**LITERATURE REVIEW**

**PAPER TITLE:**

**Wind speed forecast using random forest learning method**

Abstract— Wind speed forecasting models and their application to wind farm operations are attaining remarkable attention in the literature because of its benefits as a clean energy source. In this paper, we suggested the time series machine learning approach called random forest regression for predicting wind speed variations. The computed values of mutual information and auto-correlation shows that wind speed values depend on the past data up to 12 hours. The random forest model was trained using ensemble from two weeks data with previous 12 hours values as input for every value. The computed root mean square error shows that model trained with two weeks data can be employed to make reliable short-term predictions up to three years ahead.

**INTRODUCTION**

Surface wind plays a crucial role in the economic development of a country as it is widely recognised as a clean, inexpensive, inextinguishable alternate source of energy. All over the world, generation of electricity from the wind has steadily been increased over the last few years, and it is estimated that, at the end of 2016 total installed wind capacity is 487 GW, and in a moderate scenario by 2020 it will reach 792 GW . In a local scenario, the total nationwide installed wind power capacity by the end of 2015 were 47 GW providing almost 67% of renewable energy connected to grid. India is also expecting 125 GW of installed capacity by 2030 with the aim of 82 million tonnes of reduction in CO2 emissions.

**RANDOM FOREST METHOD**

Random forest is a non-parametric ensemble based learning technique used for both classification and regression problem and is first suggested by Leo Beriman [12]. It is an extended version of decision tree algorithm [13] which works on a set of rules and the possible outcomes to form a tree-like structure. For an incorrect rule adds the impurity to the subsequent nodes, a high risk of error propagation is always associated with decision trees. Random forest algorithm eliminates error diffusion property inherent in decision trees by constructing multiple G. V. Drisya et al. / International Journal on Computer Science and Engineering (IJCSE) ISSN : 0975-3397 Vol. 9 No.06 Jun 2017 362 decision trees. Random samples of given data set are generated and fed to several tree-based learners to form a random forest. Splitting condition for each node in a tree is based on only the randomly selected predictor attributes which lower the error rate by avoiding the correlation among the trees. In short, random forest algorithm is an extended version of an ensemble learning method called bagging in which averaging of trees is done cleverly [14]. The successful application of random forest regression algorithm has already been reported in many fields like cheminformatics, speech recognition, bioinformatics, classification and prediction in ecology, analysis of complex remote sensing datasets etc

**RESULTS AND DISCUSSIONS**

The main objective of this paper is to investigate the performance of random forest regression model to forecast wind speed measurements. We also assess the model prediction error as time separation between train and test data increases. The data we used in this work is the measurements wind speed at 10 minutes interval at height of 80 metres at location latitude 34.98420o longitude -104.03971o from January 1, 2004 to December 31, 2016. As a first step, since the random forest regression algorithm uses vector input for model building, a matrix of time dependent chunks of data is generated using time delay information. The dependency of a value in a time series on the previous values can be estimated by autocorrelation function. Another method of estimating the period of dependence is the calculation of mutual information between delayed time series .

After training and validation of the model, we obtained 10 minutes ahead prediction for the test data set. We obtained one step ahead predictions of 2016 data points at an interval of two weeks from second month onwards upto the third year. Predictions from 4032 and 150001 data points are given in Fig. 3. Note that each forecast value was obtained using previous 72 original data points. That is, we utilized only the data of first month for training and validation of the total duration three year of the given wind speed time series. The 1-step ahead (10 minutes) prediction throughout test data of 35 months shows similar accuracy as in Fig. 3. It may be noted that at peak values model tend to underestimate the speed. The support vector prediction computed by D. Liu [9] also shows a similar trend.

**CONCLUSION**

This paper presented the results of random forest regression approach for modelling and prediction of wind speed variations. The analysis shows that with the proper time delay embedding matrix random forest algorithm offers a reliable short-term prediction with two weeks training data. The model trained with two weeks data shown remarkable prediction accuracy for test data upto three years ahead of training data set in time. The root mean square error calculated for different prediction horizons at different points show not much variation in the accuracy of the model. From the results presented here, it is clear that random forest method is a suitable candidate for an effective wind speed prediction and thereby helping an efficient, cost effective wind energy management.

**PAPER TITLE:**

**Different Models for Forecasting Wind Power Generation: Case Study**

**Abstract:**

Generation of electric energy through wind turbines is one of the practically inexhaustible alternatives of generation. It is considered a source of clean energy, but still needs a lot of research for the development of science and technologies that ensures uniformity in generation, providing a greater participation of this source in the energy matrix, since the wind presents abrupt variations in speed, density and other important variables. In wind-based electrical systems, it is essential to predict at least one day in advance the future values of wind behavior, in order to evaluate the availability of energy for the next period, which is relevant information in the dispatch of the generating units and in the control of the electrical system. This paper develops ultra-short, short, medium and long-term prediction models of wind speed, based on computational intelligence techniques, using artificial neural network models, Autoregressive Integrated Moving Average (ARIMA) and hybrid models including forecasting using wavelets. For the application of the methodology, the meteorological variables of the database of the national organization system of environmental data (SONDA), Petrolina station, from 1 January 2004 to 31 March 2017, were used. A comparison among results by different used approaches is also done and it is also predicted the possibility of power and energy generation using a certain kind of wind generator.

1. **Introduction**

The evaluation of wind potential in a region requires systematic data collection and analysis on wind speed and regime. Generally, a rigorous assessment requires specific surveys of the region where the wind farm will be placed [1–3]. There are three major markets for the field of global wind power generation: Europe, USA and China [4]. Wind energy penetration levels continue to rise, led by Denmark with a 40% use of this energy, followed by Uruguay, Portugal and Ireland with over 20%; Spain and Cyprus with about 20%; Germany with 16%; and the major markets of China, the US and Canada with 4%, 5.5% and 6% wind energy, respectively. The forecast of five years ahead is almost 60 GW of new wind power installations in 2017, rising to an annual market of 75 GW by 2021, and an accumulated installed capacity of more than 800 GW by the end of 2021 [5]. Wind energy is a clean and renewable alternative for the production of electric energy, presenting great social acceptance [6]. In the social feature, wind power plants do not cause major environmental impacts such as in hydroelectric plants and allow the compatibility between the production of electricity from the wind and the use of land for livestock and agriculture.

1. **Literature Review**

Wind Power Generation Potential The potential of electric energy produced from wind generation is obtained through the kinetic energy of the wind, which is converted into mechanical energy by a process that turns the wind force into a torque that acts on the rotor blades. The amount of energy generated by winds is a function of their speed (v) and mass (m) and is given by the kinetic energy equation [11,12]; thus, it is very important to make a good wind speed prediction. The power available in the wind, however, cannot be fully utilized by the wind turbine for the generation of electricity. According [7], to take into account this physical characteristic, an index called power coefficient Cp is introduced, which can be defined as the fraction of the available wind power that can be extracted by the rotor blades.

The potential of electric energy produced from wind generation is obtained through the kinetic energy of the winds, which is converted into mechanical energy by a process that turns the wind into torque acting on the rotor blades. According to [11], the power curve regions can be described as follows:

• Optimum constant Cp region, where increasing power with increasing wind speed;

• Limited power region, generating a constant power, even in higher winds, by decreasing the Cp rotor efficiency; and

• Region of power shutdown, where power generation is decelerated to zero, and wind speed approaches the cut-out limit.

In, it is emphasized that wind speed prediction plays a vital role in the management, planning and integration of the energy system. In previous studies, most forecasting models have focused on improving the accuracy or stability of wind speed prediction. However, for an effective forecast model, considering only one criterion (precision or stability) is insufficient. In [16], a new design methodology to predict wind farm energy production by means of a spiking neural network-based system is developed. Authors established that the calculation of the flow around wind turbines is a very complicated issue. They affirmed that turbine wakes are responsible for important power losses in wind farms and that randomness of wind speed and unexpected variations of wind speed may increase operating costs of the electricity grid as well as set potential threats to the reliability of electricity supply. A synergetic neural network (SNN)-based model for the prediction of wind farm energy production is developed. This model performs a prediction of the energy produced by one wind turbine of the wind farm, by using the wind speed and direction data coming from the three anemometric towers, during the whole day. This is a very useful model for analyzing turbines inside of a farm. Authors demonstrate that this model could accurately predict the energy produced by each wind turbine of the wind farm by means of testes conducted with experimental data that show low values of median absolute deviation (MAD), mean absolute percentage error (MAPE) and root mean square error (RMSE). In this work, the focus is limited to predictive models of time series with the use of ARIMA filters, as well as the use of neural networks, however, it will also be analyzed the auto regressive (AR) models, moving average model (MA) and auto regressive moving-average (ARMA) model to enable a better understanding of the ARIMA Model.

* 1. Types of Wind Energy Forecast

There are different methods for predicting the wind power to generate. These methods are classified according to time scales and according to different methodologies that are available in the literature. Time scales and methods for predicting wind power or energy, combining the literature can be divided into four categories [17–19]:

• Ultra-short-term forecast: From a few minutes to 1 h ahead.

• Short-term forecast: From one hour to several hours ahead.

• Medium term forecast: From several hours to one week ahead.

• Long-term forecast: From one week to one year or more ahead.

Wind Speed Prediction Models Wind forecasting models can be broadly classified into the following three categories:

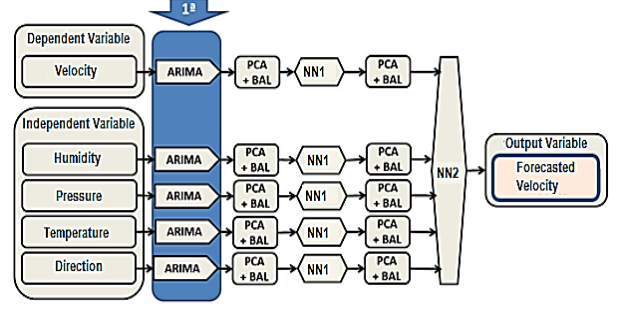
1. physical model;
2. statistical and computational model; and
3. hybrid model

The artificial intelligence approach belongs to the statistical approach. The essence of the artificial intelligence approach is to establish the relationship between input and output by artificial intelligence methods, rather than using the analytical method. The model described in this form is usually a non-linear model. Many methods of artificial intelligence are better than conventional methods and have a good perspective of development .

* 1. Statistical Models and Artificial Neural Networks

Statistical models are easy to use and cheaper to develop compared to other models. Basically, statistical methods use the previous history of wind data to perform a forecast over the next few hours, they are good for short periods of time. The disadvantage of this method is that prediction error increases as time forecasting increases, i.e., statistical time series and methods of neural networks are primarily intended for short-term predictions [23,24]. According to [25], the sub classification of this approach can be defined as: models based on time series and methods based on neural networks. These forecasting methods are generally used for small forecast horizons (mesoscale), because, in these horizons, the correlation between the velocities of the winds, and consequently the generation, are greater. The statistical models most disseminated by researchers include: auto regressive (AR), auto regressive moving average (ARMA), and auto regressive integrated moving average (ARIMA).

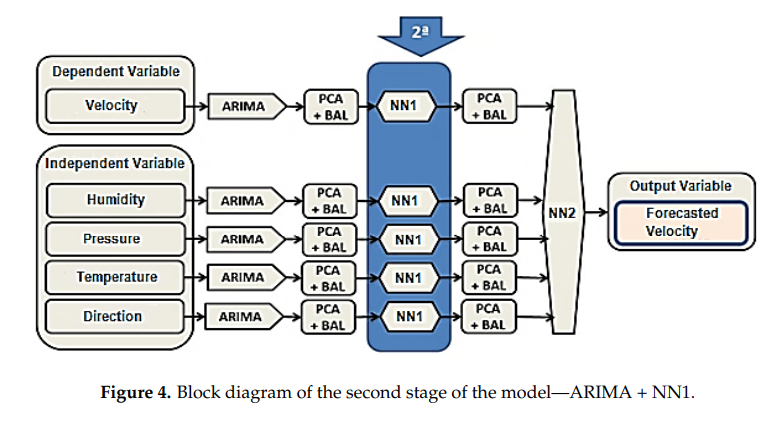
3.2. ARIMA Model The first step of the proposed algorithm illustrated in Figure 3 is the ARIMA model (Auto Integrated Regressive of Moving Average), which results from the combination of three filters: the AR Energies 2017, 10, 1976 9 of 27 component, the Integration filter (I), and the MA component. The representation of this model is done



through ARIMA notation of order (p, d, q). An ARIMA (1, 2, 0) representation indicates an order 1 for the AR (Self-Regressive) component, order 2 for component I (Integration or differentiation) and the last 0 for the MA, where: p is the number of seasonal auto-regressive terms; d is the number of seasonal differences; and q is the number of seasonal media moving. Energies 2017, 10, 1976 9 of 26 differentiation) and the last 0 for the MA, where: p is the number of seasonal auto-regressive terms; d is the number of seasonal differences; and q is the number of seasonal media moving. Figure 3. Block diagram of the first stage of the model—ARIMA. Speed is classified as a dependent variable because its predicted value in ARIMA + NN1 + NN2 depends on the values of the variables humidity, pressure, temperature and direction.

3.3. ARIMA + NN1 Model

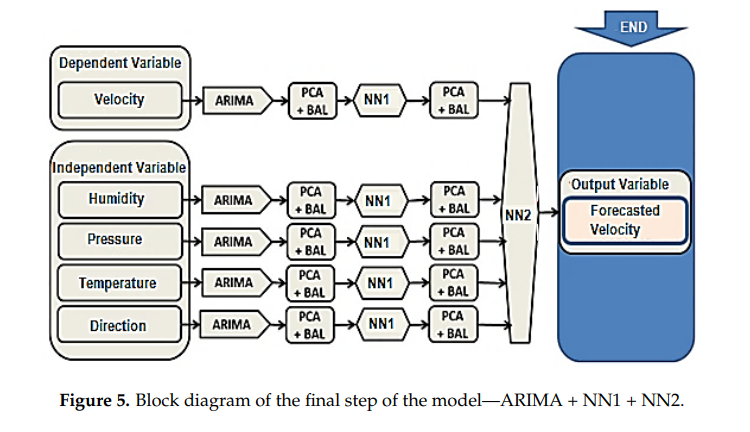
The second step of the model illustrated with the block diagram in Figure 4 is performed in the first Neural Network—NN1. This step is done to predict explanatory variables and it uses ARIMA results as input variables, which are reduced through the principal component analysis (PCA), which finds linear combinations of the input fields reducing the components for using the main variables.



The NN1 presents eight neurons in the input layers, two neurons in the hidden layer and one neuron in the output layer, configuration used for all variables. The back propagation error training algorithm was used, which adjusts the network weights in order to minimize the error between the actual values and the predicted outputs. Data partitioning was 80% for training and 20% for testing. The stopping criterion used is the maximum training time per model. The network was trained with sigmoidal tangent activation function for all neurons. The network follows a standardized programming with multilayer perceptron (MLP) with the topology 8-2-1, logistic activation function and back propagation algorithm. The network uses the data for each respective horizon, i.e., 180 for minutes, 72 for hours, 21 for days, 12 for weeks, 38 months, and 38 for years by applying the extrapolation in the data. These delays signify three cycles of each representation, and project one step forward, the recursive network re-inserts each projection at the input of the MLP and does that repetition automatically 20 times. In this step, the ARIMA + NN1 speed prediction is made and all values are analyzed based on errors, standard deviation and linear correlation.

3.4. ARIMA + NN1 + NN2 Model

The final step of the algorithm is represented by the block diagram in Figure 5. In this step, the final speed prediction is made.



e NN2 uses the outputs of the ARIMA + NN1 model as input to optimize the results, which adjusts the weights of the neural network in for minimizing the error between the actual values and the predicted outputs. Data partitioning was 80% for training and 20% for testing. The network follows a standardized programming with MLP with the topology of 11 neurons in the input layer, eight neurons in the hidden layer and one neuron in the output layer, logistic activation function and back propagation algorithm. It uses values for each respective horizon; that is, 180 for minutes, 72 for hours, 21 for days, 12 for weeks, 38 months, and 38 for years by applying extrapolation to the data. These delays signify three cycles of each representation. They project one-step forward; the recursive network re-inserts each projection at the input of the MLP and do this repetition automatically 20 times. The stopping criterion is the maximum training time per model. The network was trained with sigmoidal tangent activation function for all neurons. 3.5. Neural Networks Model The model of neural networks was used to compare the results obtained by the proposed hybrid model; the configuration of the model has as input the environmental data of real values described as input variables. The network follows a standardized programming with MLP with the topology formed with nine neurons in the input layer, seven neurons in the hidden layer and one neuron in the output layer, logistic activation function and backpropagation algorithm. It uses values for each respective horizon, that is, 180 for minutes, 72 for hours, 21 for days, 12 for weeks, 38 months, and 38 for years by applying extrapolation to the data. These delays represent three cycles of each representation, and project one step forward, the recursive network re-inserts each projection at the input of the MLP and does this repetition automatically 20 times. 3.6. Forecast of Wind Speed and Generated Power The final objective of this work is to forecast the wind speed to predict the generated power. For a given wind speed, the power generated depends on the type of generator to use. The wind turbine Figure 5. Block diagram of the final step of the model—ARIMA + NN1 + NN2. The NN2 uses the outputs of the ARIMA + NN1 model as input to optimize the results, which adjusts the weights of the neural network in for minimizing the error between the actual values and the predicted outputs. Data partitioning was 80% for training and 20% for testing. The network follows a standardized programming with MLP with the topology of 11 neurons in the input layer, eight neurons in the hidden layer and one neuron in the output layer, logistic activation function and back propagation algorithm. It uses values for each respective horizon; that is, 180 for minutes, 72 for hours, 21 for days, 12 for weeks, 38 months, and 38 for years by applying extrapolation to the data. These delays signify three cycles of each representation. They project one-step forward; the recursive network re-inserts each projection at the input of the MLP and do this repetition automatically 20 times. The stopping criterion is the maximum training time per model. The network was trained with sigmoidal tangent activation function for all neurons.

1. Result Analysis

The results obtained will be shown for each forecast universe as described above. In the case of wind speed, there were used symmetric daubechies wavelets combined with ARIMA model for the forecast.

4.1. Ultra Short Term Forecast—CPU (Minutes) The data used have a base containing 7200 rows, and a total of 36,000 data; such amount is equivalent to a universe of five days. The multi-step ahead results show that some of the used models have a significant loss of precision, the higher the prediction step, the lower the precision. The ARIMA model already starts with very large errors, the absolute mean error for example in 5-min steps is 0.795 and the MAE percentage is 15.024%. For 20-min steps, the absolute mean error becomes 1.182, and the mean absolute percentage error reaches 20.591%. However, because the ARIMA model returns the prediction values with one-step delay, it will always be worse than the model with neural networks. The ARIMA + NN1 + NN2 model in the step forecast or 5 min, obtained an absolute mean error response of 0.199 and an absolute mean error response of 3.620%. For prediction of 20 min steps, the absolute mean error goes to 0.308 and the mean absolute percentage error goes to 5.305% this result is superior to the other models, confirming its performance.

**4. Conclusions**

• Accurate and reliable wind speed prediction is vital for wind farm planning and operational planning for electrical networks. To improve the accuracy of wind speed prediction, many forecasting approaches have been proposed; however, these models typically do not account for the importance of data pre-processing and are limited by the use of individual models.

• Achieving accurate forecasts of wind speed and power is still a critical problem. Since wind power is proportional to the wind speed cubed, the wind power potential assessment is summarized as wind speed prediction.

• There are many models and their variants for predicting wind speeds, both simple and hybrid, but none of them cover the full range of forecasting possibilities from ultra-short-term forecasts to several years ahead.

• To forecast the wind speed and the possible power to be generated, four prediction models were used: ARIMA ARIMA + NN1 ARIMA + NN1 + NN2 NEURAL NETWORKS The four types of forecast were made according to the revised literature: Ultra-short-term forecasting Short-term forecasting Medium-term forecasting Long-term forecasting

• Of the models used, the hybrid model of ARIMA + NN1 + NN2 was the one that presented the best results with the smallest errors in the prediction of wind speed in all forecast horizons, as can be seen in the table and graphs presented in this paper. For the prediction for a five-step forward, the best response to the MAE was obtained for the hour horizon with a result of 0.180, and the worst response obtained was for the weeks horizon with a response of 0.292. The best response for the RMSE was obtained for the hour horizon with 0.403, and the worst response was for the weeks horizon with an error of 0.654. For the mean absolute percentage error (MAPE) responses, the best response was obtained for the month’s horizon with 2.329%, and the worst response for the weeks horizon with 3.948%. For prediction of 20 steps forward, the best response to absolute mean error (MAE) was obtained for the hour horizon with a result of 0.189, and the worst response was for the week’s horizon with an error of 0.413. The best response for RMSE was obtained for the hour horizon with 0.843, and the worst response was for the week’s horizon with a response of 1.848. For the mean absolute percentage error (MAPE) responses, the best response was obtained for the hour horizon with 2.571%, and the worst response for the week’s horizon with 5.796%. The use of hybrid models is shown to be more efficient by considering the linear and non-linear characteristics of the modeled signals. The higher the forecasting step, the lower the guarantee of results, but the market needs quick responses and the trend of this response speed feature is increasing in the world scenario. In addition to an efficient model for the need for demand, the database is crucial for results with optimum accuracy. With the prediction of wind speed, it is possible to predict the wind generation of the analyzed region, depending on the generators to be installed. These results of wind speed prediction and therefore of wind power generation potential are unprecedented in the literature, more so, with the combination of models and term predictions, which is undoubtedly a novelty. The use of wavelets does not improve the wind speed forecasting; ARIMA + WAVELETS forecasting results are very closed to ARIMA forecasting ones.

**References**

1. Nagababu, G.; Simha, R.R.; Naidu, N.K.; Kachhwaha, S.S.; Savsani, V. Application of OSCAT satellite data for offshore wind power potential assessment of India. Energy Procedia 2016, 90, 89–98. [CrossRef]

2. Gao, X.; Yang, H.; Lu, L. Study on offshore wind power potential and wind farm optimization in Hong Kong. Appl. Energy 2014, 130, 519–531. [CrossRef]

3. Adaramola, M.S.; Agelin-Chaab, M.; Paul, S.S. Assessment of wind power generation along the coast of Ghana. Energy Convers. Manag. 2014, 77, 61–69. [CrossRef]

4. Kaplan, Y.A. Overview of wind energy in the world and assessment of current wind energy policies in Turkey. Renew. Sustain. Energy Rev. 2015, 43, 562–568. [CrossRef]

5. GWEC. Strong Outlook for Wind Power. 2017. Available online: http://www.gwec.net/global-figures/ graphs/ (accessed on 2 June 2017).

6. Manwell, J.F.; McGowan, J.G.; Rogers, A.L. Wind Energy Explained: Theory, Design and Application; John Wiley & Sons: Hoboken, NJ, USA, 2010.

7. Dutra, R. Energia Eólica–Princípios e Tecnologia; Centro de Referência para Energia Solar e Eólica Sérgio de Salvo Britto: Rio de Janeiro, Brazil, 2011.

8. Fundação Estadual do Meio Ambiente (FEAM). Utilização da Energia Eólica no Estado de Minas Gerais: Aspectos Técnicos E O Meio Ambiente. 2013. Available online: http://www.feam.br/images/stories/ materia/comunicado\_tecnico2\_gemuc.pdf (accessed on 23 February 2017).

9. Wang, C.; Zhang, H.; Fan, W.; Ma, P. A new chaotic time series hybrid prediction method of wind power based on EEMD-SE and full-parameters continued fraction. Energy 2017, 138, 977–990. [CrossRef]

10. Wang, D.; Lou, H.; Grunder, O.; Lin, Y. Multi-step ahead wind speed forecasting using an improved wavelet neural network combining variational mode decomposition and phase space reconstruction. Renew. Energy 2017, 113, 1345–1358. [CrossRef]