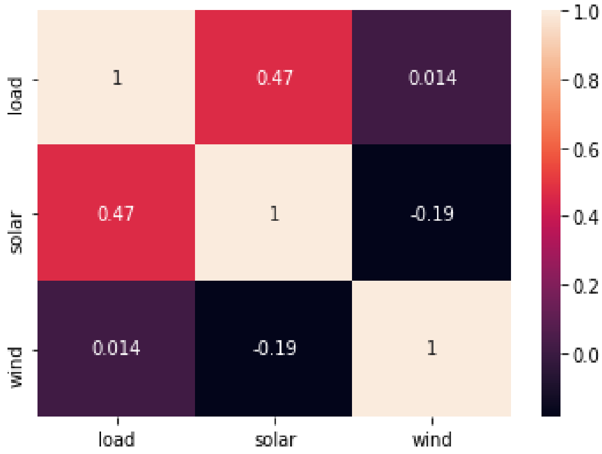

Fig 2 Heat map of Correlation for forecasted values using Spearman Rank Coefficients

Discussions

Correlation coefficients obtained from Pearson and Spearman rank correlation help us identify relations between any two variables. For the correlation between Load and Solar it can be seen from the table that, value 0.29 was obtained using Pearson and a value of 0.47 using Pearson correlation. This tells about the contribution of solar at any time in meeting load demand. Solar generation was high at any time of day when solar penetration was more or the PV cells are installed in huge numbers. Solar generation was not there during the evening. Solar generation can be integrated to the grid for meeting load demand.

For correlation between Load and Wind it can be seen from the table that, value 0.046 was obtained using Pearson and a value of 0.014 using Spearman Rank. It can be inferred that Load dependency on wind is very less. This tells about the contribution of wind at any time in meeting load demand.

For correlation between Solar and Wind it can be seen from the table that, value -0.17 was obtained using a Pearson and a value of -0.19 using Spearman Rank. It can be inferred that solar and wind have a negative relation or dependency on each other.

Electricity produced due to RES is highly varying in nature. The RES system can be integrated with the main grid and generated power can be supplied through the grid to meet the load demands. Optimization of solar and wind power plants can increase the power generation and thus the dependency or the pressure on the conventional sources to produce electricity can be reduced. This will reduce pollution and can help replenish nature.

8. CONCLUSION AND FUTURE SCOPE

To find the correlation between the load data with the solar and wind data and to find correlation between wind and solar data **Pearson's and Spearman's correlation coefficient algorithms** was used. The Pearson's and Spearman's correlation coefficient clearly shows how much energy generated from solar and wind helps in meeting the day to day load requirement. This paper aims in highlighting the contributions made by the renewable energy sources to the smart grid and how their contribution can help relieve the load on conventional energy power plants More solar and wind power plants can be incorporated into the grid (based on the geographical location) to reduce the load on conventional energy sources as stated above. This will reduce pollution and can help replenish nature.

REFERENCES

[1] I.S. Jha, Subir Sen, Kashish Bhambhani, Rajesh Kumar, "Grid Integration of Renewable Energy Sources", Power Grid Corporation of India Ltd., Gurgaon, India- 2015.
[2]. Subrata Mukhopadhyay, Sushil K. Soonee, Sushil K. Soonee,Ashok K. Rajput,"On the Progress of Renewable Energy Integration into Smart Grids in India ", Members of IEEE -2012.

[3]. Shreya Duttaa, Yanling Lia, Aditya Venkataramana, Luis M. Costac, Tianxiang Jiangc, Robert Planab, Philippe Tordjmanb, Fook Hoong Chooa, Chek Fok Fooa and Hans B. Puttgen,"Load and Renewable Energy Forecasting for a Microgrid using Persistence Technique",World Engineers Summit – Applied Energy Symposium & Forum: Low Carbon Cities & Urban Energy Joint Conference, WES-CUE 2017, 19–21 July 2017, Singapore.

[4]. Ekanki Sharma,"Energy forecasting based on predictive data mining techniques in smart energy grids.",From The 7th DACH+ Conference on Energy Informatics Oldenburg, Germany. 11-12 October 2018.

[5]. Big Data Analytics for Load Forecasting in Smart Grids: A Survey",International Conference on Cyber Security and Computer Science (ICONCS'18), Safranbolu, Turkey-2018.

[6]. Maheen Zahid , Fahad Ahmed , Nadeem Javaid , Raza Abid Abbasi , Hafiza Syeda Zainab Kazmi , Atia Javaid , Muhammad Bilal , Mariam Akbar and Manzoor Ilahi ,"Electricity Price and Load Forecasting using Enhanced Convolutional Neural Network and Enhanced Support Vector Regression in Smart Grids",- 2019

[7]. Qiang Jiang,Jia-Xiong Zhu, Min Li, Hai-Yin Qing,"Electricity Power Load Forecast via Long Short-Term Memory Recurrent Neural Networks",4th Annual International Conference on Network and Information Systems for Computers-2018.

[8]. Seunghyeon Park, Sekyung Han, Yeongik Son,"Demand Power Forecasting with Data Mining Method in Smart Grid", Korea Institute of Building Energy Technology Seoul, South Korea- 2016.

[9]. Ms. Seema Kore, Prof. V. S. Khandekar,"Residential Electricity Demand Forecasting using Data Mining",International Journal of Engineering Trends and Technology (IJETT) – Volume 49 Number 1 July -2017

[10]. Tanwalai Panapongpakorn and David Banjerdpongchai ,"Short-Term Load Forecast for Energy Management Systems Using Time Series Analysis and Neural Network Method with Average True Range", Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand - 2019.

[11]. Jian Zheng, Cencen Xu, Ziang Zhang and Xiaohua Li,"Electric Load Forecasting in Smart Grid Using Long-Short-Term-Memory based Recurrent Neural Network",Department of Electrical and Computer Engineering State University of New York at Binghamton Binghamton, NY - 2017.

[12]. Aleksandra Dedinec, Aleksandar Dedinec,"Correlation of variables with electricity consumption data",6th International Conference on Information Society and Technology ICIST-2016.

[13]. "Wind-Solar Resource Complementarity and itsCombined Correlation with Electricity Load Demand",School of Electrical and Information Engineering The University of Sydney-2006

[14]. Dipti Srinivasan and Sujana Gundam,"Correlation Analysis of Solar Power and Electric Demand", International Conference on Renewable Energy and Sustainable Energy [ICRESE'13]-2013