IPS 2018

SOLAR ENERGY FORECASTING – A PATHWAY FOR SUCCESSFUL RENEWABLE ENERGY INTEGRATION

(An ANN Based Model using NARX Model for forecasting of GHI)

Manish Kumar Tikariha, DGM(Operations), NTPC Ltd

K V S S Gupta, Manager (C&I), NSPCL

ABSTRACT

Renewable resources create new challenges in the planning and operation of the electric power grid. In particular, the variability and uncertainty in the renewable resource availability must be properly accounted for in the complex decision-making processes required to balance supply and demand in the power system. It is becoming increasingly evident that forecasting is a key solution to efficiently handle renewable energy in power grid operations. In this paper we discuss about the solar forecasting in detail with emphasis on various forecasting models and their relative usage for the grid operations particular to our country. Solar energy is dependent on the solar radiance reached by the PV panel. Reliable knowledge of solar radiation is essential for informed design, deployment planning and optimal management of self-powered nodes. Solar radiation is a highly complex and Non Linear function which is dependent on various factors, which are exogenous in nature. The problem of solar irradiance forecasting can be well addressed by machine learning methodologies over historical data set. In this study we present an Artificial Neural Network (ANN) based model which uses Non Linear Auto Regressive with exogenous inputs (NARX) for prediction of solar irradiance from the Historical data. To validate the effectiveness of these methodologies, a series of experimental evaluations have been presented in terms of forecast accuracy and Normalised root mean square error (nRMSE). The MATLAB interface has been used as simulation platform for these evaluations. The dataset from national renewable energy laboratory (NREL) has been used for experiments.

Keywords: Solar Energy, Solar Forecasting, Machine Learning.

1. INTRODUCTION:

Global warming and the energy crisis over the past century and most importantly technological innovation in the last 10-15 years have motivated the use and development of alternative, sustainable, and clean energy sources. Out of various renewable energy sources solar and wind are the most preferred ones due to their vast resource potential available with them. These are also inexhaustible and considered as the most promising renewable resources for bulk power generation. In solar energy there are various technologies available to tap the available energy, out of those Photovoltaic (PV) cells are the basic technology for converting solar energy into electric power.

Boosted by a strong solar PV market, renewables accounted for almost two-thirds of net new power capacity around the world in 2016, with almost 165GW coming online. By the end of 2016 PV capacity installed in China (78GW), Japan (41 GW), Germany (40 GW), USA (34 GW), India (10 GW)..PV power generation has introduced significant economic and environmental interests to the public social awareness, such as reducing emissions of CO_2 as well as creating employment.

Coming to Indian Perspective on Renewable energy, in its Intentional Nationally Determined Contribution (INDC), India has declared a significant scaling up of its renewable energy capacity to 175 GW by 2022.this shows that the national energy policies are undergoing a major paradigm shift from fossil fuels to more sustainable resources. This much scaling of renewable energy resources from the present capacity of 40 GW will pose a new challenges to the existing electric power system which needs to be assessed to integrate the renewable energy which calls for efficient system planning.

With the introduction of renewable energy plants, various new challenges emerge from it in the form of unpredictability, uncertainty and variability which are need to be dealt in a different way. To mitigate them it necessitates the use of accurate RE generation forecast, of suitable temporal and spatial resolution, hence its proper utilization in the relevant activities of power system can be ascertained.

PV power is reaching higher and higher penetration level in the smart grid. An important feature of the smart grid is its high ability to integrate renewable energy generation. However, as an intermittent energy source, PV generation introduces significant volatility to the smart grid, which brings severe challenges to system stability, electric power balance, reactive power compensation, frequency response, etc.

To ensure secure and economic integration of PVs into the smart grid, accurate PV power forecasting has become a critical element of energy management systems. Accurate forecasting can help improve electric power quality of the electric power delivered to the electricity network and, and thus reduce the ancillary costs associated with general volatility (Huang et al, 2014). Since PV power output is directly related to solar irradiance at the ground level, solar irradiance prediction is also equally important to energy management in the smart grid (Yang et al, 2013). Moreover, solar prediction with multiple look-ahead times is significant in that it addresses the needs of different operation and control activities, including grid regulation, power scheduling, and unit commitment in both the distribution and transmission grids. Due to the chaotic nature of weather systems and the uncertainties involved in atmospheric conditions such as temperature, cloud amount, dust and relative humidity, precise solar power forecasting can be extremely difficult. A number of forecasting models have been developed for solar resources and power output of PV plants at utility scale level in the past few years.

In recent years several power forecasting models related to PV plants have been published. The existing solutions can be classified into the categories of physical, statistical and hybrid methods. Some of these models were at first oriented to obtain solar radiation predictions while other works present models specifically dedicated to the forecasting of the hourly power output from PV plants . Nowadays the most applied techniques to model the stochastic nature of solar irradiance at the ground level and thus the power output of PV installations are the statistical methods; in particular regression methods are often employed to describe complex non-linear atmospheric phenomena for few-hours ahead forecast and specific soft-computing techniques based on artificial neural network (ANN) are used for few-hours power output forecast. Some other papers use physical methods. Some papers report the comparison of the results obtained with different models based on two or more forecasting techniques. Nowadays the most important forecasting horizon is 24 hours of the next days. Only a few papers describe forecasting models used to predict the daily irradiance or directly energy production of the PV plant for all the daylight hours of the following day.

This paper uses a model based on ANN which uses Non Linear Auto Regressive with exogenous inputs (NARX) for prediction of solar irradiance from the Historical data from the existing infrastructure and used to predict the Solar Irradiance for the next day and can be used for running an hour ahead forecasts in a real-time basis for effective scheduling incorporating the corrections in the previous hour. It has been assessed by changing the size of the training data sets in input, the number of iterations and launching single/multiple-runs. Error analysis has also been done for various other processes in order to evaluate the results.

This paper is organised as follows: section 2 deals with a brief review of existing solar forecasting methods for prediction of solar irradiance is presented. Section 3 emphasises about the method used for Solar Irradiance prediction and how the historical data is pre-processed for the analysis in later stages. Section 4 gives the details about the data collection method employed and the algorithm used in this model. In Section 5 the Solar Irradiance prediction for all the time blocks (15min time blocks) of the following day is presented, in terms of normalized mean absolute error, weighted mean absolute error, and normalized root mean square error. In the end some conclusions are stated.

2. ENERGY FORECASTING METHODS:

Solar forecasting commonly outputs solar irradiance or PV power. The properties of PV generation are essential to solar energy modelling and forecasting. PV Generation or output mainly depends on the amount of solar global irradiation incident on the panels, but that irradiation is not uniform over time and space. Solar resource variability and the uncertainty associated to forecasts are behind most of the problems that must be handled to maintain the stability of the power grid. A part of the fluctuations are deterministic and explained by the rotational and translational movements of the Earth with respect to the Sun, which are accurately described by physical equations. However, there also exists unexpected changes in the amount of solar irradiance arriving at the Earth's surface, mainly derived from the presence of clouds, which stochastically block the Sun's rays and influence PV power forecasting a certain level of uncertainty.

Coming to the development of many studies done worldwide in this area to obtain forecasts are mainly of two approaches.

Indirect Forecasts rely on predicting the solar irradiance, then uses it with the existing PV Performance Model of the plant and obtain the power produced by the plant.

Direct Forecasts directly calculate the power output of the plant.

Many studies focus mainly on the prediction of the solar irradiance since it is associated with highly uncertain, variability issues. Irrespective of the type of forecasts both of them are approached in a similar manner to obtain the forecasted values through the use of similar techniques.

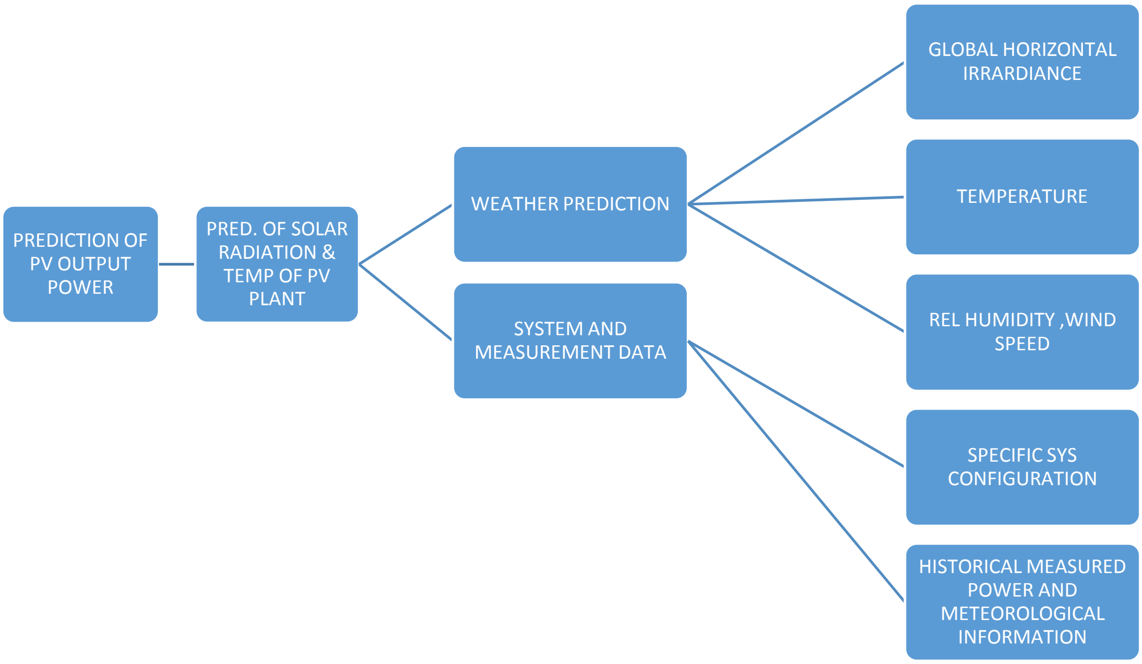


Figure 1: Typical Framework of various approaches in this type of PV Forecasting.

Conversion of GHI or Irradiance on Plane of Array (IPOA) into power output is a Physical process in itself not a forecasting technique. The forecasting effort has been previously made in the prediction of irradiance and other necessary variables, such as temperature or wind. Irradiance predictions generally follow similar patterns to those of power, in most of the comparable techniques are found depending on the time horizon. NWP models generally give us the temperature, wind speed & direction. Some of the irradiance forecasting techniques are given in various literature is given in various papers (Inman et al., 2013; Diagne et al., 2013).

2.1 Physical Models

PV performance method gains upper hand over the statistical method since it doesn't require any historical data, it is possible to obtain the power output of a plant prior to construction. From the technical specifications and the physical attributes decide how much power output can be obtained. The major disadvantage of parametric models is the high reliance on NWP, which lack sufficient spatial and temporal resolution and have been reported as one of the main sources of error of this approach. For instance, whereas an error of 1.2% was reported for the plant modelling, it increased to 10% when irradiance predictions were incorporated to the model (PVCROPS, 2015).

To minimise these type of errors Model Output Statistics (MOS) are applied to weather forecasts. The simplest MOS technique to improve temporal resolution consists in directly interpolating the values. To increase accuracy of these methods ,it requires inclusion of the clear sky index along with the solar zenith angle (Lorenz et al., 2008) or a method based on stepwise linear regression to select the variables that best represent the errors. The application of MOS requires some historical weather data, which may not be always available, which makes to decrease the advantages of PV performance models. PV performance model is sitespecific and because all technical specifications of equipment are rarely known, simplifications have to be made, impacting on the model accuracy.

These type of methods are generally called as white box approach due to the way of forecasting used by them to obtain the solar power output.

2.2Statistical models:

The core principle of statistical model relies on the past data and predicts the future behaviour from it of the system. They don’t consider the internal information about the plant and model the system. The accuracy of prediction output depends entirely on the quality of past data. In some of the application these models prove to be superior to the existing PV performance models. To obtain the forecast we need to have a historical data for which plant must have run for a while which makes it less reliant for the new PV Plants

These models will help in correcting the systematic errors associated with the measurement of various inputs. Basically these models can be broadly divided into two different types

2.2.1Regressive methods:

These techniques estimate the relationship between a dependent variable and some independent variables called predictors. Generally these methods work on a huge time series data which may be treated as linear / non-linear, stationary/ non-stationary. These models basically work on the stochastic property of the predictor and predict the output.

2.2.2 Deep Learning / Machine Learning / Artificial Intelligence Techniques:

Most used machine learning techniques in solar power forecasting are ANN’s (Figure 3). ANN’s are inspired from the neuron operation in our brain, where a group of neurons are interconnected to form a neural network (NN). Each of the Connections have numeric weights, whose final value is given while running the training phase which is an iterative process, and all together predict the output of the system. This range of techniques has proven useful when the system is difficult to model in a wide variety of situations and when we deal with a large number of inputs.

In ANN’s a large number of topologies exists. The classification by the number of hidden layers (simple perceptron or multi-layer) is the main one. Multi-Layer Perceptron (MLP) with number of hidden layers and delays incorporated is considered as a universal proximate of functions and shows wide applicability in the AI techniques. Apart from MLP, many others are also used: Time Delay NN, Adaptive Neuro-Fuzzy (ANFIS), Elman recurrent NN, Radial Basis Function NN, Adaptive Resonance Theory (ART), etc.

In the Figure 3, which gives us a brief idea about how research is being done with respect to time horizon and the type of inputs being taken into consideration while modelling the system.

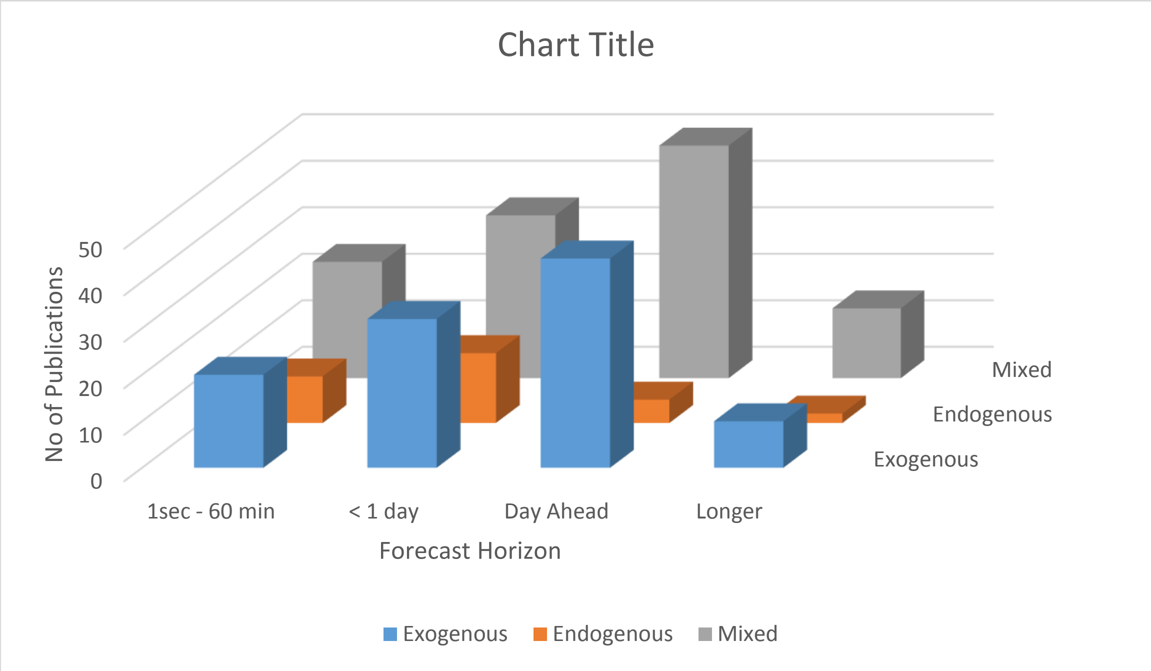


Figure 2: No of research papers Published in Solar forecasting w.r.to type of inputs & time horizon

3. ARTIFICIAL NEURAL NETWORKS

3.1 Basics of ANN

In its most general form, an ANN is a machine that models a task or function of interest, performing useful computation through a process of learning. In fact, the artificial neural network derives its computing power through its massively parallel distributed structure and its ability to learn and generalize, which means finding reasonable outputs whenever inputs are not encountered during training (learning).

The ANN consists of simple processing units, the neuron, and directed, weighted connections between those neurons. The inputs channels have an associated weight, such that the incoming information xi is multiplied by a corresponding weight wi. The network input is the result of the so-called propagation function. Here, the strength of a connection between two neurons i and j is a connecting weight wij. Experimental knowledge, acquired by the network through a learning process, is stored by massively interconnecting these units (synaptic weights). These connecting weights can be inhibitory or excitatory and by being connected with the neurons, data are transferred. The output is a function of the particular activation function chosen and a possible bias. The latter is similar to a weight, albeit it has a constant input of 1. This bias term is used by the neuron to generate an output signal in the absence of input signals. Figure.2 illustrates the nonlinear model of a neuron. (Haylin, 1999).

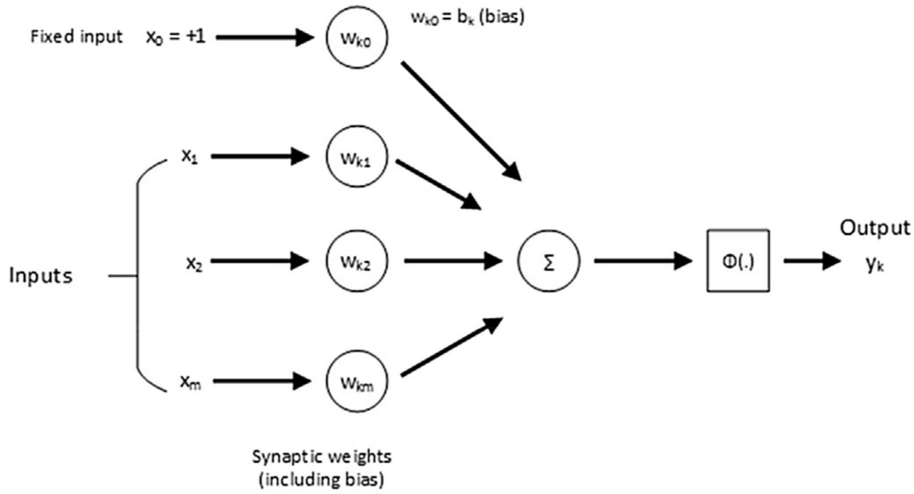


Figure 3 : Nonlinear model of a neuron (Source:Hayline,1999).

The transfer function or activation function controls the amplitude of the output of the neuron and is based on the neuron reactions to the input values and depends on the level of activity of the neurons (activation state). Essentially, neurons are activated when the network input exceeds the uniquely maximum gradient assigned value of the activation function, known as threshold. The activation function is dependent of the previous activation state of the neuron and the external input. It is also referred as a squashing function because limits the permissible amplitude range of the output signal to some finite value. Although theoretically any differential function can be used as an activation function, the identity and sigmoid functions are the most used. In fact, an important feature of the ANN theory is the need for differentiability. Typically, the normalized amplitude range of the output of a neuron is written as the closed unit interval [0, þ1] or alternatively [1, þ1]. The sigmoid function, whose graph is s-shaped, is one of the most common forms of activation used in the construction of ANNs. It exhibits a balance between linear and nonlinear behaviour. The logistic function assumes a continuous range of values from 0 to þ1, and it is easily differentiated. The hyperbolic tangent function tanh is also, often, used because of the simplicity in finding its derivatives. In this case, the activation function assumes an antisymmetric form with respect to the origin, ranging from 1 to þ1. This function is applied in the hidden layer of the network and takes the input with any value between plus and minus infinity to generate an output in the range between 1 and þ1. The identity or linear function f(x) ¼ x, where the inputs and outputs range from minus infinity to plus infinity, is a flow-through mapping of the networks' potential to its outputs.

3.2 NARX model

The input output recurrent model is illustrated in Figure. 3, with a design that follows the typical multilayer perceptron, which are neurons with adjustable synaptic weights and bias. Typically, this model has supervised learning, involving modification of the synaptic weights of the neural network by applying a set of labelled training samples. To a unique input signal there is a corresponding desired response. After being presented with an example picked from the set, the synaptic weights of the network are modified to minimize the differences between the desired response and the actual value response of the network produced by the input signal.

A tapped-delay-line (TDL) memory of q elements is applied to the model inputs. A delay line tap extracts a signal output from somewhere within the delay line and usually sums with other taps to form an output signal. Moreover, via another TDL memory with q units, the single output is also fed back to the input. Thus, the contents from both TDL memories are fed to the input layer of the multilayer perceptron.

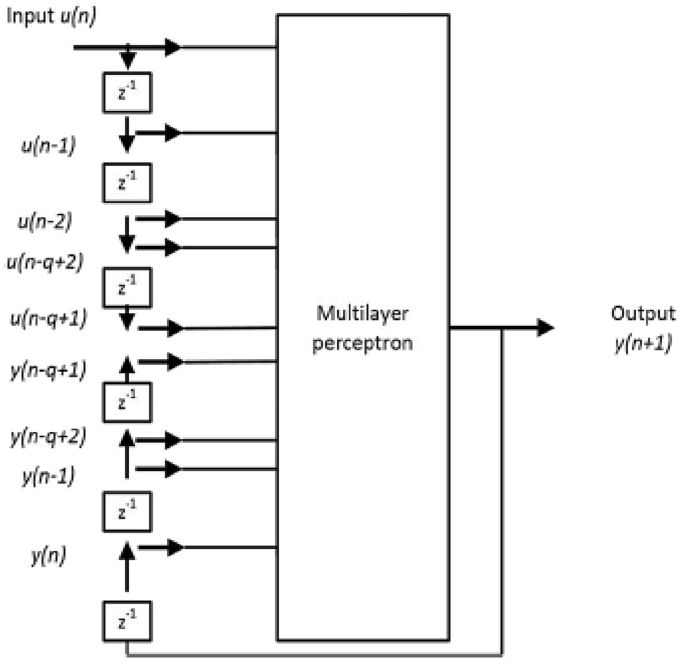


Figure 4: Non Linear Auto Regressive Model with Exogenous Inputs

In Figure. 4, u(n) denotes the present value of the model input and y(n þ 1) corresponds to the value of the model output. Hence, the present and past values of the input, which are exogenous inputs generated from outside the network, and delayed values of the output, on which the model output is regressed, are the data window of the signal vector applied to the input layer.

This recurrent network described above is also referred as nonlinear autoregressive with exogenous inputs (NARX) model. Equation (1) demonstrates the dynamic behaviour of the NARX model, where F is a nonlinear function of its arguments. The two delay line memories in the model are generally different, albeit they can have the same q size.

$$y(n+1) = F\big((y(n), \ldots, y(n-q+1)\big), (u(n), \ldots \ldots .. u(n-q+1)) \tag{1}$$

The training of the network is repeated for many examples in the set until the network reaches a steady state where there are no further important changes in the synaptic weights. In MATLAB we can specify the stopping criteria by giving different metrics like no of validation checks, Gradient increase between Iterations, and Performance improvement between each iteration, to stop the training process. During the training phase the true output is available and, therefore, is used to reduce the associated errors (series parallel architecture, see Figure. 5a). With this mechanism the ANN adopts the feedforward architecture with more accurate inputs. After the training phase the architecture changes and the output is fed back to the input of the feed forward neural network (Parallel architecture, see Figure. 5b) to perform multistep predictions, which is part of the standard NARX architecture.

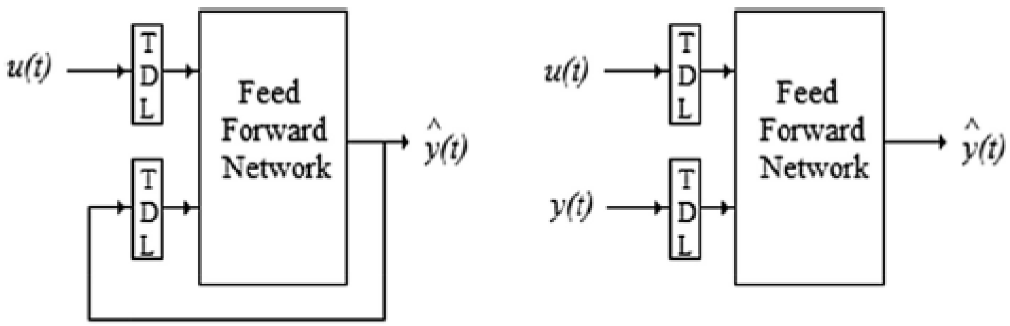


Figure 5: a) Series Parallel Architecture b) Parallel Architecture

4. METHODOLOGY

Data Collection:

In this we use an NARX using ANN for solving this highly non-linear pattern in the solar radiation (GHI). To determine the effectiveness of above mentioned models, historical solar data set including daily solar irradiance over a period of starting from January 1, 2015 to October 31, 2017, has been collected from national renewable energy laboratory (NREL). NREL is primary national laboratory in US for renewable energy that uses baseline measurement system (BMS) with latitude 39.742 north, longitude 105.18 west and elevation 1828.8 m with time zone GMT-7. The data set obtained from NREL has been sampled on per minute basis on horizontal plane with 1440 samples per day. Historical data is converted into 15 mins average which converts the data into 96 samples per day and taken into the model for training. (Source: BMS, NRE)

Along with solar radiation we have taken cloud cover (%), Relative Humidity (RH), Ambient Temp (^{0}C), Average Wind speed at 6ft height are taken into consideration for our model.

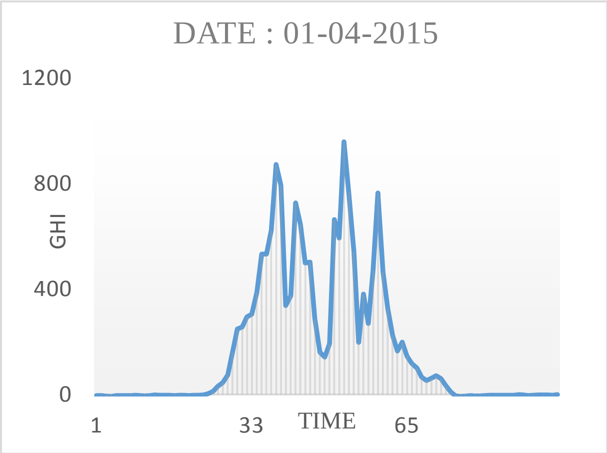
The raw energy data collected once every minute from the source was smoothed using a moving average and it was used to determine the Solar Radiation GHI with 15 min steps. Similarly all other inputs are also smoothed through moving averages and converted into 15 min time steps. All the data sets are normalised to have uniformity with the training and easy analysis.

Selecting the number of hidden neurons involves a heuristic approach. After testing a range of hidden neurons, it was found that the difference in the final results was negligible, and, therefore, the MATLAB Neural Network Toolbox's default value of 10 hidden neurons was used. The TDL of 96 were used throughout this work. The transfer functions selected for the hidden layer and the output layer are the hyperbolic tangent and linear function, respectively. Therefore, to be consistent with the transfer function being used, the input data was scaled between -1 and +1. Yet, the data was scaled back to the original dimensions (0 to 1) after the network being processed.

In this work, the Levenberge Marquardt algorithm was the algorithm used for every training process and the number of epochs is set to a maximum of 1000 (J Nocedal et al ,1999). Moreover, the number of training interactions is defined automatically as the early stopping principle was applied.

5. Simulation & Results:

A series of analysis has been done on the input data to examine which model can be used to for prediction .for this we plotted daily data of GHI for the last 3 years and observed the pattern. For example when we plot the data for two dates as given in figure 6a & 6b shows how uncertain at various conditions like cloud cover, ambient tempearture and average wind speed.



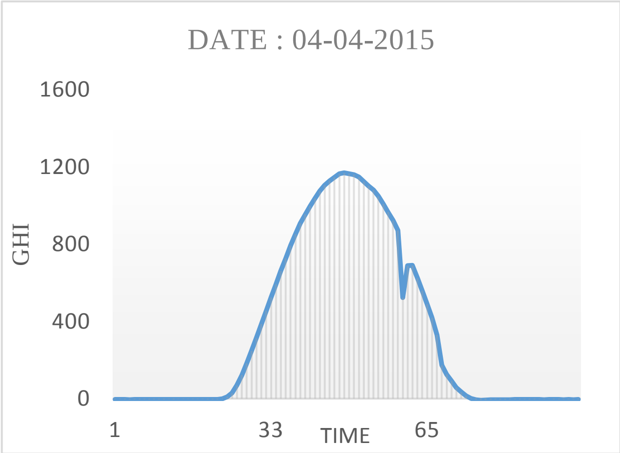


Figure 6: a) Solar Radiation on 01-04-2015 **b)Solar Radiation on 04-04-2015**

From the above figure we can see that the difference is high even though both days can be treated as sunny day. Similarly when we plot the trend of GHI for a summer season for a same time can be seen in the Figure 7, which shows there exists a huge variation which need to be modelled for prediction.

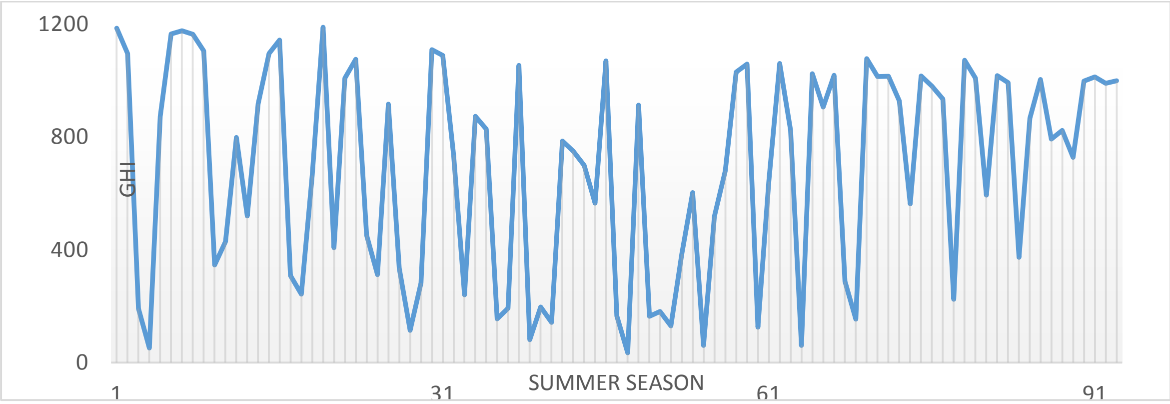
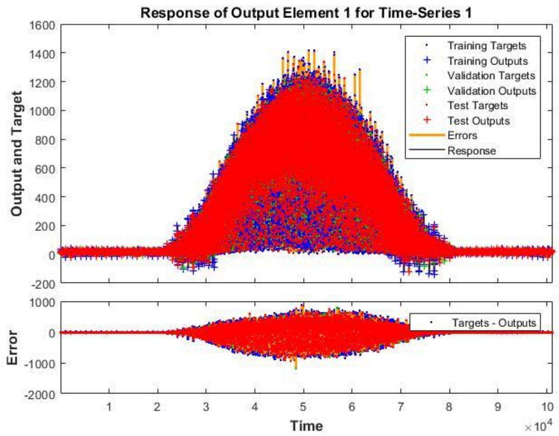


Figure 7: Variation of Solar Radiation of the 49th Time Block in the Summer Season

Since the variation is highly uncertain and the factors influencing the Solar Radiation are mainly exogenous in nature so, we used this data as an input for the NARX model and trained and the outputs are being analysed as given below in figure 8.From the figure 8 we can see that the error increases mainly when the radiation is above certain threshold level and it increases along with increase in magnitude along with solar radiation value.



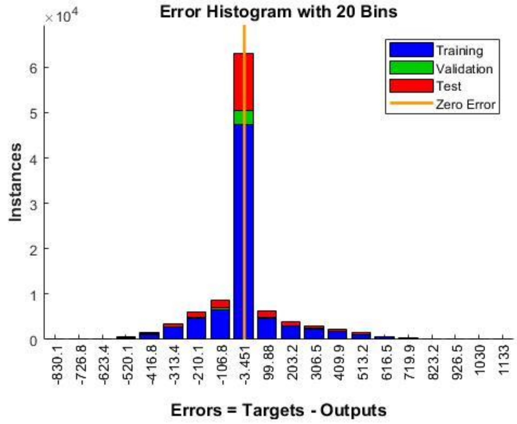


Figure 8: Response of Output w.r.to Time Series **Figure 9: Error Histogram**

From the figure 9, which shows the error distribution of the predicted model, from this we can say that 88% of the samples are predicted with accuracy of +/-100 W/m^2 which is of reasonably good accuracy.

Finally the RMSE and the R of the NARX model discussed above is 10.36% and 84.1%.

When we divide the data w.r.to various different seasons (divided w.r.to calendar year) and analyse independently we would be getting the results as shown in the Figure 10. In this we can see that the model prediction is reasonably fair in the autumn and winter season, which can be attributed to the low level of radiation availability in that season. In summer the model prediction is not that fair due to high level of radiation and cloud cover which can be attributes for the increase of RMSE % .

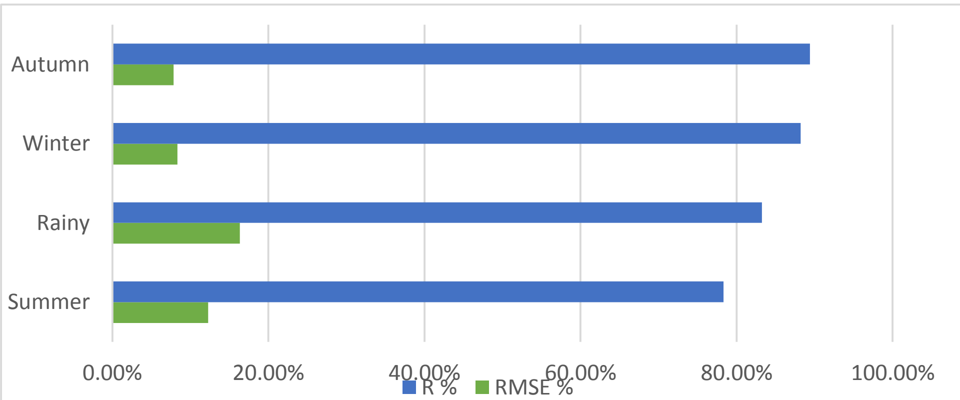


Figure 10: Comparison of the model w.r.to data fed for different seasons

6. Conclusion:

A NARX model for solar power forecasting was developed. Several input combinations of different seasons and different time blocks allowed the determination of the relevant parameters to the forecast performance of a PV system. In general the prediction of Solar GHI is fairly good. In our country if the parameters of cloud cover, wind speed available with a good resolution in the project areas, it can be fed to the model to obtain the Solar Radiation predicted and in turn used as an input for PV output forecasting depending on the Physical Model of the PV plant in intraday & day ahead basis.

The model can be improved further by introducing the advanced classification algorithms, introduction of various other meteorological inputs and the better modelling of the physical system will make this as a Hybrid model and in turn the forecasts can be improved further.

7. References:

1. Yang, H. T., Huang, C. M., Huang, Y. C., & Pai, Y. S. (2014). A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output. IEEE transactions on sustainable energy, 5(3), 917-926.

2. Yang, D., Gu, C., Dong, Z., Jirutitijaroen, P., Chen, N., & Walsh, W. M. (2013). Solar irradiance forecasting using spatial-temporal covariance structures and time-forward kriging. *Renewable Energy*, *60*, 235-245.

3. Inman, R. H., Pedro, H. T., & Coimbra, C. F. (2013). Solar forecasting methods for renewable energy integration. *Progress in energy and combustion science*, *39*(6), 535-576.

4. Diagne, M., David, M., Lauret, P., Boland, J., & Schmutz, N. (2013). Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renewable and Sustainable Energy Reviews*, *27*, 65-76.

5. PVCROPS, Estimation of the PV power that can be integrated in current EU networks. Deliverable D5.5. October 2015

6. Lorenz, E., Hurka, J., Karampela, G., Heinemann, D., Beyer, H. G., & Schneider, M. (2007). Qualified Forecast of ensemble power production by spatially dispersed grid-connected PV systems. *Measurement*.

7. S. Haylin, Neural Networks: a Comprehensive Foundation, Prentice Hall International Inc, Hamilton, Ontario, Canada, 1999

8. J. Nocedal, S.J. Wright, Numerical Optimization, Springer, New York, 1999.

Author:

MANISH KUMAR TIKARIHA, working as DGM(Operations) in NSPCL , Bhilai.He has completed his Post-Graduate Programme in Management from IIM Raipur, currently he is pursuing Executive Fellowship Programme from IIM Raipur in the area of Business Environment & Economics.

GUPTA K.V.S.S, presently pursuing EPGDM from NTPC School of Business. Graduated in 2006 from JNTU-Hyderabad in Electrical &Electronics Engineering. I am working as a Control & Instrumentation Engineer in NSPCL-Bhilai for the last 11 years. My interests include Data Analytics in Energy sector for a betterment of existing systems,