

IOP Conference Series: Earth and Environmental Science

PAPER • OPEN ACCESS

Wind speed time series reconstruction using a hybrid neural genetic approach

To cite this article: H Rodriguez *et al* 2017 *IOP Conf. Ser.: Earth Environ. Sci.* **93** 012020

View the [article online](#) for updates and enhancements.

Related content

- [Operational management of offshore energy assets](#)
A J Kolios and M Martinez Luengo
- [Comments from *physicsworld.com*](#)
- [Testing and comparison of lidars for profile and turbulence measurements in wind energy](#)
M Courtney, R Wagner and P Lindelöw

Wind speed time series reconstruction using a hybrid neural genetic approach

H Rodriguez^{1,5}, J J Flores², V Puig³, L Morales⁴, A Guerra¹ and F Calderon²

¹Instituto Tecnológico de Culiacán, División de Estudios de Posgrado e Investigación, Juan de Dios Bátiz 310 pte. Col Guadalupe C.P. 80220, Culiacán, México

²División de Estudios de Postgrado, Facultad de Ingeniería Eléctrica, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, México

³Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Carrer LLorens Artigas, 4-6, Barcelona (Spain), 08028

⁴Facultad de Ingeniería Civil, CONACYT-Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, México

E-mail: hrodriguez@itculiacan.edu.mx

Abstract. Currently, electric energy is used in practically all modern human activities. Most of the energy produced came from fossil fuels, making irreversible damage to the environment. Lately, there has been an effort by nations to produce energy using clean methods, such as solar and wind energy, among others. Wind energy is one of the cleanest alternatives. However, the wind speed is not constant, making the planning and operation at electric power systems a difficult activity. Knowing in advance the amount of raw material (wind speed) used for energy production allows us to estimate the energy to be generated by the power plant, helping the maintenance planning, the operational management, optimal operational cost. For these reasons, the forecast of wind speed becomes a necessary task. The forecast process involves the use of past observations from the variable to forecast (wind speed). To measure wind speed, weather stations use devices called anemometers, but due to poor maintenance, connection error, or natural wear, they may present false or missing data. In this work, a hybrid methodology is proposed, and it uses a compact genetic algorithm with an artificial neural network to reconstruct wind speed time series. The proposed methodology reconstructs the time series using a ANN defined by a Compact Genetic Algorithm.

1. Introduction

Energy plays a very important role in human activities. Thus, energy production has become one of the main economic and environmental issues worldwide [1]. In addition, the energy demand has increased steadily over time, causing environment damage when producing it. According to the census conducted by Observ'ER & Foundation Énergies pour le Monde in 2012, the dominant production of energy is the one produced through fossil fuels, covering more than two-thirds of total energy production (68.1%) [1]. This type of energy, besides being non-renewable, generates irreversible environmental damage to our planet. Due to the large amounts of carbon dioxide (CO₂) emitted into the atmosphere, thereby increasing greenhouse gases.

A mechanism that seeks to counteract environmental damage and reduce energy production costs is the use of renewable energy, also called clean energy. A wide range of options within the renewable



energies such as: solar, wind, biomass, geothermal, non-renewable waste and hydro, among others were found. Of the above, solar and wind energy have presented an annual growth between 2002 and 2012 of 50.6% and 26.1%, respectively [1].

Taking into account the great and continuous growth of these renewable energies, many challenges have been encountered (meeting demand, lowering production costs, improve energy plant planning, etc.). So to reach a full adoption of this type of energy is necessary to find effective solutions to these challenges. One of the main ones is the necessity of making predictions of the variables involved in the production of energy. That is, to have a certainty of the amount of energy to be produced in the next few minutes, hours, days, or months [2]. Some of the variables involved in the production of clean energy could be the following: water flows, tides and internal heat of the earth, etc. This work deals with wind speed forecast.

The prediction of these variables is usually performed in the short, medium and long term. There is no definite time frame for each of the periods [3]. But regularly medium and long-term forecasting is used for tactical and strategic planning, respectively. The short-term prediction is used at the operational level and its periodicity could be in time scale of minutes, hours, days [3,4].

The ability to forecast wind speed is essential for the correct integration of wind energy into electrical systems. According to Barber *et al* [5], the importance of wind forecasts for the wind energy industry stems from three facts:

- The aggregate power produced and consumed throughout a power system must be nearly in balance at all times to guarantee supply reliability and safety.
- The power output of a wind farm is highly variable since it depends strongly on wind speed and direction.
- An efficient and cost effective energy storage mechanism does not exist.

Recently, an estimation for the year 2030 mentions that a perfect forecast will be valued at 3 Billion US dollars annually [6], for the United States power system; the estimation was performed by the Department of Energy [5]. Therefore, wind speed forecasting still plays an important role in the task of energy supply [4].

In literature different approaches employed for this task were found [7-10]. Chang presents [4] a survey that categorizes different approaches that deal with the forecasting problem establishing Persistence, Physical, Statistical, Spatial Correlations, Artificial Intelligence, and Hybrid Methods. The Persistence method (or Naïve) is a basic forecasting method, where $Y_{t+\Delta} = Y_t$ [11]. This particular method is often used as a baseline method. The Physical methods are numerical weather predictions developed by meteorologists for a large scale weather prediction [12]. The Statistical methods (SM) uses historical data to find relationships inside the wind speed time series. Within the Statistical methods are the Auto Regressive (AR), Auto Regressive Moving Average (ARMA), Auto Regressive Integrate Moving Average (ARIMA) [13-15]. Artificial Intelligence methods (AIM) uses Artificial Neural Networks (ANN) [14,16], Support Vector Machine (SVM) [17], k-Nearest Neighbor (kNN) [18], among others as a forecast method. AIM improve in some cases the results obtained using statistical methods. Hybrid Methods are a combination of forecasting methods [14] (e.g. combination of statistical and artificial intelligence approaches).

SM (e.g. ARIMA, ARMA, Holt-Winters, etc.) or AIM (e.g. ANN, SVM, KNN) usually perform the forecast using Historical data. But, if the dataset presents missing values inside, the SM may have problems to perform an adequate forecast. In reality, wind time series present missing data due to problems with the sensors (e.g. communication problems, natural wear, lack of maintenance, etc.). The wind speed is irregular and intermittent [19] such that added to the fact that the time series are incomplete, the task of short term forecast is even harder. This work deals with the problem of the wind speed time series reconstruction. This work starts by characterizing the temporal behavior of the times series, creating a pattern database. Then, it uses a hybrid method to model the time series. This method uses an ANN as a forecasting approach and Genetic Algorithms (GA) to define the best topology of the ANN. Once, the ANN was trained, it is possible to reconstruct the wind speed time

series.

The rest of the paper is organized as follows: Section 2 describes the methodology proposed to reconstruct a wind speed time series, Section 3 presents the experiments performed with our proposal, and finally the conclusions of this work are presented in Section 4.

2. Time series reconstruction

Dealing with the problem of forecasting wind speed time series is not something new. It is a complicated task since there is a wide range of different behavior patterns in the data to model. Adding the problem that the time series could be incomplete, the problem becomes even more complicated, although, nowadays there are a number of algorithms and techniques that offers an effective forecasting [4,11]. Nevertheless, in most of the literature, the problem of wind speed forecasting is addressed assuming that the time series is complete (without any missing data). However, this is not the case in reality. Regularly, there are problems with sensors giving as a result missing data within the wind speed time series. Thus, to deal with this problem using Statistical Methods as ARIMA or Holt-Winters becomes a hard work, or even impossible.

To deal with the missing values that are within the time series, a temporality removal of the time series was proposed. Instead of having a sliding window of size k , the Temporality Removal Process creates a database where k features of the time series are recorded, associated with its respective forecasting value. With the database created, the Construct the Time Series Model (CTSM) process uses an ANN for the regression problem and instead of using the trial error method for selecting the optimal topology, a Compact Genetic Algorithm was used to define the optimal topology. Once the optimal ANN structure is defined, the Data Reconstruction Process ends the process using the ANN trained to reconstruct the missing data. Figure 1 shows the flow chart of the wind speed time series reconstruction process.

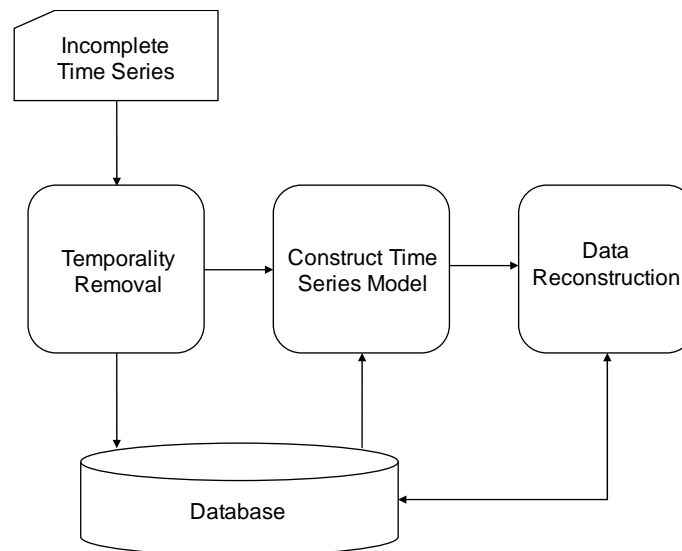


Figure 1. Flow chart of the wind speed time series reconstruction process.

2.1. Temporality removal process

Regularly, for the forecasting problem, there is a window of size k that slides through the time series. But because of the missing values inside the time series, this window can no longer slide in the same way. Therefore, a database with the time series temporal patterns is created. The database records the k (window size) characteristics and associates them with their respective values forecasting value (y_{t+1}). After creating the database, all records where missing data is present are deleted, leaving aside the time dimension.

Table 1 shows the result of creating the database and extracting the k features associated with an expected output. Each record inside the database represents the last k observations at a defined time (t), associated to the expected output y_{t+1} (for training purposes). Table 1 shows records with the presence of missing values (φ), this records (φ) are deleted. With the database defined (without missing data), it is possible to start searching for an optimal ANN that models the behavior of the wind speed time series.

Table 1. Transformation the time series into a database structure.

Inputs (K_t)						\hat{y}_{t+1}
k	y_1	y_2	y_3	...	y_k	y_{k+1}
$k+1$	y_2	y_3	y_4	...	y_{k+1}	y_{k+2}
$k+2$	y_3	y_4	y_5	...	y_{k+2}	φ
$k+3$	y_4	y_5	y_6	...	φ	y_{k+4}
$k+4$	y_5	y_6	y_7	...	y_{k+4}	y_{k+5}
$k+5$	y_6	y_7	φ	...	y_{k+5}	y_{k+6}
$k+6$	y_7	φ	y_9	...	y_{k+6}	y_{k+7}
$k+7$	φ	y_9	y_{10}	...	y_{k+7}	y_{k+8}
$k+8$	y_9	y_{10}	y_{11}	...	y_{k+8}	y_{k+9}
$k+9$	y_{10}	y_{11}	y_{12}	...	y_{k+9}	y_{k+10}
	
$n-1$	y_{n-k}	y_{n-k+1}	y_{n-k+2}	...	y_{n-1}	y_n

2.2. Construction of a time series model

In recent works, ANN has been proved to present great classification and regressions qualities. ANNs have their basis in the study of the human brain, trying to simulate the process of neuron interconnections. ANNs are thus able to learn from previous knowledge provided at the training phase, regardless of the type of task assigned to them. That is why they are useful in different fields, such as industry and science [4].

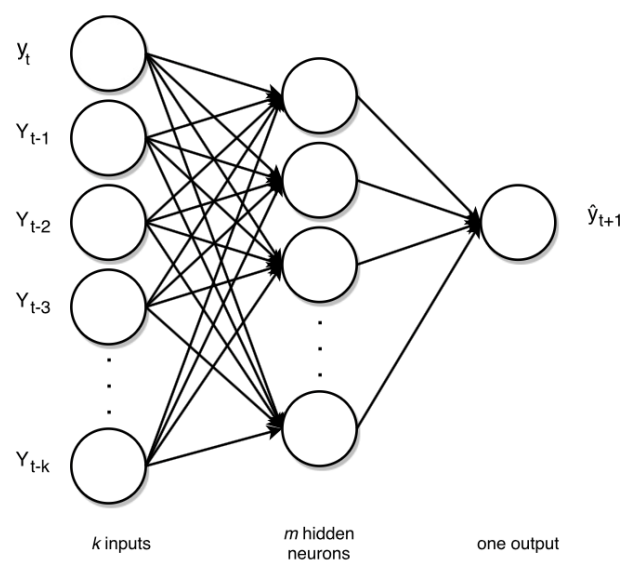


Figure 2. ANN architecture with $k+1$ past observations, m hidden neurons and a single output (\hat{y}_{t+1}).

Here, the ANN architecture proposed is a Feedforward Multilayer Perceptron trained, by gradient based methods. Figure 2 shows the architecture of the MLP ANN proposed in this work. An ANN, as a universal approximator can learn any given function. A set of k past observations are considered as the input data, the hidden layer consists of a m hidden neurons, the output layer corresponds to the forecast value \hat{y}_{t+1} ; A sigmoid is used as the activation function. The ANN that models the wind speed time series \hat{y}_{t+1} can be defined as:

$$\hat{y}_{t+i} = f_1 \left(\sum_l^{k+1} w_l x_l \right) \quad (1)$$

$$x_l = f_2 \left(\sum_j^m w_{lj} y_t - l_j \right)$$

where f_1 and f_2 are the activation functions, and w are the coefficients (also known weights connection).

In the wind speed time series modelling problem, it is necessary to provide an accurate model that defines the behavior of the wind speed. To define the accuracy of the optimal model, the Mean Square Error statistical measure is used, which is defined as

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (2)$$

This measure will be minimized during the training process. GA is an optimization technique inspired on Darwin's principle of evolution. That is, it mimics a simplistic version of the process of biological evolution, which consists of creating a population of individuals, where each individual represents a prospective solution of the problem being solved. GA modifies this population using genetic operators: selection, mutation, recombination, etc. [20]. cGA is a variant of the simple GA in which a probability distribution is used to represent the population. Therefore, it requires less memory than the simple GA. The cGA can be used to give a quick answer to define an optimal architecture of an ANN [21].

Table 2. Training algorithms.

1	Gradient descent BP
2	Gradient descent with momentum BP
3	Gradient descent with adaptive learning rate
4	Gradient descent with momentum BP and adaptive
5	Resilient BP
6	Broyden-Fletcher-Goldfarb-Shanno
7	Newton conjugate gradient
8	Conjugate gradient

Determining the best ANN architecture and training algorithm is an optimization problem. This is, defining the architecture of the ANN by determining the number of inputs, the number of neurons in the hidden layer, and and the training algorithm. This is a critical step in the training process that is usually performed by trial and error. In previous works, the use of GA has been tested yielding excellent results [13]. The difference here is the use of a cGA, which is an improvement of a simple GA, also several training algorithms are considered. Each individual (the chromosome) in the GA is defined as a binary vector that is coding the number of inputs, number of hidden neurons and the training algorithm of the ANN. Algorithm uses GA to finds the best ANN topology and describes step

by step the used process. The cGA starts by defining the probability vector, from this vector two individuals are generated and evaluated. The evaluation process is used to test the topology proposed by the cGA. The training algorithms used in this approach are defined in table 2. The result of the training algorithm corresponds to the individual's evaluation. Both individuals compete, and according to the winner individual, the probability vector is updated. This procedure repeats until a convergence criterion is achieved.

The genetic algorithm basically provides as the result of how many inputs will be optimal for the model, how many neurons must be considered for the hidden layer of the network and which training algorithm should be used. At the end of the process, the best individual (i.e., the one presenting best fitting) found during the evolution is returned as the solution of the problem (topology of the ANN). After defining the topology of the ANN, a refinement process is started. Refinement is the same process as the evaluation process, but the number of epochs is increased significantly.

2.3. Data reconstruction

For the reconstruction process, the ANN that models the behavior of the wind speed time series is used. Using this model, it is possible to start with the reconstruction process.

Algorithm 1 Finding the best topology algorithm using cGA

```

1: procedure GA( $n$  as population size,  $l$  as chromosome length)
2:   Initialize probability vector
3:   for  $i := 1$  to  $l$  do  $p[i] := 0.5$ ;
4:   Generate two individuals from the vector
5:    $a := \text{generate}(p)$ ;
6:    $b := \text{generate}(p)$ ;
7:    $\text{evaluate}(a)$ ;
8:    $\text{evaluate}(b)$ ;
9:    $\text{winner}, \text{loser} := \text{compete}(a, b)$ ;
10:  //Update the probability vector towards the better one
11:  for  $i := 1$  to  $l$  do
12:    if  $\text{winner}[i] \neq \text{loser}[i]$  then
13:      if  $\text{winner}[i] = 1$  then  $p[i] := p[i] + 1/n$ 
14:      else  $p[i] := p[i] - 1/n$ ;
15:  //Check if the vector has converged
16:  for  $i := 1$  to  $l$  do
17:    if  $p[i] > 0$  and  $p[i] < 1$  then return to step 4;
18:  //p return the final solution

```

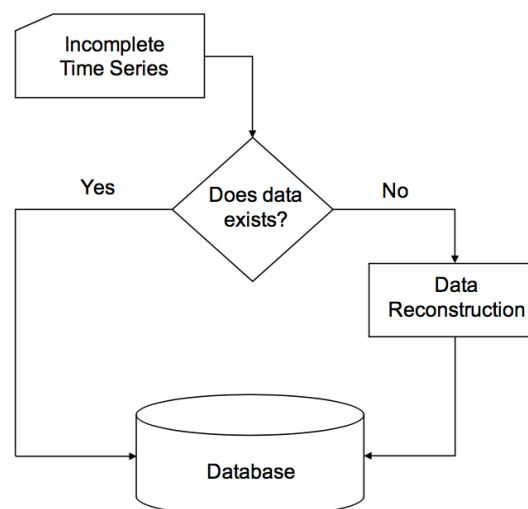


Figure 3. Reconstruction process for wind time series.

This process starts using a vector ordered observations that are evaluated sequentially. When a missing value in the time series is found, the reconstruction data process starts. This process uses the k last observations of the vector used as input. The output of the ANN already trained in previous steps is used to fill the missing value. The process continues until the end of the observation vector is reached. Figure 3 shows the flow chart of the reconstruction process.

3. Results

In order to test the suitability of our proposal, experiments were carried out with time series of the wind speed coming from anemometers located in different locations inside the the state of Michoacán, México recorded hourly. Since the sensors are susceptible to miss-functions and reading errors, the problem of obtaining time series with missing data was encountered. In some cases, this could be a quite important problem since the time series contains more that 5,000 adjacent missing data.

We perform several experiments applying the approach described in figure 1. For the lack of space, in this section, the results of one of the experiments are described. Each experiment starts by removing the temporality of the time series by creating a database following the procedure described in the previous section. Once the database has been obtained, an optimal ANN that models the behavior of the wind speed time series is defined. The optimal topology obtained at this stage is described in table 3

Table 3. The winner topology obtained by the cGA.

Input Neurons	25
Hidden Neurons	28
Training Algorithm	Conjugate

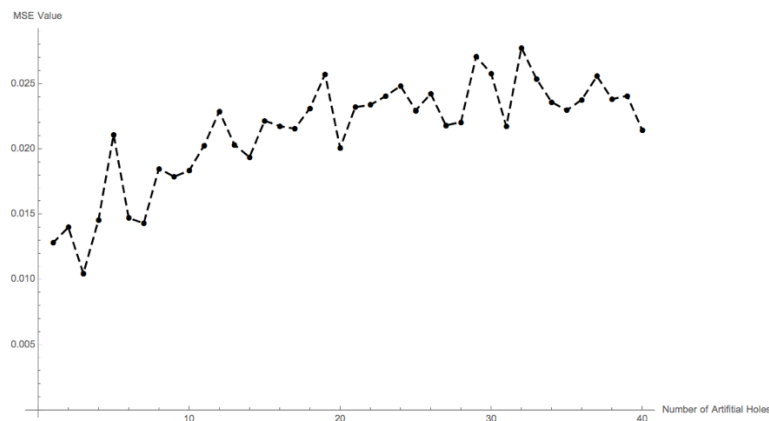


Figure 4. Experiment No. 1 MSE results for 1 to 40 artificial missing values.

The optimal topology obtained in this stage is used in a refinement process, which is only increasing the number of epochs of the evaluation function of the cGA. In this case, 3,000 epochs were used.

In order to test the accuracy of the reconstruction process proposed in this work, a new training set in which we artificially place missing values of size h was created. Once the missing values were generated, the reconstruction process is started. With this action were able to compare the results of the reconstruction with real information (reconstruction simulation). In the experiments, the proposed approach was tested by generating artificial training sets with missing data up to size 40. Figure 4 shows the MSE values for the performed experiments. Table 4 shows numerically the results. From figure 4, we observe that as the size of the missing data grows, the accuracy is gradually lost, but this accuracy loss is not very significant. The implementations have been developed under Python platform

using the library NeuroLab 0.3.5 [22].

Table 4. Experiment No. 2 time series MSE for 1 to 40 missing values.

Number of holes	MSE	Number of holes	MSE	Number of holes	MSE	Number of holes	MSE
1	0.012802972	11	0.020267953	21	0.02319457	31	0.021700934
2	0.013979674	12	0.022868016	22	0.023364267	32	0.027749779
3	0.010446051	13	0.020316086	23	0.024026512	33	0.025341943
4	0.01453436	14	0.019358147	24	0.024815368	34	0.023582028
5	0.021092693	15	0.022149374	25	0.022944667	35	0.022948563
6	0.014692938	16	0.021738649	26	0.024217165	36	0.023763214
7	0.014295101	17	0.021536263	27	0.021791693	37	0.025565695
8	0.018491887	18	0.023112223	28	0.022043973	38	0.023803695
9	0.017864363	19	0.02571133	29	0.027073025	39	0.024064536
10	0.018317789	20	0.020054946	30	0.025773363	40	0.02142727

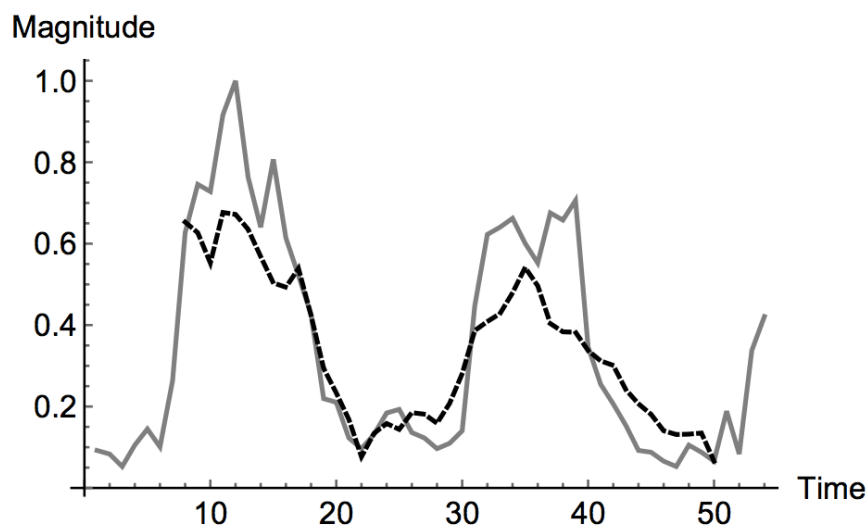


Figure 5. Reconstructed time series.

A graphical example of the reconstruction process is shown in figure 5. In the figure 5, the continuous line represents the real data, and the dotted ones represent the reconstruction.

4. Conclusions

This paper has proposed a time-series reconstruction approach based on a hybrid neural- evolutionary methodology. This methodology allows to determine the optimal ANN architectures to model incomplete time series and later to be able to reconstruct them. Experiments were performed considering multiple time series with artificially added missing data in the different phases: model generation and data reconstruction. A correlation was determined between the size of the missing data and the error obtained from the data validation, showing a positive trend between the size of the missing values present and the error obtained. The experiments were performed using the under Python platform using the NeuroLab library [22] for the ANN implementations.

References

- [1] Observ E 2013 *Worldwide Electricity Production from Renewable Energy Sources* (Paris,

- France: Ninth Inventory)
- [2] Rodriguez Rangel H, Puig V, Farias R L and Flores J J 2016 Short-term demand forecast using a bank of neural network models trained using genetic algorithms for the optimal management of drinking water networks *J Hydroinformatics* **19** 1-16
 - [3] Giacomoni M, Kanta L and Zechman E 2013 Complex adaptive systems approach to simulate the sustainability of water resources and urbanization *J Water Resour Plan Manage* **139** 554-64
 - [4] Chang W-Y A 2014 literature review of wind forecasting methods *J Power Energy Engin* **2** 161
 - [5] Barber C, Bockhorst J and Roebber P 2010 Auto-regressive hmm inference with incomplete data for short-horizon wind forecasting *Advances in Neural Information Processing Systems* (California, USA: Curran Associates, Inc.) pp 136-44
 - [6] Piwko D and Jordan G 2010 The economic value of day-ahead wind forecasts for power grid operations *2010 UWIG Workshop on Wind Forecasting*
 - [7] Cassola F and Burlando M 2012 Wind speed and wind energy forecast through kalman filtering of numerical weather prediction model output *Appl Energy* **99** 154-66
 - [8] Liu H, Tian H-Q, Pan D-F and Li Y-F 2013 Forecasting models for wind speed using wavelet, wavelet packet, time series and artificial neural networks *Appl Energy* **107** 191-208
 - [9] Foley A M, Leahy P G, Marvuglia A and McKeogh E J 2012 Current methods and advances in forecasting of wind power generation *Renew Energy* **37** 1-8
 - [10] Alexiadis M, Dokopoulos P, Sahsamanoglou H and Manousaridis I 1998 Short-term forecasting of wind speed and related electrical power *Solar Energy* **63** 61-8
 - [11] Zhao X, Wang S and Li T 2011 Review of evaluation criteria and main methods of wind power forecasting *Energy Procedia* **12** 761-9
 - [12] Ma L, Luan S Y, Jiang C W, Liu H L and Zhang Y 2009 A review on the forecasting of wind speed and generated power *Renewable and Sustainable Energy Reviews* **13** 915-20
 - [13] Flores J J, Loeza R, Rodriguez H and Cadenas E 2009 Wind speed forecasting using a hybrid neural-evolutive approach *MICAI 2009: Advances in Artificial Intelligence* (Guanajuato, México: Springer) pp 600-9
 - [14] Flores J J, Graff M and Rodriguez H 2012 Evolutive design of arma and ann models for time series forecasting *Renew Energy* **44** 225-30
 - [15] Erdem E and Shi J 2011 Arma based approaches for forecasting the tuple of wind speed and direction *Appl Energy* **88** 1405-14
 - [16] Barbounis T G, Theocharis J B, Alexiadis M C and Dokopoulos P S 2006 Long-term wind speed and power forecasting using local recurrent neural network models *IEEE Trans Energy Conversion* **21** 273-84
 - [17] Zhou J, Shi J and Li G 2011 Fine tuning support vector machines for short-term wind speed forecasting *Energy Conver Manage* **52** 1990-8
 - [18] Yesilbudak M, Sagiroglu S and Colak I 2013 A new approach to very short term wind speed prediction using k-nearest neighbor classification *Energy Conver Manage* **69** 77-86
 - [19] Hayashi M and Kermanshahi B 2001 Application of artificial neural network for wind speed prediction and determination of wind power generation output *Proceedings of ICEE* pp 12-5
 - [20] Rangel H R, Puig V, Farias R L and Flores J J 2016 Short-term demand forecast using a bank of neural network models trained using genetic algorithms for the optimal management of drinking water networks *J Hydroinformatics* doi:10.2166/hydro.2016.199.
 - [21] Harik G R, Lobo F G and Goldberg D E 1997 The compact genetic algorithm, Urbana 5161801
 - [22] <https://pypi.python.org/pypi/neurolab>