Applied Energy 105 (2013) 30–46

Contents lists available at SciVerse ScienceDirect

Applied Energy

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

Multi-decadal variability in a centennial reconstruction of daily wind

N. Kirchner-Bossi [a], L. Prieto [b], R. García-Herrera [c], L. Carro-Calvo [d], S. Salcedo-Sanz [d,*]

[a] Department of Physics of the Earth, Astronomy and Astrophysics II, Universidad Complutense de Madrid, Spain
[b] Department of Energy Resource, Iberdrola Renovables, Spain
[c] IGEO, Instituto de Geociencias, CSIC-UCM, Spain
[d] Department of Signal Processing and Communications, Universidad de Alcalá, Spain

HIGHLIGHTS

- A wind clustering methodology for wind speed reconstruction is presented.
- The method allows long-term reconstruction of daily surface wind series.
- An evolutionary algorithm and a constructive heuristic are presented.
- The method is tested in six meteorological towers at different wind farms in Spain, for the period 1871–2009.

ARTICLE INFO

Article history:
Received 15 September 2012
Received in revised form 25 November 2012
Accepted 28 November 2012

Keywords:
Wind patterns
Wind speed
Wind roses
Frequency distributions
Circulation weather types
Time series reconstructions

ABSTRACT

A wind clustering methodology capable of dynamically characterizing and long-term reconstructing daily surface wind series is introduced and tested for six meteorological towers at different wind farms in Spain, for the period 1871–2009. On this basis this paper provides for the first time a centennial surface wind reconstruction with a daily resolution without the need of numerical simulations. Thus, several soft-computing algorithms are developed, with public domain Sea Level Pressure (SLP) Reanalysis data as the only input. These algorithms are constructed by tackling an Euclidean distances' problem at the geostrophic speeds' space. Once the wind-independent classifications are obtained, the methodology is calibrated by linking the obtained classifications with observed wind data, thus allowing to estimate and characterize the daily surface wind speed and direction. A cross-validation is then performed in order to obtain several measures of goodness of the method, such as its wind speed estimation uncertainty in terms of *Mean Absolute Error* (MAE) and Pearson correlation (r) for both the wind module and vectorial values. Regarding previous approaches, this statistic downscaling shows an outstanding performance: Wind speed module estimates produce a MAE of 1.12 m/s (0.32 m/s) in some towers for a daily (monthly) scale, as r reaches values of 0.78 (daily scale) and 0.91 (monthly scale).

The wind-independent classifications allowed to perform daily surface wind speed and rose reconstructions in time periods when no wind data are available, which constitutes the main goal of this work. Thus, a 140 year daily wind reconstruction is performed and analyzed for one tower located at central Iberia. There, significant low frequency variations are detected, as well as wind speed oscillations in the 20 y band. Remarkable changes are also identified over reconstructed decadal wind speed frequency distributions and wind rose. Since long-term wind measurements are rarely available at modern wind farm sites, such an analysis on centennial reconstructed wind series can represent an appropriate tool that places the last years of observed wind speed in a climatological perspective.

1. Introduction

Understanding the wind conditions and variability in a particular location is of great importance from the point of view of climate [1,2], energy [3,4], safety [5,6] and environmental management [7,8]. Particularly, the knowledge of flow conditions at a given location and their associated synoptic situations becomes relevant when evaluating the efficiency and operability of wind energy farms [9,10].

Long-term observational wind datasets constitute a useful instrument to understand multi-decadal or lower frequency changes of wind variability. However, the observations are rarely available at locations with particularly high wind speeds. In the

* Corresponding author. Address: Department of Signal Theory and Communications, Universidad de Alcalá, 28871 Alcalá de Henares, Madrid, Spain. Tel.: +34 91 885 6731; fax: +34 91 885 6699.
E-mail address: sancho.salcedo@uah.es (S. Salcedo-Sanz).

0306-2619/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.apenergy.2012.11.072

Nomenclature

$\mathcal{A}$	set of 8 angular borders
a_i	a certain angular border
α	angle of a certain vectorial flux with respect to the North
c_κ	a certain class or WindType, $\kappa = 1, \ldots, 26$
e_i, e'_i	original element, mutated element
F, F_r, F_s	(V.) daily index of geostrophic flux intensity, index at train and test periods
f	coriolis parameter
Φ	cost function
φ, $\Delta\varphi$	longitude, longitude difference
G, G_u, G_v	(V.) daily geostrophic flux, geostrophic flux components (zonal, meridional)
i_r, i_s	a certain day of training period, test period
J	a certain grid point of the reanalysis SLP field
κ	number of class or WindType, from 1 to 26
λ, $\Delta\lambda$	latitude, latitude difference
μ	statistic mean
MAE_V, $MAE_{\lvert V\rvert}$	mean absolute error (vector, module)
N_1	normalized distribution with $\mu = 0$, $\sigma = 1$
p	probability
p_i	point of the grid, $i = 1, \ldots, 16$
$\mathcal{R}$	a 3×8 matrix containing the 3 radial borders of the 8 sectors
R	Earth radius
r_i	a radial border, $i = 1, \ldots, 3$
r	Pearson correlation coefficient
r^2	explained variance
ρ	air density
SF, SZ	southern component of F, Z
T, $T0X$	a certain tower, $X = 1, \ldots, 6$
τ, τ_r, τ_s	a certain time range, training time range, test time range
V, V_r, V_s	(V.) daily wind speed, wind speed at train and test periods
V_{c_κ}	representative of a certain class or WindType c_κ
WF, WZ	western component of F, Z
Z	daily index of geostrophic flux vorticity
Ω	Earth angular speed

case of wind industry, this circumstance becomes a real handicap when performing a mid to long-term wind farm feasibility layout. Usually, economic feasibility plans carried out by the electric industry consider a wind resource evaluation period that rarely exceeds two years of measurements. Since the atmospheric variability ranges from the daily cycle to century scales, this period of *in situ* wind measurements is clearly too short for providing information on the low frequency wind variability at the eventual wind farm, impeding a realistic estimation of the long-term wind power production and its variability. This problem is usually faced by harnessing a reanalysis [11,12] numerical model, that characterizes wind by performing dynamic simulations based on all the available previous meteorological data (meteorological stations, satellites, radiosondes, etc.). This paper provides a tool capable of characterizing and estimating daily wind over wide past time ranges through harnessing only a few years of observational wind series.

Contrary to observational wind series, information on synoptic circulation can be extended far back in time. The classification of a wide spectrum of daily synoptic circulation conditions into a specific number of patterns allows identifying empirically the underlying flow mechanisms influencing the local climate, an issue that is not possible if only local observational data are considered. Moreover, since synoptic circulation is also physically related to local surface wind, this classification allows a particularly accurate wind characterization if clustering conditions of synoptic circulation are properly established.

Since the 70s of last century, numerous studies have been conducted with the aim of classifying objectively large scale atmospheric circulation patterns, so called Circulation Weather Types (CWT). The methodology applied to obtain CWT classifications covers a wide set of possibilities. Initially, subjective manual classifications were performed on CWT [13] including those considering the Iberian Peninsula [14]. Jenkinson and Collison [15] developed an automated CWT classification based on geostrophic flow indexes derived from Sea Level Pressure (SLP) field. This methodology was applied also to the British Isles by [16] and to the Iberian Peninsula by [17,18]. Several automated CWT methods were implemented through correlation-based techniques [19,20] or [21].

The characterization of different meteorological variables, as precipitation [22,23] or temperature [24,25] among others have been possible through CWT techniques. Thus, [26] applied a *Cluster Analysis* (CA) to a *Principal Components Analysis* (PCA) classification for the obtention of rainfall-related CWT. Approaches in the study of variations in temperature and precipitation based on artificial neural networks have been also developed [27,28]. Nevertheless, with the target of better parameterizing wind conditions, wind-specific pattern (WP) classifications have been eventually induced directly from observed wind datasets. Several works have focused on this topic, such as [29], where a WP classification is performed by applying a PCA to a wind data set. In turn, [30] applied CA on temporal similarity. [31,32] implemented an automated *Complete Linkage Algorithm* (CLA) to construct a WP classification based on Euclidean distances within the wind speed space. Later, [33] included an ad hoc dendrogram algorithm. In [34] a PCA and a CA were applied to gridded wind reanalysis data. [35] performed both spatial (CA including CLA and an ad hoc k-means algorithm) and temporal (CA + PCA) similarity for achieving WP from a real wind data set, while synoptic pressure patterns were obtained from gridded data. In turn, [36] developed a monthly wind speed estimation method from the scoring of four SLP modes of variability selected through a *Canonical Correlation Analysis* in an attempt to understand multi-centennial wind speed changes and its association to the selected main circulation modes, obtaining a monthly wind speed module correlation of 0.7 with respect to observations. Some other works investigate on direct (i.e. with no classification) wind speed reconstruction processes. This has been done either through a geostrophic wind dataset obtained by interpolation of subjective forecast charts [37], or though statistics-based algorithms. To this extent, Autoregressive Moving Average Processes (ARMA [38,39]), Artificial Neural Networks (ANN [40,41]), Support Vector Machine (SVM [42,43]) and others [44,45] have been implemented.

Usually, papers on CWT methods employ pressure or geopotential field data for constructing pressure pattern classifications. On the other hand, most of the WP works obtain the wind patterns exclusively from observed wind data. In this paper we have developed a method that, although statistic, is based on the circulation atmospheric conditions that rule the climate. Specifically, a daily wind clustering methodology based on the optimization of Euclidean distances of geostrophic flow indices (exclusively derived from SLP data) is introduced and tested at six different locations in Spain. This methodology allows to identify dynamic situations that harbor common wind features, allowing the characterization of their seasonal frequencies and intensities as well as their prevailing direc-

tions. The Euclidean distances clustering problem (similarly to that employed by [31,32]) is this time computed through two different soft-computing algorithms [46],which obtain approximate, yet efficient, solutions when a hard problem with a large amount of possibilities is tackled. Regarding to this, in [47,48] a soft-computing real wind-dependent classification of SLP patterns is developed. This approach has been constructed without a discernible dynamic relationship among them but with an outstanding performance when validated. The current work departed from the necessity emerged then to develop a wind-independent classification designed through any dynamic criteria which could be reflected in the arrangement of the obtained classes.

Once the wind classifications are obtained through the introduced algorithms (training), the derived patterns are referred and assimilated to observational wind data through a calibration experiment, allowing to parameterize each obtained wind class with a specific real wind-derived vector. Since the geostrophic indices employed to classify are exclusively obtained from gridded SLP reanalysis data, the wind vector class-parametrization allows to characterize wind much beyond where observed wind data is available, as SLP series embrace a much wider temporal and spatial range than wind observations. An assessment of the performance and reliability of the wind classifications' estimation ability is conducted through a validation experiment (test) over the observational period, permitting to determine the uncertainties related to the wind speed estimates, as well as other measures of goodness as the Pearson correlation values with respect to the observations, for daily and monthly time scales. This statistic downscaling is employed to develop long term wind reconstructions beyond the wind observations' period. These reconstructions allow the characterization of low frequency multi-decadal wind variability, which constitutes the main goal of this paper. To this extent, a wind reconstruction for the last 140 years in central Iberia is presented and analyzed, in terms of both speed and directional wind features.

In the next sections an exhaustive description of the wind classification methodology (Section 2) is introduced, and details on the employed data (SLP and wind observations) are provided (Section 2.4). The results of the paper (Section 3) consider two different issues on the obtained classifications. First (Section 3.1), the reliability of the developed methodology is assessed through different measures of goodness, while (Section 3.2) presents and discusses the annual, decadal and multi-decadal variability of a centennial wind reconstruction at a meteorological tower in central Spain. Finally, the main conclusions (Section 4) are presented.

2. *F* field optimization methodology

This section explains the process followed to construct daily wind classifications by harnessing SLP gridded data, as well as the way those classifications are eventually referred to real wind speed to obtain a reliable wind reconstruction tool.

First, all the considerations leading to the optimization problem approached in this work are described. Then, specific conditions and particular features are laid out for the designed algorithms. After this, the wind statistic downscaling process is explained. These three points constitute the training part of the method. Finally, the testing part is described. There, the validation of the approach with respect to the observations is referred, so that the wind speed uncertainties as well as other measures of goodness can be calculated.

2.1. Optimization problem description

The algorithmic architecture within the new methodology departs from considering the role of the surface pressure field over surface atmospheric circulation. In this way, the atmospheric circulation features have been parameterized through two SLP-directly derived geostrophic indices, *F* and *Z*. The first one is directly proportional to the geostrophic wind speed, while the second refers to the absolute geostrophic wind vorticity. In this work both measures have been obtained through the utilization of gridded SLP data. The specific formulae which relate these indices with gridded SLP are detailed in Appendix A. Through these relationships *F* can be considered as a proper proxy [37] for the observed real wind, *V*. In turn, vorticity given by *Z* can provide some complementary information on characterizing wind conditions when the flow intensity given by *F* is weaker than a given threshold.

The procedure followed for the obtention of *F* and *Z* in a given location *J* harnesses the SLP field in a similar way to that employed in the weather classification technique developed by [15], referred hereafter as WT. There, *F* and *Z* can be obtained from interpolation of certain pressure values adjacent to *J*, when a gridded SLP field is provided (see Appendix B).

The WT technique has been applied alongside to our methodology to be used as a benchmark for our results. The WT is a classification performed by establishing a set of rules concerning *F*, *Z* and α which lead to the definition of 26 circulation types: eight of them are pure directional related to the eight wind rose main directions, two are rotational classes defining pure cyclonic and anticyclonic patterns respectively, and 16 are considered hybrid classes, produced by the mixture between pure directional and each one of the rotational classes. In order to compare our methodology with the WT approach, our geostrophic flow characterization will also result into 26 classes.

Unlike WT, the classification criteria of the present methodology are based in a problem of dispersion minimization at the *F* vectorial space. Let F_r, $r = 1, \ldots, r$, be a series of daily *F* vector calculated at a given location *J*, for a given period of time τ_r, associated to a certain class c_κ, $\kappa = 1, \ldots, n$. The new methodology performs an optimization approach in such a way that the dispersion of the values F_r associated to the classes c_κ is, in average, minimized. This minimization is implemented by a *cost function* Φ, expressed as follows:

$$\Phi_F(J) = \frac{1}{\tau_r}\sum_{\kappa=1}^{n}\sum_{i_r \in c_\kappa} |F_{c_\kappa} - F_{i_r}| \tag{1}$$

where i_r stands for a generic day of the period considered, and F_{c_κ} stands for the average of *F* values within a class c_κ. This problem is based in the same geometric concept employed by [31,32] and others to develop their wind field spatial similarity methods within their clustering processes. However, in this work the optimization for the reduction of *F* dispersion independently from observed wind data *V* allows to characterize wind conditions in those points or time series where no real observations are available.

2.2. Training: F field optimization algorithms

The *F* cost function optimization problem is faced through two soft-computing algorithms, which are performed in a given period of time τ_r (*r* for"Training") over six grid points near their corresponding meteorological towers in Spain, so called hereafter T01–T06. One of the approaches applies an evolutionary computing method [49], which allows a high computing performance when a large range of possibilities is considered. In turn, the other one consists on a greedy [50] algorithm approach, designed in order to obtain an analytical solution for the problem as its results can be easily compared to those obtained by WT. Through these algorithms, daily classifications according to the geostrophic wind conditions set by *F* can be obtained. Additionally, *Z* values are also considered to distinguish among weak flow conditions.

2.2.1. F field optimization through evolutionary computing (FE)

Evolutionary Computation (EC, [51–53]) is a subfield of artificial intelligence which considers a set of stochastic and population-based optimization techniques which are based in the concepts of genetic evolution [54]. EC tackles problems with a high amount of possibilities by evolving approximated solutions in a computer, following certain rules borrowed from natural evolution. Thus, it is expected to achieve a high performance in the problem of *F* daily values' dispersion minimization. Different evolutionary algorithms have been applied to many different optimization problems, in a wide range of applications, such as energy-related [55], short-term forecasting [56], or financial markets [57] problems.

The iterative procedure of FE is performed by means of an *Evolutionary Algorithm* (EA), which is an algorithm of the family of EC. Given an optimization problem, an evolutionary algorithm typically starts from an initial set of random solutions, group into a *population*. These solutions are submitted to a set of evolutionary operators [54], which evolve and finally retain or dismiss them. This process is applied repeatedly, in a set of loops called *generations*. Individuals are normally selected according to the quality of the solution they represent, so a fitness operator is applied to each individual of the population. Hence, the individuals with the best values of fitness are more likely of being selected for replication and survival. The selected individuals are reproduced by means of crossover and mutation operators. While crossover exchanges some genetic material between two or more individuals, mutation changes parts of individuals with a small probability, avoiding that the algorithm keeps in local minimums. By applying this iterative procedure, the EA explores the whole space of possible solutions, as it has shown to be highly efficient in extensive spaces.

The optimization problem to be tackled by FE consists on the minimization of the cost function enounced in Eq. (1). The 26 classes which will define an individual in the *F* space are sorted into eight angular sectors and three radial magnitudes per sector. Additionally, two classes with low overall *F* values are also defined. Specifically, the set of conditions to determine the 26 classes is distributed into eight angular borders $\mathcal{A} = [a_1, \ldots, a_8,]$, $a_i \in [0, 360°]$, which will define eight angular sectors. In turn, every sector is split into 4 *slots* by 3 radial magnitudes ($r_{a,1} < r_{a,2} < r_{a,3}$), so that a matrix $\mathcal{R}$ of size [3,8] is defined. The three sectorial classes keep defined by the three highest slots, as a near-to-zero area is defined bye the lowest slot of every sector. This area for low *F* values determines in turn two classes depending on wether *Z* is positive (cyclonic calm) or negative (anticyclonic calm). Eventually, all $\mathcal{A}$ and $\mathcal{R}$ elements are set to be adjusted in each iteration so that a consistent solution is obtained. Fig. 1 shows a result of this stratification on the *F* = [*WF*, *SF*] space, where *WF* and *SF* stand for the *F* zonal (from West to East) and meridian (from South to North) components. There, each dot represents the daily F vector, in terms of the origin of the flux.

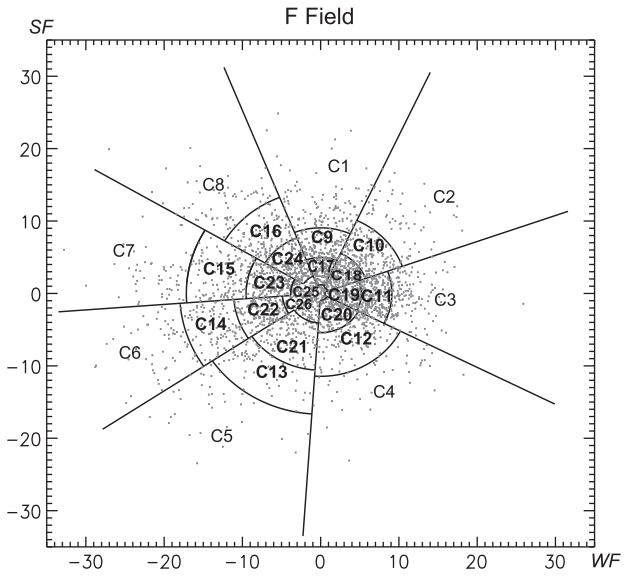

Fig. 1. Example of the class-stratification structure performed over the *F* field by the FE algorithm, where angular and radial borders are defined.

Each generation can be divided into a series of steps. At each step, the population is treated by a certain operator. Following, technical details on the FE operators implemented over a generation with a population of *N* individuals characterized by $\mathcal{A}$ and $\mathcal{R}$ are described:

(1) *Initialization*: Each individual in the population is randomly generated within the margin of the angular borders and radial magnitudes (i.e. we keep the constraints $a_i \in [0,360°]$ and $r_i \in (0,\text{inf}).$).
(2) *Repair*: In order to keep the increasing order at both angles and the radial borders in the encoding, elements are sorted from the smallest to the largest value.
(3) *Fitness*: Each individual is associated with a value of fitness (i.e. performance) obtained from the cost function (Eq. (1)).
(4) *Selection*: The algorithm selects those individuals whose *F* average class-dispersion (or bias) is lower than the population's average dispersion. The rest of individuals will not survive for the next generation, and they will be replaced by new elements created through crossover and mutation of selected individuals.
(5) *Crossover*: Two individuals from the selected population are taken at random to generate a new one. The configuration of the elements of the new individual, for both the vector $\mathcal{A}$ and matrix $\mathcal{R}$, is performed within a multi-point crossover procedure. In this crossover, each a_i an r_i from a parent has a probability of *0.5* of being transmitted to the new individual.
(6) *Mutation*: The new individuals have a small probability of being mutated. For FE this probability is 0.05. The mutation consists of modifying 50% of the elements from the vector $\mathcal{A}$ and matrix $\mathcal{R}$ as:

$$\mathbf{e}_i' = \mathbf{e}_i + 0.1 \cdot \mathbf{N_1}(\mathbf{0}, \mathbf{1}) \tag{2}$$

where e_i stands for a given element of $\mathcal{A}$ or $\mathcal{R}$ before mutation, e_i' is the element after mutation and $\mathbf{N_1}(\mathbf{0}, \mathbf{1})$ represents a normalized Gaussian distribution of random numbers with $\mu = 0$ and $\sigma = 1$. Note that we keep the constraints $a_i \in [0, 360°]$ and $r_i \in (0, \text{inf})$.

For this work, this process has been performed over 3000 generations for a population of $N = 1000$ individuals.

As FE employs random sequences at some of its operators, it works as a non-deterministic application, and it obtains a different solution each time that it is launched. In this work FE has been launched 30 times for each one of the six points *J* where our methodology has been applied. At each one of them, only the best outcomes on dispersion minimization are eventually reflected in the results.

2.2.2. F field optimization by a greedy algorithm approach (FG)

Unlike FE, the *F* field optimization by a greedy [50] Algorithm (called hereafter FG) has been conceived to produce an analytical (i.e. deterministic) solution of the cost function minimization. To achieve this, FG introduces two differences in the conditions of the algorithm structure, compared to FE:

First, the angular borders which determine the classes at the F space have been kept fixed, so that the defined sectors have been timed to coincide with the eight cardinal wind rose directions (similarly to WT). With this, the physical interpretation of the synoptic dynamics behind the obtained classes is facilitated.

Second, FG has been set to track the dispersion minimization for every sector individually. By having fixed the angular borders, this procedure can be done without implying any loss of performance. Thus, to fix the three classes at a given sector the algorithm operates by exploring all the possibilities in the combination of $r_{a,1}$, $r_{a,2}$ and $r_{a,3}$ within that sector, with a tracking accuracy of 0.1 m/s. The rest of the algorithm conditions is similar to that of FE. Hence, the algorithm conditions are similar to $\mathcal{R}$ set in FE, and a 3×8 condition matrix is obtained.

2.3. Observational wind series assimilation and wind speed accuracy test

With the implementation of these methods, a daily classification according to the criteria established to arrange F and Z daily values is obtained. Now, the 26 obtained patterns are associated to observational wind series with the aim of parameterizing each wind class with a specific real wind vector. With this calibration, each one of the obtained wind types can be characterized according to real wind in terms of wind speed and angular configuration.

To attain the speed characterization, the already performed classification of the set of days i_r into 26 classes c_κ for six different grid point locations are eventually assimilated into their corresponding daily values of observed wind speed $V_r = [u, v]$ at the six meteorological towers T01–T06, located near the considered grid points. In this context, a wind *representative* V_{c_κ} is defined for each class as the barycenter, in the space of speed $[u, v]$, formed by the wind vectors V_r corresponding to the set of days $i_r \in c_\kappa$ for a given period τ_r. Through this linkage between the daily classifications and the observed wind, the clustering capability of the obtained classifications can be assessed by measuring the resulting real wind speed dispersion per class. This can be made in the same way as the dispersion of F was measured, this is by substituting F for V in Eq. (1). Additionally, wind speed frequency distributions can be estimated, this time by considering for each class not only a representative V, but a complete wind speed distribution computed from the set of all the class-elements wind speeds at the observational period. Similarly to wind speed, the angular information obtained through this observational data assimilation allows to compute a different wind rose for each one of the obtained patterns, enabling the possibility to estimate the overall wind rose for a certain period where only SLP information is available.

The vectorial parametrization performed for the wind patterns allows estimating wind speed in periods when only SLP data are available. In order to measure the degree of accuracy of this parametrization with respect to the real wind values, a validation exercise has been performed over the considered SLP and observed wind series belonging to a period τ_s (s for "Test") not employed previously. There, the new set of days, according to their F_s and Z_s daily values, are distributed into the 26 classes previously defined at τ_r. This data crossing is needed in order to prevent the overfitting of the method [58]. This validation has been designed to determine the value of the uncertainty of the wind estimation capability of the methods. It can be determined by measuring the differences between the wind representatives V_{c_κ} defined in τ_r and the corresponding values V_s associated with a class c_κ. This uncertainty can be defined for each considered tower T through

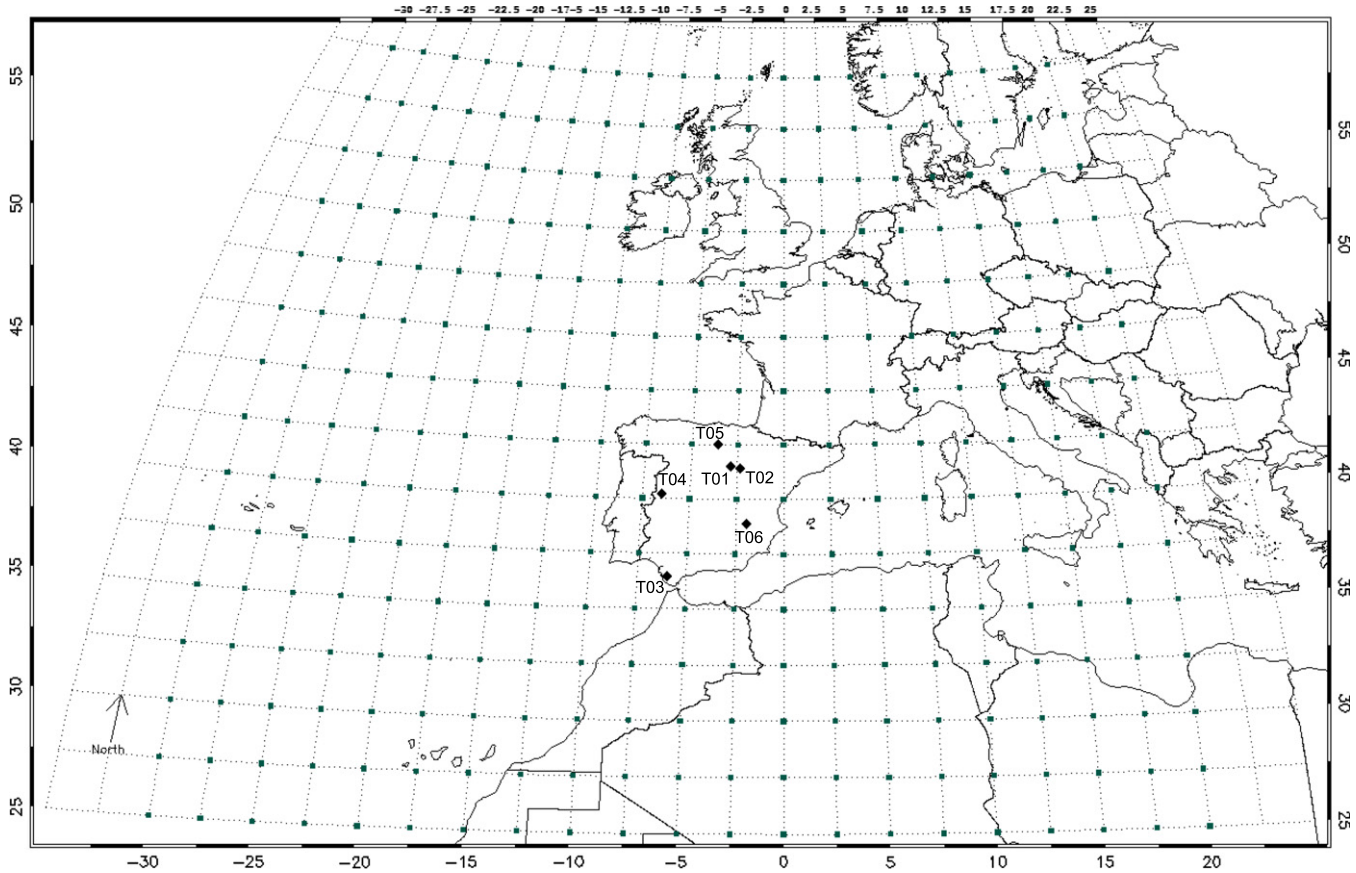

Fig. 2. Domain of the employed SLP data set. Squares represent considered grid points as diamonds stand for the position the six towers taken into account.

the *Mean Absolute Error* (MAE), which has been applied to the vector (MAE_V) and the module ($MAE_{|V|}$) values of wind, and can be calculated through these expressions:

$$MAE_{|V|}(T) = \frac{1}{\tau_s}\sum_{c_K=1}^{26}\sum_{i_s \in c_K} ||V_{c_K}| - |V_{i_s}|| \quad (3)$$

$$MAE_V(T) = \frac{1}{\tau_s}\sum_{c_K=1}^{26}\sum_{i_s \in c_K} |V_{c_K} - V_{i_s}| \quad (4)$$

where i_s stands for a generic day of the period τ_s. These expressions for the wind speed uncertainties represent a goodness measure for our methodology and reflect the degree of accuracy reached when daily wind reconstructions are performed.

By accounting for with these 26 representatives V_{c_K}, the *F* space borders configuration, and the *F*, *Z* and alpha series for a certain period where SLP field data is available, it is possible to perform a reconstruction of the wind conditions for such a period. In Fig. 3 a fluxogram of the whole process is depicted.

2.4. SLP and observed wind data

Sea level pressure gridded data have been retrieved from the National Center for Environmental Prediction/National Center for Atmospheric Research Reanalysis Project (NCEP/NCAR) [11]. They consist on daily SLP values at 1200 GMT, with a grid resolution of 2.5 × 2.5 degrees for the period 1948–2009. As it is shown in Fig. 2, a uniform grid of 15 × 21 grid points (latitude and longitude) centered on the Iberian Peninsula has been considered. Since SLP data have not been spatially interpolated, the closest grid point (*J*) to a meteorological tower location *T* has been selected as the central point on which perform the calculations.

On the other hand, wind speed and direction data from six meteorological towers distributed throughout Spain (see Fig. 2) have been considered for the period 1999–2009, so that τ_r and τ_s comprehend the period 1999–2005 and 2006–2009 respectively. They consist of ten minute frequency data taken at 40 m height. Every data element in the series comprises all wind speed and direction measured values over one day, from 00.00 GMT to 00.00 GMT. Thus, averages over 24 h have been performed to obtain daily average wind speed vectors *V*.

In order to perform a wind reconstruction back to the 19th century, an additional SLP dataset from 1871 to 1947 has been considered to complement the SLP NCEP/NCAR dataset. It has been retrieved from the second version of the "Twentieth Century Reanalysis" Project (hereafter 20*CRV*2) [12]. 20CRV2 has a spatial resolution of 2 × 2 degrees, and is based on surface and sea level pressure observations in spite of radiosonde data. It has been developed by applying an Ensemble Kalman Filter to the background "first guess" supplied by an ensemble of 56 forecasts obtained from the GFS prediction model run globally.

3. Experiments and results

This section consists of two main parts. The first one (Section 3.1) focuses on the measures of goodness calculated on the developed methodology. There, the consistency of the obtained SLP patterns (Section 3.1.1) and the class clustering capability (Section 3.1.2) are analyzed, and the wind speed estimation accuracy (Section 3.1.3) in terms of wind speed uncertainty and Pearson correlation are presented and compared to other approaches as WT. In the second part (Section 3.2) a wind speed (Section 3.2.1) and a wind rose (Section 3.2.2) reconstruction at T01 for the period 1871–2009 over central Iberia is carried out by implementing the vectorial parametrization of the obtained classifications to a 140 year daily SLP dataset.

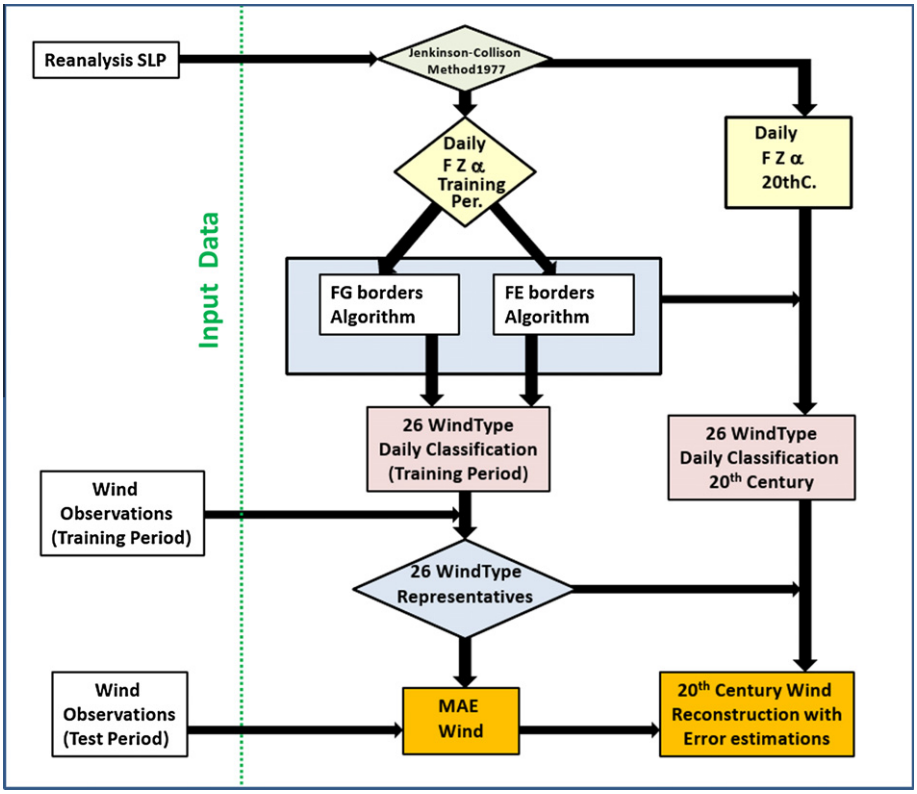

Fig. 3. Fluxogram of the overall process (for either FE or FG).

3.1. Wind characterization of the obtained classifications

3.1.1. Intra-class SLP fields consistency

In order to evaluate the consistency of the synoptic situations associated to the obtained classes, a study on the pressure dispersion per grid point has been performed to the different composites related within the classes c_κ generated by WT, FE and FG methods. This has been implemented by assessing the standard deviation σ of the obtained SLP values per class and grid point. To do this, a Monte Carlo analysis was performed by comparing the obtained σ's to those obtained by a set of 1000 artificially generated classifications whose elements within each class were randomly chosen (hereafter called RC classification).

To perform a reliable statistical analysis, RC has been configured with the same amount of elements per class than those classifications obtained from WT, FG and FE. In Fig. 4 the relative decrease of the average SLP σ obtained for each method versus the average σ of RC are represented for each location. Results show that FE and FG keep inside each class an overall higher consistency per grid point compared to WT. Specifically, FG shows the best performance in all locations, while FE is better than WT in all sites except T03, where a slight difference (0.3%) takes place. A Chi-square test has been performed to assess the statistical significance of the SLP dispersion reduction. By considering all the grid points of the domain, the average statistic significance of the SLP dispersion reduction per grid point, averaged for all six considered locations, is shown to be larger for the new methods. Thus, the amount of grid points (%) with a significant ($p < 0.95$) dispersion reduction was 48, 50 and 56 for WT, FG and FE respectively. These results show a robustness in FG and FE similar or even higher than WT to explain the synoptic circulation ruling their classes, specially when a large spatial scale is considered. These results suggest that the obtained classes are more unambiguous and better related with *V*, which implies, apart from a better characterization of wind, the possibility of developing important climatic applications, being their analysis beyond the scope of this paper.

3.1.2. Wind clustering performance

As it has been explained at Section 2, *F* represents a valid proxy for the real wind, *V*. Figs. 5 and 6 show *F* and *V* values obtained with FE (FG showed a similar behavior) for two different locations. There, each point represents the pointer of the obtained values for *F* and *V* vectors (describing the origin of the flow), at two towers with so different wind conditions as T01 and T03. While T01 can be representative for the overall behavior of geostrophic circulation with scarce orographic complexity at central Iberia, T03 stands for a location with singular and strong local effects. Each box depicts *F*'s and their corresponding *V*'s values of the set of days related to each resulting class, for each one of the 26 classes. From results at T01 it can be observed for most classes that *F* and *V* show similar components, only altered by roughness effects. Indeed, the observed differences between *F* and *V* for both FG and FE methods in all towers can be explained through *Ekman* effects due to local surface roughness and orographic causes [59]. This can explain that, with respect to *F*, *V* shows in most of the days a weaker intensity (a 24% smaller in average) and a left tending direction (31° in average) in the speed space. This behavior corresponds with the prevalence of the synoptic circulation on the flow conditions. On the other hand, in T03 *V* shows a particular behavior additionally to Ekman effects, showing some prevailing directions in the eastern classes. This particularity can be explained by the existence of local effects caused by the proximity of the Gibraltar Strait [60] to the area of study. There, the particular funnel orography exerts a big influence in the direction of flow, promoting the predominance of strong *levanters* (easterlies).

Although differences between *F* and *V* can be larger in T03 than in T01 due to local flow distortions, the turn from *F* into *V* evidences for all the towers an overall capability in both FG and FE methods to retain similar wind features within elements belonging to a same class. This clustering capability within *V* can be measured by calculating a dispersion measure as the *radius* of a class, i.e. the average Euclidean distance to the class representative V_{c_κ} in the speed space. Hence, this magnitude can be also derived through Eq. (1) by employing this time *V* instead of *F*. The average vectorial radius per class is clearly smaller for FG and FE methods, compared to WT, for the six considered towers. Thus, in average for all towers WT showed a vectorial *radius* of dispersion a 12 and a 17% compared to FG and FE respectively. In turn, FE showed the best performance in situations with complex orography as T03, with an improvement of a 18% when compared to WT, while the improvement obtained by FG reaches a 9%. This can be explained through the fact that FE is the only method which considers the angular borders as a variable to be adjusted, allowing the predominant directions to be fixed more accurately. When strictly speaking of wind speed module, performances are similar. The obtained average distances for all towers was 1.70 m/s for WT. In turn, FG and FE submitted distances of 1.40 and 1.38 m/s respectively, a clustering capability more than 23% higher compared to WT.

3.1.3. Wind speed validation test

In this section the wind speed estimation ability of the considered methods (FG and FE) is analyzed with respect to the observa-

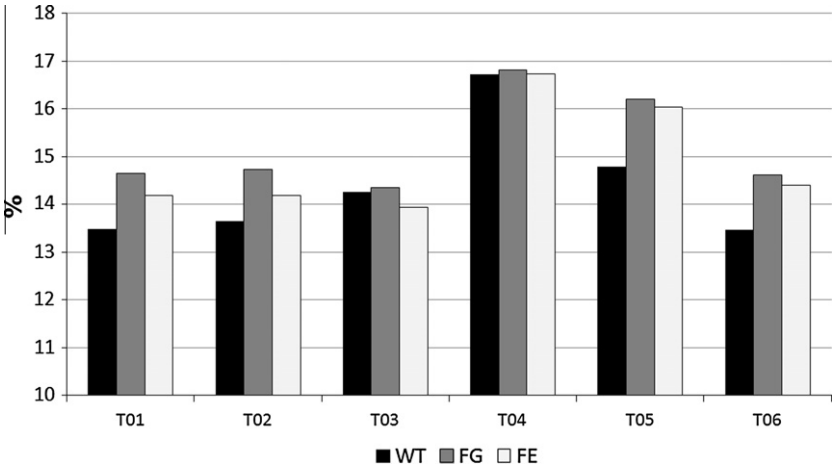

Fig. 4. Average decreasing ratio of the SLP field dispersion against a random classification. Histograms show results as spatial averages for each classification versus the mean SLP dispersion considering a random classification with same number of elements per class.

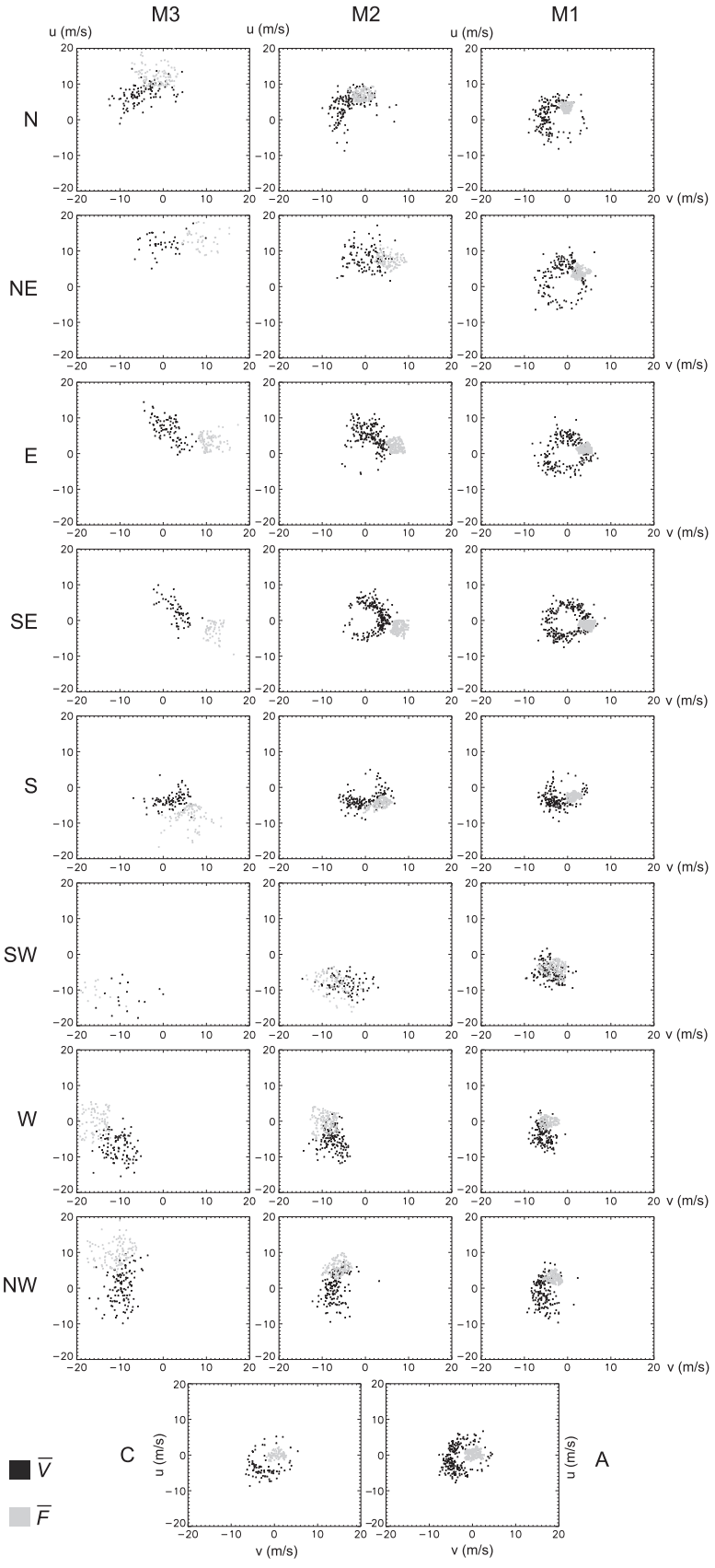

Fig. 5. *F* and *V* values for each class obtained by the FE algorithm for tower T01.

tional wind series in terms of two measures of goodness (observed wind speed error and Pearson correlation). Additionally, wind roses are also developed for the test period, so that the angular configuration accuracy can be qualitatively assessed. By following

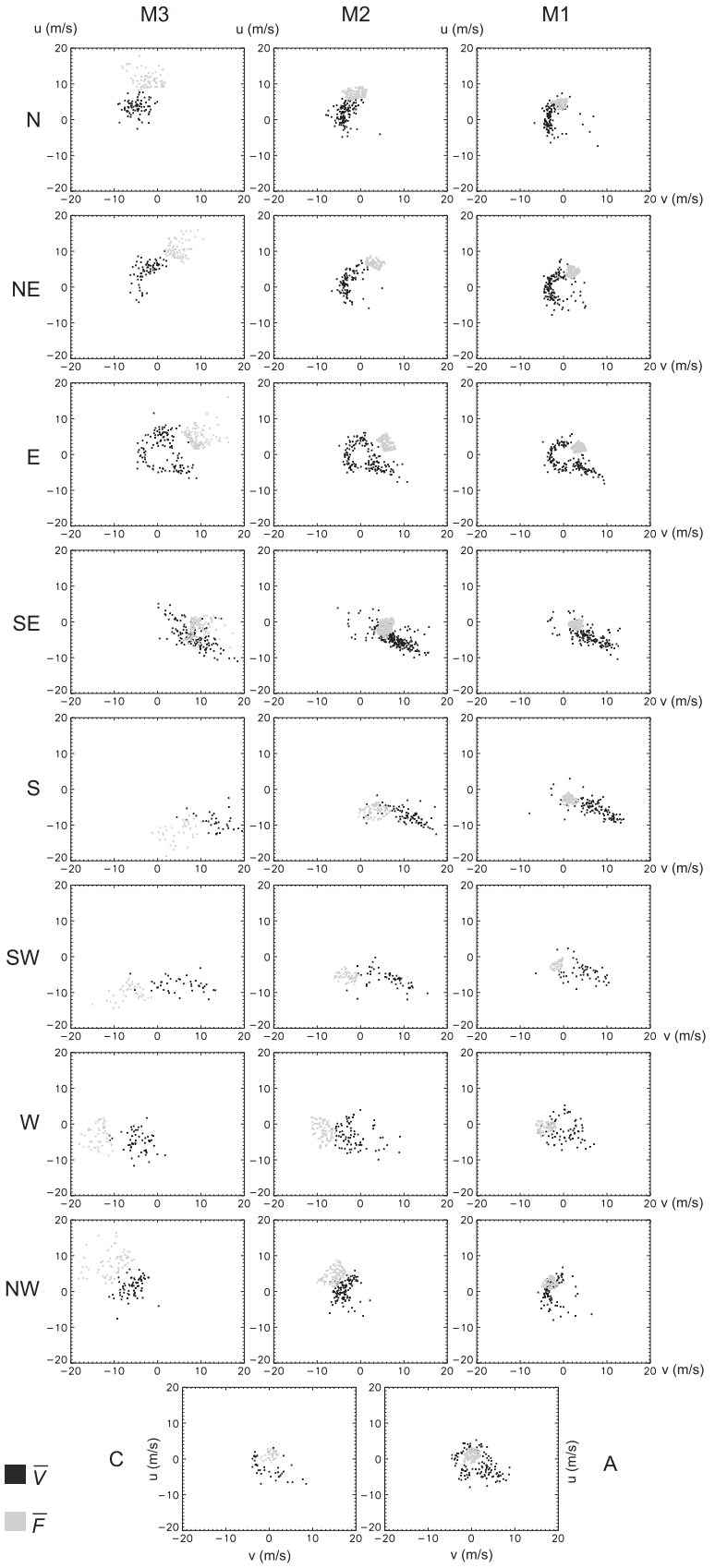

Fig. 6. *F* and *V* values for each class obtained by the FE algorithm for tower T03.

the methodology explained in (2.3), the representative class values V_{c_K} obtained in the training period τ_r and the observed wind speed at the test period τ_s have been cross-linked (to avoid overfitting) so that the bias of τ_r classification with respect to the observed wind

speed as well as its Pearson correlation could be estimated. The values for the bias have been computed in the form of the Mean Absolute Error for both wind speed ($MAE_{|V|}$) and wind vector (MAE_V) variables, which represent a measure for the wind speed uncertainty of the performed methods. In addition, Pearson Correlation Coefficients have been calculated for the comparison between real and reconstructed wind series.

The reconstruction performance of the new methods has been compared to that obtained by a *no-classification* (NC), which is an experiment where the entire wind speed series is treated as a unique class. Thus, the error obtained with this method can be considered as an upper bound of the implemented methods' accuracy. In addition, MAE values have been also calculated for the surface wind speed provided by NCEP/NCAR [11] reanalysis itself (W-NCAR), at the same grid point locations employed for WT, FG and FE. In order to make the comparison with the other methods reliable, a linear regression between W-NCAR wind speed values and the observed wind has been performed for the τ_r period, and thus the obtained linear coefficients have been applied to the test period to perform the reconstruction. This has been done since W-NCAR data sets had not been previously referred to observed wind speed, while WT FG and FE had done so in the Train-Test experiment.

In Table 1, results on the obtained MAEs are shown. Among the considered inputs, our methodology presents the smallest module and vectorial errors. Thus, FE shows the lowest values for $MAE_{|V|}$ (1.44 m/s) and MAE_V (3.77 m/s) in average for the six towers. In turn, FG presents, respectively, values just a 3 and a 5% higher than FE. Specifically, FE performs the lowest $MAE_{|V|}$ for all towers, except for towers T06 and T05, where FG is equal and slightly higher than FE respectively. These exceptions do not occur for MAE_V, where FE values are slightly lower than FG for all towers. Compared to the rest of the inputs, WT $MAE_{|V|}$ is higher than FG for all classifications in all towers, showing an average error a 25% higher than that obtained by the *F* methods, and reaching a 34% higher in T06, where the obtained $MAE_{|V|}$ scores lowest (1.12 m/s) for both FG and FE. These values can be compared to those obtained by [42], which obtains a 37% higher error, [43], with a 1.28 m/s $MAE_{|V|}$ in the best case and [38], where our performance on estimating the wind speed daily average can be compared to that obtained when forecasting wind speed with only one hour in advance.

Regarding MAE_V, WT lead to values a 13 and a 19% higher in average than FG and FE respectively. As W-NCAR approach presented high correlations (0.67) with daily real wind speed module at the considered towers (see Table 2), it could be considered as representative for them. However, as it was expected, it is not an accurate approach of the wind speed, as no classification based on dynamic features was set. The highest daily correlation with real wind is performed by FE and FG, with a similar 6 tower average value of 0.73. The biggest difference between them is observed for the T03 location (that on the Gibraltar Strait), with FE performing better (0.74) than FG (0.71). In turn, WT shows an average daily correlation of 0.53, and its highest performance is observed again on T03 (0.65). This evidences that the closest performance between WT and the *F* methodology occurs where local dynamic effects are remarkable. MAE results obtained by FE are found to be in the same order of accuracy than those obtained by [47] also for T01, where it is performed a $MAE_{|V|}$ and a MAE_V of 1.37 (a 6% higher than FE) and 3.73 m/s (a 7% lower than FE) respectively.

Table 1
Comparison of test results obtained by the Weather Types (WT), *F* field Simple Computing (FG) and *F* field Evolutionary Computing (FE) algorithms for the six daily wind speed data sets considered. Weighted averages within each entire classification (26 classes) for both $MAE_{|V|}$ and MAE_V are shown. For comparing purposes, values for W-NCAR are also depicted. In addition, the no-classification (1 class, NC) are also detailed.

Method	T01	T02	T03	T04	T05	T06	Mean
$MAE_{\|V\|}$ (m/s)							
NC	2.17	1.79	2.93	2.13	1.87	2.75	2.27
W-NCAR	2.32	2.22	2.57	2.50	2.19	2.04	2.31
WT	1.65	1.49	1.99	2.55	1.76	1.51	1.82
FG	1.32	1.22	1.83	2.00	1.36	1.12	1.48
FE	1.29	1.18	1.72	1.99	1.37	1.12	1.44
MAE_V (m/s)							
NC	6.39	6.37	7.96	6.58	6.55	6.37	6.70
W-NCAR	7.85	8.01	6.22	10.65	8.61	6.12	8.10
WT	4.65	4.05	3.97	5.09	4.58	4.26	4.48
FG	4.08	3.50	3.65	4.46	4.10	3.85	3.97
FE	3.99	3.33	3.40	4.37	3.9	3.41	3.77

Table 2
Test results for the Pearson Correlations between observed wind speed and values obtained by WT, FG and FE algorithms for the six towers daily wind speed data sets considered. Values for W-NCAR are also depicted. In addition, a no-class (NC) and a direct wind approach (DV) are also shown. The last column stands for the weighted averages for all six towers.

Method	T01	T02	T03	T04	T05	T06	Mean
Pearson correl. coeff.							
NC	0	0	0	0	0	0	0
W-NCAR	0.67	0.70	0.55	0.63	0.77	0.70	0.67
WT	0.61	0.57	0.65	0.39	0.50	0.52	0.53
FG	0.77	0.73	0.71	0.68	0.72	0.77	0.73
FE	0.78	0.74	0.74	0.68	0.71	0.77	0.73

A monthly timescale wind speed analysis has been carried out by averaging wind speed values according to their natural months. As it was expected, the uncertainty decreased in all extents with respect to the daily resolution. In overall for all towers, the average monthly Pearson correlation between observed and estimated wind speed is 0.81 and the $MAE_{|V|}$ scores 0.54 m/s, while the maximum performance is reached at T06 (with an $r = 0.91$) and in T02 (a $MAE_{|V|}$ of 0.32 m/s), both by means of FG method. Very similar results are obtained for FE, while WT showed an average results for all towers of $r = 0.55$ and $MAE_{|V|} = 0.81$ m/s, moving away from the *F* based methods [40], which computed a monthly average estimation from neural networking, performed on a similar level.

Fig. 7 shows two examples on the type of adjustment at daily and monthly scales of a test reconstruction applied on T06 through WT, FG and FE. Fig. 7a shows it along a sample time range of 100 days within τ_s period, represented and compared to the real wind speed test signal. The illustration confirms the higher reconstruction performances using FG and FE, with an overall better track than WT with respect to the observed wind. In turn, Fig. 7b shows the monthly wind speed test reconstruction this time applied throughout the whole considered instrumental time range ($\tau_s + \tau_r$). There, FG performed best, with a $MAE_{|V|}$ decreasing until 0.34 m/s.

Wind roses test reconstructions have been performed by FG and FE for the six considered locations, for τ_s period (2006–2009). The data cross-validation procedure to build the wind rose consisted on a similar procedure as for wind speed. First of all the wind rose for each obtained class (frequency and speed) is calculated in the training τ_s period. Then, the weighted average of these roses is implemented depending on the frequency of each class within the considered time range (τ_s). The obtained values are displayed in Fig. 8, where 16 sector wind rose reconstructions are compared to the real data in τ_s period. Results show a consistent similarity between real data and the reconstructed series both in wind intensity (color) and wind frequency (shape) for each sector of the wind rose for both FG and FE methods.

3.2. 1871–2009 Wind reconstruction

An immediate application for the developed classification methodology is the reconstruction of wind in time periods where

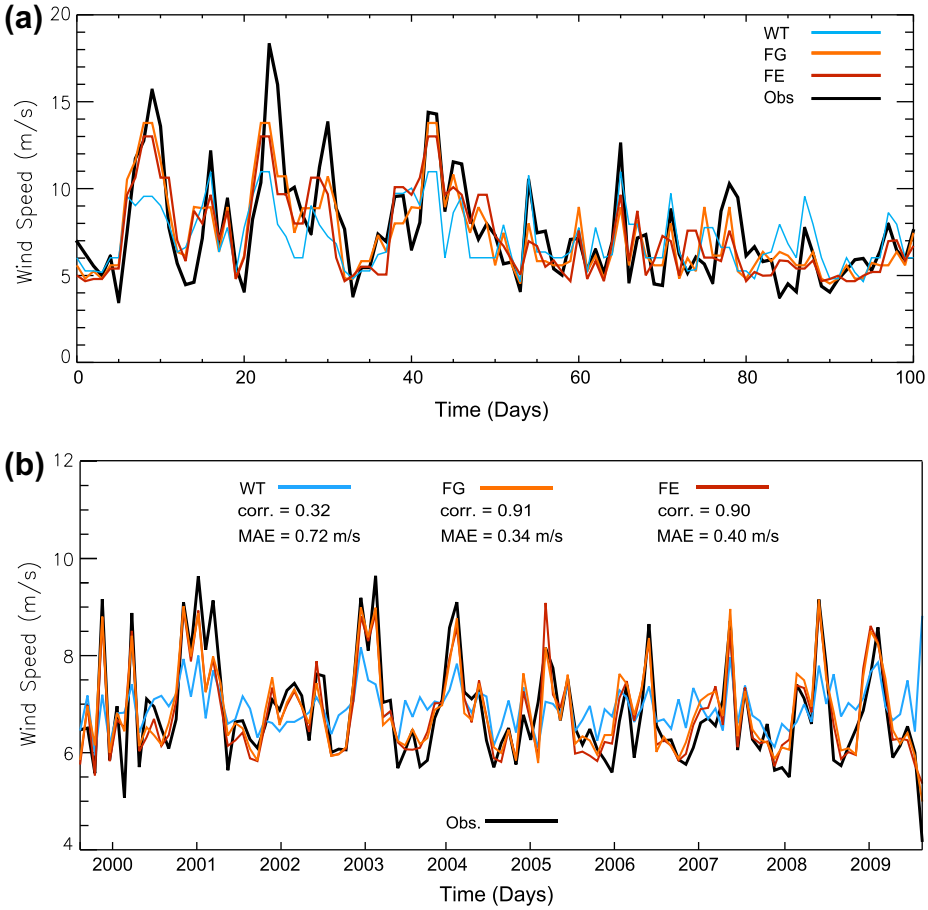

Fig. 7. 100 days sample (*a*) of the daily wind speed reconstruction at the τ_s period performed by WT, FG and FE methods for T06 compared to the observed wind speed signal (from February 9 to May 20, 2006) and monthly (*b*) observed and estimated wind speed for the period 1999–2009 by the same methods and tower.

no observations are available. In order to analyze the wind conditions in the last century in central Iberia, a daily mean wind reconstruction for the period 1871–2009 has been carried out. Since FG and FE present very similar MAEs, only FG method has been employed. The reason to use FG lies on the clear parametrization of the dynamic class characterization obtained through it. To perform the reconstruction, SLP data from 20CRV2 were employed to obtain a daily classification for the period 1871–1947, as NCEP/NCAR data was utilized in the same way to obtain those classes within the 1948–2009 period. This split into two different data inputs has been already performed by [61], arguing that NCEP/NCAR data offered a slightly higher reliability for the period where it is available. After the 140 year classification was obtained, it was again related to the observational period wind features, so that every day could be wind-characterized. Since few local effects are perceived compared to the other towers' locations, T01 can be considered an appropriate representative of the region, so that the experiment has been performed on that location.

3.2.1. Wind speed reconstruction

Throughout the daily wind speed characterization, a daily reconstruction was performed for the 1871–2009 period (Fig. 9). Additionally, the 1-year and the 11-year moving averages with their corresponding uncertainties were computed for a clearer comprehension. First of all, a spectral analysis was performed to the annual average wind speed signal through a Fourier transform, in order to detect possible low frequency multi-decadal wind variations. It revealed an statistically significant ($p < 0.975$) variability cycle within the 23 year frequency band, obtained also by considering either the 20CRV2 and the NCAR periods separately. Regarding the wind resource variability, the annual average wind speed range reaches 1.86 m/s, with a maximum annual wind speed of 7.95 m/s and a minimum value of 6.09 m/s. These annual values imply a relative inter-annual variability higher than 30%. By considering wind power empirically equivalent to the square of wind speed [62], this variability entails a wind power output variability higher than 70%. Finally, a linear fit performed over the whole considered period shows a slight but statistically significant negative linear trend of 0.1 m/s every 100 years, which implies, with a wind speed average of 6.91 m/s, a wind power decrease of about a 3% every 100 years. These results serve to set the wind farm performance reached in the last years into a multi-decadal context, so that the last period of observations (from 2005) appears located within a minimum phase of the multi-decadal variability, which could have led to underestimate the long-term wind speed resource in T01. Further work is under way in order to determine the possible causes that explain these wind speed changes.

Wind speed variability has also been analyzed by assessing its changes in frequency distribution, a useful tool, very extended in wind industry. Since it permits a clear comprehension of the wind speed behavior at different speed regimes it is employed for assessing the wind speed resource at a certain location. In this work we have computed the daily wind speed frequency distribution at decadal and annual ranges with a 1 m/s resolution for the whole considered period (1871–2009). In order to provide a bigger robustness to the reconstructed distributions, in this case the daily reconstructed wind has been characterized by a complete wind speed distribution (instead of just the *V* average), computed from

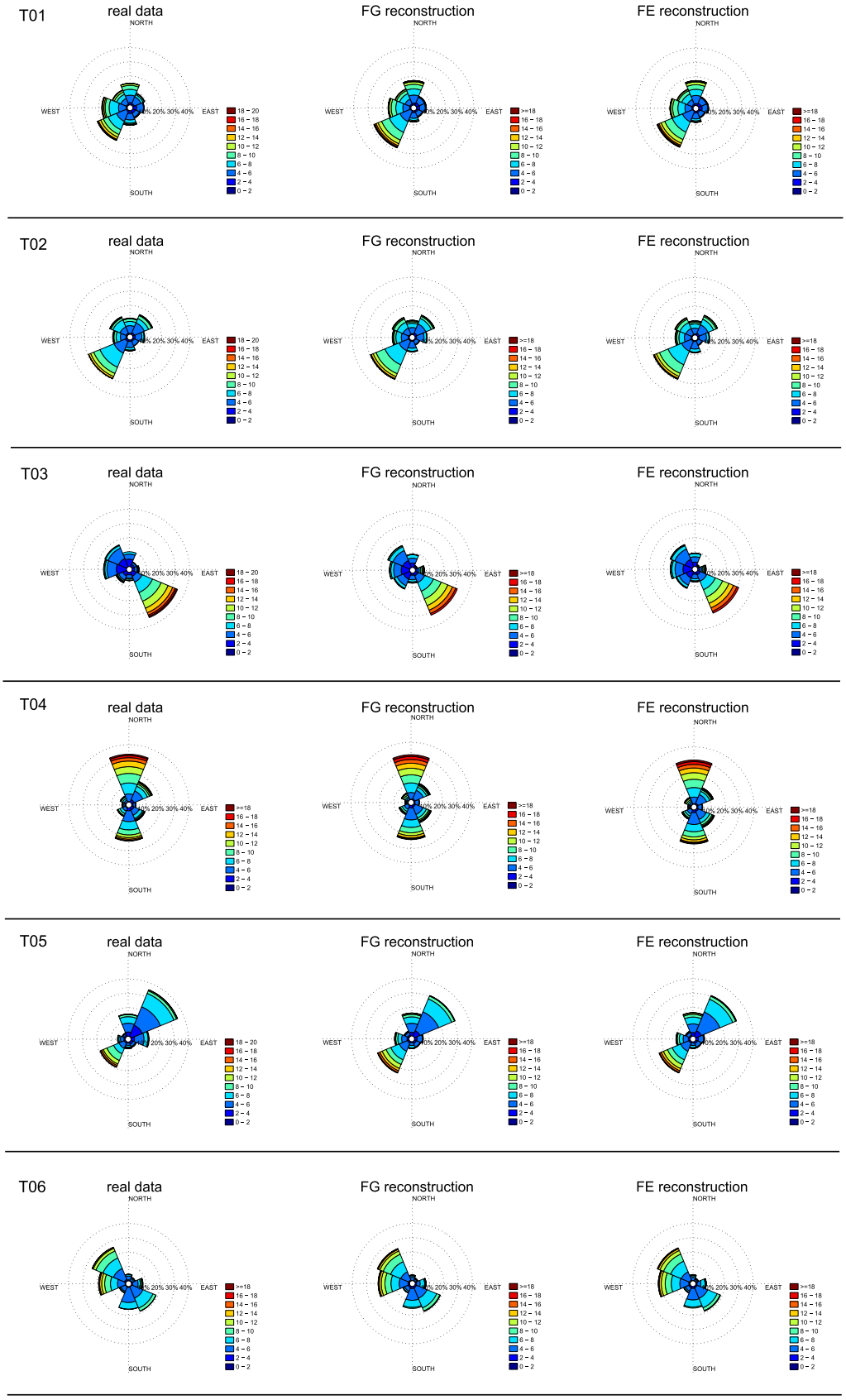

Fig. 8. Wind roses obtained by real data assimilation (left), FG reconstruction (center) and FE reconstruction (right) for the six towers available data for the test period.

the whole set of intra class-elements within the observations period. Special emphasis has been granted to wind speed frequency distribution at those decades with the highest (1926–1935) and the lowest (1979–1988) average wind speeds (10). Additionally,

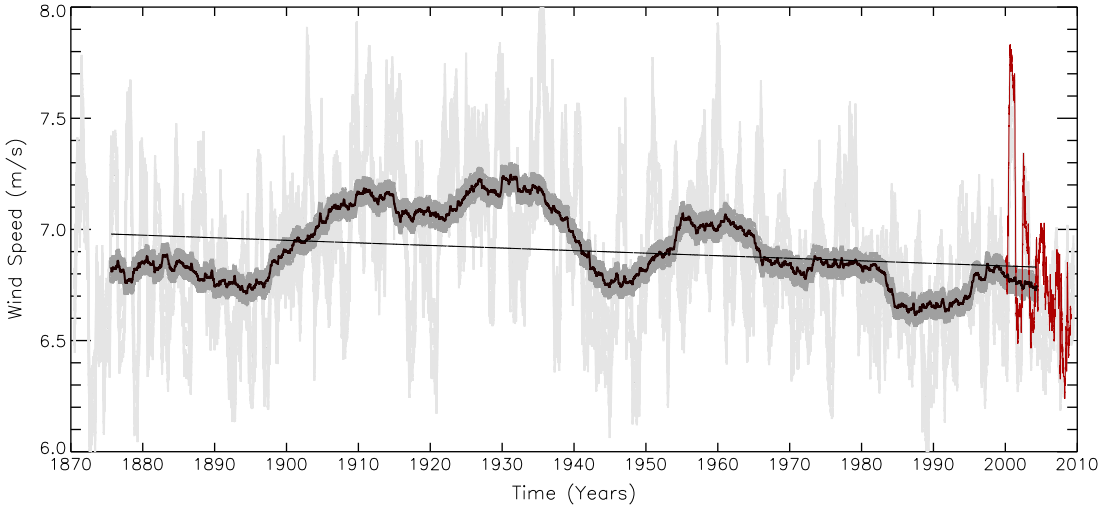

Fig. 9. 1871–2009 wind speed reconstruction for T01 performed with FG method. Light gray represents the uncertainty for the 1-yr moving average series, as dark and medium gray represent the 11-yr moving average series and its uncertainty, respectively. The linear regression for the whole period (straight line) and the observations' 1-yr moving average (red) are also depicted. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

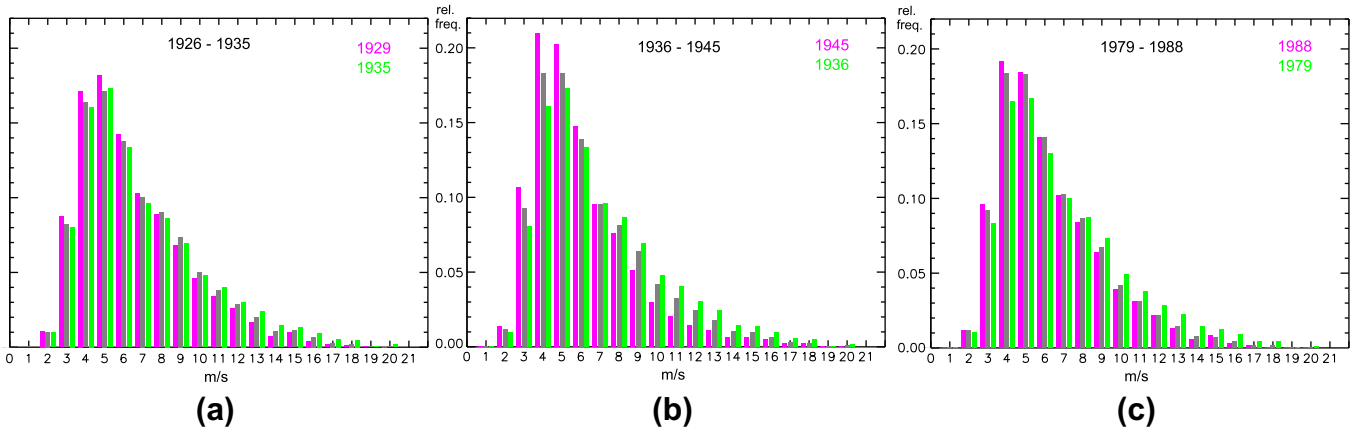

Fig. 10. Reconstructed wind speed frequency distribution for the decadal (gray bars) periods 1926–1935 (a), 1936–1945 (b) and 1979–1988 (c). Green bars represent the frequency distribution of the year with the highest wind speed average of the decade (1935, 1936 and 1978 resp.), as the magenta ones correspond to that with the lowest wind speed (1929, 1945 and 1988 resp.). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

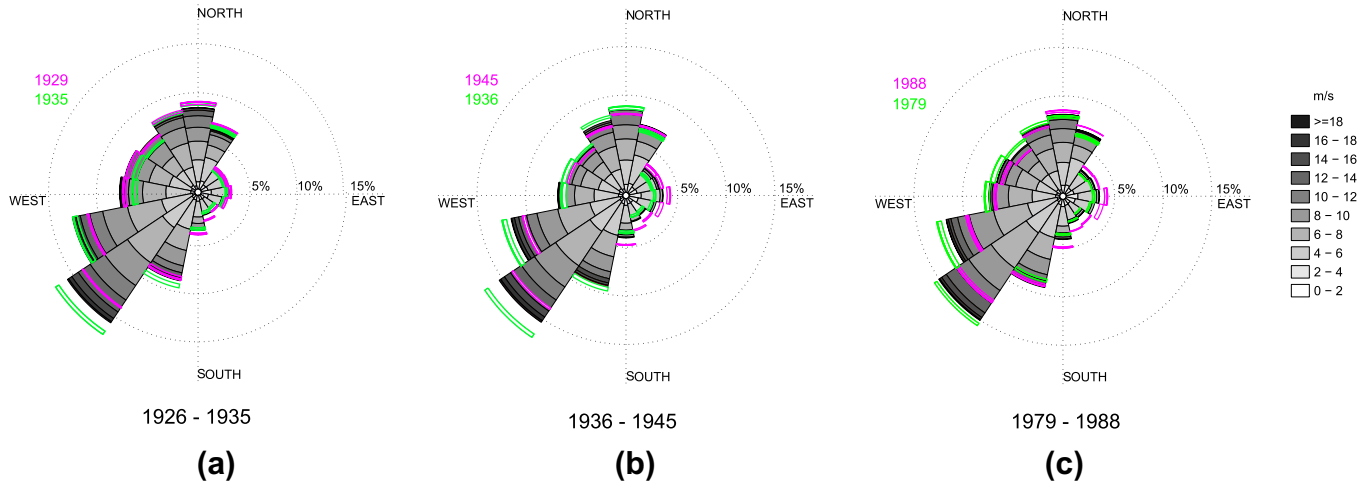

Fig. 11. Same as Fig. 10 but for the wind rose reconstructions. Again, green (magenta) profiles stand for the relative frequency per sector for the year with the highest (lowest) average wind speed of the decade. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the decade with highest annual wind speed changes (1936–1945) has been considered. For these three decades, the years with the highest and the lowest wind averages have been taken into account with the purpose of detecting possible significant changes. Fig. 10 shows that decades and years with the highest annual wind speed averages have longer right tails (i.e. higher frequencies at higher wind speed ranges) as well as lower frequencies at low speed ranges. Thus, the decadal distributions regarding the periods 1926–1935 and 1979–1988 where observed to hold statistically significant ($p < 0.95$) differences in a Chi-square test of homogeneity. Regarding the difference among different speed ranges, annual distributions with low winds (1945) versus high winds (1936) showed some remarkable differences. Thus, 1945 showed more than 30% higher frequency values at the first quartile (3–4 m/s speed range) of the distribution with respect to 1936. In turn, 1936 presented frequency values 35% higher than 1945 at the third quartile (8–9 m/s) of the distribution.

3.2.2. Wind rose reconstruction

The same inputs and method than those employed for the wind speed reconstruction where employed to perform the 16-sector wind rose reconstruction of the decades of the period 1871–2009, year per year. Additionally, single year wind roses where computed for those years with a particularly extreme (high and low) average wind speed. Results show that wind roses can vary its morphology at a decadal scale, these differences being remarkably bigger at annual scales. Specifically, the SW sector has been observed to be as the most prevailing sector throughout the whole period. Regarding the sector variability, it has been observed that when the annual average wind speed increases above the average, the SW sector frequency is restrengthened, while N and E directional quadrants become weakened. Thus, a Pearson correlation of 0.50 is obtained between SW annual wind average frequency and speed. These results can be observed also in decadal averages. In Fig. 11 the wind roses are depicted again for those decades with the highest (1926–1935), the lowest (1979–1988) and the most varying decade (1936–1945). There, the frequency of the strongest sector (SW) is 30% higher during 1936 than in 1945. The maximum annual differences within the SW sector reached 50% between annual frequencies of years 1962 (min.) and 1968 (max.).

4. Summary and conclusions

In this work a methodology based on geostrophic flow indices (strictly derived from Sea Level Pressure (SLP) reanalysis data) has been designed to develop daily surface wind classifications capable of characterizing wind in a given location as well as statistically downscaling wind speed and wind rose for six wind farm meteorological towers in Spain. To achieve this, an evolutionary algorithm (FE) and a greedy-based technique (FG) have been implemented. Both algorithms worked by minimizing the average class-dispersion of the Euclidean distance at the geostrophic speed space. The obtained patterns were then assessed through several measures of goodness, and validated by associating them with their corresponding observed wind daily series. This allowed the methodology to estimate wind speed, providing a measure for its uncertainty. This downscaling enabled the possibility of performing long term reconstructions throughout time ranges where no wind observations were available. Performances obtained by these approaches have been compared with that of a circulation weather types technique first introduced in [15] (WT), as well as others as the NCEP/NCAR directly retrieved u and v wind datasets.

After applying the classification methodology to six different locations in Spain, an analysis on the synoptic conditions ruling the obtained classes showed a statistically significant consistency on the average pressure field dispersion per grid point for the classes obtained within both FG and FE, being even bigger than those obtained by WT. This results evidence the ability of the developed methods to define synoptic circulations associated to them. Regarding an analysis on the wind clustering ability of the proposed methods, the set of elements of a given class defined only by the geostrophic wind generally maintained a similar observed surface wind speed behavior. This consistency within the obtained wind patterns occurred even when the location of study was strongly influenced by local circulation effects.

A validation test of the daily wind speed estimation capability of the introduced methodology was performed in order to measure its uncertainty through two measures of goodness, its bias and its Pearson correlation with respect to the observations considered. Regarding the first one, a wind speed accuracy 25% higher than that obtained through the WT method was reached. It implied an average bias (in terms of *Mean Absolute Error*) for the six considered locations of 1.44 m/s (FE), reaching a value of 1.12 m/s in some towers, outperforming other statistic models [38,42,43] In turn, the accuracy for the vectorial distance (*Euclidean radius*) in the speeds' space showed an average value a 34% higher than WT. Meanwhile, the Pearson correlation coefficient obtained for both new algorithms was 0.73 with respect to daily observed wind speed, a correlation explaining 25% more variance (r^2) than WT. When compared to the surface wind speed directly retrieved from the u and v variables at the NCEP/NCAR reanalysis dataset, our methodology largely reduced the speed errors, and the Pearson correlation improved a 8% the explained variance. In a monthly scale, the Pearson correlation obtained by our methodology showed an average value of 0.81, reaching 0.91 in some cases, while the average wind speed error was 0.54 m/s, with a maximum accuracy of 0.32 m/s among the considered towers. This performance is comparable to that obtained by [40]. By considering these results and the different features of the two developed methods, it can be derived that FG results more appropriated to obtain an analytic and dynamic-consistent result, while FE provides a non-deterministic but slightly better performance.

Through these new wind classifying methods, a wind reconstruction of the las 140 years at central Iberia has been performed. With a daily wind speed uncertainty of 1.29 m/s, it reflected a statistically significant variability cycle with a frequency of 23 years, as well as a slight but statistically significant negative linear trend. In this context, the obtained long-term wind speed reconstruction revealed a minimum of production in recent years. In turn, the performed reconstructions of decadal and annual wind speed frequency distributions revealed a statistically significant distribution difference between high and low wind speed decades, as well as annual frequency variations of 30% for a wind speed range of 3–4 m/s (1st quartile) and 35% for a wind speed range of 8–9 m/s (3rd quartile). Decadal and annual 16-sector wind rose reconstructions where also performed throughout the considered 140 year period. Results show some differences within the wind rose morphology when considering different wind speed performances within annual and decadal scales. In high wind speed periods the SW sector is observed to strengthen, while N and E sectors become weakened. Indeed, SW sector frequencies between different years have been observed to vary up to 30%.

Summarizing, our methodology allows to perform a realistic approach on wind clustering, characterization and statistic downscaling with daily resolution, by exclusively employing public domain SLP reanalysis data and with a very low computational cost. Since the geostrophic wind at synoptic scale has been parameterized through the designed algorithms, it constitutes, additionally, a significant instrument for analyzing the existing connections between surface wind and general circulation climatology. Thus, in this paper we have developed a downscaling experiment that, although

statistic, is based on the circulation atmospheric conditions that rule the climate. In this context, we have shown that the proposed methodology can be implemented for the reconstruction of centennial wind series at the wind farm locations, where long-term wind measures are rarely available. To the best of our knowledge, such centennial wind reconstructions, with a daily resolution and without the need of the implementation of a numerical model, have been computed here for the first time. These wind reconstructions allowed to analyze changes in the wind speed and direction low frequency variability, as well as to detect periods with significant different wind features. For all this, they allow to place the present wind farm performance in a climatic perspective, enabling to assess the current state of the wind power production with respect to its historical signal. Therefore, this new outlook provides an important contribution for the wind multi-decadal analysis on the prospects of wind performance variations in upcoming years.

Acknowledgements

This work has been partially supported by the Spanish Ministry of Education through Project ECO2010-22065-C03-02, and by Iberdrola Renovables Energía S.A. Special thanks are given to Prof. Emiliano Hernández Martín from Universidad Complutense de Madrid and to Prof. Ricardo Trigo from Universidade de Lisboa.

Appendix A

F and *Z* are related to the spatial SLP field through the theoretical statements of the geostrophic flow *G*. Thus, *F* can be determined through its relationship with the zonal ($G_{u\lambda}$) and the meridian ($G_{v\lambda}$) components of *G* in the following terms:

$$G_{u\lambda} = -\frac{1}{\rho f}\frac{\partial p}{\partial y} \approx -\frac{1}{\rho f_\lambda R \Delta\lambda} WF \tag{5}$$

$$G_{v\lambda} = \frac{1}{\rho f}\frac{\partial p}{\partial x} \approx \frac{1}{\rho f_\lambda R\, cos(\lambda)\Delta\varphi} SF \tag{6}$$

where *WF* and *SF* stand for *F* zonal and meridian components respectively, λ and φ are the latitude and the longitude of the measured grid point and $\Delta\lambda$ and $\Delta\varphi$, their differences. *f* stands for the Coriolis parameter, given by the expression $f_\lambda = 2\Omega sin(\lambda)$ (Ω represents the angular speed of the Earth), while *R* is the Earth radius and ρ is the density of the air. This shows how F can be determined by differences of pressure only.

In turn, the relationship with *Z* and G_W and hence with SLP field, as seen in (5) and (6) come out of these expressions:

$$\frac{\partial G_{u\lambda}}{\partial y} \approx \frac{1}{R}\frac{\Delta u_g}{\Delta\lambda} = -\frac{1}{R^2 \rho f_\lambda (\Delta\varphi)^2} WZ \tag{7}$$

$$\frac{\partial G_{v\lambda}}{\partial x} \approx \frac{1}{2Rcos(\lambda_\lambda)}\frac{\Delta v_g}{\Delta\varphi} = \frac{1}{R^2 \rho f_\lambda (\Delta\varphi)} SZ \tag{8}$$

where *WZ* and and *SZ* represent the *Z* components produced respectively by the West–East and the the South–North pressure differences.

Appendix B

The following expressions show how *F* and *Z*, calculated for a given grid point *J* at a latitude $\lambda = 40°$, can be set exclusively using the 16 grid point pressure values shown in Fig. 12, through the following expressions:

$$WF = \frac{1}{2}(p_{12} + p_{13}) - \frac{1}{2}(p_4 + p_5) \tag{9}$$

$$SF = \frac{1}{cos(40°)}\left[\frac{1}{4}(p_5 + 2p_9 + p_{13}) - \frac{1}{4}(p_4 + 2p_8 + p_{12})\right] \tag{10}$$

so

$$F = \sqrt{(WF)^2 + (SF)^2} \tag{11}$$

In turn, for *Z* these relationships are:

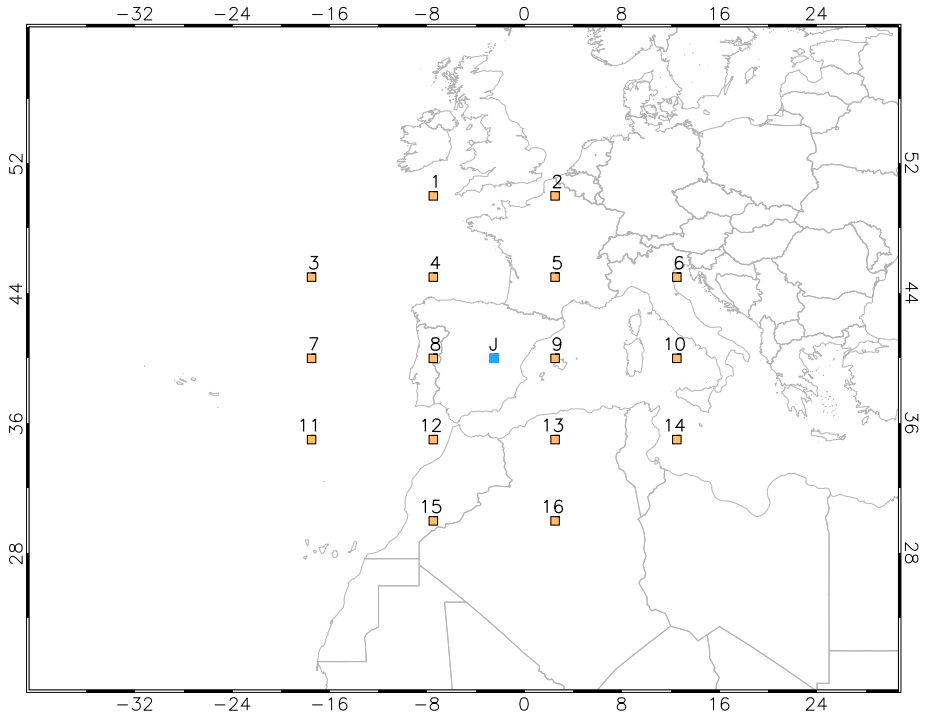

Fig. 12. Example of the considered SLP grid structure employed in the definition of the *F*, *Z* and α indexes at the introduced algorithms for a generic point *J*.

$$WZ = \frac{sin(40^\circ)}{sin(35^\circ)}\left[\frac{1}{2}(p_{15}+p_{16}) - \frac{1}{2}(p_8+p_9)\right] - \frac{sin(40^\circ)}{sin(35^\circ)}\left[\frac{1}{2}(p_8+p_9) - \frac{1}{2}(p_1+p_2)\right] \quad (12)$$

$$SZ = \frac{1}{2cos^2(\lambda_{40^\circ})}\left[\frac{1}{4}(p_6+2p_{10}+p_{14}) - \frac{1}{4}(p_5+2p_9+p_{13}) - \frac{1}{4}(p_4+2p_8+p_{12}) + \frac{1}{4}(p_3+2p_7+p_{11})\right] \quad (13)$$

Hence, as Z is a scalar value,

$$Z = WZ + SZ \quad (14)$$

Finally, a value for the angle of the *F* vector direction is obtained as following:

$$\alpha = atan\frac{SF}{WF} \quad (15)$$

References

[1] Conil S, Hall A. Local regimes of atmospheric variability: a case study of Southern California. J Climate 2006;19:4308–25.
[2] Palutikof JP, Kelly PM, Davis TD. Windspeed and climatic change. J Wind Eng 1986;10(4).
[3] Diaf S, Notton G, Belhamel M, Haddadi M, Louche A. Design and techno-economical optimization for hybrid PV/wind system under various meteorological conditions. Appl Energy 2008;85(10):968–87.
[4] Oztopal A, Sahin AD, Akgun N, Sen Z. On the regional wind energy potential of Turkey. Energy 2000;25(2):189–200.
[5] Palutikof JP, Brabson BB, Lister DH, Adcock ST. A review of methods to calculate extreme wind speeds. Meteorol Appl 1999;6(2):119–32.
[6] Traveria M, Escribano A, Palomo P. Statistical wind forecast for Reus airport. Meteorol Appl 2010;17(4):485–95.
[7] Green MC, Flocchini RG, Myrup LO. Relationship of the extinction coefficient distribution to wind field patterns in southern California. Atmos Environ 1992;26:827–40.
[8] Darby LS. Cluster Analysis of Surface Winds in Houston, Texas, and the Impact of Wind Patterns on Ozone. J Appl Meteorol 2005;44:1788–806.
[9] Knudsen T, Bak T, Soltani M. Prediction models for wind speed at turbine locations in a wind farm. Wind Energy 2011;14(7):877–94. SI.
[10] Xydis G, Koroneos C, Loizidou M. Exergy analysis in a wind speed prognostic model as a wind farm sitting selection tool: A case study in Southern Greece. Appl Energy 2009;86(11):2411–20.
[11] Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, et al. The NCEP/NCAR reanalysis project. Bull Am Meteorol Soc 1996;77:437–71.
[12] Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan R, et al. The Twentieth Century Reanalysis Project. Quart J Roy Meteorol Soc 2011;137(654):1–28.
[13] Lamb HH. British Isles weather types and a register of the daily sequence of circulation patterns 1861-1971. Geophys Memoirs 1972(116).
[14] I. Font, Climatologia de España y Portugal, Ediciones Universidad de Salamanca second edition, 422 pp., 2000.
[15] A.F. Jenkinson, F.P. Collison, ‘‘An initial climatology of gales over North Sea’’, Synoptic Climatology Branch Memorandum no. 62. Meteorological Office: Bracknell, 1977.
[16] Jones PD, Hulme M, Briffa KR. A comparison of Lamb circulation types with an objective classification scheme. Int J Climatol 1993(13):655–63.
[17] Trigo RM, DaCamara CC. Circulation weather types and their influence on the precipitation regime in Portugal. Int J Climatol 2000;20:1559–81.
[18] Spellman G. The application of an objective weather-typing system to the Iberian peninsula. Weather 2000;55:375–85.
[19] Lund IA. Map-Pattern Classification by Statistical Methods. J Appl Meteorol 1963;2:56–65.
[20] W. Kirchhofer, ‘‘Classification of European 500 mb patterns’’, Swiss Meteorological Institute, Technical Report 43, 1974.
[21] Von Storch H, Zorita E, Cubasch y U. Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. J Climate 1993(6):1161–71.
[22] Goodess CM, Palutikof JP. Developement of daily rainfall scenarios for southeast Spain using a circuation-type approach to downscaling. Int J Climatol 1998;18:1051–83.
[23] Paredes D, Trigo RM, García-Herrera R, Trigo IF. Understanding precipitation changes in Iberia in early spring: weather typing and storm tracking approaches. J Hydrometeorol 2006;7:101–13.
[24] Post P, Truija V, Tuulik J. Circulation weather types and their influence on temperature and precipitation in Estonia. Boreal Environ Res 2002;7:281–9.
[25] Jones PD, Lister DH. The influence of the circulation on surface temperature and precipitation patterns over Europe. Climate Past 2009;5(2):259–67.
[26] Romero R, Summer G, Ramis C, Genoves A. A classification of the atmospheric circulation patterns producing significant daily rainfall in the spanish mediterranean area. Int J Climatol 1999;19:765–85.
[27] Trigo RM, Palutikof JP. Simulation of daily temperaturas for climate changescenarios over Portugal: a Neural Network Model Aroach. Climate Res 1999:45–9.
[28] Trigo RM, Palutikof JM. Precipitation scenarios over Iberia: a comparison between direct GCM output and different downscaling techniques. J Climate 2001;14:4422–46.
[29] Palutikof JP, Kelly PM, Davies TD, Halliday y JA. Impacts of spatial and temporal wind speed variability on wind energy output. J Appl Meteorol 1987;26:1124–33.
[30] Green M, Flocchini RG, Myrup y LO. Use of Temporal Principal Components Analysis to Determine Seasonal Periods. J Appl Meteorol 1993;32:986–95.
[31] Weber RO, Kaufmann P. Automated classification scheme for wind fields. J Appl Meteorol 1995;34(5):1133–41.
[32] Kaufmann P, Whiteman CD. Cluster-Analysis Classification of Wintertime Wind Patterns in the Grand Canyon Region. J Appl Meteorol 1999:1131–47.
[33] Gomez-Muñoz VM, Porta-Gandara MA. Local wind patterns for modeling renewable energy systems by means of cluster analysis techniques. Renew Energy 2002(2):171–82.
[34] Burlando M. The synoptic-scale surface wind climate regimes of the Mediterranean Sea according to the cluster analysis of ERA-40 wind fields. Theor Appl Climatol 2009;96:6983.
[35] Jimenez PA, Gonzalez-Rouco JF, Montavez JP, Garcia-Bustamante E, Navarro y J. Climatology of wind patterns in the northeast of the Iberian Peninsula. Int J Climatol 2008(29):501–25.
[36] Garcia-Bustamante E, Gonzalez-Rouco JF, Navarro J, Xoplaki E, Jimenez PA, Montavez JP. North Atlantic atmospheric circulation and surface wind in the Northeast of the Iberian Peninsula: uncertainty and long term downscaled variability. Climate Dynam 2010;38:141–60.
[37] Palutikof JP, Guoa X, Hallidayb JA. Climate variability and the UK wind resource. J Wind Eng Indus Aerodynam 1992;39(1–3):243–9.
[38] Torres JL, García A, De Blas M, De Francisco A. Forecast of hourly average wind speed with ARMA models in Navarre (Spain). Sol Energy 2005;79:65–77.
[39] Morales JM, Minguez R, Conejo AJ. A methodology to generate statistically dependent wind speed scenarios. Appl Energy 2010;87(3):843–55.
[40] Fadare DA. The application of artificial neural networks to mapping of wind speed profile for energy application in Nigeria. Appl Energy 2010;87(3):934–42.
[41] Barbounis TG, Theocharis JB. Locally recurrent neural networks for long-term wind speed and power prediction. Neurocomputing 2006;69(4–6):466–96.
[42] Mohandes MA, Halawani TO, Rehman S, A Hussain A. Support vector machines for wind speed prediction. Renew Energy 2006;29:939–47.
[43] Bouzgou H, Benoudjit N. Multiple architecture system for wind speed prediction. Appl Energy 2011;88(7):2463–71.
[44] Li G, Shi J. Application of Bayesian model averaging in modeling long-term wind speed distributions. Renew Energy 2010;35(6):1192–202.
[45] Barbounis TG, Theocharis JB. A locally recurrent fuzzy neural network with application to the wind speed prediction using spatial correlation. Neurocomputing 2007;70(7–9):1525–42.
[46] Zadeh LA. Fuzzy, Logic, Neural Networks, and Soft-Computing. Commun ACM 1994;37:77–84.
[47] Carro-Calvo L, Salcedo-Sanz S, Kirchner-Bossi N, Portilla-Figueras A, Prieto L, Garcia-Herrera R, et al. Extraction of synoptic pressure patterns for long-term wind speed estimation in wind farms using evolutionary computing. Energy 2011;36:1571–81.
[48] Carro-Calvo L, Salcedo-Sanz S, Prieto L, Kirchner-Bossi N, Portilla-Figueras A, Jiménez-Fernández S. Wind speed reconstruction from synoptic pressure patterns using an evolutionary algorithm. Appl Energy 2012;89:347–54.
[49] Yao X, Liu Y, Lin G. Evolutionary programming made faster. IEEE Trans Evolution Comput 1999;3(2):82–102.
[50] W. Bednorz, ‘‘Greedy Algorithms’’, IN-TECH, 583 pp., ISBN-13: 978-953-7619-27-5, 2008.
[51] Goldberg DE. Genetic algorithms in search, optimization and machine learning. Reading, MA: Addison-Wesley; 1989.
[52] Bäck T, Schwefel HP. An overview of evolutionary algorithms for parameter optimization. Evolution Comput 1993(1):1–23.
[53] Fogel DB. An introduction to simulated evolution. IEEE Trans Neural Networks 1994:3–14.
[54] Eiben AE, Smith JE. Introduction to evolutionary computing. Springer-Verlag; 2003.
[55] Liao CC. Genetic k-means algorithm based RBF network for photovoltaic MPP prediction. Energy 2010;35(2):529–36.
[56] Jursa R, Rohrig K. Short-term wind power forecasting using evolutionary algorithms for the automated specification of artificial intelligence models. Int J Forecast 2008;24:694–709.

[57] Y. Liu and X. Yao, "Evolving neural networks for Hang Seng stock index forecast," In Proc. of the 2001 Congress on Evolutionary Computation CEC2001, pp. 256-260, 2001.
[58] Barnett TP, Preisendorfer RW. Origin and levels of monthly and seasonal forecast skill for United States air temperature determined by canonical correlation analysis. Monthly Weather Rev 1987;115:1825–50.
[59] E.L. Petersen, I. Troen, S. Frandsen and K. Hedegaard, "Windatlas for Denmark", RISO National Laboratory, R-248, 1981.
[60] Dorman CE, Beardsley RC, Limeburner R. Winds in the Strait of Gibraltar. Quart J Roy Meteorol Soc 1995;121(528):1903–21.
[61] P.D. Jones, C. Harpham and K.R. Briffa, "Lamb weather types derived from reanalysis products", International Journal of Climatology, DOI: 10.1002/joc.3498.
[62] Anderson PM, Bose A. Stability simulation of wind turbine systems. IEEE Trans Power Appl Syst 1983;102:3791–5.