International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 15 (2023) pp. 287-299

Received: 3 January 2023; Accepted: 4 May, 2023; Published: 14 June, 2023

Application of Bio-Inspired Optimization Techniques for Wind Power Forecasting

Judite Ferreira[1,2], Ricardo Puga[3], José Boaventura[3,4], A. Abtahi [2,5] and André S. Santos[1,2]

[1] ISEP, Polytechnique of Porto, Porto, Portugal
[2]ISRC - Interdisciplinary Studies Research Center
mju@isep.ipp.pt;abg@isep.ipp.pt;
[3] Universidade de Trás-os-Montes e Alto Douro (UTAD), Vila Real, Portugal

ricardo.ferreira.puga@gmail.com; jboavent@utad.pt;
[4] Centre for Robotics in Industry and Intelligent Systems, INESC TEC, Porto, Portugal
jboavent@utad.pt;
[5]College of Eng. And Comp. Science, FAU, Boca Raton, Florida, USA
abtahi@fau.ed

***Abstract*: As the need for replacing fossil and other non-renewable energy sources with renewables becomes more critical and urgent, wind energy appears to be among the two or three best choices for the short and medium time frames. The dominance of wind energy as the first choice in many regions, leads to an increasing impact of wind power quality on the overall grid. Wind energy's inherent intermittent nature, both in intensity and longevity, could be an impediment to its adoption unless utility operators have the tools to anticipate the impact and integrate wind resources seamlessly by increasing or reducing its contribution to the overall capacity of the grid. The wind forecasting science is well established and has been the subject of serious study in multiple fields such as fluid dynamics, statistical analysis and numerical simulation and modeling. With the renewed interest and dependence on wind as a major energy source, these efforts have increased exponentially. One of the areas that shows great promise in developing improved forecasting tools, is the category of "Biological Inspired Optimization Techniques. The study presented in this paper is the result of a study to survey and assess an array of forecasting models and algorithms.**

***Keywords*:** Wind Forecasting, Computational Fluid Dynamics, Support Vector Machine Method, Random Theory, Bio-inspired Optimization Methods.

I. Introduction

Wind power has seen considerable increase in all energy power systems. In the near future offshore wind power is also expected to expand rapidly as the world needs quickly to be energy independent from Russia.

Wind power is variable and intermittent since it is weather dependent, as such, wind power integration into traditional transmission lines needs additional power systems and electricity market planning and management for system balancing. This extra system balancing means that there are additional system costs associated with wind power assimilation. Wind power forecasting and prediction methods are used by system operators to plan unit commitment, scheduling, and dispatch and by electricity traders and wind farm owners to maximize profit. Accurate wind power forecasting and prediction has numerous challenges.

In this paper, different methodologies for wind power prediction are introduced and analyzed. As wind energy takes on a more significant role in fulfilling the need to replace fossil fuels, developing better forecasting tools for wind power become even more critical. The accuracy and precision of wind power forecasting are essential in making wind energy a reliable component of the overall energy supply infrastructure. The increasing ratio of wind energy supplies in relation to conventional sources, puts on more stringent requirements for its stability and reliability. Better forecasting allows increasing integration of wind power for utilities and therefore an increasing opportunity for wind power to replace fossil fuels. There are several techniques and mathematical models that allow to minimize the negative impacts of wind and power prediction problems [1-7]. Some of these models and mathematical techniques are inspired by biological algorithms based on Artificial Intelligence [8-18]. This paper is an extension of the research started by Puga, et al. [17]. To determine the best option, the following techniques were considered and analyzed: Fuzzy Logic based algorithms, Neural Network base algorithms, The Grey Model, Algorithms (GA), Empirical Mode Decomposition (EMD), Random Theory-based algorithms (RTA), Computational Fluid Dynamics-based algorithms (CFD), and Support Vector Machine Method algorithms.

This paper, it is organized as follows: Section II, Bio-Inspired Optimization Techniques, section III, a summarized Fuzzy Logic based algorithm, section IV, Neural Network base algorithms, section V, the Grey Model and section VI an integration of Neural Networks and Genetic Algorithms. Section VII presents the Empirical Mode Decomposition, section VIII, Random Theory based algorithms, section IX, Computational Fluid Dynamics based algorithms and section X, Support Vector Machine Method algorithms. In section XII a comparative analysis of the multiple algorithms is presented and in section XIII, main conclusions, along with future work suggestions are provided.

II. Bio-Inspired Optimization Techniques

The routine of an organization often involves solving a large number and diversity of optimization problems, with a significant impact on the organization's performance. Production and distribution planning, transport planning, resource allocation (raw materials, labor or time availability in machinery) and task scheduling are classes of combinatorial optimization problems common in the industrial reality.
The existence of efficient, reliable, and cheap computing processing power over the last decades, had transformed many fields of science and engineering. This context facilitates the development of new optimization algorithms that operate in a rather different way than the classical ones, and that allow practitioners to solve optimization problems where the classical optimization methods are not applicable or simply too hard (in terms of processing time and other resources) to use.
Over the past decades there has been a growing interest in the application of algorithms that somehow adopt the principles of natural processes, particularly about the biological component. The assumption that by understanding the solutions that nature employs in your daily life, we can use this knowledge acquired to solve our problems [9-11]. Such class of methods have received different names, such as Meta-Heuristics, Nature-Inspired Techniques, Soft Computing, Evolutionary Computing and Swarm Intelligence [7-11],[19-21].
Bio-Inspired Optimization Techniques form a class of approximate resolution methods, based on Artificial Intelligence concepts, specially developed to address complex combinatorial optimization problems. They consist of general research strategies, inspired by concepts from diverse areas such as classical heuristic procedures, biological evolution, neuronal systems, nature-inspired behaviors like the collective behavior of decentralized and self-organized systems, or statistical mechanisms. For example, Taboo Search, Genetic Algorithms, Neural Networks, Ant Colony Optimization, Particle Swarm Optimization, and others [7-11].

III. FUZZY LOGIC-BASED ALGORITHMS

Even if the theory of fuzzy logic had been studied since the 1920's, the term fuzzy logic was only introduced in 1965 by Lotfi Zadeh, a professor of UC Berkeley in California. Lotfi showed that conventional computer logic was inefficient when it manipulated data representing subjective or unclear human ideas. Fuzzy logic was designed to allow the computer to determine the distinctions between data which is neither true nor false. This theory can be used to the Wind Power Forecasting. Fuzzy algorithm has been applied to various areas, from control theory to Artificial Intelligence (AI).

Fuzzy logic algorithms approach the truth or false through the following procedure: 0 and 1 are set as the extremes, and in between there exist different degrees of truth. Something like the process of human reasoning.
There are largely three types of fuzzifiers:
- Singleton fuzzifier;
- Gaussian fuzzifier;
- Trapezoidal or triangular fuzzifier.

Advantages of Fuzzy Logic System:
- This system can work with inaccurate, distorted, or noisy input data;
- The structure of Fuzzy Logic is simple and clear;
- Fuzzy Logic comes with mathematical concepts from set theory and using it is simple;
- It supplies efficient solution to complex problems in all fields that resembles human reasoning and decision-making;
- Algorithms that can be characterized without much data, so less memory is needed.

Disadvantages of Fuzzy Logic Systems:
- There is no systematic approach to solve a given problem through fuzzy logic;
- Confirmation of its characteristics is difficult or impossible;
- As fuzzy logic works on precise as well as imprecise data so it is not timely accurate.

Fuzzy logic is applied in three steps: first, fuzzification, where fresh inputs are exchanged for blurred ones; second, these inputs are used to generate a blurred signal; third, defuzzification, where results are graded and there can be more than one result, each with a different grade. Fig. 1 shows the diagram of fuzzy logic.

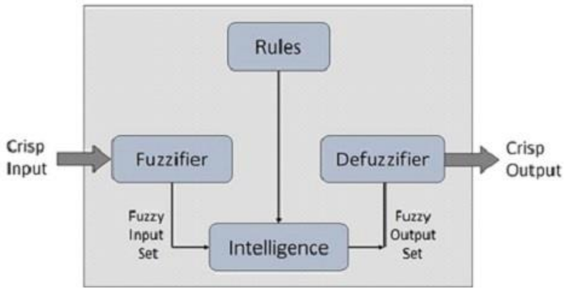


Figure 1 – Fuzzy Logic structure

Adaptive Neuro-Fuzzy Inference System (ANFIS) is an adaptive multi-layer feed forward network used for prediction of non-linear systems. This method uses past data samples to predict the future data, aligned with the self-learning ability of neural networks, in conjunction with a linguistic expression function of a fuzzy inference system.

A. *Fuzzy Rules knowledge base*

IF- THEN rules. Firstly, is consider Multi-Input, Single-Output fuzzy systems:

$$U \subset R^n \rightarrow R \qquad (1)$$

Where U is compact.

The Multi-Input, Single-Output rules are:

R(j): IF x_1 is $A_{j,1}$,

x_2 is $A_{j,2}$, and…and x_L is $A_{j,L}$;

THEN y is B_j,

Where;

j = 1,2,3,…,N;
x_i for i = 0,1,2,3, …,L;
x is the input variable and y is the output variable.
A and B are the linguistics terms characterized by the fuzzy membership functions $\mu_{Aj,I}$ and μ_{Bj}, respectively.
Every R(j) able to be seen as a fuzzy implication where:

$$A_{(j,1)} \times \cdots \times A_{(j,L)} \rightarrow B_{j,1}. \qquad (2)$$

Rule base includes the rule set and the IF-THEN conditions, provided by the experts to support the decision-making, based on linguistic information. Current developments in fuzzy theory offer several effective methods for the design and tuning of fuzzy controllers. Most of these developments reduce the number of fuzzy rules.

B. *Fuzzifier*

The concept of the fuzzifier is to transform a crisp input *x* into a fuzzy [13] sets defined in $Ux \rightarrow [0,1]$.
Summing up the fuzzification step helps to convert inputs. It allows the conversion of crisp numbers into fuzzy sets. For example, crisp inputs measured by sensors and passed into the control system for further processing, like temperature, pressure, wind speed measure, etc.

C. *Inference Engine*

It helps to determine the degree of match between fuzzy input and the rules. Based on the % match, it determines which rules need implementation according to the given input field. Afterwards, the applied rules are combined to develop the control actions. So, when data is passed given, the fuzzy rules in the first stage are inferred, and then the consequences of the intermediate variables in the inference stage are formulated before they are passed on to the next stage as data. To construct a multi-stage inference fuzzy system, the most influential parameters of input variables are usually organized in the first stage, and so on [13]. There are several models for fuzzy inference in this paper, for example, one fuzzy inference system mapped to a neural network structure with five layers.

D. *Defusser*

Defuzzification convert the fuzzy sets into a crisp value. There are many techniques available, so one need to select the most suited for the defuzzification.
One example is the mapping from fuzzy sets in U_x to a crisp point $y \in R$. This mapping is generally chosen as the centre average:

$$y_i = \sum_{i=1}^{n} \bar{y}_i^{\,l} \times \left(\mu_{B^i}\ (\bar{y}_i^{\,l})\right) / \sum_{i=1}^{n} \left(\mu_{B^i}\ (\bar{y}_i^{\,l})\right) \qquad (3)$$

Where $\bar{y}_i^{\,l}$ is the center point in R at, where $\left(\mu_{B^i}\ (y_i)\right)$ achieves its maximum value in the i^{th} inference stage.

E. *Fuzzy Logic for Wind Forecasting*

Use of renewable energy resources, mainly wind power, gathered considerable attention in a number of countries after the adoption of the Kyoto protocol and, currently, the war in Ukraine. However, despite its benefits to minimize climate change, the fluctuations in wind speed and other weather variables make the wind power output completely stochastic and different from conventional energy sources. Because of its stochastic nature, there several challenges in connecting large wind energy supplements into a power system grid. In order of increase the economic competence and acceptability of the wind energy and allow a reduction in the market overestimation or underestimation, accurate wind forecast power as well as wind speed is required. Accurate forecasting system can help the distribution operators and traders to make better decisions. Several techniques have been developed to predict wind power and speed. Existing techniques can be statistical, physical and time series models. Among these methods the is the Adaptive Neuro-Fuzzy Inference System (ANFIS) proposed in [2].

Determining the output power of wind generators is always associated with uncertainties in wind speed and other fluctuating weather conditions. Short-term forecasting is essential for its efficient operation. In [2], the author, propose a double stage hierarchical Adaptive Neuro-Fuzzy Inference System (ANFIS) for wind power forecast of a microgrid wind farm in Beijing, China. The first objective of ANFIS Numerical Weather Forecast (NWP) is to predict the wind speed at the exact location of the wind farm and turbine hub height. The second objective is to models the current wind speed and power ratios. So, the following day wind speed for the first day are applied to predict the wind energy. In order to assess the performance of the proposed model, it was compared to three other forecasting techniques and shown the most accurate.

F. *Proposed Wind Power Prediction Strategy*

The prediction performance of the wind power forecaster in this approach is highly depends on the Numerical Weather Prediction (NWP) models. In fact, the focus of the research was to study is the impact of NWP in the improvement of short-term predictions. The prediction scheme is depicted in Fig. 2. While modelling, a one-year record, provided from Supervisory Control and Data Acquisition (SCADA) historical measurements Weather Research and Forecasting (WRF) and NWP/WRF model historical weather forecasts are used to train an ANFIS that successfully can estimate a transfer function between specific patterns of input and output. Then, Back Propagation (BP) is applied to optimize the parameters of the membership functions of ANFIS. This process continues until the prediction error reaches to a suitable value [2].

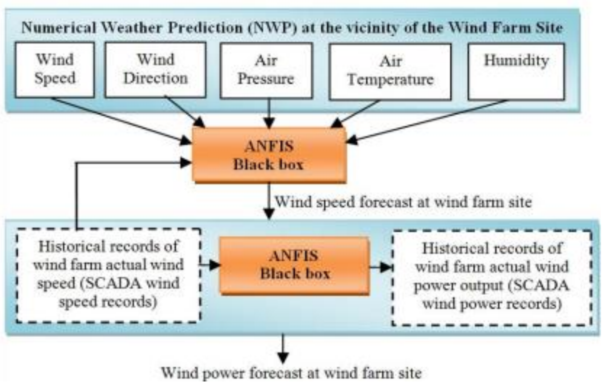


Figure 2 Double-stage prediction model using ANFIS [2]

An example of an ANFIS network is the Takagi-Sugeno fuzzy inference system mapped to a neural network structure with five layers is presented in [2].

Each layer has several nodes characterized by the node function as seen in Figure 3.

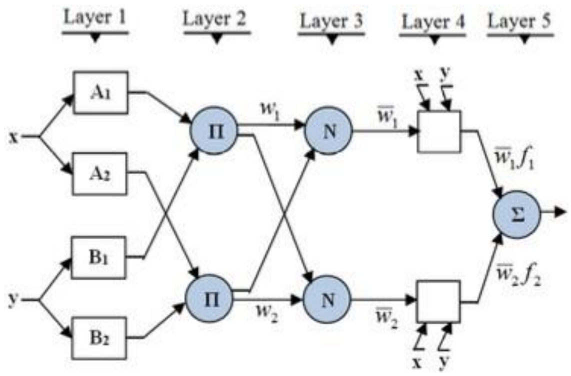


Figure 3 Development pipeline for the proposed algorithms [2]

The first stage implements the wind forecast at the turbine site; the second stage relates the wind speed to the power output of the turbine. The combination of the first two stages provides a day-ahead power generation forecast. The same study also proposes a new hybrid approach for short-term wind power prediction using ANFIS with two hierarchical stages. The results are compared for every season of the year with the following forecasting methods:

- Mean Absolute Percentage Error (MAPE);
- Sum squared error (SSE);
- Root mean squared error (RMSE);
- Standard deviation of error (SDE);
- Mean absolute error (MAE).

The average MAPE (8.1133%) for one-day ahead forecast, shows an improved when compared with the other techniques.

IV. NEURAL NETWORK-BASED ALGORITHMS

Lijie Wang study the brings wavelet transform into the time series of wind power and noted that the decomposed series all have chaotic characteristic.

A new technique of wind power prediction with Artificial Neural Network (ANN) model based on wavelet transform to predict wind power data from the Fujin wind farm and Saihanba wind farm of China were proposed in this study. The prediction results are presented and compared to the no wavelet transform method and ARMA method.

The results show that the new method based on wavelet transform ANN will be a useful tool in wind power prediction.

A. *The Wavelet Theory*

Multi-resolution approximation by wavelet basis functions is a technique for representing a function, which are formed by a scaled and translated mother wavelet. The Continuous Wavelet Transform (CWT) of a signal *x(t)* is defined as [3]:

$$WT_x(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} x(t)\,\psi^*\left(\frac{t-b}{a}\right)dt = \langle x(t), \psi_{a,b(t)}\rangle \quad (4)$$

Where *ψ(t)* is the mother wavelet, and other wavelets:

$$\psi_{a,b(t)} = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \quad (5)$$

The associated dilated and translated versions. “a” and “b” are the dilation and translation parameters respectively.

Where $a = a_0^j$, $b = ka_0^j b_0$,j, k ∈ Z.

Wavelet theories use the mathematical method Wavelet Transform, which decomposes the original signal into several time series with simpler frequency components, achieving a new level of analysis of the initial wave. The wavelet theory has a continuous (Equations 4 and 5 and a discrete form, with the latter applied in practical applications.

Fig. 4 shows a decomposition process that results in many lower resolution components of the initial wavelet.

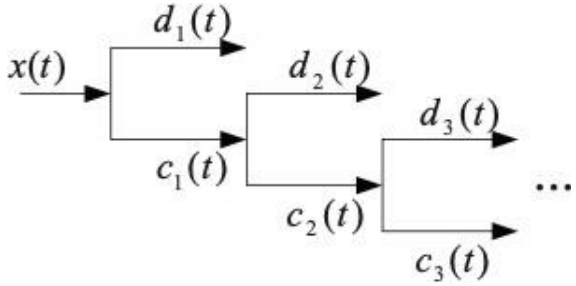


Figure 4 Mallat wavelet decomposition tree [3]

This method was used in conjunction with an ANN to predict the power output of a wind turbine. The results achieved are better than the ARMA method [3]. The maximum power output of the wind turbine during the test was 850 KW.

Model	1h ahead	3h ahead
ARMA	60.97	120.10
no WT	58.58	110.88
Hybrid	55.63	111.04

Table 1. MAE of power output prediction for WT 1-(in kW)[3]

Model	1hour-ahead	3hours-ahead
ARMA	3.20	5.22
no WT	3.15	5.35
Hybrid	2.97	5.21

Table 2. MAE of power output prediction for WT 2 (in kW) [3]

The main difference between the prediction results shown in Tables 1 and Table 2 may be explained by the difference in the size of the dataset for each case. WT-1 had 1 month of training data, and 1 month of data for testing, while WT-2 had 3 months of training data and 3 months of data for testing [3].

B. *Back Propagation Neural Network (BPNN)*

The traditional Neural Network algorithms are based on a simple feed forward processes with a training data layer and a testing data layer and no feedback loop for error assessment and minimization [4]. The Back Propagation Neural Network (BPNN) model is a multi-layer feedforward Neural Network

with an error back propagation training algorithm to approximate any nonlinear mapping [4].
The BPNN model for this study has one hidden layer, as shown in Figure 5.
The input layer has M neurons, expressed by m from 1 to M. The hidden layer has I neurons expressed by I from 1 to I. Similarly, the output layer has J neurons expressed by j from 1 to J.

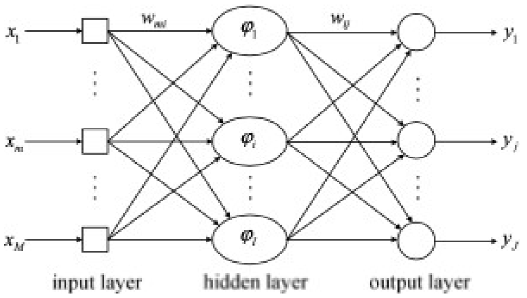


Figure 5. BP neural network with one hidden layer [4]

The synaptic weight of the input layer and the hidden layer are expressed by w_{mi} (m = 1, 2, …, M; i = 1, 2, …, I). And the synaptic weight of the hidden layer and the output layer are expressed by w_{ij} (i = 1, 2, …, I; j =1,2, …, J)
The output of each neuron is:

$$y_{kj(n)} = \sum_{i=1}^{I} w_{ij}(n)\varphi_i(\sum_{m=1}^{M} w_{mi}(n)\, x_{km}) \quad (6)$$

The error signal of each neuron of the output layer is:

$$e_{kj}(n) = d_{kj}(n) - y_{kj}(n) \quad (7)$$

The data is divided in the sets: the training set; the error set; the retraining set; and the test set. The flowchart is represented in Figure 7 below.

In order to test it, the results are compared with another prediction-based algorithm, in this case the Support Vector Machine [4]. The method shown better results in Table 3, when compared with the MAE/KW method, the Mean Relative Error (MRE) and the Root Mean Square Error (RMSE) per kW, for both the traditional and the improved prediction methods.

Method	**MAE/kW**	**MRE/%**	**RMSE/kW**	**Time/s**
Traditional pred.	44.75	30	56.48	65.34
Improved pred.	28.56	22	34.32	86.27

Table 3. Comparison between improved and normal method with BP as base [4]

The results of the BPNN show better accuracy than the conventional NN as shown in Figure 6 below:

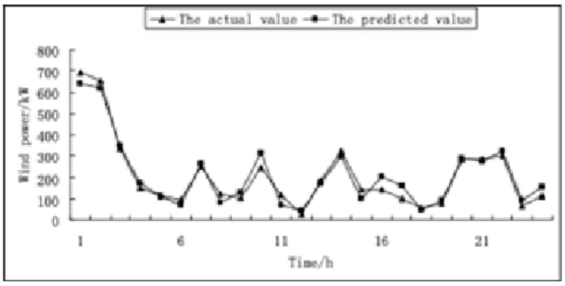


Figure 6 - Improved Method with BP as base [4].

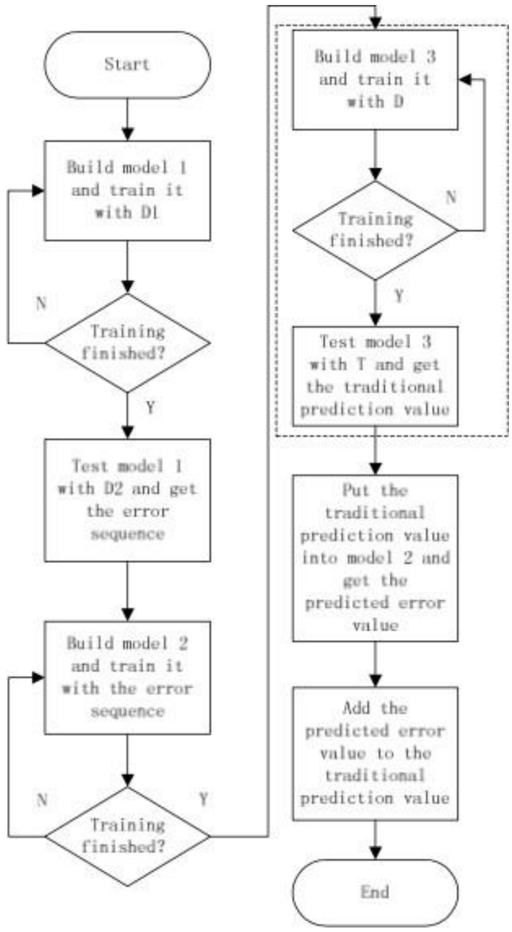


Figure 7 Improved wind power prediction process [4]

C. *Back Propagation Neural Network aligned with storage system*

The goal of the wind prediction is ultimately to achieve a better power supply to the grid from energy provenience from wind power. A way to achieve this goal is to have Energy Storage System (ESS) to be able to store energy when the production is above a certain level, as show in the figure 8. Where $P_w = P_{es} + P_g$ is based on power equilibrium. P_w is the output of wind farm, P_{es} is the observing energy of the ESS and the P_g is the power inject to grid.

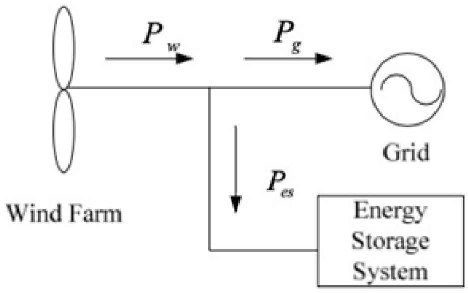


Figure 8 Scheme of wind power structure with power storage [5]

Where P_w is the output power generation as a function of time and is integrated from t_1 to t_2, when generation exceeds demand, and P_g is the net power output to the grid integrated over the time interval t_3 to t_4, when demand exceeds generation. A study based on back propagation model with six input variables and one hidden layer was created to predict the power output of a wind turbine [5].

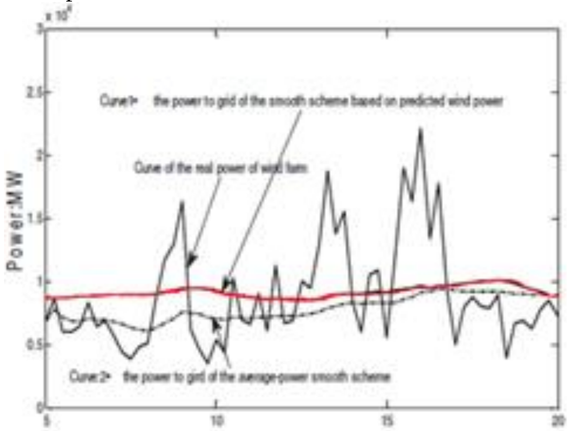


Figure 9 - Plots showing the real power as compared with predicted power [5]

As we can see in Fig. 9 the prediction of the wind power output is poor, and yet, ESS allows a smooth power output, as shown in Figure 10.

The improvement of the accuracy of the wind prediction would allow a lower energy storage by the system and a bigger power output to the grid.

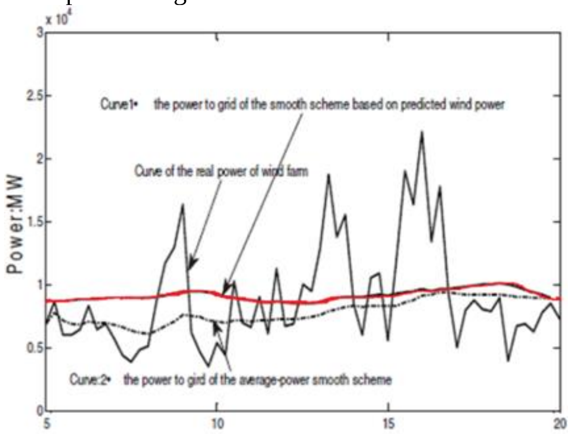


Figure 10 - Curves of power grid 错误!未找到引用源。

D. *Back Propagation Neural Network with ARMA*

The expression of the zero mean stationary sequence for ARMA (*p*,*q*) method is:

$$\varphi(B)x_t = \theta(B)a_t \tag{8}$$

Where φi, θi, $I = 1, ..., p$ and $j = 1, ..., q$.

The autoregressive polynomial is

$$\varphi(B) = 1 - \varphi_1(B) - \cdots - \varphi_p(B) \tag{9}$$

The moving average polynomial is

$$\varphi(B) = 1 - \theta_1(B) - \cdots - \theta_p(B) \tag{10}$$

In the study a BP Neural Network residual correction is combined with the ARMA method. The data is from a wind farm with a total installed capacity of 17.56 MW. Once the data is submitted the ARMA model is applied. Secondly, the BP neural network and the prediction value or ARMA is superimposed on the output of the network. Time series model flow chart of BP neural network residual correction is represented in Figure 11 [6].

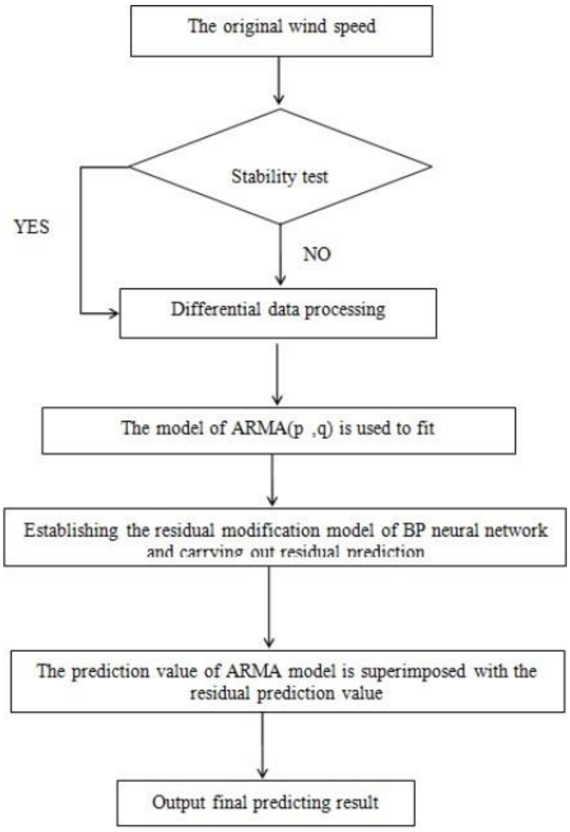


Figure 11 - Time series model flow chart of BP neural network residual correction [6].

To compare the results of these techniques, a side-by-side analysis was conducted between the full model and ARMA method. As shown in Table 4, MAE and MSE errors are 0,1880 and 0,0468 respectively for the complete method, which are an improvement of 26.83%, 27.02% and 1.42%, for MAE, MSE and MAPE respectively, over the ARMA method [6].

Error Índex	Arma (4,5)	BP-ARMA
MAE	0.4563	0.1880
MSE	0.3170	0.0468
MAPE	0.0154	0.0012

Table 4. Table with the model errors [6]

V. THE GREY MODEL -WIND SPEED-WIND POWER FITTED CURVE

The power curve of wind turbines can reflect the power generation performance under different wind speeds. At present, there are two kinds of power characteristic curves: theoretical power curve and actual operation curve. The theoretical wind speed power curve can be expressed as $P=1\ 2\ CpA\rho v^3$, where P is wind turbine output power, Cp is wind energy utilization coefficient, $A = \pi R^2$ is the area swept by the wind turbine, R is hub radius, ρ is air density, v is wind speed. The power curve is obtained by simulation under ideal condition. However, the wind turbine power is often affected by turbulence, wind shear and other factors in the environment. The measured power curve is obtained from wind speed recorded in SCADA. However, in practical application, the part of the measured scatter points is disorderly and highly dispersed.

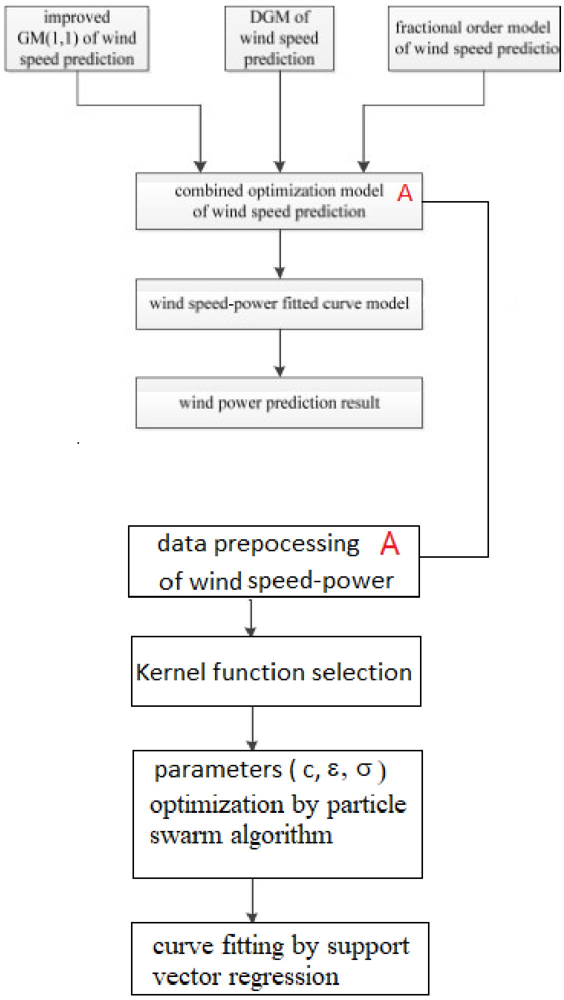


Figure 12 - Power prediction model [12]

The forecasting model shown in Figure 12 is supported by the Grey Model. The Grey Model is used for systems were with high uncertainty and poor information quality. It is ideally suited for wind speed forecasting. A disadvantage is its low prediction accuracy. For this study, an improved Grey Model, a discrete Grey Model and fractional Grey Model are presented [8].

Since the Grey Model lends itself well to self-adaptation and self-learning, a Neural Network architecture was developed to include these improvements. The results of the study were bench-marked through a comparison with the Auto-Regressive Integrated Moving Average (ARIMA).

The error of the improved Grey Model for wind power prediction are shown in Figure 13.

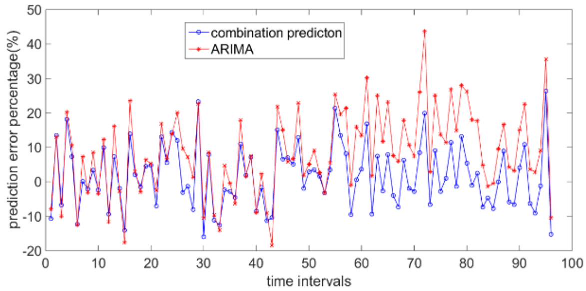


Figure 13 Results for power prediction model [12]

	MAE (kW)	MAPE	RMSE (kW)
Improved GM	76.9	15.7%	109.4
DGM	89.4	18.2%	129.6
1/ 2 order	106.6	21.1%	143.7
1/4 order	101.2	19.7%	142.3
2/3 order	104.6	20.8%	139.7
combination	43.4	9.7%	54.2
ARIMA	69.7	14.9%	82.6

Table 5. Wind speed prediction error of different models based on the power prediction model [8].

Table 5 shows that the "combination" has great results when compared with the results of ARIMA. The improvement of the power output prediction over MAE, MAPE and RMSE was 37.7%, 34.9% and 34.4%, respectively [8].

VI. COMBINATION RECURRENT NEURAL NETWORKS WITH GENETIC ALGORITHMS

A combination of Recurrent Neural Networks (RNN) with Genetic Algorithms (GA) was implemented in [13] to predict short- and medium-term wind speed.

A Recurrent Neural Network takes as an input the present and the recent past, responding the data in different ways for each feedback loop. This methodology is memory based where the weight of each loop changes with the new loop inputs [13].

A more complex RNN is Long Short-Term Model where the error is preserved, it maintains a more constant error as they learn with a goal of linking causes and effects remotely [13].

A. *Recurrent Neural Network*

A Recurrent Neural Network (RNN) is a type of ANN that uses sequential data or time series data. These deep learning techniques are commonly used for ordinal or temporal problems such as language translation, Natural Language Processing (NLP), speech recognition, and image captioning. They are built into popular apps like Siri, voice search and Google Translate. Like feedforward and CNNs, RNNs use training data to learn. They are distinguished by "memory" as they distinguish their information from inputs and outputs. While traditional deep neural networks assume inputs and outputs as independent of each other, the output of RNN depend on the previous elements of the sequence. Although future events are also useful to determine an output of a given sequence, since unidirectional RNN may not explain these events in their determinations [19]. Figure 14 shows the comparison of recurrent Neural Networks and the Feedforward Neural Networks

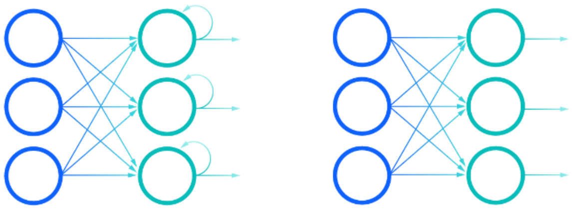


Figure 14 - Comparison of Recurrent Neural Networks (on the left) and Feedforward Neural Networks (on the right)

Recurrent Neural network inputs are the present and the recent past, feedback loop. All three combine to determine how RNN respond to new data. RNN are memory based and preserve a sequential information spanning many time steps in the hidden state as the network cascades forward finding correlation between events.

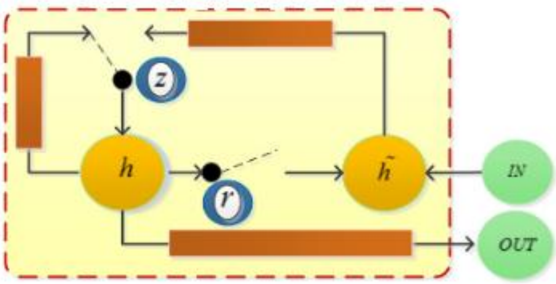


Figure 15 - Basic configuration of a GRU [19]

Figure 15 show the basic configuration of a Gated Recurrent Unit (GRU), where *z* and *r* refer to the update and reset gates, respectively. Whereas *h* and $\tilde{h}$ refer to the activation function and the candidate activation function, respectively.
Decomposing wind power time series using various Wavelet Decomposition (WD) [19] techniques can reduce the volatility of the signal by dividing it into simple parts, making it easier for prediction models to process. Recurrent Neural Networks (RNN), such as the Long Short-Term Memory (LSTM) with wavelet activation functions focus

B. *Long short-term memory*

Long Short-Term Memory (LSTM) is a form of RNN that is initiated by changing the structure of its recurrent connections [14]. As a self-connected unit, the LSTM acts essentially as a Constant Error Carousel (CEC), that considers and adapts to the flow of errors over time. It incorporates input, output, and forget gates to maintain and adapt to the content of the memory stored as part of its CEC function. This protects the other components of the LSTM from unrelated memory contents. It also enables the CEC memory to be periodically reconstructed once its contents become ineffective and superseded by updated information.
Figure 16 is a diagrammatic illustration of a signal LSTM cell.[19]. The black circles depict the three-component states: input i_t, output o_t, and forget gate f_t, and the current memory cell state *C* and new candidate values for cell along $\tilde{C}$ along with the Ψ (tanh activation function) are learned by considering information at the previous, current, and next time steps, that is, *t* -1, *t* and *t*+1.

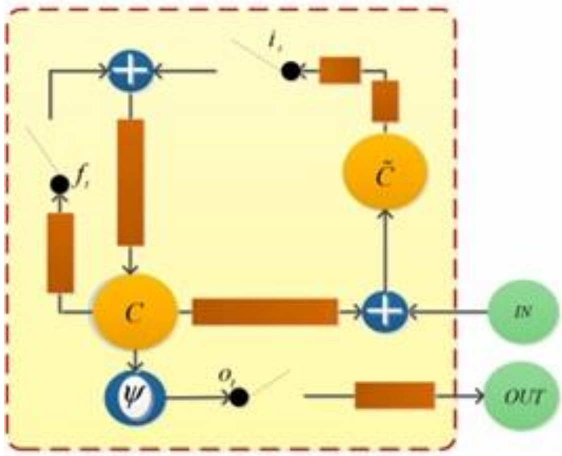


Figure 16 - Basic configuration of an LSTM model [19].

The architecture of the LSTM model comprises three gates named input, forget, and output gates by applying modulation functions to the input and output. The information at the gates typically involves current and recurrent inputs adapted through sigmoid functions (σ). On the other hand, the modulation process usually applies a hyperbolic tangent activation function (*tanh* Ψ).

For many applications, the prediction performances of Gated Recurrent Unit (GRU) and LSTM models are similar, although LSTM is computationally more time-consuming. That additional computation time enables LSTM to create deeper networks that can exploit more complex relationships between the variables and spanning more time points.
The vanishing and exploding gradient problems in machine learning models can cause consequential variations from one iteration to the next as the model weights changed by a large magnitude, leading to model instability and inefficient learning. The vanishing gradient problem can also impact RNN models. It refers to the inefficient transfer of gradient information from the model output to update the weights applied to the layers at the input end for subsequent model iterations. Careful selection of activation functions based on trial and error is required to overcome this problem in GRU and LSTM models.

C. *Methodology of LSTM-GA*

The methodology for the study started with the division of the data set in previous values and future values. The window selection involves decoding using a Genetic Algorithm based on the selection of parents, crossover, mutation, and fitness function as we can be seen in the figure 17.
In the beginning, the wind power data is separated from future wind power data over time (t). Having the data set ready, the LSTM (Long Short-Term Model) is used for training and compared with the validation set. Figure 17 shows forecast results using the GA optimized window and LSTM model for weekly variations [12].

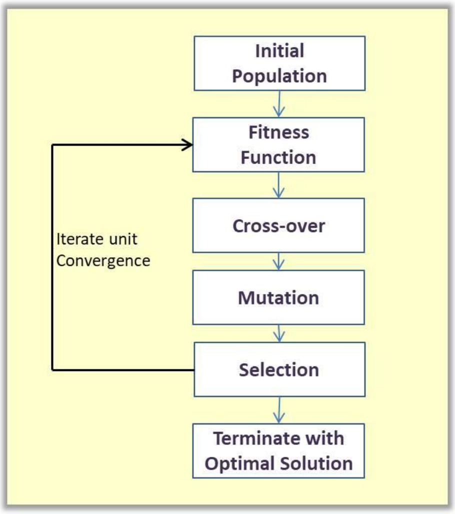


Figure 17 Steps in the GA process

For data preparation and training used the wind energy data set of the wind forecasting track a Global Energy Forecasting competition 2012 (GEFCom2012) collected for three years duration. The author of this paper [12] divided the data into training and testing in the ration 80:20. The last 20 percent of the data is retained for validation while the 80 percent of data is used for training. The temporal order of observations is maintained throughout. Test data is used for evaluating the accuracy of the proposed forecasting model. The fitness value of the window size is obtained for each data string corresponding to the Root Mean Square Error (RMSE). The GA operations on the dataset are-reproduction, crossover, and mutation to generate a new data. The new data is checked for the convergence. The process is repeated until the convergence criterion is satisfied and the process is stopped [12].

Figure 18 shows the wind pattern variation predicted using the GA optimized LSTM model. For the benefit of comparison, the actual data is also plotted in the same graph. There is a better fit of the data as the time is increased. A RMSE of 0.0957993 and 0.0929905 is obtained for the short term (week) and medium predictions (one month) indicating the effectiveness of the LSTM model with optimum window size[12].

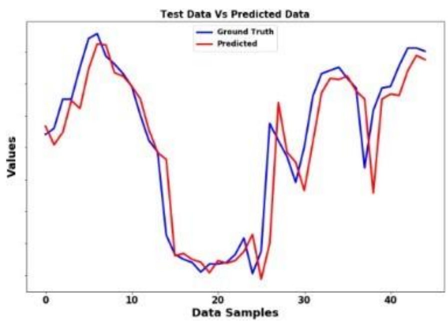


Figure 18. Results of wind speed tolerance intervals prediction with 90% confidence level 1 [12]

VII. Empirical Mode Decomposition of Wind Power

In [14], the authors, to make full use of the effective information in historical data to further improve the prediction accuracy of wind power generation. They proposed a model of Empirical Mode Decomposition (EMD) and Deep Long-Term Memory (DLSTM), constructing a multi-scale combined prediction model (EMD-DLSTM). In the process of model construction, firstly, the EMD method is used to decompose the wind power sequence into Intrinsic Mode Function (IMF) of different scales to weaken the volatility of the wind power sequence and obtain more regular components. Then, the DLSTM network is used to model the decomposed wind power sequence and its features, and then weights the predicted values of each component to obtain the true predicted wind power. The empirical analysis of a wind farm data in Laizhou [14] shows that the EMD-DLSTM method has better prediction ability than the existing forecasting methods. The study proposes prediction method that combines EMD and DLSTM model. The EMD method is used to decompose the wind power sequence into different Intrinsic Mode Functions (IMFs) and a trend term; the IMFs and the corresponding time feature are combined into a new vector; the DLSTM neural network is used to construct each feature vector. The modulo is finally weighted by the component prediction values to obtain the predicted value of the wind power. The forecasting process is shown in Fig.19.

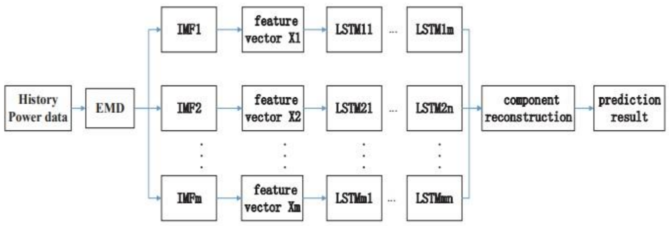


Figure 19 Predict flow chart [14]

In the case study presented in [14] the sample data was obtained from a wind farm in Laizhou from January to April in 2018. The installed capacity of the wind farm is 100MW. The data obtained is six different time series including wind direction, wind speed, temperature, atmospheric pressure, humidity, and wind power. Sampling time interval is 15min, the total data is 6000 samples. In this study, the Keras Deep Learning Framework was used for experimental design. Among them, the sigmoid function is used as activation function of the LSTM network, and the model is optimized by the Adam Optimization Algorithm. Adam Optimization Algorithm has no smooth requirement for objective function, whose loss function can change with time, and can better process the noise samples, and naturally has an annealing effect.

To verify the prediction accuracy of wind power, the average absolute percentage error η_{MAPE} and root mean square error η_{RMSE} are used as the basis for evaluation. In wind power ultra-short-term prediction, the smaller η_{MAPE} and η_{RMSE} value are, the more accurate the prediction result is η_{MAPE} and η_{RMSE} are expressed as:

$$\eta_{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| / y_i \tag{11}$$

$$\eta_{MAPE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \tag{12}$$

Where $\hat{y}_i$ and y_i are the predicted and actual values of wind power respectively; *n* is the number of predicted verification data; *i* is the predicted point sequence number.

To verify and explain the prediction performance of EMD-DLSTM network, it was compared with LSTM, DLSTM, and EMD-LSTM network, with a prediction step size is 15 minutes. The results show the prediction results of 192 time points in two days, Table 6 and figure 20 below shows the results of the error analysis [14].

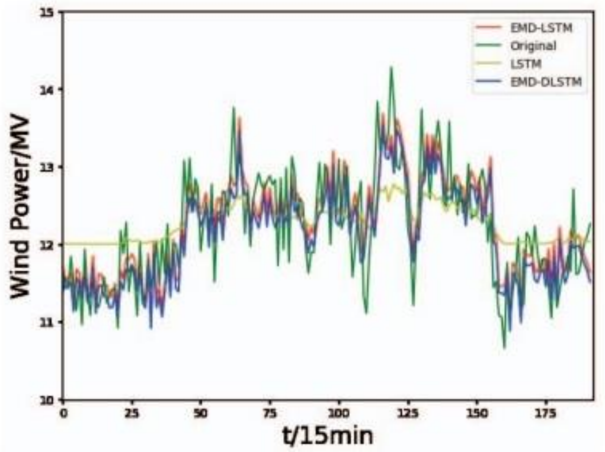


Figure 20 Error analysis

The EMD-DLSTM model had the better results between the three techniques. The error index is reduced by 9.21% (86.91-77.70) and (79.87-77.70) 2.17%, and RMSE is reduced by (31.68-19.07)12,61% and (26.40-19.07) 7,33% compared with the LSTM and EMD-LSTM network respectively [14].

	LSTM	EMD-LSTM	EMD-DLSTM
MAPE	86.91	79.87	77.70
RMSE	31.68	26.40	19.07

Table 6. Error Analysis [14]

Briefly, the authors concluded that the EMD decomposition algorithm can significantly eliminate the negative effects of non-periodic and non-stationary wind power, thereby effectively improving the usability of training data and reducing the difficulty of establishing a model. Based on DLSTM network, the authors establish a wind power prediction model, which can be model for long-term time series, and then realize the ultra-short-term prediction of wind power. The proposed method can effectively utilize multivariate information and has higher prediction accuracy than conventional machine learning algorithms such as single-layer LSTM networks.

VIII. Randow Theory-Based Algorithms

The Inductive Confidence Machine (ICM) is based on the division of the dataset in the proper training set and the calibration set. The data from the training set is used to build a prediction rule, with the derived rule a strangeness measure is aligned with the data of the calibration set [1].
Support Vector Regression is the mapping of input into high dimensional space by nonlinear function [1].
For 1 hour mean wind velocity prediction using this method the results were very accurate as we can see in the figure 20.

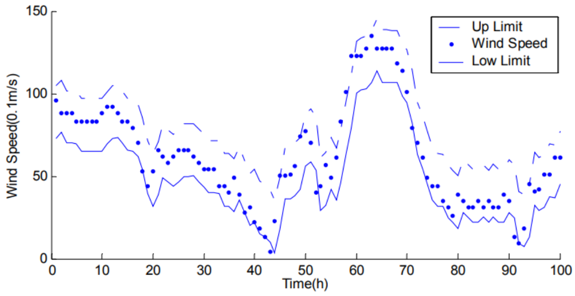


Figure 21 Results using GA-optimized window and the LSTM model for weekly variation [1].

The results of 50, 100,200, 500, 1000, 2000 and 5000 samples test with three confidence levels are shown in Table 7. The results shown that the predictive region output by ICM can cover the real wind speed value with the given confidence level.

Confidence Level	Number of Samples/result (%)						
	50	**100**	**200**	**500**	**1000**	**2000**	**5000**
90	92	89	85	89,2	89.9	91,25	89.52
95	100	94	91	94.4	94.2	95.8	94.34
99	100	100	99.5	99	98.8	98.95	98.5

Table 7. Results of high confidence intervals [1]

ICM is a novel algorithm for wind speed prediction. The experimental results show that the predictive region output by ICM meet expected level of confidence. The approach is better than the Transitional Support Vector Regression Algorithm. Confidence information provided by ICM effectively reduces the risk of the decision-making and improves the availability of the prediction algorithms. Shortening the interval length of the predictive is a way to improve the prediction method.

IX. Computational Fluid Dynamics Method-Based Algorithms

The Computational Fluid Dynamics (CFD) is a well-established field of engineering where numerical techniques based on fundamentals of fluid flow are applied to analyze fluid flow through or around objects. For this study, a CFD-based algorithm was combined with the geological characteristics of a wind farm to achieve an accurate assessment of the wind power generation.

CFD simulated studies by State Key Laboratory of North China Electric [15], provide a day-ahead forecast for 15-minute interval. The results shown in Table 8 exhibit good accuracy with an annual mean absolute error (MAE) at less than 1.0 m/s. The absolute error for more than 85% of the forecasts is less than 2.5 m/s. It is important to note that these simulations were conducted for stable wind speeds.

It must be taken into consideration that the prediction, in the case of high variation of wind speed in a short period, would not have enough quality for energy production estimations.

Fan Num ber	Measure average Wind speed (m/s)	Predicted average wind speed (m/s)	Annual MAE (m/s)	Annual RMSE (m/s)
3#	6.85	7.22	1.90	2.70
43#	7.15	7.39	1.75	2.28
76#	7.08	7.15	1.76	2.35
117#	7.67	7.11	1.88	2.49

Table 8. Results of wind speed for 4 wind turbines [15]

X. SUPPORT VECTOR MACHINE METHOD-BASED ALGORITHMS

The prediction of the wind speed for this study was based on Ensemble Empirical Mode Decomposition (EEMD) of the original wind speed sequence, each sub-sequence with similar Sample Entropy (SE) are merged to improve the prediction efficiency. Secondly the method Least Square Support Vector Machine (LSSVM) is applied and optimized with the Particle Swarm Optimization (PSO) algorithm [16] Figure 21 shows the Structure of the prediction model.

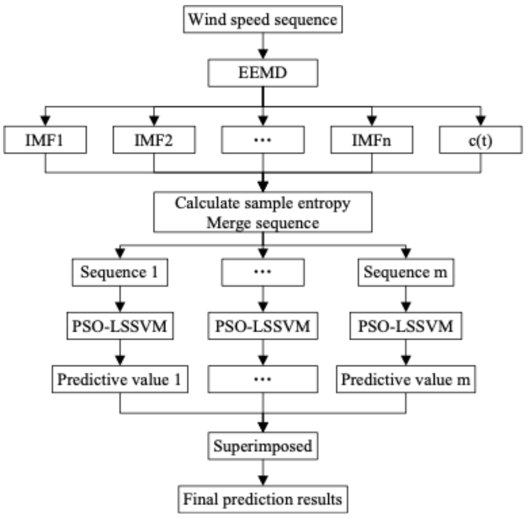


Figure 22. The Architecture of the Model [16]

The model is based on EEMD-SE-PSO-LSSVM, combined with the advantages of several methods, as shown in Fig. 21. The steps are as follows: use EEMD to decompose the original wind speed sequence to obtain the IMF components of wind speed; calculate the entropy of each series to improve the prediction efficiency; establish the PSO-LSSVM model for each sequence, and obtain the predictive values; then superimpose the predictive values of wind speed components and obtain the final wind speed predictive value; finally compare the results with actual wind speed data, and calculate the error index through error analysis [16].

EEMD is an improvement to the EMD method that adds Gaussian white noise to the original signal. It has less parameters to be selected and replace the inequality constraints with the equation constraints resulting in a reduced number of uncertain factors [16]

Figure 22 shows the prediction result of the combined model, namely the absolute error of the results.

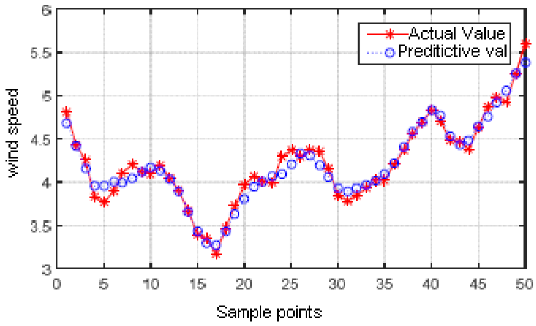


Figure 23. The results of the combined Model

Figure 23 shows the prediction result of LSSVM model [16]

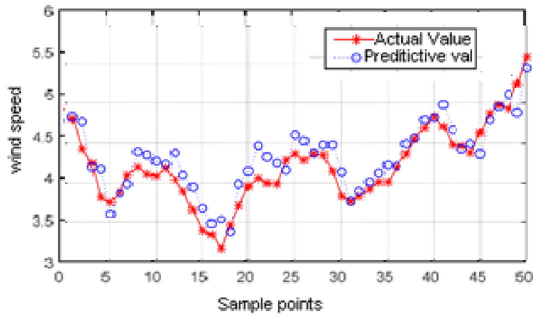


Figure 24. LSSVM prediction result [16]

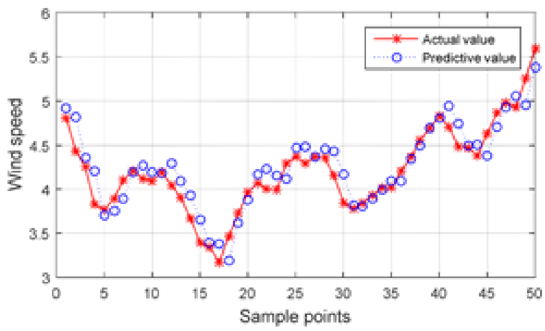


Figure 25. PSO-LSSVM prediction result [16]

Figure 24 shows the forecast results of PSO-LSSVM model. The RMSE and the MAE values for the three models, LSSVM method, PSO-LSSVM, and the proposed combined model (EEMD-SE-PSO-LSSVM) are shown in Table 9 [16].

MODEL	RMSE	MAE
LSSVM	0.2087	0.1736
PSO-LSSVM	0.1770	0.1439
Combine model	0.0941	0.0748

Table 9. Comparison of prediction errors of the three models

The Combined Model has a high practical value, with a mean square error and the average absolute error of 0.0941 and 0.0748 respectively.

XI. Comparative Analysis of the Models

In this section a comparative analysis of the prediction models is presented. A summary of the results is presented in table 10.

	Fluid dynamic	Fuzzy logic
Max measure	14.0 m/s	800 kW
RMSE	2.0 m/s	14.86 kW
Ratio	17.86 %	1.86 %

Table 10. Not Neural Network Results

Table 10 shows the great differences between Fluid Dynamics and Fuzzy Logic are Fluid Dynamics: Fuzzy Logic theory has shorter range prediction.

	Back Propagation NN ARMA	ARMA	Grey Model
Max measure	700 kW	21 m/s	1200 kW
RMSE	34.32 kW	0.216 m/s	54.2 kW
Ratio	4.62%	1.03%	4.52%

Table 11. Neural Network Results

Table 11 shows the results of three models: BPNN with ARMA, ARMA and Grey Model for the Neural Network. ARMA presents the best Ratio (1.03%) as compared with similarly poor results by the other 2 methods (4.62%, 4.52%).

	Empirical Mode Decomposition	Support Vector Machine
Max measure	14 kW	5.5 m/s
RMSE	19.07 kW	0.0941 m/s
Ratio	0.14 %	1.71 %

Table 12. Neural Network Results

Table 12 shows the clear advantage of the Empirical Mode Decomposition over the Support Vector Machine with a relative error of only 0.14%, an order of magnitude improvement.

XII. Conclusion

In the current context, its importance in the production of energy using renewable energy is undeniable. However, this importance sometimes misunderstood by politicians. No one doubts that a country's economic independence is related with energy independence. Countries are not only the obligation to comply with the Kyoto treaty for climate change, but the war in Ukraine and its energy implications, made the need for energy independence more visible.

Renewable energies are weather dependent, and their prediction is not easy. Namely the wind forecast to produce wind energy. Many researchers apply different models and mathematical forecasting techniques to predict wind characteristics, such as: wind speed, direction etc.

This paper made a brief compilation of wind prediction techniques.

A wide range of fuzzy logic-based algorithms were introduced and discussed in this paper. A proposed new hybrid approach for short-term wind power forecasting was implemented using ANFIS with two hierarchical stages. The results were compared with the Mean Absolute Percentage Error (MAPE), the Sum Squared Error (SSE), the Root Mean Squared Error (RMSE), and the Standard Deviation Error (SDE) for every season of the year. The average MAPE for this approach at 8.1133% for a day-ahead prediction, is a promising improvement.

In another approach, the wavelet theory aligned with an artificial neuronal network-based and phase-space reconstruction method was used to predict the power output. The results achieved are better than the ARMA method.

We can also conclude that a combination of the Grey Model, along with the discrete-Grey Model, and the fraction-Grey Model show improved results. Specifically, when compared with ARMIA as the benchmark, power output prediction compared with MAE, MAPE and RMSE was 37.7%, 34.9% and 34.4%, respectively.

The wind power forecasting has been modeled using the Recurrent Neural Networks, but the difficulties are learning long range dependencies [13]. LSTM networks have been used to overcome the limitations of RNN networks. In section IX a Support Vector Machine Method Based Algorithms was presented. It obtained the best results to prediction.

We can finally conclude that, adding a storage energy unit in combination with a good prevision power forecasting algorithm, provides an optimal solution for a stable and continuous power supply from the wind farm system to the power grid.

References

[1] Guo-rui Ji, Ze Dong, Dong-Feng Wang, Pu Han, Da-Ping Xu,” wind speed conformal prediction in wind farm based on algorithmic random- ness theory” july 2008.

[2] T. Eseye, J. Zhang, D. Zheng, H. Ma and G. Jingfu, "A double-stage hierarchical ANFIS model for short-term wind power prediction," 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing, China, 2017, pp. 546-551, doi: 10.1109/ICBDA.2017.8078694.

[3] Wang, Lijie & Dong, Lei & Hao, Ying & Liao, Xiaozhong. (2009). Wind power prediction using wavelet transform and chaotic characteristics. WNWEC 2009 - 2009 World Non-Grid-Connected Wind Power and Energy Conference. 1 - 5. 10.1109/WNWEC.2009.5335780.

[4] M. Mao, Y. Cao and L. Chang, "Improved fast short-term wind power prediction model based on superposition of predicted error," 2013 4th IEEE International Symposium on Power Electronics for Distributed Generation Systems (PEDG), 2013, pp. 1-6.

[5] Zhang Xinyu and Dong Lei, "A smooth scheme of wind power generation based on wind power prediction," Proceedings 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE), 2011, pp. 958-961

[6] Z. Chenhong, W. Penghui, Z. Yuan and Z. Yagang, "Wind speed prediction research based on time series model with residual correction," 2017 2nd International Conference on Power and Renewable Energy (ICPRE),

[7] Chengdu, China, 2017, pp. 466-470

[8] Patrick Siarry, e Zbigniew Michalewicz, Advances in Metaheuristics for Hard Optimization, Springer, 2008.

[9] El-Ghazali Talbi, Metaheuristics: From Design to Implementation, ISBN, Wiley, 2009.

[10] A. Madureira, C Ramos, SC Silva, Toward dynamic scheduling through evolutionary computing. WSEAS Transactions on Systems 3 (4), pp.1596-1604, 2004.

[11] A. M. Madureira, Meta-heuristics for the single-machine scheduling total weighted tardiness problem, Proceedings of the 1999 IEEE International Symposium on Assembly and Task Planning (ISATP'99), IEEE publisher, DOI: 10.1109/ISATP.1999.782992.

[12] A Madureira, Meta-heuristics application to scheduling in dynamic environments of discrete manufacturing, Ph. D. Thesis, Minho University, 2003 (in Portuguese).

[13] Yi ZHANG,Hexu Sun, KYingjun Guo, "Wind Power Prediction Based on PSO-SVR and Grey Combination Model" IEEE Open Access Journal, October 2019.

[14] S. S. Vaitheeswaran and V. R. Ventrapragada, "Wind Power Pattern Prediction in time series measurement data for wind energy prediction modelling using LSTM-GA networks," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019, pp. 1-5.

[15] B. Zhou, B. Sun, X. Gong and C. Liu, "Ultra-short-term prediction of wind power based on EMD and DLSTM," 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), Xi'an, China, 2019, pp. 1909-1913.

[16] Li Li, Yimei Wang, Yongqian Liu," Wind Velocity Prediction at Wind Turbine Hub Height Based on CFD Model", 2013, International Conference on Materials for Renewable Energy and Environment (ICMREE).

[17] Y. Yue, Y. Zhao, H. Zhao and H. Wang, "Short-term wind speed combined prediction for wind farms," 2017 IEEE International Conference on Mechatronics and Automation (ICMA), 2017, pp. 18-22.

[18] Ricardo Puga, José Baptista, José Boaventura, Judite Ferreira and Ana Madureira, "State of the Art of Wind and Power Prediction for Wind Farms" Proceedings of 11th World Congress on Information and Communication. Technologies (WICT 2021), online, December 16-17, 2021.

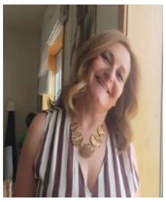
[19] Extended fuzzy Petri net for multi-stage fuzzy logic inference",VolumShahid Farah, Wood David A, Nisar Humaira, Zameer Aneela, Eger Steffen ,"Short-term multi-hour ahead country-wide wind power prediction for Germany using gated recurrent unit deep learning, Renewable and Sustainable Energy Reviews", pag. 167, 2022,112700, ISSN 1364-0321.

[20] A Madureira, I Pereira, N SousaSelf-organization for scheduling in agile manufacturing, 2011 IEEE 10th International Conference on Cybernetic Intelligent Systems, 2011

[21] A Madureira, B Cunha, JP Pereira, S Gomes, I Pereira, JM Santos, Using personas for supporting user modeling on scheduling systems, 4th International Conference on Hybrid Intelligent Systems, 279-284, 2014.

[22] B. Cunha, A Madureira, B Fonseca, J Matos, Intelligent scheduling with reinforcement learning, Applied Sciences 11 (8), 3710, 2021.

Author Biographies



Judite Ferreira was born in Portugal in 1969. She received the M.S. degree in Electrical and Computers Engineering 2014, and Ph.D. in same area in 2017. She is Professor in Electrical Engineering Department at the Institute of Engineering (ISEP/P.PORTO) since 1999, and has a significant experience in power system. She has been involved in R&D projets. Her research interests include energy market, power systems, Smart Grid control. Hydrogen Market, LMP hydrogen.

Ricardo Puga was born in Portugal in 1994. He received M.S. degree in Electrical and Computers Engineering, in 2018 and beginning of a Ph.D. in the same area. He has two years of technology consulting experience in Salesforce and one year in the Microsoft Dynamics Area.

He is passionate about developing projects with social and economic impact within a team, but he can face projects on his own as well.

J. Boaventura-Cunha Born in Vila Real, Portugal, 24 November 1961, graduated in Electronics and Telecommunications Engineering from the University of Aveiro, Portugal, 1985. He received the Ph.D. degree in Electrical and Computer Engineering, University of Trás-os-Montes e Alto Douro (UTAD), Vila Real, Portugal, 2002. Presently he is an Associate Professor with Habilitation at the School of Science and Technology, UTAD University, and he is a senior researcher at the INESC TEC, Porto Portugal. His research interests include modeling, system identification and adaptive control applied to agroforestry and industry.

André Serra e Santos was born in Porto. He received the B.S. in Electrical and Computers Engineering, M.S. in Electrical and Computers Engineering, M.S. in Mechanical Engineering and a Ph.D. in Production and Systems. He is Professor in Mechanical Department at the Institute of Engineering (ISEP/P.PORTO) and his interests are Production Management, Scheduling of Operations, Approximate Techniques and Meta-Heuristics.