Energy 36 (2011) 1571–1581

Contents lists available at ScienceDirect

Energy

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

Extraction of synoptic pressure patterns for long-term wind speed estimation in wind farms using evolutionary computing

L. Carro-Calvo [a], S. Salcedo-Sanz [a,*], N. Kirchner-Bossi [b], A. Portilla-Figueras [a], L. Prieto [c], R. Garcia-Herrera [b], E. Hernández-Martín [b]

[a] Department of Signal Theory and Communications, Universidad de Alcalá, 28871 Alcalá de Henares, Madrid, Spain
[b] Department of Physics of the Earth, Astronomy and Astrophysics II, Universidad Complutense de Madrid, Spain
[c] Department of Energy Resource, Iberdrola Renovables, Spain

ARTICLE INFO

Article history:
Received 3 September 2010
Accepted 4 January 2011
Available online 4 February 2011

Keywords:
Pressure patterns extraction
Wind speed
Wind farms
Wind speed series reconstruction
Evolutionary algorithms

ABSTRACT

In this paper we present an evolutionary approach for the problem of discovering pressure patterns under a quality measure related to wind speed and direction. This clustering problem is specially interesting for companies involving in the management of wind farms, since it can be useful for analysis of results of the wind farm in a given period and also for long-term wind speed prediction. The proposed evolutionary algorithm is based on a specific encoding of the problem, which uses a dimensional reduction of the problem. With this special encoding, the required centroids are evolved together with some other parameters of the algorithm. We define a specific crossover operator and two different mutations in order to improve the evolutionary search of the proposed approach. In the experimental part of the paper, we test the performance of our approach in a real problem of pressure pattern extraction in the Iberian Peninsula, using a wind speed and direction series in a wind farm in the center of Spain. We compare the performance of the proposed evolutionary algorithm with that of an existing weather types (WT) purely meteorological approach, and we show that the proposed evolutionary approach is able to obtain better results than the WT approach.

1. Introduction

Long-term wind speed prediction and wind speed series reconstruction are two important problems currently faced by companies exploding wind farms. Although both problems seem quite different, actually they are usually solved using very similar techniques: in both cases a model for characterizing the wind speed is constructed based on previous real wind measures, and then, applied to future values of time in the case of long-term wind speed prediction, or to values in the past in order to reconstruct wind speed series. Different techniques have been used to obtain these wind speed models, such as statistical methods [1,2], neural networks [3–5], support vector machines [6] etc.

The majority of the existing techniques to construct long-term wind speed models are exclusively based on past wind speed data, and some of them include other atmospheric variables such as local temperature, radiation or pressure at the measuring point. Intuitively, this approach is correct, since almost all the models are developed for a specific geographic point (usually a wind farm), where several years of past data are available. The question that arises is whether we could obtain a reasonably accurate model for long-term wind speed prediction (or wind speed series reconstruction), based on synoptic information (pressure), instead on only local information. In fact, this idea has been successfully applied to rainfall or pollution prediction in the last few years [7–9]. In these papers, rainfall or pollution measures are explained depending on different pressure synoptic patterns. The objective of this paper is to do something similar with a measure of wind speed in a given point, i.e. obtaining the pressure patterns (pressure clusters) which better explain a wind speed measure in a wind farm. As has been mentioned before, this system would be really interesting for companies which manage wind farms, since it allows a good analysis of the wind farm production results, and can also be used to long-term production prediction.

Specifically, in this paper we state the problem as a clustering problem in a search space formed by a grid of synoptic pressure measures. The objective is to obtain N groups of synoptic pressure situations which produce the most similar wind vectors (wind speed and direction) in a given point (usually a wind farm project). The objective of the paper is twofold, since we propose the use an

* Corresponding author. Tel.: +34 91 885 6731; fax: +34 91 885 6699.
E-mail address: sancho.salcedo@uah.es (S. Salcedo-Sanz).

0360-5442/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.energy.2011.01.001

Evolutionary Algorithm (EA) to solve this problem. Thus, we show that a soft-computing technique such an EA can be successfully applied to this clustering problem. In the paper, we describe in detail the problem formulation, algorithm encoding, and the different operators that must be included in our approach to improve its efficiency in this problem. We compare the performance of the EA with that of a previous pure meteorological approach for the problem, and we show that the proposed EA obtains better results in terms of the objective functions defined to measure the quality of the solutions obtained.

The structure of the rest of the paper is the following: next section provides the necessary concepts about clustering problems and evolutionary computation techniques needed to follow the rest of the paper. Section 3 presents the problem formulation. Section 4 describes in detail the main points of the proposed EA: encoding, crossover operator, mutation operator and selection operator. Section 5 presents the experimental part of the paper, where we provide the main results obtained and a comparison with an existing algorithm, purely meteorological, described in a subsection of this part of the paper. Section 6 closes the paper giving some final conclusions.

2. Background

2.1. A brief review of clustering algorithms

Clustering is an important subgroup of unsupervised learning techniques consisting in grouping data objects into disjoint clusters [10–12]. The classification into clusters must be done in such a way that objects in the same cluster are similar with respect to a given measure, and different from the objects in the other clusters, with respect to the same measure. Clustering has been applied to a wide variety of problems in many different fields such as pattern recognition, bio-engineering, image quantization, renewable energy prediction [13–16], etc.

In Ref. [17], four major class of clustering algorithms are identified: first, algorithms based on the idea that neighbor data should share the same cluster. Classical clustering algorithms such as density-based approaches [18] belong to this first group. As pointed out in Ref. [17], these kind of algorithms are robust to detect clusters of any shape, but they fail to locate clusters when there is small spatial separation between clusters. The second set of clustering algorithms is formed by those approaches which consider intra-clusters variation (intra-clusters points or centroids) to form the final solution. This category of algorithms includes the well known K-means [17,19,20], and other approaches such as model-based clustering [21]. Following [17], the third category includes a simultaneous row–column clustering known as bi-clustering algorithms [22]. Finally, the four group of clustering algorithms includes approaches that optimize different characteristics of the data set. This group includes the multi-objective clustering algorithms [14,23] and also clustering ensembles approaches [24].

In the last few years, evolutionary computing algorithms (EAs) have been widely applied to clustering problems, due to their capacity to be applied to very different problems with very few changes, and also because these algorithms are able to manage constraints in an efficient way. Thus, EAs have been applied to improve the K-means approach, for example in [25], obtaining the genetic K-means algorithm, which is known to be more effective that the K-means in hard clustering problems. Similar approaches have been used in different applications, such as color quantization [15] or bio-engineering [17]. EAs have also been applied to other clustering problems [26,27], and also recently to bi-clustering problems [14,28]. There are different types of evolutionary approaches that have been studied in clustering problems, such as evolutionary programming [29], evolutionary algorithms [30] or particle swarm optimization [31].

In the last few years a lot of applications of clustering in atmospheric sciences have been tackled, in the majority of cases in climatology and meteorology applications, but also in energy-related problems. For example, the prediction of the maximum power point of photovoltaic systems is tackled in [38] as a clustering problem, solved by a genetic K-means algorithm. The classification and track of storms or weather systems can also be stated as a clustering approach [32]. Moreover, the climatological analysis about relations between winds, precipitation or even pollution and pressure patterns is a classical problem that can be solved as a clustering problem [7–9,33–35]. In the majority of these approaches, pure meteorological solutions are adopted, in which the equations of the atmospheric dynamics are used to obtain criteria for the classification of the systems.

2.2. An introduction to evolutionary-based algorithms

In this section we summarize the basic principles of Evolutionary Computation (EC). EC is a set of population-based, stochastic and iterative optimization techniques, based on the concepts of natural evolution [36]. EC approaches tackle difficult problems by evolving approximate solutions in a computer, following certain rules borrowed from Darwin's evolution theory. Usually, any algorithm based on EC is called an Evolutionary Algorithm (EA). EAs have been applied to many different optimization problems, in a huge range of applications, including energy-related problems [37–41].

Given an optimization problem, EAs typically start from an initial set, called *population*, of random (candidate) solutions (individuals). These solutions are evolved by the repeated application of a set of evolutionary operators mainly selection, crossover and mutation. Individuals are typically selected according to the quality of the solution they represent. To measure the quality of a solution, a fitness function is assigned to each individual of the population. Hence, the better the fitness of an individual, the more possibilities the individual has of being selected for reproduction and the more parts of its genetic material will be passed on to the next generations. This is the principle of any selection mechanism incorporated to an EA. The selected individuals are reproduced by means of crossover and mutation operators. In simple terms crossover exchanges some genetic material between two or more individuals, while mutation changes a small part of the genetic material of an individual to a new random value.

By applying these operators in a loop fashion, as EA explores the space of possible solutions of an optimization problem. EAs have been shown to be efficient in searching in huge spaces.

3. Problem formulation

The problem in this paper may be summarized as follows: having a given synoptic-scale pressure situation, can we approximate the wind speed that we will have in a given point (wind farm)? Is there a direct relationship between the wind speed in the point and the synoptic-scale pressure situation (and can we extract such a relationship or pattern)? These problems are of major importance for all the companies involved in the management of wind farms. The first question is related to the prediction in the long-term mean wind speed and wind rose. The second question is related to the analysis of production of wind farms. Both problems are in the core of this research, and also the problem formulation we are interested in, depends much on them. We have structured the problem formulation specifically for the case of a given wind farm in Spain, with the corresponding synoptic pressure measures over the Iberian peninsula, however, the problem can be easily extended to any other region or peculiarity.

Let d_t, $t = 1 \ldots, T$, be a series of daily wind speed real vector (module and direction), measured in a given point (a wind farm in this case), for a given period of time T. Let P_t, $t = 1 \ldots, T$, be a series of daily synoptic-scale pressure measures in a grid. In our case, each component of P_t is a matrix of 14×13 surface pressure values (182 values), measured in a grid surrounding the Iberian Peninsula (Fig. 1). The problem of pressure pattern extraction consists of forming a set of N clusters (centroids) in the space of pressure (space P_t), in such a way that the dispersion of the associated values of d_t in each cluster is minimized, i.e., in such a way that the following total measure is minimized:

$$f_1(x) = \frac{1}{T}\sum_{i=1}^{N}\sum_{t \in \gamma_i}\left|d_t - d^i\right| \tag{1}$$

where x is a vector representing a given synoptic pattern assignment of length T (we consider a series of T pressure patterns to be assigned, in N clusters or centroids), γ_i stands for the set of days belonging to a given class i, and di stands for the mean value of the wind speed within class i.

We can also consider a second objective function based on the difference of modules between each wind vector and the mean value of each class:

$$f_2(x) = \frac{1}{T}\sum_{i=1}^{N}\sum_{t \in \gamma_i}\left|\left|d_t\right| - \left|d^i\right|\right| \tag{2}$$

Note that f_1 is a measure of Mean Absolute Error (MAE), and tries to explain both wind speed and wind direction. Note that using f_1 we can reconstruct the average wind rose from the pressure maps. On the other hand, f_2 is only focused on the wind speed module, without taking into account the wind direction. This point is sometimes important, since wind farms are designed to be optimum for a medium wind coming from a giving direction. In addition, this study with f_2 allows a reconstruction of the historic wind variability in a given zone, since there exist pressure maps since about 1950 which can be used.

Note that this is a problem of clustering in the space of matrices P_t, with a function of evaluation in the space of wind speed vectors d_t. These separate working spaces make the processing of clustering computation and evaluation difficult. Also, note that the high dimension of the pressure space P_t is an extra difficulty. We have

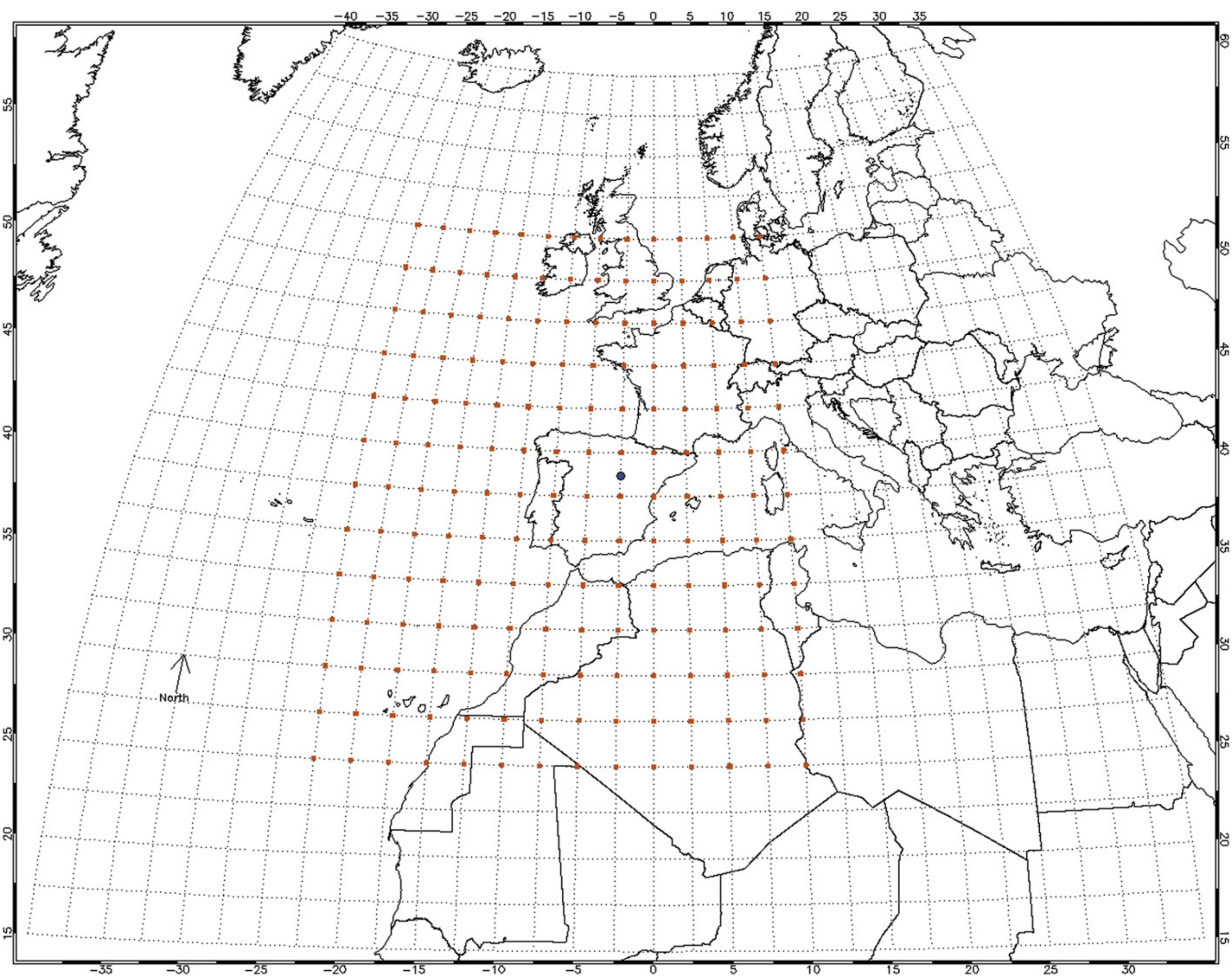

Fig. 1. Pressure measurements grid and location of associated wind speed measurements (in Lambert projection).

Initial couple:

8 15 107 146 3 27 99 17 | 1.173 0.742 .....

16 114 33 2 20 7 100 76 | -1.001 2.543 ....

Template: 1 1 0 1 0 0 1 1

Couple after crossover:

16 114 107 2 3 27 100 76 | 1.173 0.742 .....

8 15 33 146 20 7 99 17 | -1.001 2.543 ....

Fig. 2. Crossover applied to the integer part of the individuals.

tackled the problem by means of an evolutionary-based systems, which will be described and fully analyzed in the next sections.

4. An evolutionary algorithm for discovering significative pressure patterns

In the following sections we describe the main particularities of the problem tackled in this paper, and how we can adapt an EA to look for pressure patterns in an efficient way. We also describe the specific operators implemented in this work to improve the search of the algorithm in this particular problem.

4.1. Problem encoding

The first important task to face this problem is to find a simplest way to encode matrices P_t. An intuitive and easy form is to reduce the number of points in the grid: instead of using the information of all the points in the grid, we can condense somehow the information by using differences of pressure between points in the grid. Of course, different number of differences can be used. In this case, we have encoded the information of the grid by using a set of four pressure differences Dp (eight points in the grid). This way we reduce the space of matrices P_t to a space of four dimensions (space of differences of pressure). Note that we do not fix the points of the grid involved in the calculations of the differences, but the algorithm must locate the optimal points which provide the best possible encoding of the synoptic situation P_t. Thus, the first part of the encoding in the proposed evolutionary algorithm is a set of eight integer numbers $\mathcal{P}_i \in [1, \ldots, 182], i = 1, \ldots, 8$, (number of points in the grid considered), representing four pressure differences, in the following way:

$$Dp_k = P_t(\mathcal{P}_{2k-1}) - P_t(\mathcal{P}_{2k}), k = 1, \ldots, 4. \quad (3)$$

Then, once we have obtained an efficient representation and encoding of matrices P_t (using the four differences Dp_k), we need to encode the different N clusters in the new space of differences of pressure. This can be easy done by encoding each centroid of the cluster in the space, as a four-dimensional vector of real values (one dimension representing each difference of pressure). Thus, each centroid will be represented by a string of 4 real numbers, and the complete set of centroids can be therefore encoded in a vector of $4N$ real numbers.

The final encoding of the problem in the proposed evolutionary algorithm will be therefore the following:

$$[\mathcal{P}_1, \ldots, \mathcal{P}_8 | c_{11}, c_{12}, c_{13}, c_{14}, \ldots, cN_1, cN_2, cN_3, cN_4] \quad (4)$$

where we have separated the integer part of the encoding from the real part. Note that we apply different operators to the integer and real part of the individuals in the algorithm, as we will describe in the next sections.

4.2. Crossover operator

The crossover operator is the core of any evolutionary search. For this problem, we have implemented a mixed crossover approach, different for the integer part of the individual and for the real part. In the case of the integer part (first of the individual in Eq. (4)), we implement a multi-point crossover. After forming couples with the individual in the population, we implement the multi-point crossover by means of a randomly generated binary template of length 8 (length of the integer part of the individual). A 1 in the template means that the corresponding genes of the couple will swap, whereas a 0 means that the genes will not swap. A different template will be generated for all the couples in a generation. Fig. 2 shows a small example of this crossover procedure for the integer part of the individuals.

The crossover operator for the real part of the individual is carried out also implementing a multi-point crossover approach, but in two different modes: first a normal mode, in a similar way as the previous operator defined above, but in the real part of the individual (see Fig. 3). Note that, in this case, the template has length $4N$. Also, we consider a second crossover mode, in which we interchange only parts of the individual belonging to a certain centroid (no part of centroids are allowed to be swapped). The length of the template in this second running mode of the crossover is N. Fig. 4 shows an example of this second crossover mode for the real part of the individuals. Basically, the first crossover mode generates new centroids by combining existing ones, and the second mode interchanges two centroids from different individuals. Both crossover modes are interesting and have an important role in the evolution of the population. Note that we apply the crossover operator a number of times necessary to obtain an offspring population of the same size as the initial (parents) one (L).

One important thing to be taken into account is that after applying the crossover operators, there may be situations in which a given centroid has no pressure matrix assigned. When one of these cases occur the void centroid is erased and its components reassigned to the proximity of a valid centroid (one with pressure matrices assigned). This procedure is carried out by assigning the coordinates of the centroid (differences of pressure) and slightly

Initial couple

17 2 12 153 21 93 19 66 | 0.17 -2.23 3.46 4.18 -1.25 -2.26 3.14 5.56 ...

11 110 70 42 17 180 21 1 | -4.10 3.12 0.16 -1.21 2.19 6.18 -1.23 -0.27 ...

Template: 1 0 0 1 0 1 1 0 ...

Couple after crossover

17 2 12 153 21 93 19 66 | -4.10 -2.23 3.46 -1.21 -1.25 6.81 -1.23 5.56 ...

11 110 70 42 17 180 21 1 | 0.17 3.12 0.16 4.18 2.19 -2.26 3.14 -0.27 ...

Fig. 3. Crossover applied to the real part of the individuals (parts of centroids interchanged).

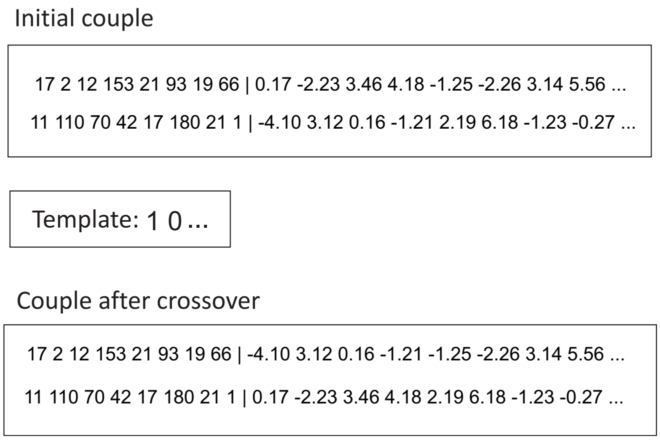

Fig. 4. Crossover applied to the real part of the individuals (complete centroids interchanged).

modifying them by adding a Gaussian noise to each coordinates. The pressure matrices are then reassigned to their nearest centroid. With this easy procedure we avoid the presence of void centroids in the evolutionary algorithm, which distorts somehow the efficiency of the search.

4.3. Mutation operator

Mutation operator is applied with a very low probability ($P_m = 0.01$) to each individual in the offspring population. Once a given individual is going to be mutated, the procedure of mutation is divided into two different versions, depending on whether it is applied on the integer part or to the real part of the individual. The mutation of the integer part is carried out by means of an integer randomized substitution of the current values of the individual, by different integers, in the interval [1,182]. The mutation in the real part of the individual is carried out by adding samples of uniform noise in the interval [-5,5], to a number of randomly chosen values of the real part in the mutated individual.

4.4. Selection operator

In this paper we use a tournament selection, which will be applied to the joint population formed from merging the initial and offspring populations. The result of the selection operator will be a single population, of size L, which will be the parents of the next generation of individuals. Basically, once the complete joint population of parents and offspring is formed, the standard tournament selection, as described in [42], has two main steps:

- Conduct pairwise comparison over the union of parents and offspring: for each individual, p opponents are chosen uniformly at random from all the parents and offspring. For each comparison, if the individual's fitness is better than the opponent's, it receives a "win".
- Select the L individuals out of the union of parents and offspring that have the most "wins" to be parents of the next generation.

Using this easy procedure, the remaining L individuals act as the parents of the next generations, and the crossover and mutation operators are applied again in a loop fashion, until the maximum number of generations are reached.

5. Experimental part

This section presents the experimental part of this study. It is structured in three different parts: first, we present the data available, and the methodology of the experiments carried out. An alternative approach for comparison is then presented in Section 5.2. Finally, we present the results obtained by the proposed

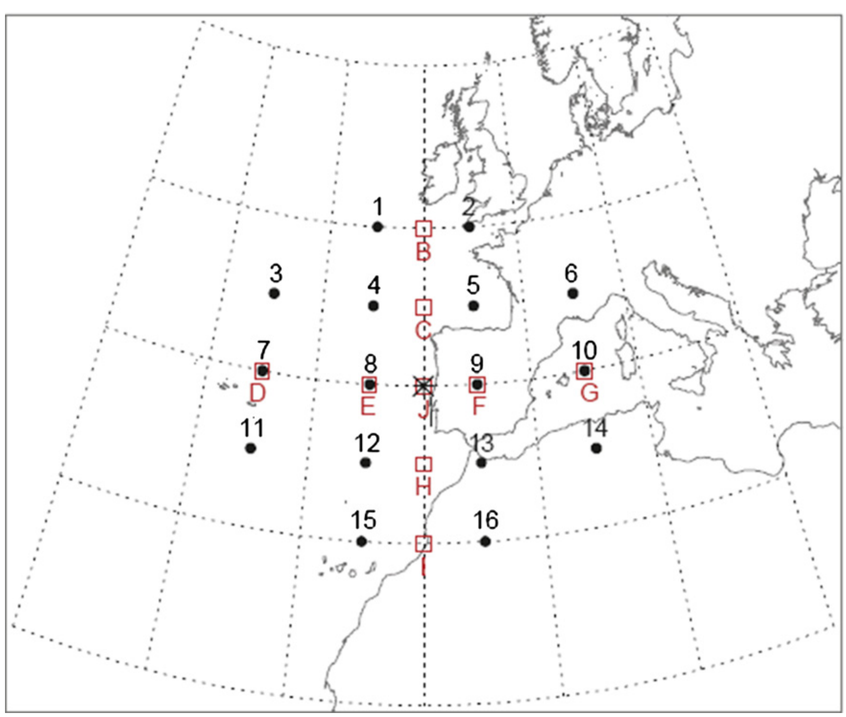

Fig. 5. Example of grid used in the definition of the WT algorithm's equations.

evolutionary algorithm and the comparison with the existing methodology.

5.1. Data available

Wind data from 10 years (1999–2008) of a meteorological tower close to a wind farm in the north of Guadalajara (Spain) are available for this study. They consist of wind speed and direction data, taken in the tower at 40 m of height every 10 min. Averages over 24 h are considered to obtain daily data vectors d_t. On the other hand, average daily pressure maps for the same period have been obtained from the National Center for Environmental Prediction/ National Center for Atmospheric Research Reanalysis Project (NCEP/NCAR) [44,45], which are public data profusely used in climatology and meteorology applications. As previously mentioned, we have considered a uniform grid in latitude and longitude, shown in Fig. 6, with 182 measurement points. Recall that the proposed evolutionary algorithm uses this value as a parameter of the encoding (in the differences of pressure, integer part of the encoding). We have set the 2/3 of the data for training and the final 1/3 as a test set, where we will measure the quality of the compared algorithms.

5.2. A weather types (pure meteorological) approach for comparison purposes

In this paper we consider an existing method recently developed and applied in the meteorology field for weather type classification. The method was first introduced in [43] and [7], as a first approach to obtain an objective weather type classification for the British islands and Portugal, respectively. More recent works have extended and improved the original method in order to obtain a robust method for synoptic pressure pattern classification, also known as weather-type classifier or weather-type circulation patterns in the meteorological field, applying the method to the prediction of rainfall in the Iberian Peninsula [8].

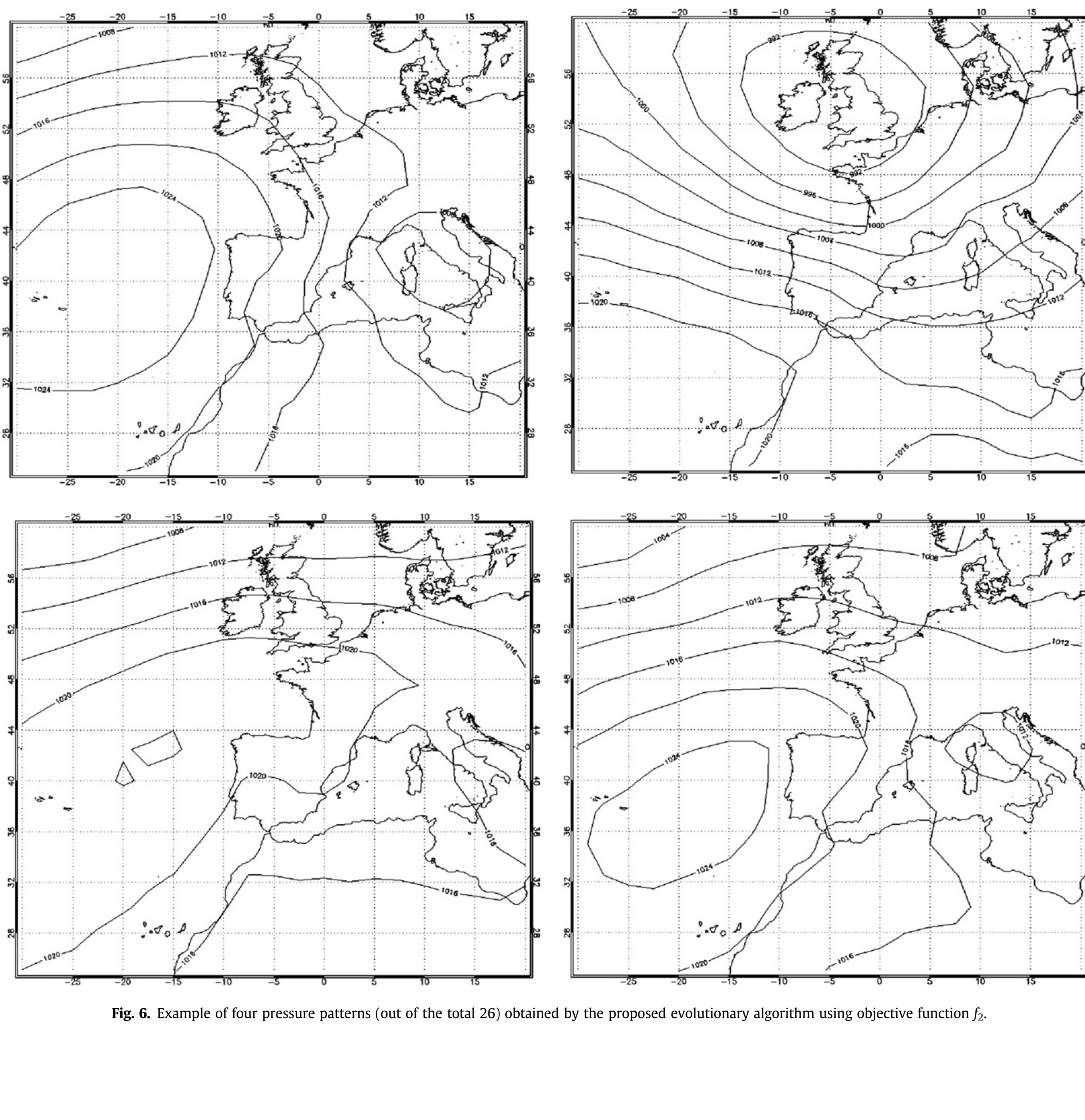

Fig. 6. Example of four pressure patterns (out of the total 26) obtained by the proposed evolutionary algorithm using objective function f_2.

The method considered is basically described in [8], and it is based on the equations of the atmospheric dynamics considering a given point of the Earth and some neighbors points. In order to obtain a classification of pressure pattern, it considers the value of two indexes flux and vorticity, F and Z respectively. F is related to the geostrophic flux, whereas Z is directly related to the absolute vorticity. The following equations show how to obtain the values of F and Z from a grid of pressure measurement such as the one given in Fig. 5 (it shows and example for a location in Lisbon, in any other location the obtained equations would be equivalents).

This method sets off from the equations for the geostrophic wind, given by:

$$u_{g40^\circ} = -\frac{1}{\rho f}\frac{\partial p}{\partial y} \approx -\frac{1}{\rho f_\lambda}\frac{[p_H - p_C]}{\Delta y} = -\frac{1}{\rho f_{40^\circ}}\frac{F_y}{R\Delta\lambda} \tag{5}$$

$$v_{g40^\circ} = \frac{1}{\rho f}\frac{\partial p}{\partial x} \approx \frac{1}{\rho f_\lambda}\frac{[p_F - p_E]}{\Delta x} = -\frac{1}{\rho f_{40^\circ}}\frac{F_x}{R\cos(\lambda_{40^\circ})\Delta\lambda} \tag{6}$$

where λ is the latitude, φ stands for the longitude, and f stands for the Coriolis parameter, given by the expression $f = 2\Sigma sin(\lambda)$. We can also calculate the variation of the geostrophic wind components with x and y, yielding:

$$\frac{\partial v_{g40^\circ}}{\partial x} \approx \frac{1}{2R\cos\left(\lambda_{40^\circ}\right)}\frac{\Delta v_g}{\Delta\varphi} = \frac{1}{2R^2\rho f_\lambda(\Delta\varphi)^2\cos\left(\lambda_{40^\circ}\right)}[(p_G - p_F) - (p_E - p_D)] \tag{7}$$

$$-\frac{\partial u_{g40^\circ}}{\partial y} \approx -\frac{1}{R}\frac{\Delta u_g}{\Delta\lambda} = \frac{1}{R^2\rho f_{40^\circ}(\Delta\varphi)^2}\left[\frac{\sin\left(40^\circ\right)}{\sin\left(45^\circ\right)}\left(p_J - p_H\right) - \frac{\sin\left(40^\circ\right)}{\sin\left(35^\circ\right)}\left(p_I - p_J\right)\right] \tag{8}$$

Now, using Eqs. (5) and (6) an index for the flux from the south (FS), flux from the west (FO) and total flux F, can be constructed in the following way:

$$FO = \left[\frac{1}{2}(p_{12} + p_{13}) - \frac{1}{2}(p_4 + p_5)\right] \tag{9}$$

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

$$F = \sqrt{(FO)^2 + (FS)^2} \tag{11}$$

In the same way, using Eqs. (7) and (8) index for south and west vorticity, and total vorticity can be calculated:

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

$$ZO = \frac{\sin\left(40^\circ\right)}{\sin\left(35^\circ\right)}\left[\frac{1}{2}(p_{15} + p_{16}) - \frac{1}{2}(p_8 + p_9)\right] - \frac{\sin\left(40^\circ\right)}{\sin\left(35^\circ\right)}\left[\frac{1}{2}(p_8 + p_9) - \frac{1}{2}(p_1 + p_2)\right] \tag{13}$$

$$Z = ZS + ZO \tag{14}$$

Table 1
Comparison of the results obtained by the Weather Types (WT) algorithm in terms of objective functions f_1 and f_2. The class acronyms stand for: N (north), NE (north east), E (east), SE (south east), S (south), SW (south west), W (west), NW (north west), H (high pressure), L (low pressure).

Class	WT (f_1)	WT (f_2)
N	6.1032	2.1610
NE	4.3350	2.8467
E	4.6294	1.6812
SE	3.3555	1.0684
S	4.2521	1.8131
SW	4.6509	2.5693
W	4.5786	2.5731
NW	5.0759	2.3238
H	4.9692	1.3034
L	4.8432	2.3045
HN	5.4200	2.3921
HNE	4.4914	2.5811
HE	3.8727	1.5669
HSE	2.9818	0.9793
HS	3.5171	1.0155
HSW	3.6246	1.3747
HW	2.9900	1.2288
HNW	3.7248	1.5307
LN	5.6579	1.5555
LNE	5.9747	1.3312
LE	4.0312	1.9162
LSE	4.4421	1.1468
LS	6.3267	2.7714
LSW	6.2862	4.9242
LW	4.7166	2.3264
LNW	5.8609	2.5247
Average	4.6771	1.8490

Thus, in [8] it is shown how to obtain a relationship between the values of F and Z and a set of weather types by using a set of rules: if $|Z| < F$, then flux is considered strait, and then this produces 8 weather types, which coincide with the directions of the wind rose.

Table 2
Comparison of the results obtained by the proposed EA algorithm in terms of objective functions f_1 and f_2.

Class	EA (f_1)	EA (f_2)
1	4.0948	0.9219
2	3.5823	1.9341
3	2.8869	0.9451
4	6.0664	1.7223
5	3.9031	1.6298
6	2.8654	1.3354
7	4.5260	1.1268
8	2.8933	1.0975
9	2.7333	2.1016
10	4.0927	2.7232
11	3.7860	1.2755
12	3.6864	1.2671
13	4.1649	1.4865
14	4.7601	1.2354
15	3.7261	1.6495
16	2.8397	1.9007
17	5.2507	2.4052
18	3.1603	2.2761
19	2.9757	2.8951
20	3.7177	0.9425
21	3.2553	1.7015
22	5.4534	2.9094
23	4.6713	1.1181
24	3.7745	1.1165
25	3.6191	1.3853
26	4.6443	1.7821
Average	3.7372	1.3755

If $|Z| > 2F$, the pattern is considered as purely cyclonic, if $Z > 0$, or purely anticyclonic if $Z < 0$. In the case that $F < |Z| < 2F$, the flux is considered hybrid, and it is characterized by its direction and its circulation, producing 16 different weather types. Considering all these possibilities, in [8] 26 weather types (classes) were identified. Moreover, in [8] this classification was used to successfully predicting rainfall in the Iberian Peninsula. Posterior works [34,35], also used this classification method in the estimation of wind patterns in the Iberian Peninsula.

Summarizing, the weather types (WT) algorithm first proposed by [43] and [7], and recently improved by [8], is the most successful existing algorithm for pressure pattern recognition. It has been chosen therefore as a reference for comparison of our proposal. In order to make possible the complete comparison, we have run our evolutionary approach setting 26 classes. The results obtained by the two algorithms are shown in the next subsection.

5.3. Results

Tables 1 and 2 show the results obtained by the proposed EA, and the weather types (WT) approach in [8], respectively. These tables summarize the results grouped by classes, in terms of the Objectives Functions (1) and (2). Note that the classes obtained by each algorithm are different, and thus, it is not possible to compare class by class the objective function values. However, it is possible to compare the average values obtained by the different algorithms in the whole test set. Note that the EA approach outperforms the WT approach using both objective functions. Note that the WT algorithm provides best quality solutions when its performance is characterized with the objective function f_2 (module of wind speed), whereas the results in terms of wind speed vector (function f_1) is not so accurate. In average of the 26 classes, the EA optimizing function f_2 obtained the best quality, with an average value of 1.37 m/s. The EA optimizing function f_1 obtains a worse result, with an average value of 3.73 m/s. Note, however, that this result is again better than the one obtained by the WT in terms of the f_1 objective function.

Fig. 6 shows an example of some of the pressure classes found by the EA using objective function f_2 (for simplicity we show 4 classes out of the total 26). Within the pressure classes found by the EA there are several which include high pressures in different positions of the Atlantic ocean, some other classes are characterized by very low pressures over the English islands and Ireland, there are also pressure classes in which it is easy to detect high pressures in central Europe and finally, a group of classes with a predominant flow from the west. In general the proposed EA groups the pressure patterns in classes which are easily recognizable as *average* meteorological situations what is, intuitively, a good point of reference.

The good performance of the approach can also be appreciated in Figs. 7 and 8, which show the total 26 different patterns in terms of the wind series. The first figure provides the results of the wind classes by using objectives functions f_1, whereas the second figure shows the results obtained by the EA algorithm optimizing objective function f_2. The centroid in each class is notated with a red cross, and the module of the centroid is depicted by a green circle (all the points in the circle have 0 error in module with respect to the centroid of the corresponding class). Note that in the solutions

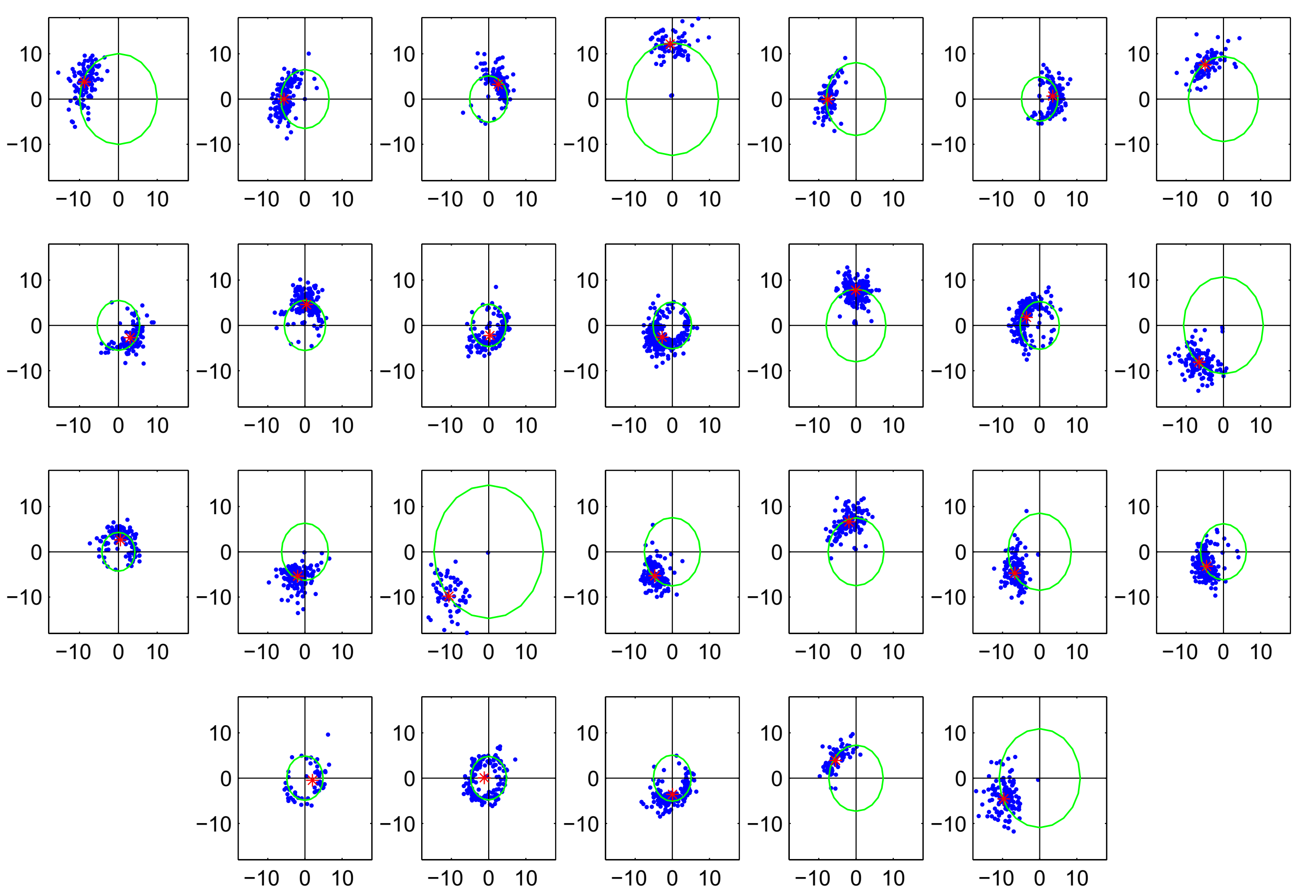

Fig. 7. Wind patterns obtained by the proposed evolutionary algorithm using objective function f_1.

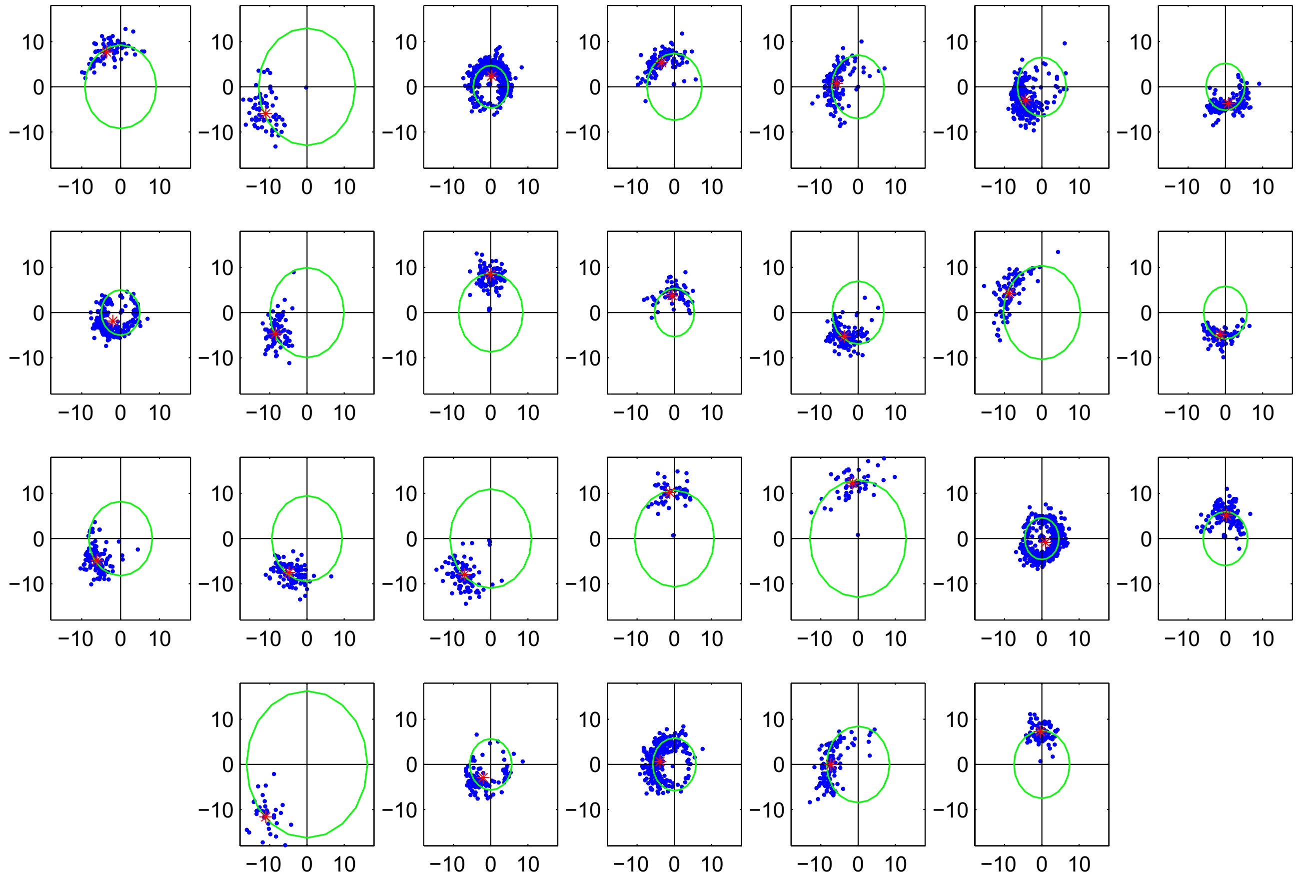

Fig. 8. Wind patterns obtained by the proposed evolutionary algorithm using objective function f_2.

provided by the proposed EA, the wind samples within each class are compacted, as expected in a good quality solution.

Fig. 9 shows a histogram of the results obtained by the EA (over 30 runs), using as objective function f_1 (given by Eq. (1)), or f_2, (given by Eq. (2)). As can be seen, when the objective function mainly optimizes the module (function f_2), the values obtained by the EA are more compacted, i.e. their variance is smaller that when both module and wind direction are optimized. Note that in these figures it can also be seen that the results obtained with the f_2 objective function are, in average, better than using objective function f_1, as previously shown in Tables 1 and 2.

Finally, we will give an example of wind speed series reconstruction based on the pressure clustering obtained with the EA and the WT. Fig. 10 shows these examples: Fig. 10 (a) shows the wind speed series reconstruction from the pressure clustering obtained with the proposed EA, whereas Fig. 10 (b) shows the wind speed reconstruction obtained with the WT approach. The first sixty samples of the test set have been chosen to show this comparison. As can be seen, the wind speed series reconstruction obtained with the pressure clustering given by the EA is more accurate than the one obtained with the WT, as expected. Note that the wind speed series reconstruction using the EA follows quite well the wind speed trend in the whole period of 60 d. This indicates that the pressure clustering obtained with the EA is significant, and we have been able to extract a correct relationship between the pressure classes and the wind speed series.

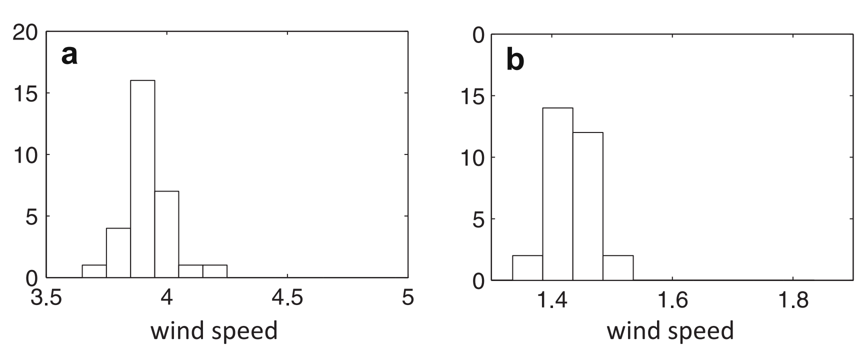

Fig. 9. Histogram (over 30 runs of the algorithm) of the results obtained by the proposed EA; (a) using f_1 as objective function; (b) using f_2 as objective function.

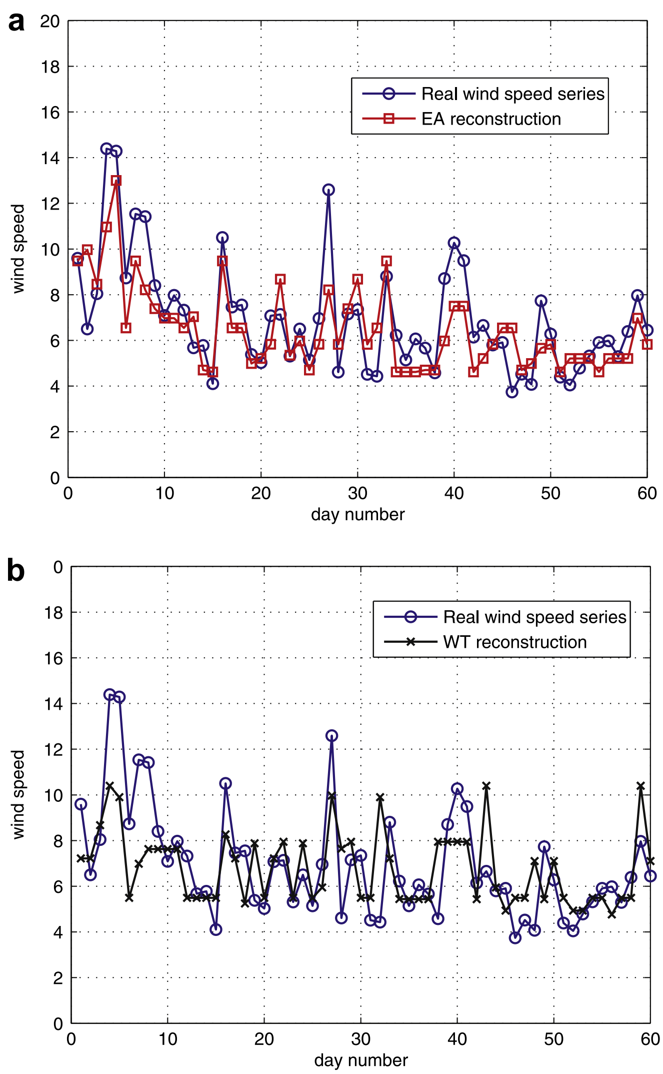

Fig. 10. Reconstruction of the 60 first wind speed values of the test series; (a) Real wind speed and EA reconstruction; (b) Real wind speed and WT reconstruction.

6. Conclusions

In this paper we have presented an evolutionary algorithm for a new problem of clustering in a search space of atmospheric pressures in a grid. The problem's objective function is related to a measure of wind speed in a given point near a wind farm. It is well known the relationship between the synoptic pressure measures and wind speed measures in related geographical zones. Thus, the accurate obtention of this relationship could be of great interest to companies involved in the management of wind farms, since this problem is in the core of different applications such as long-term wind speed prediction and also seasonal analysis of production results. In addition, note that there are not many clustering algorithms to extract pressure patterns under accurate wind speed measures so, in this sense, this work covers, from the perspective of artificial intelligence, a niche of important research that had only been tackled from the meteorological point of view. Thus, to our knowledge, this is the first work which incorporates an evolutionary algorithm to the extraction of these synoptic pressure patterns. We have developed a specific encoding in the evolutionary algorithm to deal with the problem, and also ad-hoc mutation and crossover operators that improve the performance of the evolutionary approach in this problem. The performance of the evolutionary algorithm proposed has been shown to be very good in a real example in the Iberian Peninsula, outperforming the results of a previously existing (purely meteorological) algorithm.

Acknowledgment

This work has been partially supported by Spanish Ministry of Industry, Tourism and Trading, under an Avanza 2 project, number TSI-020100-2010-663.

References

[1] Torres JL, García A, De Blas M, De Francisco A. Forecast of hourly average wind speed with ARMA models in Navarre (Spain). Solar Energy 2005;79:65–77.
[2] Khashei M, Bijari M, Raissi-Ardali G. Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs). Neurocomputing 2009;72(4–6):956–67.
[3] Mellit A, Kalogirou SA, Hontoria L, Shaari S. Artificial intelligence techniques for photovoltaic applications: a review. Progress in Energy and Combustion Science 2009;34(5):406–19.
[4] Barbounis TG, Theocharis JB. Locally recurrent neural networks for long-term wind speed and power prediction. Neurocomputing 2006;69(4–6):466–96.
[5] Li G, Shi J. Application of Bayesian model averaging in modeling long-term wind speed distributions. Renewable Energy 2010;35(6):1192–202.
[6] Mohandes MA, Halawani TO, Rehman S, Hussain AA. Support vector machines for wind speed prediction. Renewable Energy 2004;29:939–47.
[7] Trigo RM, DaCamara CC. Circulation weather types and their influence on the precipitation regime in Portugal. International Journal of Climatology 2000;20:1559–81.
[8] Paredes D, Trigo RM, García-Herrera R, Trigo IF. Understanding precipitation changes in Iberia in early spring: weather typing and storm tracking approaches. Journal of Hydrometeorology 2006;7:101–13.
[9] Chen ZH, Cheng SY, Li JB, Guo XR, Wang WH, Chen DS. Relationship between atmospheric pollution processes and synoptic pressure patterns in northern China. Atmospheric Environment 2008;42(24):6078–87.
[10] Jain AK, Murty MN, Flynn PJ. Data clustering: a review. ACM Computing Surveys 1999;31(3):264–322.
[11] Xu R, Wunsch D. Survey of clustering algorithms. IEEE Transactions on Neural Networks 2005;16(3):645–78.
[12] Liao TW. Clustering of time series data: a survey. Pattern Recognition 2005;38:1857–74.
[13] Wong KY, Yip CL. Identifying centers of circulating and spiraling vector field patterns and its applications. Pattern Recognition 2009;42:1371–87.
[14] Mitra S, Banka H. Multi-objective evolutionary biclustering of gene expression data. Pattern Recognition 2006;39:2464–77.
[15] Scheunders P. A genetic c-means clustering algorithm applied to color image quantization. Pattern Recognition 1997;30(6):859–66.
[16] Gomez-Munoz VM, Porta-Gándara MA. Local wind patterns for modeling renewable energy systems by means of cluster analysis techniques. Renewable Energy 2002;2:171–82.
[17] Chang DX, Zhang XD, Zheng CW. A genetic algorithm with gene rearrangement for K-means clustering. Pattern Recognition 2009;42:1210–22.
[18] M Ester, HP Kriegel and J Sander. A density-based algorithm for discovering clusters in large spatial data bases with noise, in Proc. of the 2nd International Conference on Knowledge Discovery and Data Mining. 1996. pp. 226–231.
[19] Kanungo T, Mount D, Netanyahu NS, Piatko C, Silverman R, Wu A. An efficient K-means clustering algorithm: analysis and implementation. IEEE Transactions on Pattern Analysis and Machine Intelligence 2002;24(7):881–92.
[20] Likas A, Vlassis N, Verbeek JJ. The global K-means clustering algorithm. Pattern Recognition 2003;36:452–61.
[21] Mclachlan GJ, Basford KE. Mixture models: inference and applications to clustering. New York: Marcel Dekker; 1988.
[22] Madeira SC, Oliveira AL. Biclustering algorithms for biological data analysis: a survey. IEEE Transactions on Computational Bioinformatics 2004;1(1):24–45.
[23] Dehuri S, Ghosh A, Mall R. Genetic algorithms for multi-criterion classification and clustering in data mining. International Journal of Computing and Information Systems 2006;4(3):143–54.
[24] Hong Y, Kwong S, Chang YC, Ren QS. Unsupervised feature selection using clustering ensembles and population based incremental learning algorithm. Pattern Recognition 2008;41:2742–56.
[25] Krishna K, Murty MN. Genetic K-means algorithm. IEEE Transactions on Systems 1999;29:433–9. Man and Cybernetics, Part B.
[26] Murthy CA, Chowdhury N. In search of optimal clusters using genetic algorithms. Pattern Recognition Letters 1996;17:825–32.
[27] Tseng LY, Yang SB. A genetic approach to the automatic clustering problem. Pattern Recognition 2001;34(2):415–24.
[28] Divina F, Aguilar-Ruiz J. Biclustering of expression data with evolutionary computation. IEEE Transactions on Knowledge & Data Engineering 2006;18(5):590–602.

[29] Sarkar M, Yegnanarayana B, Khemani D. A clustering algorithm using an evolutionary programming-based approach. Pattern Recognition Letters 1997;18:975–86.
[30] Zhu Q, Cai Y, Liu L. A multiple hyper-ellipsoidal subclass model for an evolutionary classifier. Pattern Recognition 2001;34:547–60.
[31] Yang F, Sun T, Zhang C. An efficient hybrid data clustering method based on K-harmonic means and particle swarm optimization. Expert Systems with Applications 2009;36:9847–52.
[32] Parikh JA, DaPonte JS, Vitale JN, Tselioudis G. An evolutionary system for recognition and tracking of synoptic-scale storm systems. Pattern Recognition Letters 1999;20:1389–96.
[33] Romero R, Summer G, Ramis C, Genoves A. A classification of the atmospheric circulation patterns producing significant daily rainfall in the Spanish Mediterranean area. International Journal of Climatology 1999;19:765–85.
[34] Burlando M, Antonelli M, Ratto CF. Mesoscale wind climate analysis: identification of anemological regions and wind regimes. International Journal of Climatology 2008;28:629–41.
[35] Jiménez PA, Gonzalez-Rouco JF, Montavez JP, García-Bustamante E, Navarro J. Climatology of wind patterns in the northeast of the Iberian Peninsula. International Journal of Climatology 2009;29:501–25.
[36] Eiben AE, Smith JE. Introduction to evolutionary computing. Springer-Verlag; 2003.
[37] Kusiak A, Zheng H. Optimization of wind turbine energy and power factor with an evolutionary computation algorithm. Energy 2010;35(3):1324–32.
[38] Liao CC. Genetic K-means algorithm based RBF network for photovoltaic MPP prediction. Energy 2010;35(2):529–36.
[39] Jursa R, Rohrig K. Short-term wind power forecasting using evolutionary algorithms for the automated specification of artificial intelligence models. International Journal of Forecasting 2008;24:694–709.
[40] Vaisakh K, Srinivas LR. A genetic evolving ant direction DE for OPF with non-smooth cost functions and statistical analysis. Energy 2010;35(8):3155–71.
[41] Toffolo A, Lazzaretto A. Evolutionary algorithms for multi-objective energetic and economic optimization in thermal system design. Energy 2002;27(6):549–67.
[42] Yao X, Liu Y, Lin G. Evolutionary programming made faster. IEEE Transactions on Evolutionary Computation 1999;3(2):82–102.
[43] Jenkinson AF, Collison FP. An initial climatology of gales over North Sea. Synoptic Climatology Branch Memorandum no. 62. Bracknell: Meteorological Office; 1977.
[44] Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, et al. The NCEP/NCAR reanalysis project. Bulletin of the American Meteorological Society 1996;77:437–71.
[45] <http://www.esrl.noaa.gov/psd/data/reanalysis/reanalysis.shtml>