Renewable Energy 92 (2016) 202–211

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Time series prediction using artificial wavelet neural network and multi-resolution analysis: Application to wind speed data

Boubacar Doucoure, Kodjo Agbossou, Alben Cardenas*

Department of Electrical and Computer Engineering - Hydrogen Research Institute, University of Quebec at Trois Rivières (UQTR), Trois Rivières, Quebec, Canada

ARTICLE INFO

Article history:
Received 23 December 2014
Received in revised form
22 December 2015
Accepted 2 February 2016
Available online 15 February 2016

Keywords:
Wind speed forecasting
Adaptive wavelet neural network
Multi-resolution analysis

ABSTRACT

The aim of this work is to develop a prediction method for renewable energy sources in order to achieve an intelligent management of a microgrid system and to promote the utilization of renewable energy in grid connected and isolated power systems. The proposed method is based on the multi-resolution analysis of the time-series by means of Wavelet decomposition and artificial neural networks. The analysis of predictability of each component of the input data using the Hurst coefficient is also proposed. In this context, using the information of predictability, it is possible to eliminate some components, having low predictability potential, without a negative effect on the accuracy of the prediction and reducing the computational complexity of the algorithm. In the evaluated case, it was possible to reduce the resources needed to implement the algorithm of about 29% by eliminating the two (of seven) components with lower Hurst coefficient. This complexity reduction has not impacted the performance of the prediction algorithm.

1. Introduction

The renewable energy sources (RES) are emerging as one of the best alternatives for sustainable electricity generation. The transition of the traditional energy systems towards renewable sources is required to reduce green-house gas emissions, and consequently to decelerate the global warming [1,2]. Different types of RES technologies are nowadays available for electricity generation. The wind turbines and the photo-voltaic modules are profitable alternatives for areas with high electricity costs, and are promoted by governments to reinforce the future smart grid. In fact, these sources have experienced strong energy market growth in the past few years. The optimal integration of renewable sources, as they are intermittent, exhibit fundamental challenges including the energy storage, the power conversion and the prediction of the available power [3,4].

The performance and efficiency of the renewable energy systems are thoroughly related to the power flow management between all components of the system. The complete energy storage system for renewable sources is not still technologically and economically well developed. However, the overproduction and the demand management could be better performed if the RES power profile is known in advance. Nonetheless, the electricity production from wind energy has some particularities derived from the intermittent behavior of wind speed, e.g. the production profile can not be adjusted satisfactorily to the one of the load demand; a balance of production and demand is then required which can be achieved using power reserves from other energy sources and/or a storage system support [5,6]. Scientific and industrial research efforts are unceasing to develop more accurate and reliable forecasting tools to mitigate the problem of irregularity in the RES power production [7]. The aim of those efforts seeks specially short term forecast and their applications on wind power management; they would enable e.g. the scheduling of the energy requirements for a given period (planning and delivering), solving more accurately and safely the micro-grid constraints and the schedule of maintenance [8–10].

Profiles of the available power of wind and solar sources depend on the geographic location, the season (or time of the year), the time of the day and other physical parameters. Wind speed modeling using time series is usually used to analyze wind profiles to obtain the predicted values. In the context of local energy management, the study of time series by means of a predictability analysis can be very helpful. Predictability analysis of time series

* Corresponding author.
E-mail address: Alben.cardenasgonzalez@uqtr.ca (A. Cardenas).

http://dx.doi.org/10.1016/j.renene.2016.02.003
0960-1481/

has been introduced by Hurst [11] and can facilitate the selection of the prediction method which is essential for energy management.

In the literature, various approaches have been developed for wind speed prediction in renewable energy systems. Different prediction horizons have been studied; these methods have strengths and weaknesses and vary according to the context or environment of prediction. The framework prediction horizons are usually defined as follows: long-term (from one day to one week ahead), medium-term (from 6 h to one day ahead), short-term (from 30 min to 6 h ahead) and very short-term (few seconds to 30 min ahead) [12].

Three main categories of wind forecasting are frequently related in the literature: the statistical, the physical and the hybrid approaches [12,13]. Statistical approaches consist of time-series prediction based on historical data. These techniques have been the most popular methods in practice. Normally, the statistical analysis does not consider the meteorological data. In this group there are included the simple persistent model, the autoregressive models including the Auto Regressive (AR) process, the Auto Regressive Moving Average (ARMA), the Auto Regressive Integrated Moving Average Model (ARIMA) and the artificial neural networks (ANN). In Ref. [14] the ARMA model is used to predict the wind speed and direction. In Ref. [15] an ARMA-ARCH structure is proposed to predict the wind speed. A second subgroup can be identified in the statistical methods, usually involving artificial intelligence (AI) techniques. Other examples of wind speed prediction methods are the work related in Ref. [16] which proposes an AWNN, the Neural Network proposed in Refs. [17], the Wavelet and Neural Network model proposed in Ref. [18] and the Wavelet-PSO-ANFIS in Ref. [19].

The physical approaches are also known as Numerical Weather Prediction (NWP) methods; these techniques predict the wind speed by using meteorological (temperature, pressure, humidity, etc.) and topological information. The main disadvantages of NWP are their complexity of construction and operation, and their inaccuracy for short term forecast. The NWP combined with Gaussian Processes and Empirical Model has been employed to predict wind power in Refs. [20] and [21], respectively.

The last group of forecasting methods is the hybrid model, based on the combination of the two previous forecasting groups. Hybrid methods also give good results in short-time predictions. However, these methods are generally more complex and expensive to achieve. Some examples are the hybrid model proposed in Refs. [22]; and the hybrid model with Ensemble Empirical Mode Decomposition (EEMD) and the Support Vector Machine (SVM) presented in Ref. [23].

Long-term forecasting relies on NWP models, the statistical models used there focus on the short-term forecasts that can be useful in the hour-ahead utilities. The complex and nonlinear nature of wind speed and the ampleness of historical data put forward that AI-based techniques would be suitable candidates for wind speed and power production prediction.

This paper investigates the predictability analysis of the time-series of wind power and its application in a forecasting method based on the adaptive wavelet neural network (AWNN). The Hurst predictability and multi-resolution analysis (MRA) of time-series decomposition is applied to the wind power profile; this study shows that the wind speed behavior can be decomposed in fairly, mildly and definitely not predictable time series. A forecast strategy for the wind power is also proposed using a modified version of the adaptive wavelet neural network with the maximum overlap discrete wavelet transform (AWNN-MODWT). Knowing the decomposition predictability, the size of the forecasting tool can be optimized without accuracy degradation compared to the conventional methods proposed in the literature. This tool estimates the predictability score of each MRA component, with this information, the respective forecast AWNN can be eliminated. Consequently, a better generalization in the prediction network is obtained, avoiding over-fitting due to a large set of parameters. The proposed approach can easily be extended to the demand forecast. It can also be used for the identification of stochastic components in the production and consumption profiles, facilitating the implementation of control algorithms.

This document is organized in 5 sections. The theoretical aspects about the forecasting of time-series, the multi-step ahead forecasting concepts and the MRA-AWNN architecture are discussed in Section 2. The Hurst predictability analysis of wind speed is investigated in Section 3, the results are discussed in Section 4. Finally some concluding remarks are presented in Section 5.

2. Theoretical background

The time-series is a sequence of numeric observations from a particular measured variable. These observations are usually made on a regular basis (days, months, years), but sampling may be irregular or not. To represent the time-series prediction process, consider a sequence $X_N = \{x_1, \ldots, x_N\}$ comprising N observations. The multi-step ahead prediction consists of predicting $\{x_{k+1}, \ldots, x_{k+h}\}$ values of X_N, where $k = 1, 2, \ldots, N$ is the index of x and $h > 1$ is the index corresponding to the number of points between the present and the future, usually called the prediction horizon [24].

A time-series analysis consists of building a model that represents a time-series; using this model the future values are predicted only from past values. The data are described by a possible linear or non-linear autoregressive process of the form:

$$x(k+h) = f[x(k), x(k-1), x(k-2), \ldots, x(k-n-1)] \quad (1)$$

were f is a function describing the relationship between the past values of x and the present.

Currently, multi-step ahead prediction consists of predicting the h next values of the time-series. This task is achieved by two deferent ways. The first one, called independent value prediction, consists of training a direct model to predict $x(k+h)$. The second strategy, called iterative method, consists of repeating one-step ahead predictions to the desired horizon; this principle is shown in Fig. 1. The iterative prediction only uses one model to forecast all the horizons needed; the objective is to analyze a short sequence of data, and try to predict the rest of the data sequence until a predefined time-step is reached [25]. The main drawback of this approach is that the error in nearest horizons can be transmitted to others horizons.

2.1. Time-series prediction using AWNN

The wavelet networks calculate a linear combination function of

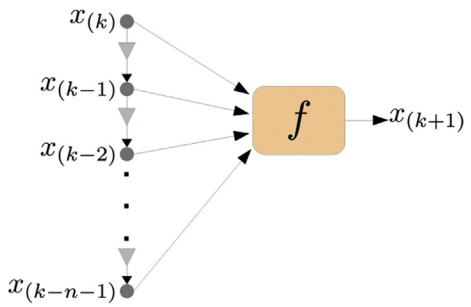


Fig. 1. Basic principle for one-step-ahead time series prediction ($h = 1$).

adjustable parameters. The changes in the parameters can be nonlinear including translations or dilations. Some characteristics relative to the wavelet networks are: the wavelets have a rapid decrease and go to zero as opposed to the back-functions of the neuron networks; and the shape of the wavelet is bonded to two structural parameters [16,26,27].

The wavelet analysis is a tool to obtain a representation, decomposition and reconstruction of signals, presenting abrupt changes in the components of time-frequency instantly, through MRA with windows of varying length. The MRA is defined as a signal details and approximations explained by discrete wavelet coefficients and scale signal. The algorithm proposed by S. Mallat for calculating the coefficients is recursive and uses linear digital filtering operations. Given the properties of the subspaces and their bases, two quadrature mirror filters are employed to obtain the details of the signal [28,29]. The objective of the MRA is to highlight the information contained in the signal so detailed. MRA is based on the decomposition of the signal into multiple components, each one with its particularity. The contribution of MRA is to allow that the network wavelet individually processes each signal component. This method has the advantage of determining the optimal parameters of the wavelet networks for each of the signals according to their specificity.

A mother wavelet is a basic function that can be translated and dilated, which tracks time-frequency plane of the analyzed signal; the mother wavelet must be a wave satisfying

$$\int \psi(x)dx = 0 \tag{2}$$

$$\psi_{a,b}(t) = \psi\left(\frac{t-b}{a}\right) \tag{3}$$

where $\psi_{a,b}(t)$ is a function obtained by translation and/or dilation of the mother wavelet. In the equation (3), a is the scale factor and b is the position factor. In the case of a signal, the frequency and the time characteristics are related with a and b, respectively.

Let the function $x(t)$ of a real variable x. The continuous wavelet transform (CWT) of $x(t)$ can be written as

$$g(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi_{a,b}(t) \tag{4}$$

with a nonzero.

The inverse wavelet transform, which permits to reconstruct the function $x(t)$ by means of the function $g(a,b)$, can be defined as

$$x(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{a^2} g(a,b)\psi_{a,b}(t)dadb \tag{5}$$

where $\psi_{a,b}(t)$ is the mother wavelet and C_ψ is a constant, which depends on the type of wavelet. Then, it is sufficient to know the function $g(a,b)$ a countable number of values to reconstruct $x(t)$ [30,31].

The high complexity in the numerical treatment of CWT, mainly due to variability in the continuous parameters of scale and translation, imposes the necessity of a discrete version of the wavelet transform. The continuous mapping is represented, in the discrete version, by a finite range or set of values, through the change of the integral by an approximation with summations. The discretization is used to represent the signal in terms of elementary functions accompanied by coefficients. The Wavelet Series is just a sampled version of CWT and its computation may consume significant amount of time and resources, depending on the needed resolution.

Let $S = [s(0), \ldots, s(N-1)]^T$ a digital signal (defined by a time-series) with $k = 0, 1, \ldots, N-1$ and $N = 2^J$. The discrete wavelet transform (DWT) can be written as (6), (7):

$$W_\varphi(j_0, n) = \frac{1}{\sqrt{N}} \sum_k s(k)\varphi_{j_0,n}(k) \tag{6}$$

$$W_\psi(j, n) = \frac{1}{\sqrt{N}} \sum_k s(k)\psi_{j,n}(k) \tag{7}$$

where $j_0 = 0, j = 0, 1, \ldots, J-1$ and $n = 0, 1, \ldots, 2^j - 1$. The elements ψ and φ are the wavelet function and the scaling function, respectively. It is to remark that the continuous time interpretation of the wavelet decomposition algorithm is based on the fundamental notion of scaling function [32]. The inverse discrete wavelet transform is written as

$$s(k) = \frac{1}{\sqrt{N}} \sum_n W_\varphi(j_0, n)\varphi_{j_0,n}(k) + \sum_{j_0=j}^{\infty} \sum_n W_\psi(j, n)\psi_{j,n}(k) \tag{8}$$

The method proposed in this work employs the Daubechies wavelet decomposition [33], and the Maximal Overlap Discrete Wavelet Transform (MODWT). MODWT is known as shift invariant Wavelet Transform, and is a highly redundant version of the DWT considered ideal for time-series analysis [34,35]. MODWT has the advantage of avoiding the loss of information generated by sub-sampling in DWT. Any sample size can be managed in MODWT, while the size in DWT is limited to a multiple of 2^{Jp} [35], where $J_p th$ is the order of DWT. This property is then important for a real application because the data of the wind speed can not be limited by the order of DWT.

2.1.1. Algebraic decomposition

The basic principle of calculating the decomposition is a result of convolution of a signal s_j with wavelet low-pass g_j and high-pass digital filters h_j [36]. This principle is described in equations (9) and (10), and illustrated in Fig. 2(a).

$$s_{j+1}(k) = \sum_p h_{p-2k}s_j(p) = s_j{*}\overline{h}(2k) \tag{9}$$

$$d_{j+1}(k) = \sum_p g_{p-2k}s_j(p) = s_j{*}\overline{g}(2k) \tag{10}$$

2.1.2. Algebraic reconstruction

The principle of reconstruction is performed similarly to the decomposition; digital reconstruction filters, complementary to decomposition filters, are employed as presented in equation (11) and illustrated in Fig. 2(b).

$$\begin{aligned} s_j(k) &= \sum_p h'_{p-2k}s_{j+1}(p) + \sum_p g'_{p-2k}s_{j+1}(p) \\ &= s_{j+1}{*}h'(k) + s_{j+1}{*}g'(k) \end{aligned} \tag{11}$$

2.2. Multi-dimensional adaptive wavelet network structure

The general structure of an adaptive wavelet network comprises an input layer, a hidden layer wavelet and an output layer. The algebraic representation of the network structure of the wavelet is

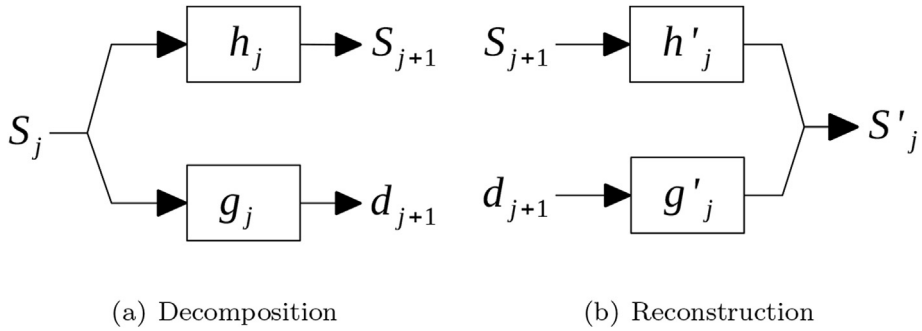


Fig. 2. Principle of the discrete wavelet decomposition and reconstruction.

presented in equations (12) and (13). Where y and x_i are the output and the input of the network, respectively. The weight of each wavelet, the bias value, and the dimension of the hidden layer are denoted by w_j, g and m, respectively.

$$y = \sum_{j=1}^{m} \omega_j z_j(x) + g \tag{12}$$

$$z_j = \prod_{i=1}^{n} \Psi_{a_{ij},b_{ij}}(x_i) \tag{13}$$

where $z_j(x)$ is a multi-dimensional wavelet, which is calculated as the product of n scalar wavelets. The structure of the wavelet network and the multi-dimensional wavelet are illustrated in Figs. 3 and 4, respectively.

The mother wavelet ($\Psi_{a,b}$) used in this application is the second derivative of the Gaussian function, which is commonly called Mexican Hat and can be represented by

$$\Psi_{a,b} = \left[1 - \left(\frac{x_i - b}{a}\right)^2\right] e^{-0.5\left(\frac{x_i-b}{a}\right)^2} \tag{14}$$

where a is the scale factor and b the position factor.

2.3. Training the WNN

The training process permits to optimize settings of the wavelet network used in the forecasting method. More specifically, after the system initialization is performed by adjusting the parameters, these parameters are used by the prediction algorithm. The weight of the drive process is based on minimizing the cost function defined by

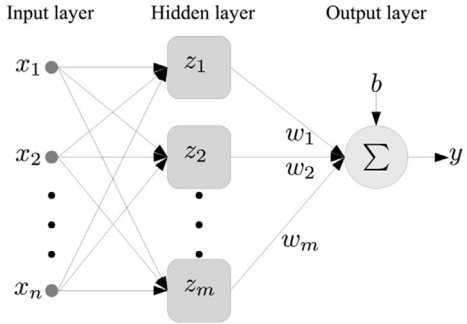


Fig. 3. WNN structure.

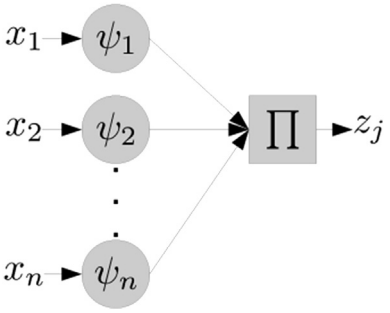


Fig. 4. Multi-dimensional wavelet for the hidden layers.

$$E = \frac{1}{2N} \sum_{p=1}^{N} e(p)^2 \tag{15}$$

$$e(p) = y_d(p) - y(p) \tag{16}$$

where $y_d(p)$ is the value to be predicted or the desired value and $y(p)$ is the predicted value. The goal of this optimization is to determine the best parameters w_j, a_{ij}, b_{ij} and g of the prediction algorithm. That means to calculate the partial derivative of the cost function E, according to each of the variables to be optimized, as defined by equations (17)–(20)

$$\Delta w_j = \frac{\partial E}{\partial w_j} = \frac{\partial E}{\partial y}\frac{\partial y}{\partial w_j} = e z_j \tag{17}$$

$$\Delta g_j = \frac{\partial E}{\partial g_j} = \frac{\partial E}{\partial y}\frac{\partial y}{\partial g_j} = e \tag{18}$$

$$\Delta a_j = \frac{\partial E}{\partial a_j} = \frac{\partial E}{\partial y}\frac{\partial y}{\partial \psi_{ij}}\frac{\partial \psi_{ij}}{\partial z_j}\frac{\partial z_j}{\partial a_j}$$

$$= e w_j z_j \left(\frac{1}{a_{ij}}\right)\left(\frac{x_i - b_{ij}}{a_{ij}}\right)^2 \left[3 - \left(\frac{x_i - b_{ij}}{a_{ij}}\right)\right] e^{-0.5\left(\frac{x_i-b_{ij}}{a_{ij}}\right)^2} \tag{19}$$

$$\Delta b_j = \frac{\partial E}{\partial b_j} = \frac{\partial E}{\partial y}\frac{\partial y}{\partial \psi_{ij}}\frac{\partial \psi_{ij}}{\partial z_j}\frac{\partial z_j}{\partial b_j}$$

$$= e w_j z_j \left(\frac{1}{a_{ij}}\right)\left(\frac{x_i - b_{ij}}{a_{ij}}\right) \left[3 - \left(\frac{x_i - b_{ij}}{a_{ij}}\right)\right] e^{-0.5\left(\frac{x_i-b_{ij}}{a_{ij}}\right)^2} \tag{20}$$

The parameters of the network wavelet (w, a, b, g) are updated by using

$$w_j(l+1) = w_j(l) + \eta\Delta w_j(l) + \alpha\Delta w_j(l-1) \tag{21}$$

where α is a dynamic parameter determined by the trial-and-error method, which can accelerate the convergence to the minimum by damping the oscillations. The parameter η (learning rate) is used to control the speed of the convergence of the gradient of the cost function. l is an integer determining the number of times that the data pass through the algorithm. The parameter η is adjusted depending on the sign of the error by using the equation (22); this adjustment improves the convergence to zero error [16,37].

$$\eta(n+1) = \begin{cases} 1.05\eta(n) & \Delta E(n) > 0 \\ 0.7\eta(n) & else \end{cases} \tag{22}$$

The structure used for the adjustment of parameters is shown in Fig. 5.

3. Predictability analysis

Hurst [11] proposed a method for the quantification of the long-term memory. The exponent or Hurst coefficient is used as a measure of the long-term memory in a time-series. It relates to the autocorrelation of the time-series and is still one of the most popular methods of research of fractal series of different nature [38–40]. The values of the Hurst coefficient vary between 0 and 1 depending on the predictability (or chaotic behavior in an opposite way) of the time series. According to their Hurst coefficient the time series can be classified into two main categories: the anti-persistent series with H lower than 0.5 and the persistent series with H higher than 0.5. A series which exhibits a value of $H = 0.5$, as in the case the random of walk series, is normally distributed and its mean value is predictable; a series with a value of H close to 0, as in the case of white noise, has more chaotic behavior than the random of walk series and hence very low predictability.

The Hurst exponent, H, is defined in terms of the asymptotic behavior of the rescaled range as a function of the process:

$$\mathrm{E}\left[\frac{R(n)}{S(n)}\right] = Cn^H \text{ as } n \to \infty \tag{23}$$

$$R(n) = \max(x^{cum}(t,n)) - \min(x^{cum}(t,n)) \tag{24}$$

$$S(n) = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}[x(t) - \overline{x}]^2} \tag{25}$$

where $R(n)$ is the range of the cumulative deviation series, $S(n)$ is the standard deviation of the time-series, $E[x]$ is the expected value, n is number of data points in a time-series and C is a constant. $R(n)$ and $S(n)$ are obtained from the mean of the time-series($\overline{x}(n)$), the mean centered series ($y(t)$), and the cumulative deviation of the series ($x^{cum}(t,n)$).

$$\overline{x}(n) = \frac{1}{n}\sum_{k=1}^{n} x(k) \tag{26}$$

$$y(t) = x(t) - \overline{x}(n) \tag{27}$$

$$x^{cum}(t,n) = \sum_{k=1}^{n} y(k) \tag{28}$$

The predictability of a time-series depends on the correlation between the past and the future data. This time line to be purely random or not depends on the intrinsic characteristics of each time-series. The existence of this relationship defines the deterministic nature of a time-series. Time-series generally are of two types: deterministic and random [40,41]. In this study, the Hurst coefficient has been calculated for deterministic analysis (with high potential of predictability), moderately deterministic and random series. The prediction algorithm has been evaluated for each type of time series allowing the performance evaluation of the prediction algorithm and also determining the influence of the nature of data on the prediction results [38,41].

3.1. Deterministic time-series

A time-series of Mackey-Glass [42], represented by equation (29), has been used in this study as an example of the deterministic series.

$$x(t+1) = x(t) + \frac{bx(t-\tau)}{1 + x^c(t-\tau)} - ax(t) \tag{29}$$

The Hurst coefficient has been computed for the Mackey-Glass series as 0.9702; this result confirms that these series are highly predictable. Fig. 6 represents the results of the prediction.

3.2. Random walk time-series

A second test of the prediction algorithm has been carried out using a series of random walk type defined by equation (30).

$$R_t = R_{t-1} + a_t \tag{30}$$

where a_t is a random variable defined for t = 1,2, …, N. The

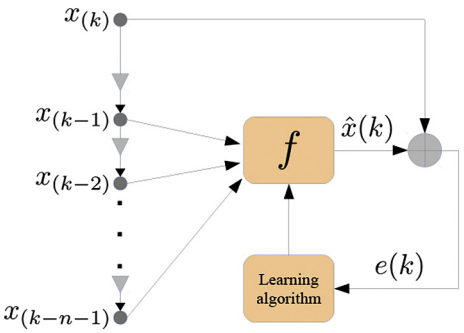


Fig. 5. Adjustment of model parameters using a gradient descent algorithm.

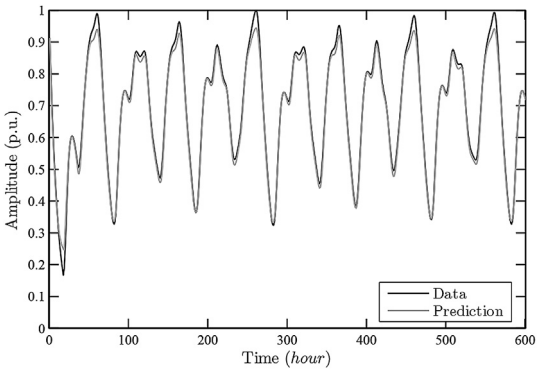


Fig. 6. Prediction of Mackey -Glass series.

prediction is fairly accurate according to the test results presented in Fig. 7. The Hurst coefficient of this series is 0.5173.

3.3. White noise time-series

Fig. 8 presents the results of prediction of a white noise time-series with zero mean and variance 1. As expected, this time-series has near-to-zero value of Hurst coefficient ($H = 0.0305$) which implies a very low predictability.

4. Validation and results

This section presents (i) the predictability analysis of the details (or components) of the wind data in order to assess the contribution of each component of the data in the improvement or deterioration of the accuracy of the prediction; (ii) the results of validating the proposed algorithm with all data components and eliminating the components with low predictability; and (iii) a discussion of the obtained results.

Hourly wind speed data have been used in this study to test the approach of the MRA-AWNN algorithm with Hurst predictability analysis. This wind speed data was used as the input of the prediction algorithm in the training and in the test phases. The main characteristics of the dataset are presented in Table 1. As illustrated in Fig. 9, the dataset of the wind speed has been divided in two parts: the first one, corresponding to 4320 h, was used for the training and the second one for the test.

The adopted strategy to test the prediction method is based on three essential parts: the decomposition of the data; the predictability analysis; and the test of the prediction algorithm. The tests of the prediction algorithm carried out were in two different ways: using data with all components, and data without the components showing low predictability.

4.1. Decomposition and analysis of predictability

The input data (wind speed profile) was decomposed in its approximation (A_0) and details signals (D_0 to D_N). The results of the decomposition of the wind speed profile are plotted in Fig. 10. Fig. 11 shows the coefficients of Hurst for the first seven components of the input signal (with non-zero Hurst exponent). As observed in these results the low frequency components have coefficients of Hurst close to one; therefore, they present a good potential of predictability; those of high frequencies are close to zero having a very low potential of predictability. In order to analyze the contribution of data components with low potential of predictability on the prediction, those data components could be removed. It is to remark, that only the first five components have been considered because they are the most significant concerning the predictability among the seven components with non-zero Hurst exponent. Without the predictability analysis, the number of components is commonly obtained by eliminating the components with near-to-zero amplitude.

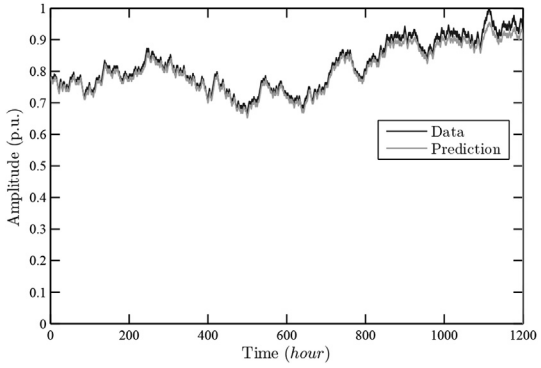


Fig. 7. Prediction of random walk series.

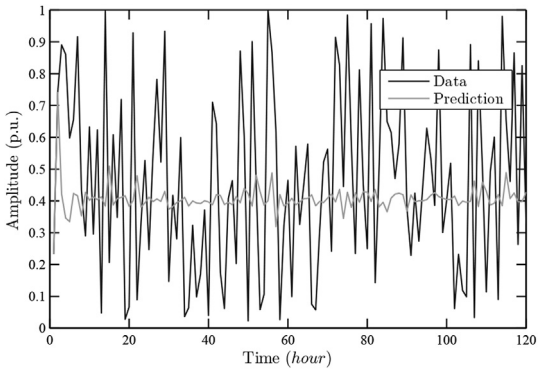


Fig. 8. Prediction of random series.

Table 1
Dataset characteristics.

Description	Value
Weather station	Trois-Rivieres(QC), Canada
Latitude/Longitude	46,350/-72,520
Measurement period	01/01/2013 to 01/01/2014
Measured variable	Wind speed [km/h]
Sampling period	1 h
Number of samples	8784
Measurement errors	152 (1.73%)
Quality control	Missing and erroneous data are corrected by using linear interpolation or by using correlation with a nearby station.
Database	SIMEB-Hydro-Québec (https://www.simeb.ca/)

4.2. Wind speed prediction

The results of the prediction of the wind speed are plotted in Fig. 12. These results show that the prediction error is similar in both cases, the one using all components and the other without the two components with low predictability (H with near-to-zero values). It is to be noted that the decomposition of the training signal has seven non-zero components, so the prediction and training algorithm process seven components. By eliminating the two components with worst predictability, the computational resources decrease by 29%, which represents a real advantage if one looks for an implementation of the prediction algorithm in a digital processor.

4.3. Prediction error analysis

Several methods for calculating the prediction error are available and are commonly used. Even if none of the commonly used error measures exhibit a superiority in all aspects (reliability, sensitivity, relationship to decisions) [43], the most widely used for the prediction are: the Symmetric Mean Absolute Percent Error (SMAPE), the Mean Absolute Percent Error (MAPE), the Mean Absolute Error (MAE), the Mean Squared Error (MSE) and the Root Mean Square Error (RMSE) [43], [44]. These individual errors are presented in equations (31)–(35). As observed in these equations, the SMAPE and MAPE can produce biased results and conclusions if

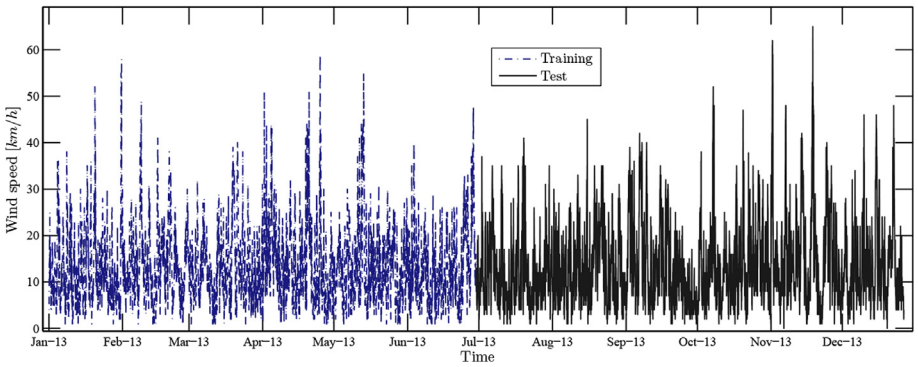


Fig. 9. Wind speed data used for validation.

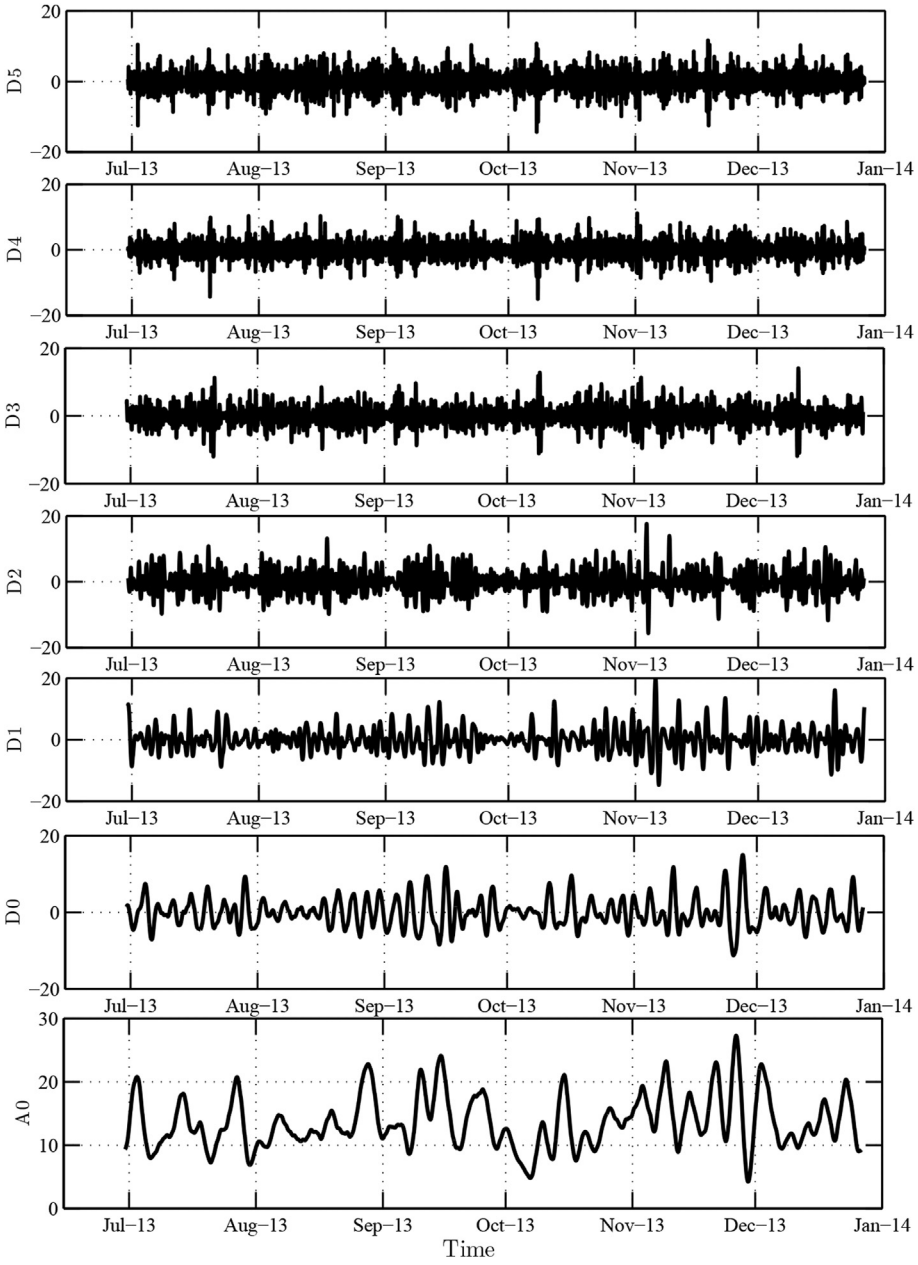


Fig. 10. Decomposition of test data.

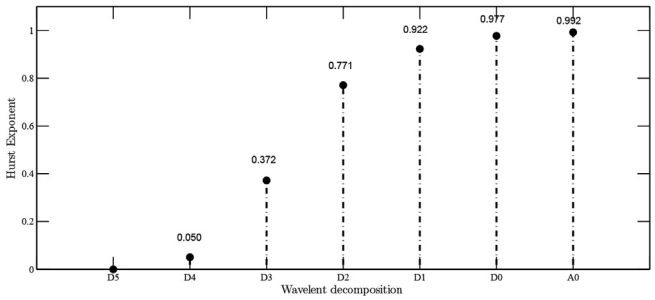


Fig. 11. Hurst coefficient of the decomposition.

the studied series contains data with very low or near-to-zero values. As a matter of facts, the RMSE is commonly preferred, by its reliability, as the measure of error in the analysis of forecasting methods.

$$SMAPE = \frac{2}{N}\sum_{k=1}^{N}\frac{|F_k - A_k|}{F_k + A_k} \tag{31}$$

$$MAPE = \frac{1}{N}\sum_{k=1}^{N}\frac{|F_k - A_k|}{|A_k|} \tag{32}$$

$$MAE = \frac{1}{N}\sum_{k=1}^{N}|F_k - A_k| \tag{33}$$

$$MSE = \frac{\sum_{k=1}^{N}(F_k - A_k)^2}{N} \tag{34}$$

Table 2
Prediction error of MRA-AWNN with all data components at different horizons (1–5 h).

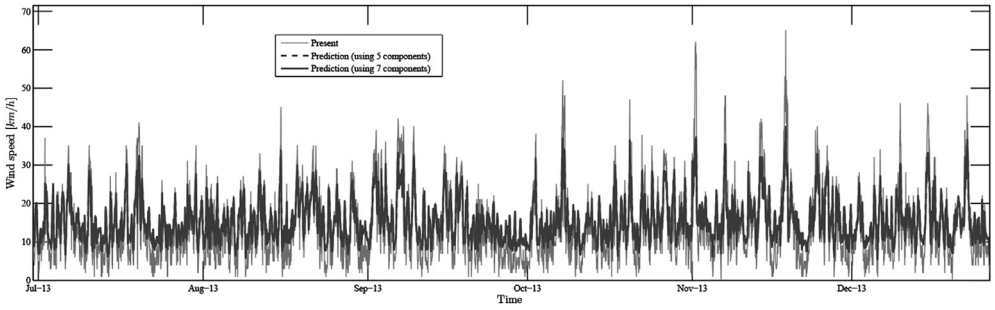
Horizon	1	2	3	4	5
SMAPE	16.8982	22.2368	23.9919	26.5795	28.5882
MAE	5.7822	8.3133	9.4731	11.0669	12.4035
MSE	0.0050	0.0092	0.0116	0.0153	0.0190
RMSE	7.0679	9.6041	10.7660	12.3678	13.7680

Table 3
Prediction error of MRA-AWNN without random data at different horizons (1–5 h).

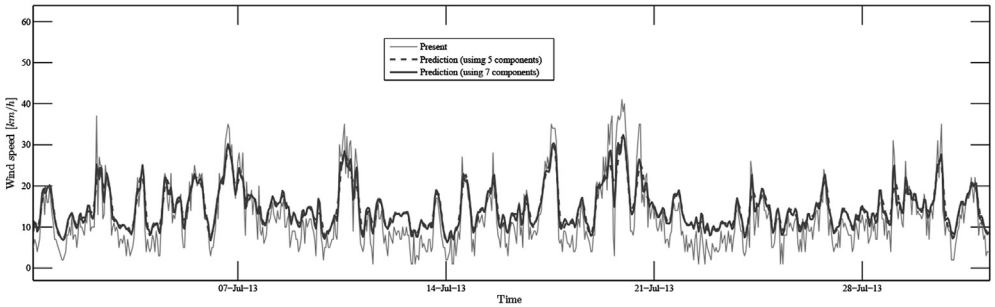
Horizon	1	2	3	4	5
SMAPE	16.8924	22.2363	24.0232	26.5788	28.5894
MAE	5.7808	8.3130	9.4892	11.0647	12.4031
MSE	0.0050	0.0092	0.0116	0.0153	0.0190
RMSE	7.0620	9.6085	10.7873	12.3678	13.7703

Table 4
Prediction error of persistence model at different horizons (1–5 h).

Horizon	1	2	3	4	5
SMAPE	16.8924	22.2363	24.0232	26.5788	28.5894
MAE	5.7808	8.3130	9.4892	11.0647	12.4031
MSE	0.0050	0.0092	0.0116	0.0153	0.0190
RMSE	7.0620	9.6085	10.7873	12.3678	13.7703



(a) Test period



(b) Zoom-in for july 2013

Fig. 12. Forecast of wind speed.

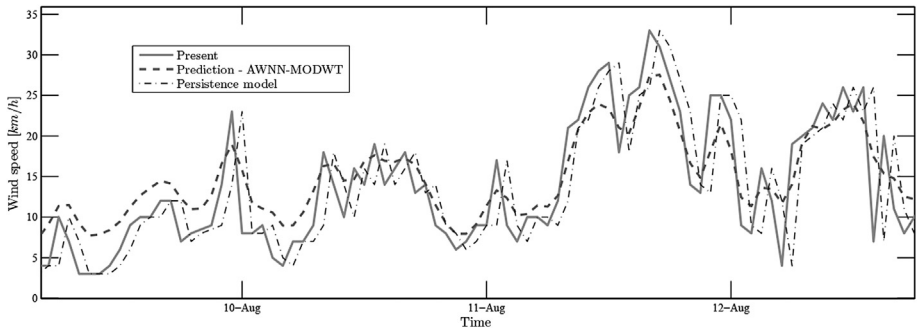


Fig. 13. Comparison of prediction with MRA-AWNN and persistence model.

$$RMSE = \sqrt{\frac{\sum_{k=1}^{N}(F_k - A_k)^2}{N}} \tag{35}$$

The estimated error calculations of the wind speed prediction using the entire data and without using the random data are presented in Tables 2 and 3, respectively. This evaluation considers different prediction horizons (from 1 to 5 h). As expected, the results presented in the two Tables are similar; e.g. for a prediction horizon of 1 h the difference between the two methods is 0.0059, 0.0014 and 0.0058 for SMAPE, MAE and RMSE, respectively; and there is not difference in MSE.

The proposed method has been compared with the well known persistence model. This comparison has permitted to confirm that the MRA-AWNN presents lower RMSE for different horizons. Table 4 presents the error obtained by using the persistence model for different prediction horizons. Fig. 13 shows a zoom-in of this comparison for a horizon of 1 h. It is to remark that the difference in phase introduced by the persistence model would be undesirable in power management systems.

5. Conclusions

A prediction method based on wavelet networks using MRA-AWNN was implemented and evaluated by numerical simulations using real wind speed profiles. The proposed prediction method includes the multi-resolution and predictability analysis of the components of the studied time-series with the aim of optimizing the prediction algorithm. In fact, the predictability analysis of each component of the decomposed signal permits to eliminate the components with low potential of predictability. This operation allows to optimize the complexity (order) of the prediction system without reduction in the accuracy. In the case studied in this paper, it was possible to reduce the computational resources to about 29%, maintaining similar performance. These significant savings in complexity are very important when the implementation of an embedded prediction system is envisaged using digital processors.

References

[1] L. Freris, D. Infield, Renewable Energy in Power Systems, A John Wiley & Sons, Ltd, Publication, 2008.
[2] Sophia Antipolis (Banque européenne d'inverstissement), Etude sur le changement climatique et énergie en Mediterranée, Tech. rep., Plan Bleu Centre d'Activités Régionales, 2008.
[3] P.K. Ruud Kempener, (IRENA), a. H. U. Of Colorado), Smart Grids and Renewables a Guide for Effective Deployment, Tech. Rep. November, International Renewable Energy Agency, U.S, 2013.
[4] R. Dixon, M. Eckhart, D. Hales, G. Thompson, Rapport mondial 2012 sur les énergies renouvelables 2012, Tech. rep., Renewable Energy Pollicy Network for the 21st Century, 2012.
[5] K. De Vos, J. Driesen, Balancing management mechanisms for intermittent power sources — A case study for wind power in Belgium, in: 2009 6th International Conference on the European Energy Market, IEEE, 2009, pp. 1–6, http://dx.doi.org/10.1109/EEM.2009.5207190.
[6] E. Goutard, Renewable energy resources in energy management systems, in: 2010 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe), IEEE, 2010, pp. 1–6, http://dx.doi.org/10.1109/ISGTEUROPE.2010.5639003.
[7] I. Bandyo, Control and management of renewable generation — necessity of look ahead study, in: 2013 IEEE Innovative Smart Grid Technologies-asia (ISGT Asia), IEEE, 2013, pp. 1–5, http://dx.doi.org/10.1109/ISGT-Asia.2013.6698726.
[8] S. Kélouwani, K. Agbossou, Nonlinear model identification of wind turbine with a neural network, IEEE Trans. Energy Convers. 19 (3) (2004) 607–612.
[9] A. Cardenas, C. Guzman, K. Agbossou, Development of a FPGA Based Real-Time Power Analysis and Control for Distributed Generation Interface, IEEE Trans. Power Syst. 27 (2012) 1343–1353.
[10] C. Guzman, A. Cardenas, K. Agbossou, Load sharing strategy for autonomous AC microgrids based on FPGA implementation of ADALINE&FLL, IEEE Trans. Energy Convers. 29 (3) (2014) 663–672.
[11] H.E. Hurst, Long term storage capacity of reservoirs, Trans. Am. Soc. Civ. Eng. 116 (1951) 509–516.
[12] S.S. Soman, H. Zareipour, O. Malik, P. Mandal, A review of wind power and wind speed forecasting methods with different time horizons, in: North American Power Symposium 2010, IEEE, 2010, pp. 1–8, http://dx.doi.org/10.1109/NAPS.2010.5619586.
[13] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, Z. Yan, A review on the forecasting of wind speed and generated power, Renew. Sustain. Energy Rev. 13 (4) (2009) 915–920, http://dx.doi.org/10.1016/j.rser.2008.02.002.
[14] E. Erdem, J. Shi, ARMA based approaches for forecasting the tuple of wind speed and direction, Appl. Energy 88 (4) (2011) 1405–1414, http://dx.doi.org/10.1016/j.apenergy.2010.10.031.
[15] S. Gao, Y. He, H. Chen, Wind speed forecast for wind farms based on ARMA-ARCH model, in: 2009 International Conference on Sustainable Power Generation and Supply, IEEE, 2009, pp. 1–4, http://dx.doi.org/10.1109/SUPERGEN.2009.5348142.
[16] M. Bhaskar, A. Jain, N. Venkata Srinath, Wind speed forecasting: Present status, in: 2010 International Conference on Power System Technology, 2010, pp. 1–6, http://dx.doi.org/10.1109/POWERCON.2010.5666623.
[17] R. Blonbou, Very short-term wind power forecasting with neural networks and adaptive Bayesian learning, Renew. Energy 36 (3) (2011) 1118–1124, http://dx.doi.org/10.1016/j.renene.2010.08.026.
[18] R.R.B. de Aquino, M.M.S. Lira, J.B. de Oliveira, M.A. Carvalho, O.N. Neto, G.J. de Almeida, Application of wavelet and neural network models for wind speed and power generation forecasting in a Brazilian experimental wind park, in: 2009 International Joint Conference on Neural Networks, IEEE, 2009, pp. 172–178, http://dx.doi.org/10.1109/IJCNN.2009.5178791.
[19] J.P.S. Catalao, H.M.I. Pousinho, V.M.F. Mendes, Hybrid wavelet-PSO-ANFIS approach for short-term wind power forecasting in Portugal, IEEE Trans. Sustain. Energy 2 (1) (2010) 50–59, http://dx.doi.org/10.1109/TSTE.2010.2076359.
[20] N. Chen, Z. Qian, I.T. Nabney, X. Meng, Wind power forecasts using Gaussian processes and numerical weather prediction, IEEE Trans. Power Syst. 29 (2) (2014) 656–665, http://dx.doi.org/10.1109/TPWRS.2013.2282366.
[21] E. Pelikan, K. Eben, J. Resler, P. Jurus, P. Krc, M. Brabec, T. Brabec, P. Musilek, Wind power forecasting by an empirical model using NWP outputs, in: 2010 9th International Conference on Environment and Electrical Engineering, IEEE, 2010, pp. 45–48, http://dx.doi.org/10.1109/EEEIC.2010.5490019.
[22] J. Shi, Z. Ding, W.-J. Lee, Y. Yang, Y. Liu, M. Zhang, Hybrid forecasting model for very-short term wind power forecasting based on grey relational analysis and wind speed distribution features, IEEE Trans. Smart Grid 5 (1) (2014) 521–526, http://dx.doi.org/10.1109/TSG.2013.2283269.
[23] J. Hu, J. Wang, G. Zeng, A hybrid forecasting approach applied to wind speed time series, Renew. Energy 60 (2013) 185–194, http://dx.doi.org/10.1016/j.renene.2013.05.012.
[24] S. Ben Taieb, A. Sorjamaa, G. Bontempi, Multiple-output modeling for multi-step-ahead time series forecasting, Neurocomputing 73 (10–12) (2010) 1950–1957, http://dx.doi.org/10.1016/j.neucom.2009.11.030.

[25] L. Herrera, H. Pomares, I. Rojas, A. Guillén, A. Prieto, O. Valenzuela, Recursive prediction for long term time series forecasting using advanced models, Neurocomputing 70 (16–18) (2007) 2870–2880, http://dx.doi.org/10.1016/j.neucom.2006.04.015.
[26] Y. Bennani, Apprentissage par réseaux de neurones artificiels, Tech. rep., Université de Paris 13, Ecole de Printemps sur l'Aprentissage arTificiel, Paris (, 2014.
[27] K. Bhaskar, S.N. Singh, AWNN-assisted wind power forecasting using feed-forward neural network, IEEE Trans. Sustain. Energy 3 (2) (2012) 306–315, http://dx.doi.org/10.1109/TSTE.2011.2182215.
[28] M. Pouliot, La détermination des coefficients des ondelette de Daubechies, Université de Laval, 2009. Ph.D. thesis.
[29] S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, Pattern Anal. Mach. Intell. IEEE I (7) (1989) 674–693.
[30] S. Dubois, Décomposition spatio-temporelles pour l'étude des textures dynamiques, Université de la Rochelle, 2010. Ph.D. thesis.
[31] T. Ho, Étude de la méthode de la transformation en ondelette et application à la compression des images Remerciements, Tech. rep., Rapport final de TIPE, Luaong Hòng Viêt, 2005.
[32] M. Unser, T. Blu, Wavelet theory demystified, IEEE Trans. Signal Process. 51 (2) (2003) 470–483, http://dx.doi.org/10.1109/TSP.2002.807000.
[33] I. Daubechies, Orthonormal bases of compactly supported wavelets, Commun. Pure Appl. Math. 41 (1988) 909–996.
[34] S. Khalighi, T. Sousa, D. Oliveira, G. Pires, U. Nunes, Efficient feature selection for sleep staging based on maximal overlap discrete wavelet transform and SVM., Conference proceedings : annual international conference of the IEEE engineering in medicine and biology society, in: IEEE Engineering in Medicine and Biology Society. Annual Conference 2011, 2011, pp. 3306–3309, http://dx.doi.org/10.1109/IEMBS.2011.6090897.
[35] S. Ahmad, A Temporal Pattern Identification and Summarization Method for Complex Time Serial Data, University of Surrey, 2007. Ph.D. thesis.
[36] J.-d. Gretillat, Modification de l'algorithme de la transformée en ondelettes descrète pour l'obtention d'une représentation invariante sous rotation, Université Laval, Québec, 2006. Ph.D. thesis.
[37] Y. Oussar, Réseaux d'ondelettes et réseaux de neurones pour la modélisation statique et dynamique de processus, VI, Université Paris, 1998. Ph.D. thesis.
[38] B. Qian, K. Rasheed, Hurst exponent and financial market predictability, in: IASTED International Conference on Financial, 2004, 2004, pp. 1–7.
[39] D. Delignières, L'analyse des processus stochastiques, Sport, Performance, Santé", EA, 2001, pp. 1–12 doi:EA2991.
[40] M. Sánchez Granero, J. Trinidad Segovia, J. García Pérez, Some comments on Hurst exponent and the long memory processes on capital markets, Phys. A Stat. Mech. Appl. 387 (22) (2008) 5543–5551, http://dx.doi.org/10.1016/j.physa.2008.05.053.
[41] C. Eom, S. Choi, G. Oh, W.-S. Jung, Hurst exponent and prediction based on weak-form efficient market hypothesis of stock markets, Phys. A Stat. Mech. its Appl. 387 (18) (2008) 4630–4636, http://dx.doi.org/10.1016/j.physa.2008.03.035 arXiv:arXiv:0712.1624v1.
[42] M. Duan, Time Series Predictability, Marquette University, Milwaukee, Wisconsin, 2002. Ph.D. thesis.
[43] J. Armstrong, F. Collopy, Error measures for generalizing about forecasting methods: empirical comparisons, Int. J. Forecast. 8 (1) (1992) 69–80, http://dx.doi.org/10.1016/0169-2070(92)90008-W.
[44] M. Shcherbakov, A. Brebels, A survey of forecast error measures, World Appl. Sci. J. 24 (4) (2013) 171–176, http://dx.doi.org/10.5829/idosi.wasj.2013.24.itmies.80032.