Wavelet cross-correlation analysis of wind speed series generated by ANN based models

Grégory Turbelin*, Pierre Ngae, Michel Grignon

Laboratoire de Mécanique et d'Energétique d'Evry (LMEE), Université d'Evry Val d'Essonne, 40 rue du Pelvoux, 91020 Evry Cedex, France

A B S T R A C T

To obtain, over medium term periods, wind speed time series on a site, located in the southern part of the Paris region (France), where long recording are not available, but where nearby meteorological stations provide large series of data, use was made of ANN based models. The performance of these models have been evaluated by using several commonly used statistics such as average absolute error, root mean square error, normalized mean square error, and correlation coefficient. Such global criteria are good indicators of the "robustness" of the models but are unable to provide useful information about their "effectiveness" in accurately generating wind speed fluctuations over a wide range of scales. Therefore a complementary wavelet cross coherence analysis has been performed. Wavelet cross coherence, wavelet cross-correlation and spectral wavelet cross-correlation coefficients, have been calculated and displayed as functions of the equivalent Fourier period. These coefficients provide quantitative measures of the scale-dependence of the model performance. In particular the spectral wavelet cross coherence coefficient can be used to have a rapid and efficient identification of the validity range of the models. The results show that the ANN models employed in this study are only effective in computing large-scale fluctuations of large amplitude. To obtain a more representative time series, with much higher resolution, small-scale fluctuations have to be simulated by a superimposed statistical model. By combining ANN and statistical models, both the high and the low-frequency segments of the wind velocity spectra can be simulated, over a range of several hours, at the target site.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

The planetary boundary layer is the lowest region of the atmosphere, directly influenced by the presence of the underlying earth surface. In the lowest part of this area (surface layer) wind represents a clean, abundant and inexhaustible energy source. Therefore, in the renewable industry, large series of wind speed data are needed: to identify potential wind turbine sites, to assess the wind energy resource of selected locations, to evaluate, by numerical simulations, the power values produced by wind energy conversion systems [1] or to estimate external load situations that drive the design of the structural components. As measured data, or records of sufficient duration, are not available for most sites, appropriate wind speed time series may be estimated from relationship of wind speed among several instrumented sites, submitted to the same climatic constraints.

In practice, the long-term trend of the time series, dominated by macro meteorological influences (with time scales of the order of a week), can be successfully estimated with Artificial Neural Networks (ANN) based models [2,3]. The short-term turbulent variations (with time scales of the order of a minute), due to local topographic and terrain effects, can be obtained by using statistical models [4]. In the medium term range, i.e. for considerations over several hours to a few days, the wind speed fluctuations are influenced by the prevailing large-scale motions as well as by micro-scale phenomena such as turbulence. As a result, two models must be combined, one for the large-scale fluctuations and one for the small [5]. Therefore, when ANN based models are used to estimate wind speed, in the medium term range, their performances must be evaluated for all the scales between the macro- and micro-scales. Normally, only a few global statistical performance measures, such as correlation coefficient (R) or root mean square error (RMSE), are employed. But such statistics are not good indicators of the "effectiveness" of a model in terms of its ability to accurately predict low, medium, and high frequencies fluctuations. Since no appropriate statistic, to quantify the time-scale quality of generated wind speed series, has been found in the literature, a wavelet cross-correlation analysis has been performed. In this work, wavelet based coefficients have been used to assess the overall applicability of ANN based models used to

---

* Corresponding author. Tel.: +33 169477513; fax: +33 169477599.
E-mail addresses: gregory.turbelin@iup.univ-evry.fr (G. Turbelin), pierre.ngae@iup.univ-evry.fr (P. Ngae), michel.grignon@iup.univ-evry.fr (M. Grignon).

estimate wind speed data at a given test site, located in Evry (France).

2. Wind model

In the atmospheric boundary layer wind is an incompressible, unsteady, air flow composed of eddies interacting with each other. It consists of a wide spectrum of eddy size ranging from large eddies, corresponding to large atmospheric phenomena, to small eddies, corresponding to atmospheric turbulence. In the near surface layer, velocity measured at a given point M, fluctuates over time. Over time periods spread, in practice, between 10 min and 1 h, in a classical stationary approximation, wind velocity can be separated into a constant mean component and a fluctuating component (turbulence) considered as a stationary random process, with normal distribution. Over larger time-period, i.e. over periods of hours and days, 10-min (or hourly) mean wind is subject to a significant temporal trend and the instantaneous time-varying wind speed, $\vec{u}(M,t)$, can be expressed by a superposition of high-frequency oscillation of small amplitude $\vec{u}'(M,t)$ around a much more slowly varying sustained speed in a prevailing direction, $\vec{U}(M,t)$ [6]

$$\vec{u}(M,t) = \vec{U}(M,t) + \vec{u}'(M,t) \tag{1}$$

Compared to the stationary wind speed model, which assumes that the first term is a constant mean speed, this model introduces a slow time-varying component, which is the sum of the overall time-mean value $\overline{\vec{U}}(M)$ and low-frequency fluctuation $\widetilde{\vec{U}}(M,t)$ (see Fig. 1)

$$\vec{U}(M,t) = \overline{\vec{U}}(M) + \widetilde{\vec{U}}(M,t) \tag{2}$$

A strong relationship between meteorological scales and frequency of wind speed fluctuations is assumed: low-frequency fluctuations are dominated by macro meteorological phenomena (large-scales flows). Thus, at a target site, slowly time-varying wind speed series can be successfully computed with models using, as input arguments, data recorded in meteorological stations submitted to the same climatic constraints (i.e. located within a radius of a few kilometres). On the contrary, the high-frequency fluctuations, i.e. with typical time scales ranging from seconds-to-few minutes, depend on the neighbouring environment of the place where they are observed. On this scale, random behaviour can be assumed and fluctuations can be obtained by using suitable statistical models with turbulence intensity, standard deviations and spectrum, which can be characterised by few local parameters, as basic inputs [7]. Since several models could be used for generating accurate wind series, the validity range of each model must be clearly defined.
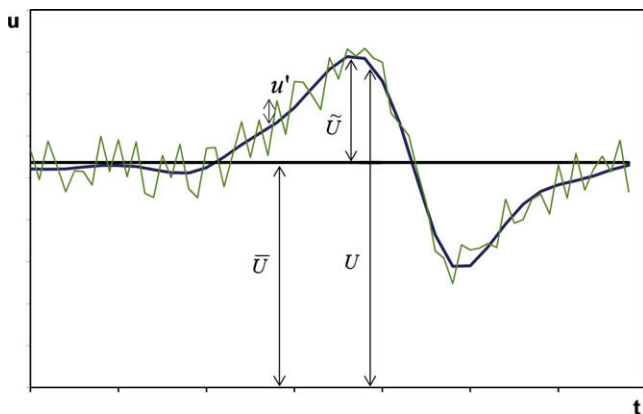

Fig. 1. Wind speed components.

3. Site, methodology and models description

3.1. Site and methodology

In the southern part of the Paris region (France), in autumn, stable conditions are dominant, the prevailing winds (gentle to moderate breeze) are westerly, or south-westerly, and correspond to air masses coming from the Atlantic and crossing the country from west-to-east. During this season, suitable ANNs were set up to obtain wind speed data, over medium term periods, at a typical suburban site, located in Evry, using data from three airport meteorological stations (Orly, Brétigny, Melun-Villaroche), see Fig. 2. The development of such models involves the following steps [2,3,8,9]: (I) wind velocity measurement at the test site, selection of the training, test and validation data sets, (II) identification of the input and output variables, (III) selection of the network's architecture, training of the models, (IV) validation of the models using various global evaluation statistics, and (V) assessment of the time-scale quality of the generated wind speed series, as explained in this paper.

3.2. Selection of data sets

The available data, at the reference stations, were the 10-min average horizontal wind speeds, simultaneously recorded, at 10 m above the ground, every 6 min, by the French meteorological office and noted $\vec{U}_{h}(M,t)$. This wind speed can be split into two perpendicular wind components U and V. U is the component of the horizontal wind from the east (a positive U component represents wind blowing from the east), and V is the component of the horizontal wind from the north. To collect concurrent data at the test site, use was made of a cup anemometer and of a potentiometer driven wind direction indicator. Data were sampled on a PC and then processed to obtain appropriate values. The measurements were performed in November. Two separate data sets have been chosen: a training set, which contained 9000 representative meteorological data, and a verification set, recorded one year later. The training data set, in which the number of high wind speed (>7 m/s) was significantly less than the number of low and medium wind speed, has been split into two smaller sets and used to adjust and test the parameters of the ANN. The verification data set covered a period of 7 days and consisted of relatively low wind speed (<3 m/s, see Fig. 8). It has been used to check the quality of the generated data. Both of them corresponded to meteorological situations typical of the considered site, in the autumn months.

3.3. Identification of the input and output variables

Comparison of the measurements made at Evry (built-up suburb) with those made at the airport stations (flat terrains) indicated that, even when there was an enormous speed difference, a shape similarity in wind profiles exists. To get a quantitative measure of the relatedness of the data, a classical cross-correlation analysis has been performed. This analysis has confirmed that the measurements, at the test site, were related to those made simultaneously at the 3 surrounding stations (even when one of them is leeward). But it is clear that such an analysis is biased towards the large-scale dynamic processes of large amplitude. Therefore, to obtain useful information concerning the behaviour of the intermediate and small-scale dynamics, the P previous measurements have also been used as inputs. P was determined using a trial and error procedure: the input vector was modified by successively adding antecedent values. A new model was developed each time, the best one was selected based on the analysis of the normalized mean square training error (see below). We concluded that, to
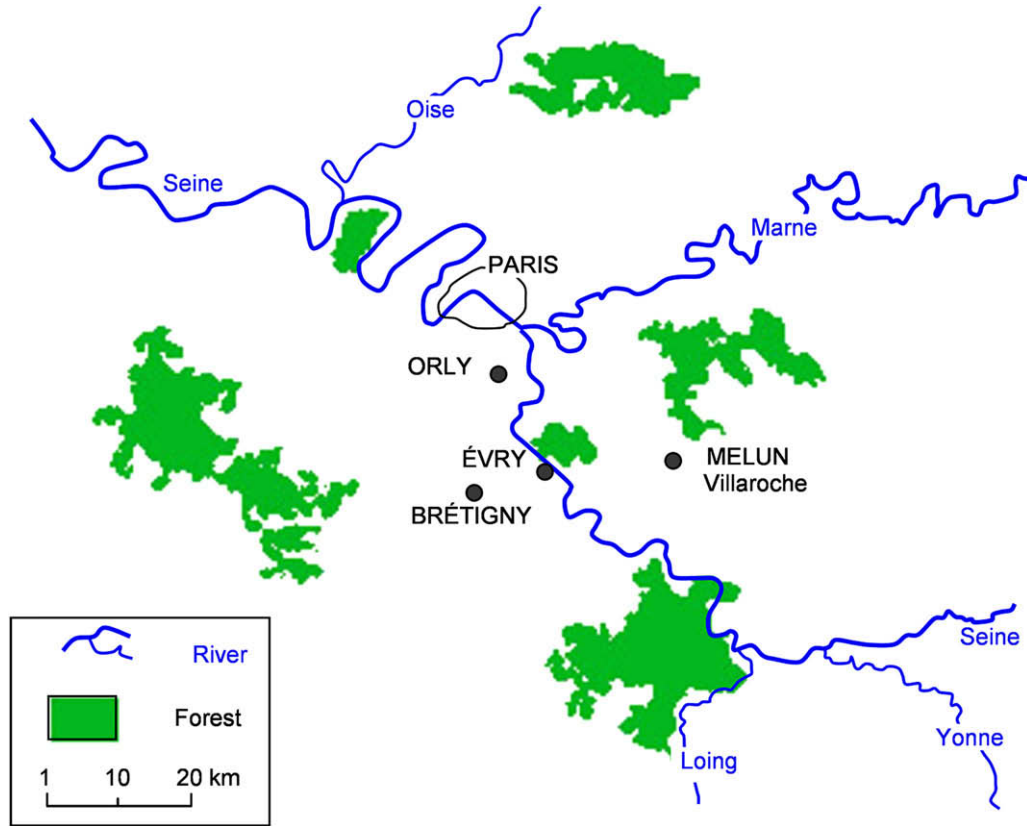
Fig. 2. Geographical localisation of the test site.

estimate the wind speed components, at Evry, at time t (network's output), the most recent 4 values ($P = 3$), from $t - 3$ backwards, of both the U and V components, observed in the three stations, were of interest.

3.4. Selection of the network architecture, training of the models

We set up a feed forward artificial neural network for each speed component. Multilayer feed forward networks were chosen because they "have been widely used as time series forecasters" [10] and allow the prediction of wind speed at one site based on measurements at other sites [3]. Since the networks structure is not a key issue in this paper, the networks parameters will be outlined briefly, for further details the reader is referred to Ref. [9]. The architecture that was ultimately selected was composed of a single hidden layer. Activation functions were hyperbolic tangent for the hidden layer and linear in the output layer. The number of hidden neurons was determined using a trial and error procedure. This number was varied from 2 to 20. For each value, the Levenberg Marquardt training algorithm was used to minimize the normalized mean square training error, at the output layer. The optimum number of hidden neurons was determined to be 12 for the U component, and 10 for the V component., see Table 1

4. Evaluating model performance

4.1. By using global statistics

The performance of a model can be defined as its ability to reproduce series of data which were unknown during the building phases. Therefore our models can be assessed by performing a term-by-term comparison between the experimental verification data, represented by r_t ($t = 1$ to N), and the computed data, represented by c_t ($t = 1$ to N). At each time t, the difference between the computed and observed values, $e_t = c_t - r_t$, can be calculated. By averaging the errors over the full data set, the average absolute error (AAE) [11] may be computed using the following expression:

$$AAE = \frac{1}{N} \sum_{t=1}^{N} |e_t| \tag{3}$$

It is clear, from this definition, that small values of this statistics indicate good estimation capabilities of the model. As performance measures, other well accepted criteria are root mean square error (RMSE) [12] and normalized mean square error (NMSE), which is also known as the average relative variance (ARV) (the Nash–Sutcliffe efficiency (E), used in some references [13], can be derived from the NMSE, $E = 1 - NMSE$). The first one is simply defined as the square root of the mean square error, the second is the ratio of the mean square error over the natural variance of the measured variable σ_r^2:

Table 1
Characteristics of the ANN based models

	Model for U	Model for V
Activation functions for the input layer	Linear	
Activation functions for the hidden layer	Hyperbolic tangent	
Activation functions for the output layer	Linear	
Inputs (i = Orly, Brétigny, Melun)	$U_i(t), U_i(t-1),…, U_i(t-3)$	
	$V_i(t), V_i(t-1),…, V_i(t-3)$	
Output	$U_{evry}(t)$	$V_{evry}(t)$
Number of neurons in the hidden layer	12	10

$$RMSE = \left(\frac{1}{N}\sum_{t=1}^{N} e_t^2\right)^{1/2} \tag{4}$$

$$NMSE = \frac{1}{\sigma_r^2}\frac{1}{N}\sum_{t=1}^{N} e_t^2 \tag{5}$$

A value of NMSE = 0 indicates perfect prediction while a value of NMSE = 1 corresponds to the prediction of the statistical average [14]. The correlation coefficient (R) quantifies the degree of similarity between the estimated and the measured values. It is defined as:

$$R = \frac{\frac{1}{N}\sum_{t=1}^{N}(r_t - \bar{r})(c_t - \bar{c})}{\sigma_r \sigma_c} \tag{6}$$

where $\bar{c}$ is the mean value, σ_c is the standard deviation of the computed wind series, $\bar{r}$ and σ are the mean value and standard deviation of the observed series. The values of (R) close to 1 indicate good model performance. In a traditional evaluation approach, these statistics are good indicators of the "robustness" of the model and provide a quantitative measure of its "efficiency" in capturing the complex relationships between wind data values collected on several locations. However, such error statistics are not good indicators of the "effectiveness" in accurately estimating non-stationary wind speed data with both high and the low-frequency fluctuations: the square terms in RMSE and NMSE places a bias on the low-frequency fluctuations (large-scale atmospheric motions) which are of large amplitude, therefore the errors in estimating the high-frequency oscillation of small amplitude will not get sufficient representation. Moreover, the correlation coefficient becomes invalid if the series include non-stationary components and/or are "characterised by highly variable processes occurring over a wide range of scales" [15].

4.2. By using wavelets analysis

So, to measure the accuracy of the model, at each scale of fluctuations, we have developed a technique based on wavelet cross-correlation analysis. During the last decade, wavelets have been extensively employed as a tool to analyse measured data, but also, in the area of wind engineering, to analyse wind effects on structures [16] or to evaluate the quality of synthetic wind speed signals [4]. The aim of a wavelet analysis is to determine a time-frequency representation of a series and to assess the temporal variation of the different frequencies involved. The continuous wavelet transform (CWT) of a signal $c(t)$ is given as

$$CWT[c(t)] = W_c(a,b) = \int_{-\infty}^{+\infty} c(t)\psi_{a,b}^*(t)\mathrm{d}t \tag{7}$$

$\psi_{a,b}(t) = (1/\sqrt{a})\psi((t-b)/a)$ is the analysis wavelet (mother wavelet), it can be real or complex and must satisfy strict mathematical conditions, * denotes the complex conjugate, a is the dilatation parameter (or scale) and b is the translation parameter, used to adjust the shape and localisation of the wavelet, respectively in scale and time domains. Low parameters a compress the wavelet, in order to analyse high frequency of the series, while high values of a dilate the wavelet so as to analyse low-frequency components, in the neighbourhood of b. The wavelet analysis results in a set of wavelet coefficients which indicate how similar the time series is to the analysis wavelet at different scales and positions. It is clear that the results strongly depend on the choice of the wavelet function and that this function must reflect the characteristics of the series. In this study, the Mexican hat wavelet (MHW), or Maar wavelet, has been chosen. This real-value function has been successfully

employed for the identification of events such as maxima, slow rise or sudden drop in temporal series and has been used in atmospheric studies [17]. It is the second derivative of the Gaussian function, given as

$$\psi(t) = (1 - t^2)\mathrm{e}^{-t^2/2} \tag{8}$$

The wavelet power spectrum, which is also known as the scalogram, can be defined as

$$SW_c(a,b) = W_c(a,b)W_c^*(a,b) = |W_c(a,b)|^2 \tag{9}$$

A cross wavelet power spectrum, or coscalogram, can be defined between two time series, $c(t)$ and $r(t)$, by replacing the square coefficient term with the product of the coefficients of each series

$$SW_{cr}(a,b) = W_c(a,b)W_r^*(a,b) \tag{10}$$

When the analysis wavelet is complex, $SW_{cr}(a,b)$ is complex, and can be separated into real and imaginary parts, called wavelet co- and quadratur-spectrum respectively [18]. The wavelet cross-correlation function [19] for a given scale a and a given time lag τ can be written as

$$WC_{cr}(a,\tau) = \overline{W_c(a,b)W_r(a,b+\tau)} \tag{11}$$

where the overbar indicates a time average. If the mother wavelet is complex, this function consists of a real part $RWC_{cr}(a,\tau)$ and an imaginary part $IWC_{cr}(A,\tau)$. For this study, another function, which we had defined in a previous paper [20], could be used: the spectral wavelet cross-correlation function that can be written as

$$SWC_{cr}(a,\tau) = \overline{SW_c(a,b)SW_r(a,b+\tau)} \tag{12}$$

All these functions permit the definition of complementary wavelet correlation coefficients, with values ranging from 0 to 1. The cross wavelet coherence coefficient has been introduced [18] to measure the "intensity coherence" of the series

$$CWCF(a,b) = \frac{2|SW_{cr}(a,b)|^2}{|W_c(a,b)|^4 + |W_r(a,b)|^4} \tag{13}$$

The wavelet cross-correlation coefficient has been defined "to overcome the limitations of classical cross-correlation analysis" [19]

$$WR_{cr}(a,\tau) = \frac{RWC_{cr}(a,\tau)}{\sqrt{RWC_{cc}(a,0)RWC_{rr}(a,0)}} \tag{14}$$

The spectral wavelet cross-correlation coefficient [20] has been defined in order to use the concept of cross-correlation to detect spectral similarity

$$SWR_{cr}(a,\tau) = \frac{SWC_{cr}(a,\tau)}{\sqrt{SW_c(a,b)^2 SW_r(a,b+\tau)^2}} \tag{15}$$

To highlight the presence of significant cross-correlation between the computed and observed series, but also to detect the scales mainly involved in such cross-correlation the wavelet cross coherence can be displayed as a function of scale only, by averaging the wavelet cross coherence over the entire time duration. It should be noted that, in atmospheric sciences, the results are interpreted in the time-frequency (or time-period) domain, rather than in the time-scale domain natural to the wavelet transform. The relationship between the equivalent Fourier frequency f_a (or period T_a) and the wavelet scale can be derived analytically for a particular

wavelet function. Formulas for several wavelet functions are given in Ref. [21]. The global wavelet coherence provides a global estimator of the coherence at a given scale:

$$\overline{CWCF}(a) = \frac{1}{N} \sum_{n=0}^{N-1} CWCF(a, b_n) \qquad (16)$$

This coefficient lies between 0 and 1, a near zero coherence indicates no relationship between the series, coherence close to 1 indicates strong relationship between the series. Moreover, to have a rapid and efficient identification of the correlation between the two series, at a given scale, the time lag which gives the maximum wavelet cross-correlation, or spectral correlation, τ_m, can be calculated for each scale. This parameter provides an estimator of the temporal synchronisation between two series at a given scale. By plotting the peak wavelet cross-correlation coefficient $WR_{cr}(a, \tau_m)$, or the peak spectral wavelet cross-correlation coefficient $SWR_{cr}(a, \tau_m)$, and the relative time lag (τ_m/T_a) on the same graph, we can check how the model behaves for all the scales: good synchronizations correspond to small relative time lag values, high correlations correspond to peak correlation coefficients value 1, whereas value 0 indicates a lack of correlation. At a given scale, good synchronisation and high correlation indicate good model performance.

5. Results and discussion

5.1. Global statistics

The results are shown on Figs. 3 and 4. The observed and estimated wind speed series appear to display a quite satisfactory match. The results in terms of various global statistics are presented in Table 1. From these statistics, it can be said that the ANN model predictions are in good agreement with the experimental data: the values of (R) exceed 0.9 (0.95 for the second model) and the ($NMSE$) values are inferior to 0.09. These good results quantify the "effectiveness" of the models in terms of their ability to accurately predict wind speeds at the target site using data from neighbouring reference locations. Therefore, this gives a good indication of the usefulness of the models in connecting wind data values collected on several locations. However, by using these classical performance measures, only the overall performances are considered. To measure the
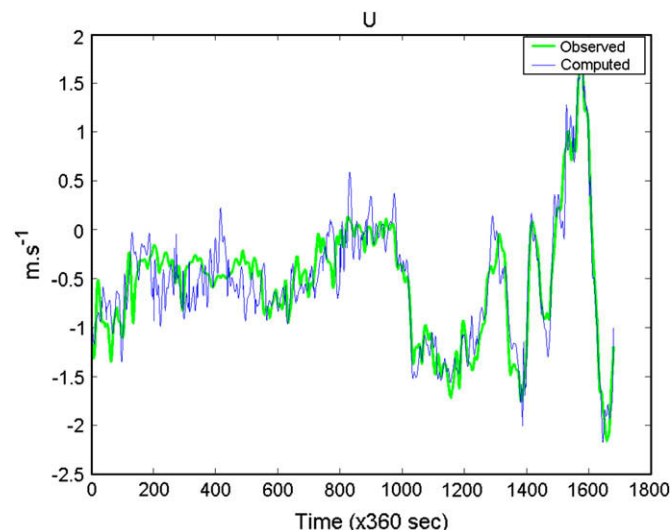

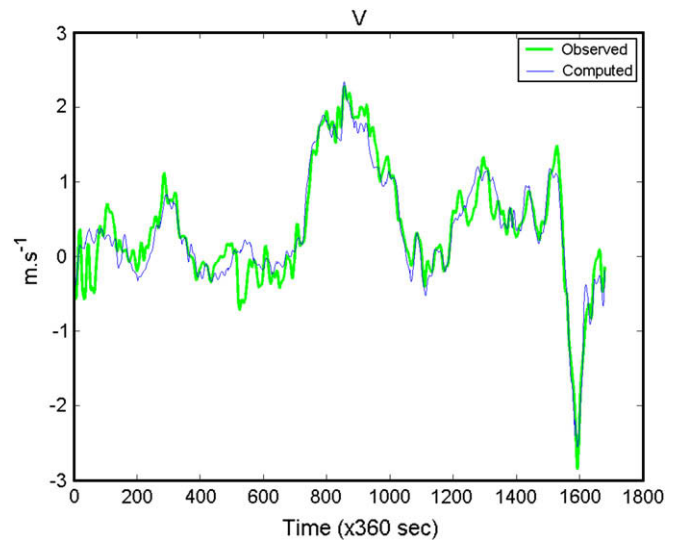**Fig. 3.** Estimated and observed U component.

Fig. 4. Estimated and observed V component.

accuracy of the models at each scale of fluctuations, a complementary wavelet coherence analysis is needed.

5.2. Wavelet coherence

The mean wavelet cross coherence coefficients are depicted in Fig. 5(a) and (b), as functions of the equivalent Fourier period, for the U and V component models respectively. For the U component, the mean wavelet coherence increases with scale, reaches a step value close to 0.5, and significantly increases for periods superior to 2.10^4 s, which correspond to low velocity fluctuations. Strong mean wavelet coherence, close to 0.8, is found at larger scales, while low level of coherence is observed at periods smaller than 10^4 s. For the V component, the mean coherence significantly increases for periods longer than 10^4 s and exceeds 0.8 for periods longer than 10^5 s. It should be pointed out that these coefficients only provide a global estimator of the wavelet coherence. However they indicate that the model performances increase with scales and that the trained models provide satisfactory results only at larger scales.

5.3. Wavelet cross-correlation

A wavelet cross-correlation analysis is also performed. The peak wavelet cross-correlation coefficients and the relative time lag are depicted in Fig. 6(a) and (b), as functions of the equivalent Fourier period, for the U and V component models respectively. These results confirm the wavelet coherence results. For both components, the peak wavelet cross-correlation coefficients increase with increasing scale, are greater than 0.8 for periods longer than 2.10^4 s and exceed 0.9 for the periods longer than 5.10^4 s, while the relative time lags are close to zero. The results provide a quantitative measure of the inefficiency of the models for the generation of high-frequency fluctuations and of their accuracy to generate the low-frequency fluctuations. But, since the coefficient continually increases with increasing scales, it seems difficult to define the "reasonable level" to which the correlation becomes "significant". Thus it is not possible to clearly define the range of validity of the models.

5.4. Spectral wavelet cross-correlation

To complete this wavelet cross-correlation study, the peak spectral wavelet cross-correlation coefficients, and relative time lag, are depicted in Fig. 7(a) and (b), as functions of the equivalent Fourier period, for the U and V component models respectively. For
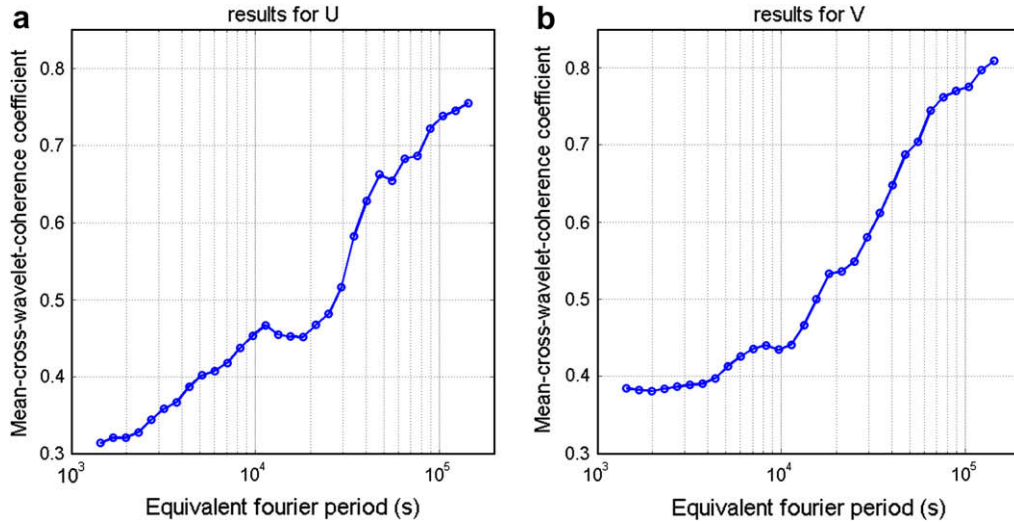
Fig. 5. Mean wavelet cross coherence coefficients between the estimated and observed wind speed components.

both components, for periods shorter than 10^4 s, the peak spectral wavelet cross-correlation coefficients are low and/or the relative time high. Then, the peak correlation coefficients significantly increase as the scale increases, and reach unit value in the period range longer than 2.10^4 s, while relative time lag are close to zero. It should be noted that the correlation values observed for periods shorter than 2.10^3 s, for the V component, are not representative since the relative time lag is close to 1. An examination of these graphs shows that the peak spectral wavelet cross-correlation coefficients reach step value close to 1 for high spectral correlations. As a result, by using these coefficients, it is possible to derive the range of validity of the models. Thus, one may conclude that the tested models, based on artificial neural networks, to define the time-varying wind speed, can be applied only to obtain large-scale components of the time series with low-frequency fluctuations and typical time periods longer than 2.10^4 s. This is a logical result since long-term regional wind data do not contain enough information for accurate computation of short-term local phenomena.

6. Combining ANN and statistical models

Time-varying wind speed can be expressed by a superposition of a rapidly varying component of small amplitude superimposed on a much more slowly varying one, see Fig. 1. The results presented in the previous section have shown that the ANN based models can be used to generate the slowly varying component of large amplitude. To generate high-frequency fluctuations, that occur in the form of wind gusts, over a typical time interval $[t,(t+T)]$, with T less than 2.10^4 s, it is necessary to use another synthetic data generation technique. Over such an interval, in a classical approximation, we assumed that the mean flow, close to surface, is directed parallel to the ground, therefore the wind velocity field is represented by the sum of a constant horizontal mean component, with constant direction, and a turbulent component considered as zero-mean stationary random process, with normal distribution

$$\vec{u}(M,t) = \overline{U}_{\mathrm{h}}(M)\,\vec{i} + u'(M,t)\,\vec{i} + v'(M,t)\,\vec{j} + w'(M,t)\,\vec{k} \quad (17)$$

$\overline{U}_{\mathrm{h}}(M)$ is the mean horizontal wind velocity over the time interval $[t,(t+T)]$, $u'(M,t)$, $v'(M,t)$ and $w'(M,t)$ are the longitudinal, lateral, and vertical turbulence components, $\vec{i}$ is the unit vector in the mean wind direction directed parallel to the ground, $\vec{k}$ is vertical and directed upwards. Since the wind series given by the ANN appear to be locally stationary (their statistical properties are slowly varying over time), they can be used in order to estimate the mean wind speed velocity and direction over the designated time interval: the wind components given by the ANN are converted to
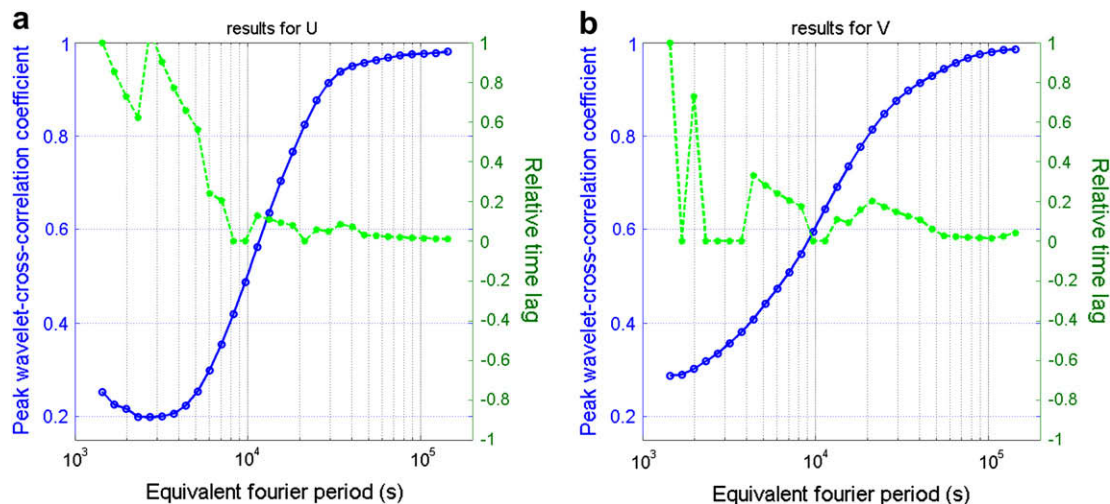

Fig. 6. Peak wavelet cross-correlation coefficients (–○–), and relative time lag (– –●– –), between the estimated and observed wind speed components.
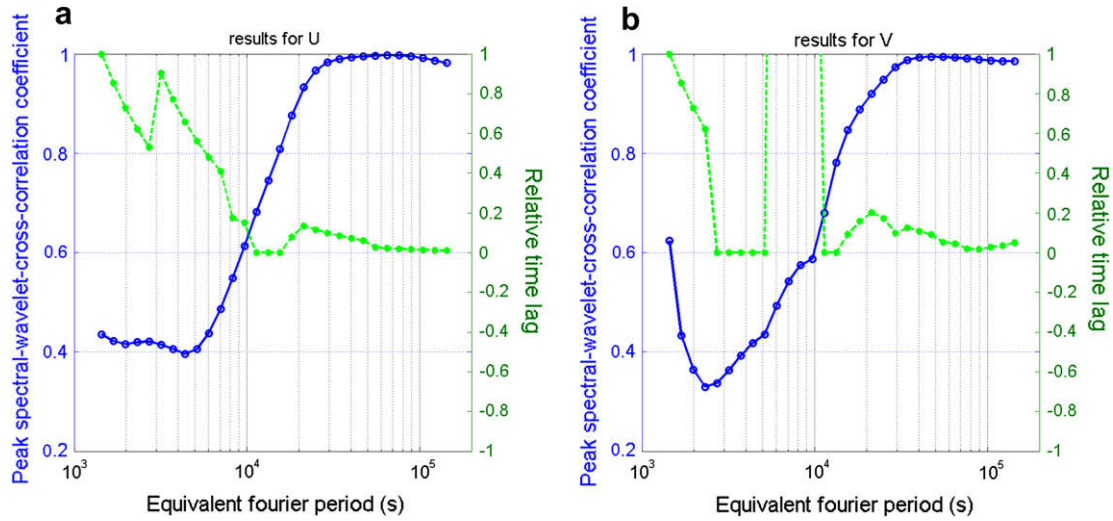
Fig. 7. Peak spectral wavelet cross-correlation coefficients (–○–), and relative time lag (– ● –), between the estimated and observed wind speed components.
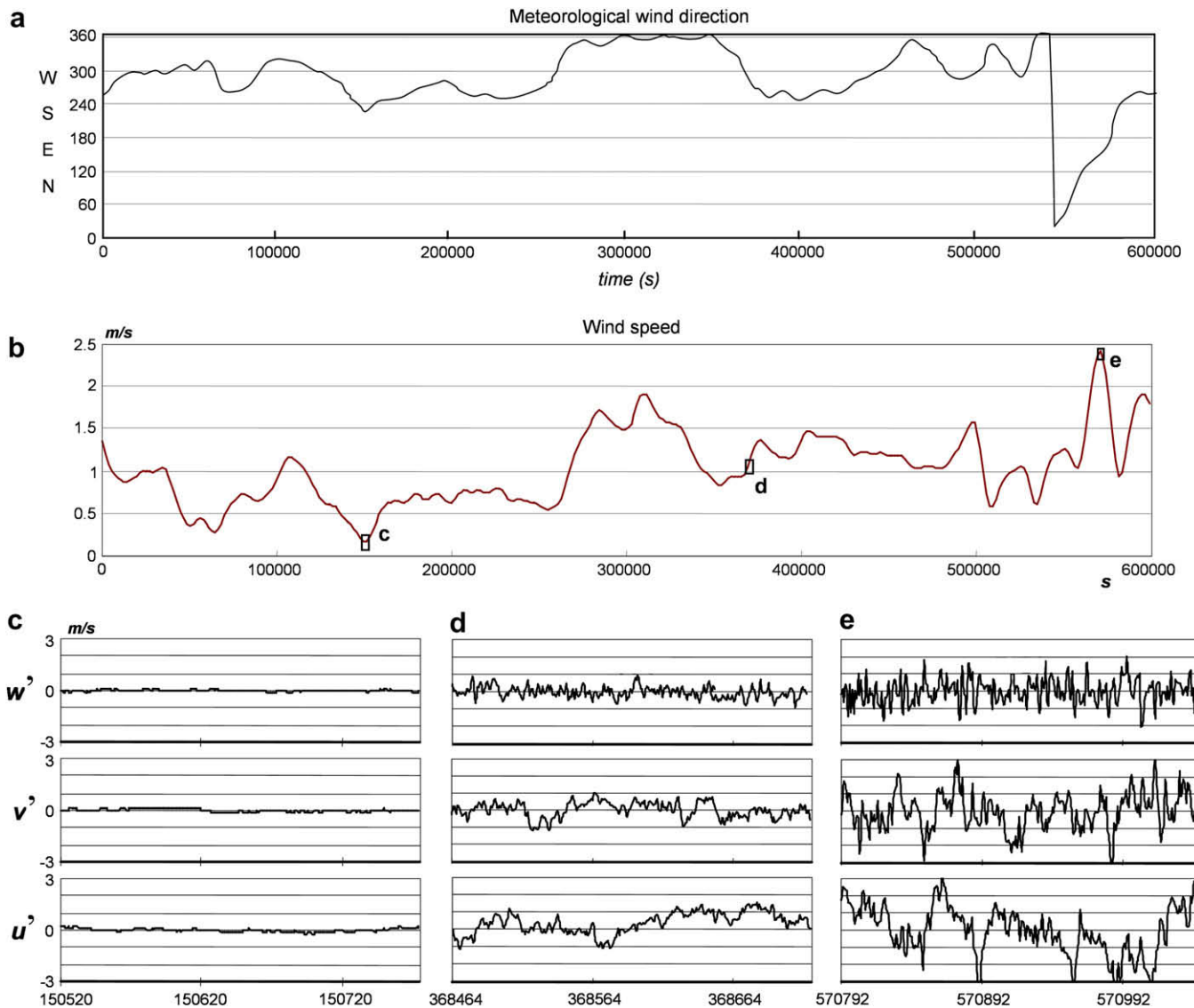

Fig. 8. Wind speed and direction derived from the ANN and velocity fluctuations over 3 characteristic intervals.

speed and direction. The direction from which the wind is blowing is given by

$$\Phi(M, t) = \frac{180}{\pi} \tan^{-1}\left(\frac{U(M, t)}{V(M, t)}\right) \tag{18}$$

It increases clockwise from north, when viewed from above, and is expressed in degrees (N: $0°$, E: $90°$, S: $180°$, W: $270°$). The horizontal wind speed is

$$|U_{\mathrm{h}}(M, t)| = \sqrt{U^2(M, t) + V^2(M, t)} \tag{19}$$

The slowly varying wind speed and direction are plotted in Fig. 8(a) and (b) respectively. The mean wind speed $\overline{U}_{\mathrm{h}}(M)$ and mean wind direction (the $\vec{i}$ vector direction) are obtained by averaging these quantities over the time interval T. Then, by using a suitable model to describe the statistical properties of turbulence, fluctuating components can be obtained by auto regressive moving average (ARMA) models or by wave superposition method. A full review of these techniques has been made in Ref. [4]. The conclusion is that the latter method "appears to give the best results in terms of overall quality of the generated signal". Following this assumption the generation could be carried out using inverse FFT techniques. The basic generation parameters are the statistical characteristics of the fluctuating wind speed components, such as their power spectral density function (PSDF), denoted as $S_\varepsilon(n)(\varepsilon = u', v', w')$, which describes the wind energy distribution over frequency n, their variance σ_ε^2, which quantifies the turbulence intensity I_ε, and their integral length scale in the mean wind direction, L_ε, which defines the position of the turbulence spectra content. In practice, these parameters vary with height (z is the height of the point M above the ground), with the roughness length z_0, and with the mean wind velocity. Their choice can be made through the use of empirical, semi empirical and theoretical statistical equations. A lot of data exist both in literature and in the building codes. In this study, the unified model, proposed in Ref. [7] and used for Eurocode 1 [22], has been chosen.

$$\frac{nS_\varepsilon(z, n)}{\sigma_\varepsilon^2(z)} = \frac{\mathrm{d}_\varepsilon n L_\varepsilon(z)/\overline{U}_{\mathrm{h}}(z)}{(1 + 1.5\mathrm{d}_\varepsilon n L_\varepsilon(z)/\overline{U}_{\mathrm{h}}(z))^{5/3}}$$

with $\mathrm{d}_{u'} = 6.868$ and $\mathrm{d}_{v'} = \mathrm{d}_{w'} = 9.434. \tag{20}$

$$\sigma_\varepsilon(z) = I_\varepsilon \overline{U}(z) \quad \text{with } I_{v'} = 0.75 I_{u'} \text{ and } I_{w'} = 0.5 I_{u'} \tag{21}$$

$$L_\varepsilon(z) = 300\lambda_\varepsilon\left(\frac{z}{200}\right)^{0.67+0.05\ln(z_0)}$$

with $\lambda_{u'} = 1$. $\lambda_{v'} = 0.25\lambda_{u'}$ and $\lambda_{w'} = 0.1\lambda_{u'}. \tag{22}$

Measurements on the site of Evry give a roughness length $z_0 = 0.75$ m and a longitudinal turbulence intensity $I_{u'} = 0.3$, typical of suburban terrain. It should be noted that, since the wind speed variance and PSDF depend strongly on the mean wind speed, the statistical properties of turbulence will change in accordance with the changing "mean" wind speed derived from the ANN. For example, velocity fluctuations given by the inverse Fourier transform method are plotted in Fig. 8(c), (d), and (e) over three time

intervals of 256 s. Their statistical properties are well functions of the slow varying "mean" wind speed: in particular, standard deviation significantly increases with increasing mean values.

7. Conclusion

Wavelet coherence, wavelet cross-correlation and spectral wavelet cross-correlation coefficients can be employed as useful performance estimators of ANN based models. These coefficients provide quantitative measures of the cross-correlation relationships, in terms of scale and time lag, between the computed and experimental verification data sets. If the data are characterised by fluctuations occurring over a wide range of scales, the spectral wavelet cross-correlation coefficient, initially defined to calculate the velocity of a thermal field [20], allows the determination of the scale below which the model fails to represent the fluctuations. In practice, the peak spectral wavelet cross-correlation coefficient and the relative time lag must be plotted on the same graph as functions of the equivalent Fourier period: in the validity range of the model, the peak spectral correlation reaches unit value, while relative time lag is close to zero. This method provides a simple yet effective way to obtain a quantitative measure of the performance of a model at any scale. As a practical application, wavelet cross-correlation analysis has been used to evaluate the performance of ANN based models used to estimate the local wind speed, on a selected site, from wind speed data recorded at neighbouring meteorological stations (Table 2). In our case, it confirms that only the largest scales of motion (low speed fluctuations) can be estimated from wind speed data recorded at nearby sites. The smallest scales (high speed fluctuations) have to be simulated by a superimposed short-term statistical model. Based on the above method, both the high and the low-frequency segments of the wind velocities spectra can be simulated, over a range of several hours, at the target site.

References

[1] Castino F, Festa R, Ratto CF. Stochastic modelling of wind velocities time series. J Wind Eng Ind Aero 1998;74–76:141–51.
[2] Kalogirou SA. Artificial neural networks in renewable energy systems application: a review. Renew Sust Energ Rev 2001;5:373–401.
[3] Öztopal A. Artificial neural network approach to spatial estimation of wind velocity data. Energy Convers Manage 2006;47:395–406.
[4] Rossi R, Lazzari M, Vitaliani R. Wind field simulation for structural engineering purposes. Int J Numer Meth Eng 2004;61:738–63.
[5] Welfonder E, Neifer R, Spaimer M. Development and experimental identification of dynamic models for wind turbines. Control Eng Pract 1997;5(1):63–73.
[6] Chen J, Xu YL. On modelling of typhoon induced non stationary wind speed for tall buildings. Struct Design Tall Spec Build 2004;13:145–63.
[7] Solari G, Piccardo G. Probabilistic 3-D turbulence modelling for gust buffeting of structures. Probabilistic Eng Mech 2001;16:73–86.
[8] Kalogirou S, Neocleous C, Paschiardis S, Schizas C. Wind speed prediction using artificial neural networks. In: Proceedings of the European Symposium on Intelligent Techniques ESIT'99. Crete (Greece); 1999.
[9] Ba N. Passage d'un modèle de connaissance régional à un modèle de connaissance local des vitesses et direction du vent, Ph. D. thesis. Université d'Evry, France; 2000.
[10] Sfetsos A. A novel approach for the forecasting of mean hourly wind speed time series. Renew Energ 2002;27:163–74.
[11] Alexiadis MC, Dokoupulos PS, Sahsamanoglou HS, Manousaridis IM. Short term forecasting of wind speed and related electric power. Sol Energy 1998;63(1):61–8.
[12] Sfetsos A. A comparison of various forecasting techniques applied to mean hourly wind speed time series. Renew Energy 2000;21:23–35.
[13] Srinivasulu S, Jain A. A comparative analysis of training methods for artificial neural network rainfall – runoff models. Appl Soft Comput 2006;6:295–306.
[14] Cholewo TJ, Zurada JM. Sequential network construction for time series prediction. In: Proceedings of the IEEE International Joint Conference on Neural Networks. Houston, Texas, USA; June 9–12, 1997. p. 2034-9.
[15] Labat D. Recent advances in wavelet analyses: part 1 a review of concepts. J Hydrol 2005;314(1–4):275–88.
[16] Kareem A, Kijewski T. Time-frequency analysis of wind effects on structures. J Wind Eng Ind Aero 2002;90:1435–52.

Table 2
Performance of the ANN based models

	Model for U	Model for V
AAE (m/s)	0.19809	0.1794
RMSE (m/s)	0.24959	0.2368
NMSE	0.089488	0.072926
R	0.92251	0.95487

[17] Krusche N, De Oliveira AP. Characterization of coherent structures in the atmospheric surface layer. Boun-Lay Meteorol 2004;110:191–211.

[18] Sello S, Bellazzini J. Wavelet cross-correlation analysis of turbulent mixing from large-eddy-simulations. In: eighth European turbulence conference. Barcelona, Spain; 27–30 June 2000.

[19] Li H, Nozaki T. Application of wavelet cross-correlation analysis to a plane turbulent jet. JSME Int J B-Fluid T 1997;40(1):58–66.

[20] Ngae P, Grignon M, Poloniecki JG. A new convection velocity based on the wavelet decomposition of a thermal field. Int J Therm Sci 1999;38:331–9.

[21] Torrence C, Compo GP. A practical guide to wavelet analysis. Bull Amer Meteor Soc 1998;79(1):61–78.

[22] European Prestandard. Eurocode 1-Actions on Structures, Part 1–4 General actions – wind actions. European Committee for Standardisation; 1996.