A Review of Wind Speed & Wind Power Forecasting Techniques

Harsh S. Dhiman
Department of Electrical Engg
Adani Institute of Infrastructure Engineering
Ahmedabad, Gujarat
Email: harsh.dhiman@aai.ac.in

Dipankar Deb
Department of Electrical Engg
Institute of Infrastructure Technology
Research and Management
Ahmedabad, Gujarat
Email: dipankardeb@iitram.ac.in

Abstract—Forecasting a particular variable can depend upon temporal or spatial scale. Temporal variations that indicate variations with time, reflect the stochasticity present in the variable. Spatial variation usually are dominant in climatology and meteorology. Temporal scale for a variable can be modeled in terms of time-series. A time series is a successively ordered sequence of numerical data points, and can be taken on any variable changing with time. Wind speed forecasting applications lie majorly in the area of electricity market clearing, economic load dispatch and scheduling, and sometimes to provide ancillary support. Thus, a proper classification based on the prediction horizon i.e. the duration of prediction becomes important for various transmission system operators.

I. INTRODUCTION

Sustainable energy sources lead to reduction in carbon footprint and thus increase the reliability of a system [1], [2], [3], [4]. With the spurt in the installation of renewable energy sources, the fossil fuel usage is put into restricted use. Most of the industries in power sectors prefer renewable energy generation owing to its negligible carbon footprint. With wind available in abundant form, tapping power from wind is a specialized task. Considering wind as a stochastic variable, its accurate prediction can yield numerous benefits to the plant operators. Prediction involves errors, similarly in this case the error processing of forecasted wind speed/power and actual wind speed/power plays a crucial role in selecting the appropriate forecasting algorithm.

Wind forecasting plays an important role when it comes to clearing day ahead market scenarios. Given there is a market situation to be cleared, an accurate wind forecasting scheme is helpful in such situations. Wind forecasting schemes are broadly categorized as (i) Weather based prediction methods (ii) Statistical or time series based prediction methods [5]. While we consider weather based prediction models, the wind forecast accuracy depends highly on the topology of the land where the wind turbines are erected. Given the topology of the land, wind speed measurements at an appropriate height from the ground, the temperature of the ambient air, air pressure etc hold important factors to take into consideration. On the other hand statistical methods solely depend on the past measurements of the wind speed to predict the future values.

This article draws down attention towards the recent wind energy forecasting schemes that have been in use for accurately predicting the wind power. It also lists down several conventional statistical prediction models like persistence algorithm, ARMA and ARIMA models. Given the time scale for the forecasting, the methods can be classified based on the basis of horizon span i.e. very short-term forecasting, short-term forecasting, long-term forecasting and very-long term forecasting. Forecasting plays an important role not only in predicting accurate values of wind power but also to clear day-ahead electricity markets. It was observed that numeric weather prediction methods which are more suitable to predict wind speed for long-term durations could yield up to an average saving of 20 percent in fossil fuel [5]. Since the wind power can provide the ancillary support to the nearby power stations, forecasting accurate wind power can also help to stabilize any abnormal operation of the interconnected power plant.

This paper provides full insight to the recent and advance wind energy forecasting schemes that can lead to optimal wind dispatch and aim to improve the economies of scale of the large power systems. The paper is organized in five sections. Section I gives an introduction to the forecasting methods, section II briefs about classification of forecasting schemes based on prediction horizon. In section III various statistical models for forecasting are discussed with persistence and ARIMA models forming the benchmark for comparing accuracy. Section IV gives insight to advance machine learning forecasting methods along with hybrid methods as well. Section V discusses various conclusions.

II. CLASSIFICATION BASED ON PREDICTION HORIZON

Wind speed forecasting applications lie majorly in the area of electricity market clearing, economic load dispatch and scheduling, and sometimes to provide ancillary support. Thus a proper classification based on the prediction horizon i.e. the duration of prediction becomes important for various transmission system operator’s (TSO’s). The time-scale classification of the wind energy forecasting methods is given as follows:

- Very short-term prediction (few seconds to 30 min)
- Short-term prediction (30 mins to 6hrs)
- Medium-term prediction (6hrs to 24hrs)
- Long-term prediction (24hrs to 72 hrs)
• Very long-term prediction (72 hrs and longer)

The above classification not only simplifies the study but also helps to choose the accurate method depending on the type of its application. Majority of the time one may find that short-term and medium term prediction methods are more in use due to its accuracy and robustness. Various performance parameters that are being used to evaluate and compare forecasting methods [5]. Given the prediction horizon the error between the forecasted value and actual value of wind speed/power can be quantified and compared using some standard error definitions such as mean absolute percentage error (MAPE), mean absolute error (MAE), mean squared error (MSE) and root mean square error (RMSE). The mathematical expression for the mean absolute percentage error (MAPE), mean absolute error (MAE), mean squared error (MSE) is given as follows

\[
\text{MAPE} = \frac{100}{N} \sum_{i=1}^{N} \frac{|p_f - p|}{p_f} \tag{1}
\]

\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |p_f - p| \tag{2}
\]

\[
\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (p_f - p)^2 \tag{3}
\]

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (p_f - p)^2}{N}} \tag{4}
\]

Where in the above equations (1-4), N is the prediction horizon, \(p_f\) is the forecasted wind power and \(p\) is the actual wind power. MAPE criterion for statistically analyzing the forecast accuracy has proven out to be a better parameter. The following subsections explain the various forecasting schemes (numeric weather or statistical model) that are being adopted depending on the prediction horizon, also we look the various performance evaluation parameters of the schemes.

A. Very Short-term Forecasting Schemes

The time-scale classification proves out to be a brownie point in terms of application of these forecasting algorithms. For any forecasting method to be accurate its performance at different sites with varied atmospheric conditions plays a crucial role in forecasting future values. Thus a parameter must be assigned that not only calculates how accurate the method is but it also determines its fitness over certain conditions. Such type of optimization studies are often done to choose the accurate forecasting method. Very short-term forecasting methods, as its name suggests predicts the future values up to a short span of time. Usual time-scale followed for this method is from a few seconds to 30 minutes. Of many methods described in literature, like spatial correlation where the 1-second ahead forecasting is done [6]. Artificial neural network (ANN)-Markov chain (MC) model [7] where a variable set of 175 min is taken and wind speeds for the next 7.5s ahead are predicted. Bayesian structural break model [8] are used to predict 1 min and 1 hr ahead forecasts. Data mining approach [9]. Apart from these methods, several intelligent learning methods like support vector machines (SVM), hybrid methods combining Empirical Wavelet Transform (EWT) with neural networks (NN’s) also form core of the very short-term wind forecasting. Recent advances in forecasting domain show evolution of machine learning algorithms, with decomposition forecasting algorithms (DFA) being most in use. The essence of a decomposition algorithm lies in breaking the time-series of a variable i.e. wind speed/power here and analyzing individual units for the forecasting and combining them to obtain the resultant series. One such decomposition algorithm is used in [11] where the time-series data of wind is broken or decomposed into several units and on each unit a feature construction process is performed and those with best features are chosen for prediction. The prediction can be carried out with standard ARIMA model of order (p,d,q) or artificial neural networks (ANN) or support vector machines (SVM). Further a forecasting technique based on Hilbert-Huang transform (HHT), that uses the decomposition technique for the non stationary and non-linear models, is also used in [12].

B. Short-term Forecasting Schemes

Short-term wind forecasting is the most used forecasting category with many day-ahead markets need to clear the market scenarios by the end of the day. Among these the most used forecasting methods are a combination of two or more machine learning methods combined with a time-series model (AR, MA, ARMA,ARIMA,ARMAX). ARMAX stands for autoregressive exogenous moving average, is a non-linear model that captures all the uncertainties that are related to the stochastic nature of the wind. Among various statistical models ARMA model is a popular forecasting method. In ARMA, AR stands for auto-regressive and MA stands for moving average. Thus a combined autoregressive (of order p) and moving average (of order q) forms a ARMA (p,q) model. The order of ARMA models i.e. \(p\) for AR and \(q\) for MA denote the lag between present and past values of the variable under test. Thus a generic ARMA model can be mathematically expressed as follows:

\[
y_t - \mu_x = \alpha_1(y_{t-1} - \mu_x) + \alpha_2(y_{t-2} - \mu_x) + ... + \alpha_p(y_{t-p} - \mu_x) + \xi_t - \beta_1 \xi_{t-1} - \beta_2 \xi_{t-2} - ... - \beta_q \xi_{t-q} \tag{5}
\]

where \(\xi_t\) is an independent process with mean zero and \(\alpha_1, \alpha_2, ..., \alpha_p\) and \(\beta_1, \beta_2, ..., \beta_q\) are the parameters of the AR and MA process respectively.

In [13], a variant of ARIMA model, i.e. \(f\)-ARIMA is used to predict the day ahead wind forecasts. \(f\) stands for fractional, where the value of differencing parameter \(d \in (-0.5, 0.5)\). Results of [13] show that \(f\)-ARIMA model was 42% more efficient than persistence model. Apart from these standard ARIMA models, a hybrid model i.e. wind forecasting by wavelet transform and neural networks used in [14], to predict
the wind power for 3 hr ahead. Data from previous 12 hrs with 15 min time step was taken as input to the ANN's input layer. The original wind power series is decomposed using wavelet transform (specifically D-WT), and the resulting series is fed to the neural network where the future values are forecasted. Wavelet transform also find their use in forecasting the load and electricity prices for a power plant particularly in deregulation market [15].

Another short-term forecasting method that involves empirical mode decomposition (EMD) and feature selection was studied in [11], where the wind series was broken down into several subsequent series. Each of these series, an intrinsic mode function (IMF) was computed. The IMF of a decomposed signal represents the irregularity and frequency components of the signal. Once the series is decomposed the appropriate forecasting tool i.e. ANN or SVM is chosen to forecast the wind power values. A generic expression of EMD based decomposition can be represented as follows:

\[ y_i = \sum_{i=1}^{N} c_i(t) + r_n(t) \]  

where in the above equation \( c_i(t) \) are the different IFM’s and \( r_n(t) \) is the final sum of all the residuals.

In [16], support vector machine technique for wind prediction is used. The idea behind carrying out such procedure is to map out time-series data of any variable into higher dimension space (for e.g. hilbert space) and carrying out regression analysis. Mathematically SVM can be expressed with the following set of equations:

\[ f(x) = \sum_{i=1}^{D} w_i \phi_i(x) + b \]  
\[ R[w] = \frac{1}{N} \sum_{i=1}^{N} |f(x_i) - y_i| + \lambda \|w\|^2 \]  

In the above equations \( \phi_i(x) \) are features and \( w_i \) and \( b \) are the coefficients to be computed from the data sets. The coefficients \( w_i \) can be found out by minimizing the function \( R[w] \). Also the results of the prediction obtained from SVM were compared with multi-layered perceptron (MLP). The performance measures used in [16] were MSE and RMSE. It was found that SVM performed better than MLP as former had a mean squared error of 0.78% compared to 0.9% of latter. Many studies have also been carried out that compare the performance of MLP with standard AR process of order \( p \). Another method where a combination of wavelet transform, support vector machines and genetic algorithm was used in [17]. As mentioned in previous methods [11], a wind series is decomposed using wavelet transform. The process of genetic algorithm helps to select the input parameters for the SVM. It is necessary to have optimized input in order to choose the best candidates for the forecast. The results showed that the MAPE obtained with WT-SVM-GA was around 14.79% and that with persistence was 22.64%. Following figure shows the results of forecasting with WT-SVM-GA method [17]. In the fig 1 below it can be seen that the technique used tracks the actual data sets accurately and precisely.

C. Long-term & Very Long-term Forecasting Schemes

Long-term forecasts find their major application in the field of unit commitment decisions, maintenance scheduling etc. In [18], a hybrid method combing particle swarm optimization (PSO) and adaptive-network-based fuzzy interference system (ANFIS) i.e. PSO+ANFIS was used for forecasting 1-day ahead in the time steps of 15 min. The PSO algorithm was chosen to find out the best parameters for neuro-fuzzy systems. Another method [19], where wavelet decomposition (WD) and artificial bee colony optimization (ABCO) based relevance vector machine (RVM) is used. Here a wind signal is decomposed into different sub-series of different frequency ranges and then further the forecasting of different models is done by RVM. The kernel parameters of RVM were chosen by a meta-heuristic algorithm known as artificial bee colony optimization (ABCO). The forecast model for relevance vector machines can be mathematically expressed as follows:

\[ h_n = y(x_n, w) + \varepsilon_n \]  
\[ y(x, w) = \sum_{i=1}^{N} w_i K(x, x_i) + w_0 \]  

In the above equations \( x_n \), \( x_n \) is the input vector of the decomposed wind series, \( \varepsilon_n \) is added noise in the process, \( K(x, x_i) \) is the kernel function which can be expressed as follows:

Apart from the above mentioned forecast techniques, many numeric weather methods such as Global forecast system [20], fifth generation mesoscale model (MM5) with neural network was used. The predictions from the global forecast systems (GFS), plus the atmospheric conditions of the topography concerned are used as boundary conditions for the MM5 model. Its output is given as an input to the neural network. From the global forecast method, the physical downscaling of the wind data is done, then using the MM5 model, the statistical downscaling is achieved.
III. STATISTICAL MODELS FOR WIND FORECASTING

Wind speed/power forecasting plays an important role for a TSO in order to guarantee certain amount of power transfer to the grid. In literature among various prediction models, time-series model [21] outperform numeric weather prediction model and its general mathematical expression could be written as \( y_t = y_{t-1} + \varepsilon \) i.e. the future values of a variable \( y \), here wind speed/power in our case, depend on its past values \( y_{t-1} \) [5]. The most simple and effective method for wind speed/power prediction is persistence method where the data from the past is used to predict the variable under test. It also important to note that wind power forecasting yields better forecast results than wind speed forecasting. The wind speed is mapped into wind power by following the power-speed curve for a particular wind turbine. For any new forecasting algorithm developed, persistence method acts as benchmark for accuracy determination.

A. Regressive Models for Wind Energy Forecasting

Statistical models have proved out to be more efficient and useful when it comes to short-term and medium term forecasting. A lot of data set-points are to be needed to forecast the future values of wind speed/power. The most commonly used statistical approach is AR, MA, ARMA & ARIMA models. These models have performed decently for short-term and medium-term forecasts. ARIMA models have been in use while considering hybrid or combinational methods for short-term forecasts. In [22], linear and non-linear regressive models have been discussed for short-term forecasting. The general expression for linear auto-regressive moving average with exogenous input (ARMAX) model can be represented by the following equation:

\[
A(q)y(t) = B(q)u(t) + C(q)e(t) \tag{11}
\]

\[
A(q) = 1 + a_1q^{-1} + a_2q^{-2} + \ldots + a_pq^{-p} \tag{12}
\]

\[
B(q) = 1 + b_1q^{-1} + b_2q^{-2} + \ldots + b_fq^{-f} \tag{13}
\]

\[
C(q) = 1 + c_1q^{-1} + c_2q^{-2} + \ldots + c_lq^{-l} \tag{14}
\]

In the above equations [11][14], \( A(q), B(q), C(q) \) are the AR, exogenous and MA parts respectively. Whereas \( p, f, l \) are orders of the AR, exogenous and MA parts. Whereas non-linear regressive models [22] can be expressed as follows:

\[
y(t) = f(y(t-1), y(t-2), \ldots y(t-p), u(t-1), u(t-2), \ldots u(t-f), e(t-1), e(t-2), \ldots e(t-l))
\]

In the above equation 15, \( y(t) \) are the future values, and \( e \) represents white noise component. ARMAX models involve exogenous variable, its origin is from the meteorological changes in the wind movement [23]. The non-linear ARMAX models have been derived from various machine learning algorithms i.e. support vector machine (SVM), MSR and bagging. Support vector machines transform the data-sets from one dimension to higher dimension. Let \( x \) be the input space vector then, \( z = \phi(x) \) is the feature space vector and \( \phi \) is the function that maps \( x \) to \( z \).

In [22], linear regression models were built in MATLAB using 10-min data sets from past 15 days. The results showed that the ARMAX models with wind direction as exogenous input (for e.g wind speed at different heights, wind direction, temperature and solar radiation) performed better than rest of the models. The performance criterion used here were MAE, RMSE and MAPE. The figure 2 below shows the wind speed prediction using ARMAX(2,3,2) non-linear regression model based on SVM, where \( p = 2, f = 3 \) and \( l = 2 \) are the orders of AR, exogenous input and MA part respectively. Two datasets have been taken, one from Sotaventa Galicia Plc a wind farm experiment supported by regional autonomous government and dataset 2 from M.S Puram, Madurai which is one of the monitoring station of Centre of Wind Energy Technology (CWET), Chennai.

![Fig. 2. 10-min ahead wind speed forecast using ARMAX(2,3,2) [22].](image)

B. ARMAX based Forecasting Scheme

Autoregressive moving average exogenous input (ARMAX) based linear regressive models are often tested for wind speed/power forecasting. One such forecasting was done for Ashcroft, one of the villages in British Columbia, Canada. The historic hourly data set of first 15 days of October’17 was taken and ARMAX model was formulated for the same. The ARMAX model chosen was ARMAX(2,2,2) i.e. with reference to equation [12][14] the orders of AR, MA and exogenous part were 2,2,2 respectively. The equation below describes the discrete-time ARMAX model for the same:

\[
A(z)y(t) = B(z)u(t) + C(z)e(t) \tag{15}
\]

\[
A(z) = 1 - 0.3659z^{-1} + 0.09601z^{-2} \tag{16}
\]

\[
B(z) = \frac{0.3659}{z} + \frac{0.09601}{z^2} \tag{17}
\]

\[
C(z) = 1 - 0.03659z^{-1} + 0.03175z^{-2} \tag{18}
\]

In the above equations, \( y(t), u(t), \) and \( e(t) \) represent the output variable wind speed, exogenous variable and added white noise respectively. The MSE using ARMAX(2,2,2) was found to be 8.805%.

IV. MACHINE LEARNING BASED FORECASTING ALGORITHMS

This section gives an insight to the various machine learning based forecasting schemes that are being employed globally
for accurate wind forecasts. Machine learning methods like Artificial Neural Networks (ANN’S), Support vector machine (SVM), Empirical mode decomposition (EMD), Support vector regression (SVR) and Extended machine learning (ELM). Ensemble methods [24]. The subsections below describe each of the above mentioned methods in detail.

A. Forecasting with Artificial Neural Networks (ANN’s)

Wind speed forecasting helps in utilizing the available wind resource to its optimum capacity. This not only reduces the carbon footprint in the ecosystem but also facilitates the possibility of interconnection of power systems. In terms of forecasting, various machine learning algorithm have come up that attempt to solve the problem of non linearity in prediction models. Artificial Neural Networks or just neural networks are one such technique. Given any data set, ANN’s learn from the experience of the past and predict future values. ANN’s find their application not only in wind forecasting but also in wind turbine control [25]. The basic working of an artificial neural comes from the working of brain i.e. communication via neurons [26]. The figure 4 shows the multi-layer architecture for an ANN. It contains an input layer, hidden layer(s) and an output layer. The input layer of the ANN has a weight associated with it. Mathematically ANN’s can be expressed as follows:

\[ f(x) = \phi(\sum_{i=1}^{N} w_i x_i + b) \]  
\[ f(x) = \phi(w^T x + b) \]  
\[ \phi(g) = \frac{1}{1 + e^{-g}} \]

In the above equations, \( w_i \)’s are the weights associated with each neuron present in the input layer, \( x_i \)’s are the inputs and \( b \) is the bias term. \( \phi \) in equation (21) denotes the activation or transfer function (here sigmoid function). The output from the weighted sum of the neurons is then sent to the activation function block where a mathematical function is used to scale down the output to zero or one. Several combinations of ANN architecture’s are possible by keeping different number of hidden layers, different number of neurons in each layer and the choice of activation function. ANN’s find their application mostly in short-term wind energy forecasting [25]. The following are the hybrid systems of ANN’s
- Genetic Algorithm (GA) & Neural Networks
- Particle Swarm Optimization (PSO) & Neural Networks
- Wavelet Neural Networks (WNN’s)
- Fuzzy Neural Networks (FNN’s)

The following sections describe the application part of ANN’s to various categories of wind speed/power forecasting as mentioned in section II i.e. based on prediction horizon.

1) ANN’s in Short-term wind forecasting: Artificial neural networks are heavily used to forecast wind speed/power for a duration ranging from 30-mins to 6 hrs. Among ANN’s the most commonly used topology is multi-layered perceptron (MLP). In [14], short-term wind power forecasting for wind farms in Portugal is achieved via a hybrid method of wavelet-transform and neural networks (NN). Here first the wind speed data is broken down into sub-series via discrete-wavelet transform and later is fed to NN for the training part. As mentioned earlier the number of neurons in each layer can affect the performance, the neurons in each layer can be chosen by trail or error method. Once the training algorithm is through, the data sets are sent for learning stage where the error minimization takes place between input values and desired values. Usually backpropagation is used as learning algorithm [27]. However backpropagation is a slower technique, it is replaced by Levenberg-Marquardt algorithm. Catalao et al. [14] have compared the NNWT approach with ARIMA (1,2,1) and NN where the forecast for 3-hr ahead is done by taking historical data of previous 12-hr. The MAPE value using NNWT approach was found to be 6.97%.

2) Wind Forecasting using Recurrent Neural Networks: In [28] talks about the architecture of a recurrent neural network. Most of the neural network based forecasting schemes employ MLP architecture to forecast wind power. However in RNN, there are connections of processing elements (PE) i.e. neurons from output layer to preceding layer (hidden layer), or subsequent layers and every connection has a weight associated with it. MLP is a special case of RNN where the weight coefficients are zero. The figure 4 below gives a diagrammatic representation of recurrent neural networks. Mathematically RNN’s can be described as follows:

In [29] a comparative analysis of univariate and multivariate ARIMA models is done with RNN’s. The wind speed data was acquired from Wind Engineering Research Field Laboratory (WERFL) at Texas Tech University. The wind speed was recorded at 5 different heights (8,13,33,70,160 ft) can the
forecasting was done. It was found that the forecast accuracy improved at higher heights from ground. Various models like univariate ARIMA, univariate RNN, multivariate ARIMA and multivariate RNN were examined and it was found that RNN’s outperform ARIMA models by and large.

B. Support Vector Machines

Support vector regression (SVR) is a type of Machine learning regression which is associated with learning algorithm equipped to analyze historical data for classification and regression. SVR works on the principle of structural risk minimization (SRM) from statistical learning theory \[30, 31\]. The core idea of this theory is to construct a hypotheses \( h \) that yields lowest true error for the unseen and random sample testing data \[32\]. Apart from SVR, a universal machine intelligent technique like Artificial neural network (ANN) with applications in character recognition, image compression and stock market prediction, is studied \[33\]. Shirzad et al. have compared the performance of ANN and SVR to predict the Pipe Burst Rate (PBR) in Water Distribution Networks (WDNs) \[34\], and found ANN to be a better predictor than SVR, but generalization is not consistent with physical behavior. SVR has an advantage over ANN with respect to the number of parameters involved in training phase. The computation time is another important factor for carrying out regression analysis.

Consider a set of training data (historical data) \((x_1, y_1), \ldots, (x_n, y_n) \subset X \times \mathbb{R}\), where \( X \) denotes the input feature space of dimension \( \mathbb{R}^n \). Let \( Y = (y_1, y_2, \ldots, y_i) \) denote the set representing the training output or response, where \( i = 1, 2, \ldots, n \) and \( y_i \in \mathbb{R} \).

\( \varepsilon \)-support vector regression: This type of SVR uses an \( \varepsilon \)-insensitive loss function that intuitively accounts for sparsity similar to SVRby ignoring errors less than \( \varepsilon \).

\( \varepsilon \)-SVR aims to find a linear regressor

\[
f(x) = w^T x + b, \text{with } w \in X, b \in \mathbb{R},
\]

for prediction, where \( x \in X \) is the input set containing all the features, \( w \) is the weight coefficient related to each input \( x_i \) and \( b \) is the bias term.

The objective is to find the \( f(x) \) with maximum deviation \( \varepsilon \) from the respective feature sets while being as flat as possible. In order to achieve the flatness of the desired regressor, the square of the norm of weight vector \( w \) needs to be minimized, and the SVR problem is structured in the form of a convex optimization problem \[35\] given as

\[
\begin{align*}
\min & \frac{1}{2} \| w \|^2 + C(\varepsilon^T \chi + \varepsilon^T \chi^*), \\
\text{subject to} & \ y - w^T x - e^b \leq \varepsilon + \chi, \chi \geq 0, \\
& w^T x + e^b - y \leq \varepsilon + \chi^*, \chi^* \geq 0,
\end{align*}
\]

where \( C \) is the regularization factor that reflects the trade-off between the flatness of regressor \( f(x) \) and the maximum tolerable deviation \( \varepsilon \). The value of \( \varepsilon \) introduces a margin of tolerance where no penalty is imposed on the errors. The larger \( \varepsilon \) is, the larger are the errors. The parameter \( C \) controls the amount of influence of the error. The variables \( \chi, \chi^* \) are the slack variables introduced as a soft margin to the tolerable error \( \varepsilon \) and \( e \) is a vector of ones of appropriate dimensions \((n \times 1)\).

In machine learning regression, the problem of over-fitting persists which results in less error in training phase and high error in testing phase. Commonly used regularization techniques include \( \mathcal{L}_1 \) and \( \mathcal{L}_2 \) regularization. Mathematically, these are expressed as

\[
\begin{align*}
\mathcal{L}_1: & \ \arg \min_w \text{loss function} + \lambda \sum_{i=1}^{n} |w_i| \\
\mathcal{L}_2: & \ \arg \min_w \text{loss function} + \lambda \sum_{i=1}^{n} w_i^2
\end{align*}
\]

The formulation of SVR is diagrammatically explained in Figure 6 where the “stars” represent the support vectors, green solid line shows the SVR regressor and blue dashed lines are the hyperplanes with soft limit on the tolerance error \( \varepsilon \).

However, this is not the case always, as the feature sets might not be linearly separable. To handle such nonlinearities in the feature sets, kernel trick or often called as kernel functions are used to transform data to a higher dimensional space. After transformation via suitable mapping function \( \phi : \mathbb{R}^n \rightarrow \mathbb{Z} \), the data becomes linearly separable in the target space (high dimensional space), that is, \( \mathbb{Z} \). The inner product \( \langle w^T, \phi(x) \rangle \) in the target space can be represented by using kernel function. Kernel functions are similarity functions which satisfy Mercer’s theorem such that \( k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \), are the elements of the kernel matrix \( K \). Several kernel functions are
available in literature like linear, polynomial with degree \(d\), Gaussian, Radial Basis Function (RBF) with bandwidth of the function \(\sigma\) and exponential function.

Figure 7 illustrates the kernel trick used when the input vectors are not linearly separable. This transformation makes the computation of weights and bias vector much easier. Now, we look at the dual form of the optimization problem as

\[
\begin{align*}
\min \frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i - \alpha_i^*)^T k(x_i, x_j)(\alpha_j - \alpha_j^*) + \epsilon^T \sum_{i=1}^{n} (\alpha_i + \alpha_i^*) \\
- \sum_{i=1}^{n} y_i (\alpha_i - \alpha_i^*)
\end{align*}
\]

\[\text{s.t. } \epsilon^T \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) = 0, \quad 0 \leq \alpha, \quad \alpha^* \leq C,\]

where \(\alpha\) and \(\alpha^*\) represent positive and negative Lagrange multipliers such that \(\alpha_i \alpha_i^* = 0, \ i = 1, 2, \ldots, n\). The regressor \(f(x)\) can be written as

\[f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) k(x, x_i) + b.\]

The complexity of this regressor is independent of the dimensionality of the feature set but only depends on the number of support vectors which are nothing but the data points which separate the feature sets from each other. However, the performance of the SVR also depends on the choice of kernel function and helps in reducing the computation time of the regression. A flowchart with a step-by-step implementation of the \(\epsilon\)-SVR algorithm is described in Fig. 8.

V. HYBRID ALGORITHMS FOR WIND FORECASTING

We have seen the various forecasting schemes each with their own set of advantages and being superior to others in terms of performance metrics i.e. RMSE, MSE and MAPE. However, the focus has been shifted recently on the hybrid forecasting schemes which combine two or more prediction method and are aggregated thereafter. Hybrid methods like ANN-ARIMA [36], PSO-ANFIS [37], Wavelet transform-Neural Network (WT-NN) [14], Kalman filter-ANN (KF+ANN) [38], Wavelet-Support vector machine optimized by genetic algorithm (WT-SVM-GA) [39], Empirical mode decomposition-support vector machine (EMD+SVM) [11], Fast Ensemble Empirical mode decomposition-regularized extended machine learning (FEEMD-RELM) [40] etc.

Hybrid forecasting methods have been in trend recently owing to its additional advantages over conventional single forecasting methods. Wind being a stochastic variable, its non-stationary and non-linear characteristics cause difficulties in its prediction [41]. Also statistical models have outperformed numerical weather predictions models like meso-scale modeling with MM5 and Global Forecast System [42]. With the growing energy demand, operators these days need to ensure that their forecast errors are minimum, which not only calls for the need of accurate methods but also captures the non-linearity of the wind speed. Decomposition algorithms like empirical mode decomposition and wavelet transform [14] in combination with time-series models like ARIMA and ARMAX have been in use recently. Machine learning methods like ANN and SVM have also found to be beneficial to address the non-linearity. Empirical mode decomposition involves decomposing the original wind speed data sets into subsequent subsets, performing forecasting on each of the subsets and further aggregating the results for final forecast. Individual forecasts could be based on time-series prediction models or machine learning algorithms. In [11] the author uses decomposition algorithm on original wind speed series. Further each sub series is forecasted using ANN or SVM. The figure below gives a diagrammatic representation of forecasting process using EMD and feature selection performed on a wind speed or power time series.
The framework for the given decomposition process can be stated as follows:

1) The wind speed series is decomposed by Empirical Mode Decomposition to get IMF’s and a residual.
2) The feature construction takes place that collects information from all the IMF’s and the residual.
3) The optimal feature subset is selected from feature combinations.
4) The forecasting model is built using ANN or SVM from the feature subsets.

EMD and feature selection technique [11] is used to select the best features from the resulting sub series. Given a wind speed series it can be expressed after decomposition as follows:

\[ x = [c_1(t), c_2(t), \ldots, c_n(t), r_n(t), c_1(t-1), c_2(t-1) \ldots c_n(t-1), r_n(t-1), c_1(t-p+1), c_2(t-p+1) \ldots c_n(t-p+1), r_n(t-p+1)] \] (29)

In the above equation the original time series \( y_t \) is decomposed using EMD and the linear regression of order \( p \) is applied to obtain the feature vector \( x \). The study is done on three wind farms in China (Jiangsu, Ningxia and Yunnan). Data sets for one month during are taken to evaluate the forecasting model. The following table shows the error metrics for the three wind farms.

| Wind Farm | RMSE (m/sec) | MAPE(%) |
|-----------|--------------|---------|
| Jiangsu   | 0.80         | 17.98   |
| Ningxia   | 0.80         | 10.29   |
| Yunnan    | 0.82         | 11.25   |

In [43], wind forecasting is done with an improved version of EMD i.e. ensemble EMD (EEMD), complementary ensemble EMD (CEEMD) and complete ensemble EMD with adaptive noise (CEEMDAN). However authors report that though EMD appeared to outperform various conventional methods (Persistence & ARMA), there appeared to be a problem of mode mixing [44] i.e. multiple IMF’s contain signals in similar frequency band. EEMD algorithm is similar to EMD except that the input time series has additional finite gaussian white noise. Other versions of EMD like CEEMD and CEEMDAN proved out to be better than its parent version EMD. Results of [43] show that EMD based hybrid SVR methods outperformed EMD based hybrid ANN methods for 1-hr, 3-hr and 5-hr ahead forecasting. Another hybrid method [14] is tested for short-term wind forecasting where wavelet transform and neural networks are used for 3-hr ahead prediction. Wavelet transform converts a wind series into several sub series, each sub series has a better performance behavior than the original one owing to the filtering effect of wavelet transform. Wavelet transform has two categories i.e. continuous wavelet transform and discrete wavelet transform (DWT). DWT can be mathematically defined as:

\[ W(m,n) = 2^{-(m/2)} \sum_{t=0}^{T-1} f(t) \phi \left(\frac{t-n2^m}{2^m}\right) \] (30)

In the above equation \( T \) represents the length of the signal \( f(t) \), where \( m \) and \( n \) are the integer variables, and \( t \) is the discrete time index [43].

Research in the field of wind forecasting has increased tremendously for the hybrid models based on machine learning [46]. Candenas and Rivera demonstrated ARIMA-ANN based forecasting model where for a fixed prediction horizon, wind forecasting is carried out [47]. Liu et al. presented a Support vector machine and Genetic algorithm (GA) based hybrid forecasting method using Wavelet decomposition transform for fragmenting the wind speed time-series to eliminate any potential stochastic variation [17]. Zhang et al. presented a hybrid technique based on gaussian process regression (GPR) and auto-regression (AR) and compared their results with SVM, ANN and persistence algorithm [11]. Mi et al. explored a hybrid method that encompasses wavelet decomposition transform, extreme learning machine and outlier correction technique to forecast multi-step wind speed [48]. Wavelet and wavelet packet decomposition eliminates noisy component from the wind series and extreme learning machine provides multi-step forecast on the sub-signals obtained from the decomposition technique.

Li et al. presented a combined method based on constant weight and variable weight for short-term wind speed prediction [49]. Jiang et al. proposed a novel method for short-term wind prediction based on modeling the fluctuations caused by adjacent wind turbines and the selected inputs are given to a v-SVM model [50]. Azimi et al. presented a feature selection model based on k-means cluster and a multilayer perceptron neural network for predicting short-term wind speed [51]. Jiang et al. presented correlation based...
discrete wavelet transform (DWT), least-square support vector machine (LSSVM) and generalized autoregressive conditional heteroscedastic (GARCH) method for short-term wind speed prediction. Correlation coefficients among different sub-series are used to assess the inputs to the LSSVR models for wind speed prediction [52].

Liu et al. discussed a modified Broyden-Fletcher-Goldfarb-Shanno (BFGS) neural network and wavelet transform based signal processing technique for short-term wind speed prediction and validated the same for four wind speed datasets [53]. Correlation coefficients are determined for each sub-series obtained after wavelet decomposition for assessing their relative importance. Tian et al. presented a multi-objective forecasting algorithm [54] wherein data pre-processing technique is based on complementary ensemble empirical mode decomposition (EEMD), variational mode decomposition and sample entropy. The proposed model is further validated for eight wind speed datasets and results reveal the superiority of the model when compared to benchmark models. However, the only limitation of this model is the time consumed in intermediate stages. Three variables, wind speed, electrical load and electricity price are predicted using an Elman neural network (ENN) whose weights and hyperparameters are optimized by a modern dragonfly algorithm as presented by Wang et al. [55]. Since the time-series for the three variables is highly non-linear, signal processing techniques such as wavelet transform, empirical mode decomposition and ensemble empirical mode decomposition are widely applied in the allied areas of forecasting.

Du et al. presented a multi-step ahead prediction based on a Whale optimization algorithm (WOA)-LSSVR technique and have applied the same to forecast wind speed, electricity price and electrical load [56]. The proposed technique is validated for six datasets from Singapore, China and Australia. Existing benchmark regressors like Generalized regression neural network (GRNN) and Back propagation neural network (BPNN) are used for a comparative analysis. In terms of performance metrics like root mean squared error and mean absolute error the proposed WOA-LSSVR model outperforms GRNN and BPNN.

Wang et al. presented a hybrid wavelet neural network (WNN) model based multi-objective sine-cosine algorithm (MOSCA) optimization technique [55]. The optimization problem is based on multi-objective sine-cosine functions. The candidate solutions are first initialized with some value and are allowed to converge or diverge in a given search space. A modified complementary ensemble empirical mode decomposition (MCEEMD) is implemented in order to solve the issues posed by simple pre-processing techniques such as ensemble empirical mode decomposition (EEMD). The model proposed by Wang et al. is also evaluated for assessing the robustness and stability for wind speed prediction as a highly non-linear time-series can cause difficulties in accurate prediction. The model based on MOSCA-WNN is applied to predict each sub-series and the results are aggregated. The proposed model is compared with WNN, GRNN, ARIMA and Persistence model.

VI. CONCLUSIONS

In this manuscript, we discuss primarily machine learning and hybrid wind forecasting techniques and its variants for regression analysis. These supervised learning models are useful for wind speed time-series datasets from different places around the globe to assess the performance of each model. In terms of accuracy, machine learning model based on supervised learning give good generalization performance which is the ability of a machine learning model to adapt to an unseen data. However, the need for ensemble based forecasts arrives from the problem of over-fitting that occurs in models like support vector regression, multiple linear regression and neural networks. Further, the superiority of hybrid forecasting models has led to an increased research in the field of wind speed forecasting.

REFERENCES

[1] H. S. Dhiman, D. Deb, and A. M. Foley, “Bilateral gaussian wake model formulation for wind farms: A forecasting based approach,” Renewable and Sustainable Energy Reviews, vol. 127, p. 109873, Jul. 2020.
[2] H. S. Dhiman, D. Deb, and J. M. Guerrero, “Hybrid machine intelligent SVR variants for wind forecasting and ramp events,” Renewable and Sustainable Energy Reviews, vol. 108, pp. 369–379, Jul. 2019.
[3] H. S. Dhiman, D. Deb, and A. M. Foley, ”Lidar assisted wind redirection in wind farms: A data driven approach,” Renewable Energy, Jan. 2020.
[4] H. S. Dhiman and D. Deb, Decision and Control in Hybrid Wind Farms. Springer Singapore, 2020.
[5] I. Okumus and A. Dinler, “Current status of wind energy forecasting and a hybrid method for hourly predictions,” Energy Conversion and Management, vol. 123, pp. 362–371, Sep 2016. [Online]. Available: https://doi.org/10.1016/j.enconman.2016.06.053
[6] M. Alexiadis, P. Dokopoulos, and H. Sahsamanoglou, “Wind speed and power forecasting based on spatial correlation models,” IEEE Transactions on Energy Conversion, vol. 14, no. 3, pp. 836–842, 1999. [Online]. Available: https://doi.org/10.1109/60.790962
[7] S. P. Kant and M. Aridhali, “Very short-term wind speed prediction: A new artificial neural network–markov chain model,” Energy Conversion and Management, vol. 52, no. 1, pp. 738–745, Jan 2011. [Online]. Available: https://doi.org/10.1016/j.enconman.2010.07.053
[8] Y. Jiang, Z. Song, and A. Kusiak, “Very short-term wind speed forecasting with bayesian structural break model,” Renewable Energy, vol. 50, pp. 637–647, Feb 2013. [Online]. Available: https://doi.org/10.1016/j.renene.2012.07.041
[9] M. Negnevitsky and P. Johnson, “Very short term wind power prediction: A data mining approach,” in 2008 IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, IEEE, Jul 2008. [Online]. Available: https://doi.org/10.1109/pes.2008.4596604
[10] Q. Hu, R. Zhang, and Y. Zhou, “Transfer learning for short-term wind speed prediction with deep neural networks,” Renewable Energy, vol. 85, pp. 83–95, Jan 2016. [Online]. Available: https://doi.org/10.1016/j.renene.2015.06.034
[11] C. Zhang, H. Wei, J. Zhao, T. Liu, T. Zhu, and K. Zhang, “Short-term wind speed forecasting using empirical mode decomposition and feature selection,” Renewable Energy, vol. 96, pp. 727–737, Oct 2016. [Online]. Available: http://dx.doi.org/10.1016/j.renene.2016.05.025
[12] Z. Liang, J. Liang, L. Zhang, C. Wang, Z. Yun, and X. Zhang, “Analysis of multi-scale chaotic characteristics of wind power based on hilbert–huang transform and hurst analysis,” Applied Energy, vol. 159, pp. 51–61, Dec 2015. [Online]. Available: https://doi.org/10.1016/j.apenergy.2015.08.111
[13] R. G. Kavasseri and K. Sentharaman, “Day-ahead wind speed forecasting using f-ARIMA models,” Renewable Energy, vol. 34, no. 5, pp. 1388–1393, May 2009. [Online]. Available: https://doi.org/10.1016/j.renene.2008.09.006
[14] J. Catalão, H. Pousinho, and V. Mendes, “Short-term wind power forecasting in portugal by neural networks and wavelet transform,” Renewable Energy, vol. 36, no. 4, pp. 1245–1251, Apr 2011. [Online]. Available: https://doi.org/10.1016/j.renene.2010.09.016
[55] J. Wang, W. Yang, P. Du, and Y. Li, “Research and application of a hybrid forecasting framework based on multi-objective optimization for electrical power system,” *Energy*, vol. 148, pp. 59–78, apr 2018.

[56] P. Du, J. Wang, W. Yang, and T. Niu, “Multi-step ahead forecasting in electrical power system using a hybrid forecasting system,” *Renewable Energy*, vol. 122, pp. 533–550, jul 2018.