Renewable and Sustainable Energy Reviews 90 (2018) 728–741

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

Feature selection in machine learning prediction systems for renewable energy applications

S. Salcedo-Sanz[a,*], L. Cornejo-Bueno[a], L. Prieto[b], D. Paredes[b], R. García-Herrera[c,d]

[a] Department of Signal Processing and Communications, Universidad de Alcalá, 28805 Alcalá de Henares, Madrid, Spain
[b] Department of Energy Resource, Iberdrola, Spain
[c] Department of Earth Physics and Astrophysics, Universidad Complutense de Madrid, Spain
[d] IGEO, Instituto de Geociencias (UCM-CSIC), Madrid, Spain

ARTICLE INFO

Keywords:
Renewable energy
Feature selection problems
Prediction problems
Coral reefs optimization algorithm
Extreme learning machines

ABSTRACT

This paper focuses on feature selection problems that arise in renewable energy applications. Feature selection is an important problem in machine learning, both in classification and regression problems. In renewable energy systems, feature selection appears related to prediction systems in the most important sources such as wind, solar and marine resources. The objective of the paper is twofold: first, a review of the most important prediction systems for renewable energy applications involving feature selection is carried out. Analysis and discussion of different feature selection problems in prediction systems are considered. We show that wrapper FSP approaches are those mostly used due to their higher performance. They include a diversity of algorithms, prevailing fast-training approaches. The lack of an uniform framework for FSP and the diversity of tackled problems impede a systematic assessment of the performance and properties of the applied methods. Thus, the simultaneously use of several global search mechanisms should be the preferred option. In a second part of the paper, we explore this possibility, by introducing a novel approach for feature selection based on a novel meta-heuristic, the Coral Reefs Optimization algorithm with Substrate Layer. This approach is able to combine different search mechanisms into a single algorithm, providing a global search procedure of high quality. We use an Extreme Learning Machine for prediction within this novel approach. The performance of the system is evaluated in a problem of wind speed prediction from numerical models input, using real data from a wind farm in Spain, where comparison with alternative regression algorithms is carried out. Improvements up to 20% in hourly and daily wind speed prediction are obtained with the proposed system versus the algorithms without the feature selection process considered.

1. Introduction

There is a global agreement that the increased use and support to renewable energy sources is critical for climate change mitigation policies [1]. In fact, the renewable energy sector is key in the new green economies in developed countries [2,3], and also for achieving a sustainable and environmental-compatible growth in developing ones [4]. It is estimated that over 23% of electricity production at global level came from renewable energy sources in 2015, and the expectations clearly show that this trend will continue in future years. This data is supported by the fact that renewable energy sources represented approximately 70% of net additions to global power capacity in 2015, with wind, solar photovoltaic, and hydro power being the most important sources in terms of installed power [1]. Alternative sources such as marine energy or bio-power are also starting to be taken into account, with a significant growth in many regions.

In spite of the benefits inherent to the sustainable sources of energy, there are different technical reasons that limit their use and higher penetration in the power system. The intermittence of the resource is mainly the most important issue that currently prevents renewables to play a more important role in the energetic-mix. This resource intermittence makes very important the developing of accurate prediction systems to estimate power produced by renewable sources [1].

Prediction is one of the keys in renewable energy management, mainly in the most important ones such as wind and solar. In fact, resource prediction is carried out at all levels: in large facilities such as wind and solar farms, and also in small or medium ones, such as micro-grids with small or limited generation resources [5]. The research associated with prediction in renewable energy sources is huge, with hundred of different models and systems proposed and analyzed in

* Corresponding author.
E-mail address: sancho.salcedo@uah.es (S. Salcedo-Sanz).

https://doi.org/10.1016/j.rser.2018.04.008
Received 19 December 2016; Received in revised form 22 March 2018; Accepted 3 April 2018
1364-0321/

Nomenclature

Algorithms

ANFIS	Adaptive Neuro-fuzzy Inference System
BSA	Backtracking Search Algorithm
CRO	Coral Reefs Optimization algorithm
CRO-SL	Coral Reefs Optimization with Substrate Layer algorithm
DNN	Deep Neural Network
DE	Differential Evolution
ELM	Extreme Learning Machine
EMD	Empirical Mode Decomposition
GA	Genetic Algorithm
GGA	Grouping Genetic Algorithm
GP	Gaussian Process
GS	Gravitational Search algorithm
HMCR	Harmony Memory Considering rate
HS	Harmony Search
MLP	Multi-Layer Perceptrons
MLR	Multi-Linear Regression
k-NN	k-Nearest Neighbors
NARX	Nonlinear Auto-Regressive eXogenous model
NN	Neural Network
MSVR	multi-output Support Vector Regression
PCA	Principal Component Analysis
PSO	Particle Swarm Optimization
RF	Random Forest
RSVR	Reduced Support Vector Regression
SVR	Support Vector Regression

Models and systems

ARIMA	Auto-Regressive Integrated Moving Average model
GFS	Global Forecasting System
FSP	Feature Selection Problem
MCP	Measure Correlate Predict approaches
WEKA	Waikato Environment for Knowledge Analysis
WRF	Weather Research and Forecasting system

Administrations

AFWA	Air Force Weather Agency (USA)
FAA	Federal Aviation Administration (USA)
FSL	Forecast Systems Laboratory (USA)
NCAR	National Center for Atmospheric Research (USA)
NDBC	National Data Buoy Center of the USA
NCEP	National Center for Environmental Prediction (USA)
NOAA	National Oceanic and Atmospheric Administration (USA)
SURFRAD	NOAA Surface Radiation Network (USA)

terms of their performance for specific technologies. Wind power production is the technology where more prediction models have been proposed [6], since it is the most developed renewable source. However, models for solar [7] or marine energy sources [8] are becoming more frequent in the last years.

In this paper we are interested in studying an important part of prediction systems for renewable energy sources: the selection of the most important variables or input parameters which improve the prediction capability of a given system. This problem is well-known in computer science and machine learning fields, where it receives the name of *Feature Selection Problem* (FSP). Note that feature selection deals with keeping the best set of features in a prediction problem, i.e. those with the maximum information to solve the problem. There are alternative approaches, such as feature extraction or dimensionality reduction approaches (such as PCA), which could also provide good results in improving the prediction capability of a machine learning system, however, they are less interpretable than the solutions from a FSP in most cases. With this in mind, the paper objective is twofold: first, we define the problem in terms of the prediction capability of regressors or classification algorithms, which are the most used techniques applied to prediction of renewable energy resource. We also include a review of the most important FSP approaches directly applied to the main renewable sources (wind, solar and marine energy). This first analysis shows that there is a very important variety of search algorithms involved in feature selection proposals, without a systematic analysis for choosing a given technique. In the second part of the paper, we discuss a novel approach which is able to combine different search patterns into a single algorithm. The approach is the CRO-SL, which provides a high-quality search based on the competitive co-evolution of different global search algorithms. We show the performance of this approach in a case study of feature selection in wind energy prediction involving real data of wind farms in Spain. The proposed CRO-SL is mixed with a neural network for regression, the ELM, and compared with the performance of alternative algorithms such as multi-linear regression or support vector regression algorithms, showing an excellent behaviour.

The rest of the paper is structured in the following way: next section describes the FSP, focused in prediction problems solved with machine learning approaches. Section 3 reviews the most important works dealing with the FSP in renewable energy prediction problems. It is structured in wind energy, solar energy, marine energy prediction problems and finally FSP for energy-related applications. Section 4 presents the case a FSP in wind energy prediction. Finally, Section 6 closes the paper by giving some concluding remarks.

2. Feature selection in machine learning

Feature selection is an important task in Machine Learning related problems because irrelevant features, used as part of the training procedure of different prediction systems, can increase the cost and running time of the system, and make its generalization performance much poorer [9,10]. In its more general form, the FSP for a learning problem from data can be defined as follows: given a set of labeled data samples $(\mathbf{x}_1, y_1), ..., (\mathbf{x}_l, y_l)$, where $\mathbf{x}_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$ (or $y_i \in \{\pm 1\}$ in the case of classification problems), choose a subset of m features ($m < n$), that achieves the lowest error in the prediction of the variable y_i.

Thus, note that there are many different algorithms which can be used to solve a FSP. In general, they can be structured in two different families or paradigms:

- The *wrapper approach* to the FSP was introduced in [11]. The feature selection algorithm conducts a search for a good subset of features using the classifier/regressor itself as part of the evaluating function. Fig. 1(a) shows the idea behind the wrapper approach: the classifier/regression technique is run on the training dataset with different subsets of features. The one which produces the lowest estimated error in an independent but representative test set is chosen as the final feature set. For further reading on wrappers methods, the following classical works can be consulted [12–14]. In the case of the wrapper method, the FSP admits a mathematical definition as follows: The FSP consists of finding the optimum n-column vector $\boldsymbol{\sigma}$, where $\sigma_i \in \{1, 0\}$, that defines the subset of selected features, which is found as

$$\boldsymbol{\sigma}^o = arg\min_{\sigma,\alpha}\left(\int V(y, f(\mathbf{x}^*\boldsymbol{\sigma}, \boldsymbol{\alpha}))dP(\mathbf{x}, y)\right), \tag{1}$$

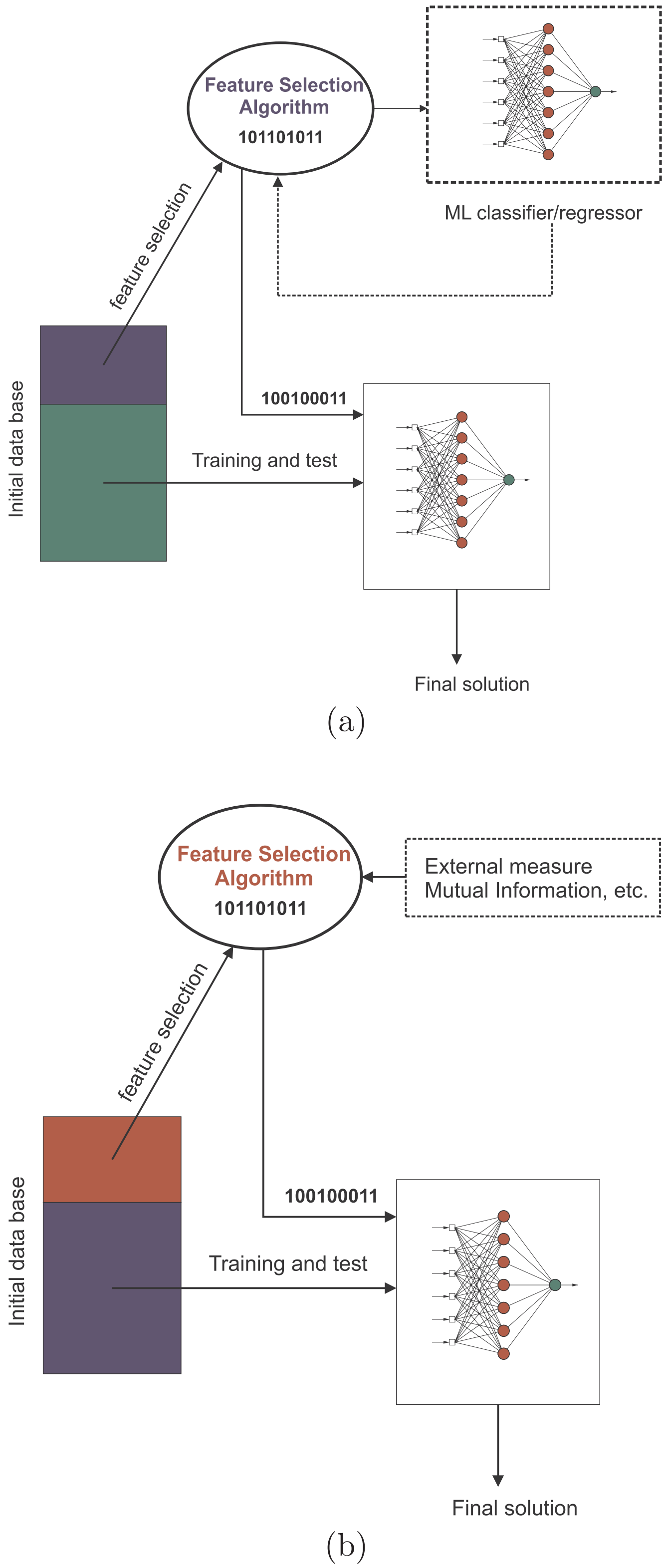

Fig. 1. (a) Outline of a wrapper method; (b) Outline of a filter method.

where $V(\cdot,\cdot)$ is a loss functional, $P(\mathbf{x}, y)$ is the unknown probability function the data was sampled from and we have defined $\mathbf{x}*\boldsymbol{\sigma} = (x_1\sigma_1, \ldots, x_n\sigma_n)$. The function $y = f(\mathbf{x}, \boldsymbol{\alpha})$ is the classification/regression engine that is evaluated for each subset selection, $\boldsymbol{\sigma}$, and for each set of its hyper-parameters, $\boldsymbol{\alpha}$.

- In the *filter approach* to the FSP, the feature selection is performed based on the data, ignoring the classifier algorithm. An external measure calculated from the data must be defined to select a subset of features. After the search, the best feature subset is evaluated on the data by means of the classifier algorithm. Note that filter algorithms performance completely depends on the measure selected for comparing subsets. Fig. 1 (b) shows an example of how a filter algorithm works. Filter methods are usually faster than wrapper methods. However, their main drawback is that they totally ignore the effect of the selected feature subset on the performance of the classification/regression algorithm during the search. So, usually their performance is poorer than wrapper approaches. Further analysis and application of filter methods can be found in [15,16].
- There are different works which have combined both wrapper and filter methodologies to build hybrid approaches. They have shown very good performance in specific applications [17–20].

For both wrapper and filter methods, a binary representation can be used for the FSP, where a 1 in the i_{th} position of the binary vector means that the feature i is considered within the subset of features, and a 0 in the j_{th} position means that feature j is not considered within the subset. Note that using this notation is equivalent to encode the problem as the vector $\boldsymbol{\sigma}$ included in expression (1). Note also that there are 2^n different subsets (being n the total number of features), and the problem is to select the best one in terms of a certain measure, which can be either internal (wrapper methods) or external (filter methods) to the classifier.

3. A review of FSP methods in renewable energy prediction problems

This section presents a review of the main works dealing with FSP in renewable energy prediction problems.

3.1. Feature selection in wind energy prediction

Wind energy is the most developed renewable energy technology, and, thus, research on its prediction has been carried out during more than 20 years. FSPs have also been tackled in wind energy during more than a decade. For example, Jursa in [21], first proposed a feature selection in wind power prediction systems. Specifically, a PSO approach was introduced in order to obtain the optimal set of features which provided the best prediction. This work was further completed in another paper by Jursa and Rohrig, [22], where NN and k-NN algorithms were used as prediction models, and two wrapper feature selection models were considered. A large number of features were used in that work as inputs for the prediction methods, such as previous wind speed and power samples (historic data) and predicted weather variables. Two different global search mechanism were used to form the wrapper approaches: a PSO algorithm and a DE approach. The encoding of the FSP into the PSO and DE algorithms consisted of different values of the predictive variables delayed in time, until a maximum delay d was reached. The evaluation of the systems was carried out in 10 wind farms in northwestern Germany, obtaining good performance in the prediction of the wind power in each wind farm. A comparison of the wrapper PSO and DE algorithms with a NN and k-NN algorithms was carried out against a manually constructed solution, which included values of wind speed, wind power and a cycle time series to capture the intra-daily variability. Persistence was also included as a baseline (reference) algorithm. It was clearly shown that the results obtained by the wrapper approaches outperform the manually constructed features, and also the persistence algorithm in the 10 wind farms considered, indicating that the feature selection is positive in problems of short-term wind speed prediction. Improvements around 5% (depending on the wind farm considered) were obtained when using the PSO and DE algorithms for feature selection versus the neural network working without feature selection mechanism. The improvement of including feature selection against persistence is higher, reaching values near 15% depending on the wind farm studied, with a mean improvement over 10% with the hybrid PSO-NN algorithm.

Gupta et al. [23], proposed a hybrid wrapper approach, formed by a GA and a NN as predictor. The database consisted of 12 different variables, which were measured in a wind farm in Jaipur, India. The input variables were a combination of atmospheric variables and previous wind speed measurements. The system for feature selection improved the NN with all the variables as inputs for the prediction, obtaining an improvement over 5% in prediction accuracy. The authors show that the best result was obtained with 8 features out of the total 12. Results showed an improvement up to 5% when the feature selection was included in the system. Note that, in this case, the number of predictive variables involved in the problem was small, so the use of a meta-heuristic to carry out the search was not necessary. In fact, given a partition of the search space and fixing all the NNs parameters, the best solution would be found by examining all the possible combinations of features (a total of 4095 feasible combinations, without the 0). Note that the use of meta-heuristic approaches is recommended for FSP problems where the search space is extremely large.

More advanced computational methods have been recently applied to different feature selection problems in wind energy estimation. In [24], Kou et al. proposed an online feature selection method with a wrapped GP as predictor, for a problem of probabilistic wind power forecasting. Experiments in real data from several wind farms in Baotou region, China, showed the performance of the proposed methodology. Short term wind speed prediction (15 min and 1 h time-horizon predictions) was considered. In this case, a large number of predictive variables from direct measurements in the wind farm and also from numerical weather modes were taken into account. Regarding observational variables, that work considered 384 inputs, from 4 measurement points, including data until four days before the prediction. The input variables from the numerical weather model were obtained from the GFS system (NOAA), with a resolution of 2.5 degrees. These variables were linearly projected to the four measurement points. The numerical variables considered to improve the performance of the prediction system were measurements of 10 m wind speed. The feature selection was carried out by means of a greedy algorithm, the sequential forward greedy search, which starts from an empty set of variables, then incorporates selected features one by one, and finally evaluates the set. The algorithm keeps the best set of features found in the search. Note that this search is not as effective as applying a meta-heuristic approach, in which different sets of features are evaluated together. However, it is a faster procedure, and ensures a good enough solution with limited computational resources. The results obtained in the paper showed improvements of 6% with respect to persistence, and smaller improvements (less than 1%) versus non-linear systems such as NNs or SVRs.

In [25], Salcedo et al. proposed a novel hybrid FSP wrapper approach formed by a CRO algorithm for feature selection with an Extreme Learning Machine as regressor, for wind speed prediction in wind farms. A large number of features were considered: temperature, wind speed and direction at different heights, etc. The predictive variables were obtained from numerical weather models in the surroundings of the wind farm, and also included synthetically constructed variables from the original ones. The wind speed prediction process was carried out by using an ELM [26]. The ELM is a extremely fast training procedure for MLP-type neural networks. It is specially well-suited to construct wrapper approaches for feature selection involving meta-heuristic algorithms, since it alleviates the computational burden of the fitness evaluation associated to the meta-heuristic. In that work, the evaluation of the CRO-ELM approach was carried out in real data from a wind farm in Oregon, USA, and compared to a different wrapper approach guided by an evolutionary algorithm. The results showed the goodness of the proposed CRO-ELM approach in comparison to the EA-ELM, with an improvement around 2%. This work was completed in [27], where the same authors proposed an improved version of the CRO algorithm including operators from a HS algorithm. In this case, instead of using a two-points crossover in the CRO, the search was carried out by including a HMCR operator to obtain new solutions in the CRO. The rest of the algorithm was not changed, but it was shown that the HMCR operator was able to improve the performance of the CRO, by obtaining an additional 2% improvement with respect to the previous system based on the original CRO meta-heuristic.

In [28], Carta et al. analyzed different feature selection methods within NNs as MCP approaches. MCP tries to predict the wind speed at a given point taking into account measurements in alternative (usually closely situated) sites. Thus, the features (input data) involved in the prediction problem are usually wind speed and direction in neighbour locations. In this case, filter and wrapper methods were tested as feature selection mechanisms, in a problem of mean hourly wind speeds and directions recorded in 2003 and 2004 at five weather stations in the Canary Islands. As mentioned, the input data were the wind speed and direction at different previous times in 4 measurement stations, and the objective is to obtain the wind speed prediction in the objective station. This work discussed the performance of a filter method for feature selection, based on correlation between input variables. In this case, the features selection worked by ranking the variables depending on their linear correlation coefficient, and keeping the less correlated ones in order to serve as inputs for the NN. In a second round of experiments, a wrapper approach for feature selection was investigated by applying an exhaustive search, which used the NN to provide the accuracy of each features set. A comparison with the case of no feature selection was carried out. The results obtained in the wind speed reconstruction in the Canary Islands, showed that the feature selection mechanisms leaded to an effective improvement. The results obtained were analyzed in terms of statistical tests, where it was shown that the feature selection methods were more effective when there is a low correlation between the objective and the input measurements sites. On the other hand, the statistical tests showed no improvements after applying feature selection if the objective measurement tower is highly correlated with the predictive towers.

In [29], Kong et al. presented a wind speed prediction system based on a Support Vector Regression approach. A modified version of the algorithm which selects a subset of data as support vectors and solves a smaller optimization problem (RSVR) was used. A problem of very short-term wind speed prediction (few minutes time-horizon) was tackled in this case. Wind speed and direction, temperature and pressure were the input variables. A PCA of the data was used as a feature extraction method to determine the most important factors affecting the wind speed. Note that, in this case, the SVR works on a different space (given by the PCA), so the initial input space should be transformed previously to the prediction. In general, feature extraction is possible, but less popular than feature selection, since the importance of each real feature cannot be evaluated using, for example, the PCA transform. This is a general issue with feature extraction methods, which impede the physical interpretation of the features. The evaluation of the proposed RSVR was carried out in data from a real wind power plant located in Neimenggu, China. Data from September 2009 were used, and the experiments consisted of comparing the performance of the RSVR with PCA against those by the SVR and a RBF network. It was shown that the performance of the RSVR with PCA is better than that of the comparing algorithms, improving over 8% the performance of the best counterpart.

Kumar and López, in [30], proposed a NARX approach to tackle a problem of wind speed prediction from meteorological series as input variables. NARX is a recurrent dynamic network, with feedback connections enclosing several layers of the network [31]. It has performed better than alternative NNs in highly non-linear problems. A feature selection was carried out previously to the prediction step by applying the ReliefF filter method. This method is a filter approach for feature selection based on estimating the quality of attributes according to how well their values distinguish between instances close to each other. In this case, a very short-term wind speed prediction problem from previously measured values (inputs) was tackled. Results in data from a meteorological tower at the University of Waterloo, Canada, were shown, though there was not a comparison with the performance of an alternative similar algorithm, what difficult its assessment. The results obtained showed that the feature selection applied was able to significatively improve the performance of the NARX algorithm over a 20%, compared to the case when no feature selection algorithm was considered.

In [32], Zhang et al. proposed two hybrid models which combine EMD, a feature selection process and machine learning regressors (NNs and SVRs), for a problem of short-term wind speed prediction. The idea is decomposing the original wind speed time series into a set of sub-series by applying the EMD method. Then, a feature selection process is introduced to identify the relevant and informative features for all the sub-series. A linear regression method completes this process, in which different subsets of features (sub-series serving as inputs for the wind speed estimation) were evaluated. The system was finally completed by applying a predictive algorithm (NN or SVR) to the final set of features obtained from the previous step. The performance of this system has been evaluated using real data from three different wind farms in China, located at Jiangsu, Ningxia and Yunnan. The complete system with EMD and feature selection was compared to the performance of the NN and SVR approaches, showing the advantages of the proposed approach, with improvements up to 34% when the EMD and feature selection was considered. Another work dealing with EMD and feature selection is due to Jiang and Huang [33], who presented a hybrid ensemble EMD approach with feature selection and error correction for a problem of short-term wind speed prediction. This system starts by decomposing the wind speed time series into a number of subseries by applying the EMD algorithm. Two feature selection mechanisms based on filter methods are then applied, including kernel density estimation-based Kullback-Leibler divergence and energy measure. After this process, two different SVR algorithms and an auto-regressive model are used to obtain the wind speed prediction, and also to correct the resulting error whether possible. Two different experiments in wind speed data from Colorado and Minnesota (USA) showed the effectiveness of this approach when compared with its different components on their own.

In [34], Zheng et al. merged PSO, GS and a ELM in a single hybrid algorithm for wind speed prediction. The meta-heuristics algorithms were applied in two different ways: a real version of both PSO and GS approaches were used to tune the parameters of the ELM, whereas a binary version of both algorithms tackled a FSP related to the wind prediction. The approach was tested in a now-casting (10 min prediction time-horizon) wind speed prediction problem at two locations of the NREL. A similar approach is presented by Zhang et al. in [35], where also a hybrid model ELM with different meta-heuristics is built. In this case, the authors use a BSA hybridized with an ELM. A real-valued version of the BSA is exploited to search for the optimal combination of weights and bias of the ELM, while a binary version of the algorithm is applied to obtain the best set of features for the prediction system. This hybrid model was tested in two different wind speed predicton problems: a half-hour wind speed observation data from two wind farms in Inner Mongolia (China), and now-casting (10-min. wind speed data) from the Sotavento Galicia wind farm.

Finally, Feng et al. in [36], developed a multi-model methodology based on the combination of machine learning techniques for a problem of short-term wind speed prediction. A NN, SVR, RF and Boosting machine were applied as an ensemble with a previous step of feature selection. A PCA together with a Granger causality test, auto-correlation analysis and recursive feature elimination were sequentially applied to reduce the number of features involved in the wind speed prediction. The proposed ensemble was able to provide a good quality wind speed prediction, including confidence intervals for the prediction. Comparison with alternative machine learning approach without feature selection showed the impact of the FSP in the wind speed prediction at seven locations of the SURFRAD network (USA).

3.2. Feature selection in solar energy

In [37], Fu and Cheng proposed a system for solar irradiance very short-term prediction (minutes time-horizon). The work used a solar irradiance prediction scheme with features extracted from all-sky images. The idea is to obtain proper features from all-sky images derived from an all-sky camera located in Taiwan, apply a feature extraction algorithm to the images, and then use a regression technique to predict a clearness index from them. In a second step, the clearness index was used to calculate the desired solar irradiance together with the extraterrestrial solar irradiance value, which only depends on astronomical variables. The features considered for the clearness index estimation were related to the all-sky images obtained: number of cloud pixels, where a RBR threshold method was applied to decide cloud pixels or not. Frame difference between images in t and $t-1$ moments, which is important to locate moving clouds. Gradient magnitude, which provides information corresponding to edges in an image. In this case, edges are often related to cloud boundaries in all-sky images. Intensity level, since in the all-sky images the brightness level of clear skies and cloudy skies are very different. Accumulated intensity along the vertical sun line. The idea of this feature is that if the sunlight is strong, the line caused by the sun would cross the entire image, whereas in cloudy situations, the vertical line of the sun would not cross the image. Number of corners, since corners in all-sky images correspond to the details of the clouds that have edges with at least two directions in a local patch of the sky. Using these features, the system applied a filter feature selection to different images consisting in ranking the features with a higher correlation to the clearness index. The most relevant features were then used in a linear regression system to estimate the clearness index from the image. The results in real data from all-sky images in Taiwan collected in August 2011 showed that the proposed system was an accurate tool to estimate short-term solar irradiation prediction locally. The experimental results showed an improvement in the short-term prediction of solar irradiance of about 4% in comparison to the estimation of the solar irradiance directly from weather variables.

In [38], Yadav et al carried out a study of the main influencing input parameters for solar radiation prediction with NNs in different locations of India, by using WEKA software. Different variables such as daily average temperature, minimum temperature, maximum temperature, altitude, sunshine hours and site location were considered. The study used a previous wrapper method with a regression tree implemented in WEKA to select the best set of features. After this process, several NNs implementations from WEKA were evaluated on this best set of features. Improvements over 13% were obtained after the FSP process (comparing the NNs without feature selection pre-processing) in some of the locations considered. Note that also in this case, the small number of features involved in the problems made an exhaustive search algorithm possible, which was implemented in WEKA software.

Wang et al. [39] applied a feature extraction method to select the best set of input parameters for a SVM classifier, in order to reconstruct a database of weather types. These weather types are directly connected with the photovoltaic power generation accuracy, so the most useful features were extracted from photovoltaic data, and they served as inputs for a SVM to reconstruct the database of weather types. This is a classification problem used as a previous step in the estimation of photovoltaic power generation prediction for buildings. Alternative classification problems and algorithms in renewable energy are described in [40].

In [41], Rana et al. forecasted the electricity power generation by a solar photo-voltaic system. Short-term prediction (from 5 to 60 min ahead) was considered. Input variables at different previous times were considered: solar irradiance, temperature, humidity and wind speed. A total of 4200 variables were finally available as inputs to predict the photo-voltaic generation in the next hour, in intervals of 5 min. In this case the correlation-based feature selection, which selects the best set of variables with higher correlation with the objective variable, was used as filter method. After this feature selection, two machine learning algorithms were applied to generate the system's prediction: an ensemble of NNs and a SVR approach. Experiments in power data collected from the St. Lucia campus of the University of Queensland in Brisbane, Australia, showed that the NN ensemble is able to outperform modestly the SVR, obtaining solutions around 1% better after applying the filter feature selection method proposed.

In [42] Mohammadi et al. applied an ANFIS system to select the most influential variables in a daily horizontal diffuse solar radiation prediction problem. Relevant variables were considered in order to study how different groups predict solar radiation: daily diffuse and global solar radiation on a horizontal surface, sunshine duration, minimum air temperature, maximum air temperature, average air temperature, relative humidity, water vapor pressure, daily maximum possible sunshine duration, solar declination angle and extraterrestrial solar radiation on a horizontal surface. Four years of measured data from the Iranian Meteorological Organization (January 2009 to December 2012) were used. Analysis of the best combination of 3 variables were carried out, showing the percentage of improvement when considering 1, 2 or 3 variables. The best result obtained (3 variables) was over 30% better than the solution of the system considering just 1 variable. No comparison results with alternative approaches were shown in the work, which makes difficult a further analysis of the proposal's performance.

In [43], Will et al. used a hybrid niching GA – linear regression approach to estimate global solar radiation in El Colmenar, Argentina. Data from 14 different weather stations were used in this work. Climatic variables such as daily average temperature, air humidity, atmospheric pressure, cloudiness, and sunshine hours were considered. The idea was to reconstruct the global solar radiation from the data in other 13 measurement stations. A niching GA with binary encoding was considered, where a 1 stands for including a given variable in the prediction, and a 0 stands for not including it. The linear regression was used due to its good computational complexity performance and its interpretability, though the authors admit that NNs could obtain better results. The prediction obtained was only compared with the niching GA with different number of individuals and generations, with the best prediction result obtained with 200 individuals and 150 generations, improving a 3% the solution of the algorithm with 50 individuals and 35 generations. A complete comparison with alternative approaches was not provided.

In [44], Aybar et al. applied a GGA to select the optimal set of features that maximizes the performance of an ELM for solar radiation prediction. As input variables, the system used the output of a numerical weather meso-scale model (WRF), i.e. variables predicted by the WRF at different nodes nearby the objective station. Experiments in real data from the Radiometric Observatory of Toledo (Spain) showed the good performance of this approach, in different prediction time-horizons, from hourly to 3 h ahead prediction. Comparison with the results of an ELM showed that the feature selection procedure was able to improve the performance of the system over a 10% in terms of RMSE. In a related approach, Salcedo et al. in [45], proposed a novel CRO with species algorithm for a FSP related to global solar energy prediction in Spain. In this case, the CRO with species was able to select the optimal number of input features for an ELM algorithm. This approach was compared against different alternative regressors, an ELM, a hybrid GA-ELM and also the GGA-ELM, in the global solar radiation Toledo's data, showing a significant improvement (over 21%) of performance versus those alternative approaches.

In [46], García et al. evaluated three non standard multivariate feature selection approaches for a problem of solar energy prediction in Spain. Novel methods which automatically select the most adequate features were compared, adapting strong regression algorithms such as SVMs or DNNs for feature selection. Comparison with classical feature extraction approaches such as PCA or Lasso was carried out. Performance improvements over 5% were reported when using the machine learning algorithms with feature selection versus the standard techniques.

Finally, note that Yadav et al., in [7], offered a first general review of some works dealing with relevant parameters selection in solar energy prediction problems, in a larger framework of solar energy prediction with NNs.

3.3. Feature selection in marine energy

There are not many works dealing with feature selection problems in marine energy applications, mainly because this is the least developed renewable energy technology among those revisited in this paper. Only very recent papers have been published on this topic. In [47], Hashim et al. presented an study focused on finding the sequence of the most influential parameters that affect the offshore significant wave height. This is a pure FSP problem in the search space of atmospheric and ocean variables related to significant wave height for marine wave-based energy systems. An ANFIS system with Takagi-Sugeno-based rules was chosen as the regressor mechanism for variable selection. Note that this type of systems produces interpretable rules which can be used to rank the importance of the variables in a given prediction problem. The data for this study, formed by five variables (significant wave height, sea surface wind speed, wind direction, air temperature and sea surface water temperature) were obtained from three NDBC buoys, located in the northern Atlantic ocean. The results obtained showed that the ANFIS was able to select a reduced number of features which provided the best significant wave height prediction.

In [48], Cornejo et al. proposed a novel procedure for feature selection based on a GGA. The GGA approach is able to group the different features affecting the problems in groups with specific behaviour in terms of prediction performance. In this case, the proposal consisted in a wrapper approach involving a GGA and an ELM network. Each group of features in the GGA was evaluated in terms of its accuracy for prediction by running the ELM. The fitness of the best group of features was used to guide the search towards the best overall set of features. The GGA-ELM system was applied to reconstruct significant wave height and wave energy flux from neighbour buoys in the west coast of California, USA (data from three buoys of the NDBC). The problem consisted in reconstructing the significant wave height and energy flux in a given buoy, from data collected in the other two buoys in the zone. The hybrid GGA-ELM showed better results than the ELM without feature selection, with an improvement over 6%, and also than alternative algorithms such as SVR approaches, where the inclusion of feature selection produced a much higher improvement, around 40%.

Finally, Cuadra et al. in [8] carried out a description of different computational intelligence methods in wave energy. In addition, this work presented a specific case study focused on the application of a hybrid CRO-ELM approach (wrapper for features selection) for significant wave height reconstruction, using data from NDBC buoys in the Gulf of Mexico. The results of the CRO-ELM were compared to those by a hybrid GA-ELM approach, finding solutions over 3% better when the CRO is used as global search approach for feature selection.

3.4. Feature selection in energy-related problems

The last works discussed in this review are focussed on problems which are related to energy, but are not direct applications for predicting the main renewable sources. For example, in [49], Ahila et al. classified power system disturbances with hybrid system including Extreme Learning Machines and PSO. In this case, the PSO approach was used to select the best features to serve as inputs of the classifier (ELM), and also the number of hidden nodes to enhance the performance of a multi-linear regression algorithm. Therefore, this system exploits a wrapper approach for classification disturbances in power systems. The experimental results showed that the proposed algorithm is faster and more accurate than alternative approaches in discriminating power system disturbances, producing improvements around 8% in classification accuracy. The database used included ten different power disturbances (10 classes classification problems), such as voltage sag, voltage swell, interruption, harmonics, flicker, oscillatory transient, sag with harmonics, swell with harmonics, notch and spike.

There are some recent works dealing with FSPs applied to electricity load forecasting. In [50] Koprinska et al. applied different filter feature selection methods to a problem of short-term electricity load from previous samples. The filter methods considered were based on autocorrelation of samples, Mutual Information, RReliefF and correlation-based techniques. The features selected were previously applied to different regressors such as NNs, model tree rules or linear regressors, in order to obtain accurate prediction of electricity loads. Two years of Australian electricity load data were used to show the goodness of the filter methods for feature selection, obtaining improvements between 3% and 4% depending on the method considered.

In [51], Jurado et al. applied different machine learning methodologies to forecast hourly energy consumption in buildings. In all cases, hybrid methodologies combining filters for feature selection (based on entropy measurements) with machine learning methods such as Fuzzy Inductive Reasoning, RF and NNs, were used. Experiments with actual data were carried out in Catalonia, Spain, where the different methods have been successfully compared with a traditional statistical technique (an ARIMA model for energy consumption prediction). The results showed that artificial intelligence methodologies outperformed classical techniques, by far when the feature selection step is considered, with improvements over 20%.

Hu et al., in [52], solved a problem of mid-term electricity loads prediction by using a MSVR approach and a memetic algorithm for feature selection. This is a wrapper proposal, where the SVR forecasts the electricity load, whereas the memetic algorithm looks for the best set of features for the problem. A memetic algorithm is a hybrid approach composed by a global search algorithm and a local search procedure to enhance the search. In this case, the global search algorithm was a firefly algorithm, and a local search based on a ranking of the best solutions was applied. The performance of this framework to predict daily interval electricity demands was tested in real data from North America and Australia, where the MSVR approach obtained important improvements (over 30%) over alternative algorithms based on NNs and classical SVR without feature selection mechanisms.

Finally, also related to the production of renewable energy in wind farms, Jiang et al. [53] tackled a problem of feature extraction in wind turbine fault diagnosis (vibration signal analysis of wind turbines). A denoising method based on adaptive Morlet wavelet and Singular Value Decomposition was applied to feature extraction, producing variables which are much more interpretable than the previous ones.

3.5. Discussion and remarks

As discussed above, the application of feature selection mechanisms improves the performance of machine learning prediction in renewable energy-related problems. The improvement fully depends on the specific application considered, but could be very important, in the range 5–40% of the prediction accuracy, just by including the feature selection algorithm. In all cases, feature selection implies an extra computational burden for the prediction systems. Filter methods, which use an external measure previous to the application of the prediction algorithm are the less computationally demanding approaches. On the other hand, wrapper approaches may suppose a heavy burden for the algorithm, and they are usually combined with fast training prediction approaches. This can be seen in Table 1, which shows the most important articles applying feature selection and related methods to renewable energy problems. A first analysis of the table indicates that wrapper approaches are most used, mainly in wind energy applications. Filter feature selection methods have had some impact in solar and energy-related applications. However, since they usually have less accuracy than wrapper approaches, they are also less used. The second interesting conclusion which can be drawn from this table is the large amount of algorithms used in wrapper FSP. Regarding prediction approaches, it is interesting that fast-trained approaches are more used, as previously suggested. In this sense, there are several works which bet for linear regression approaches, though in the last years, fast-trained neural networks like ELMs have been used more frequently.

Another important aspect to be discussed is related to the global search algorithms for wrapper feature selection. As can be seen in Table 1 there is a large amount of techniques applied, such as different types of GAs, PSO, DE, HS, CRO, greedy approaches or exhaustive search. The latter is only possible in those FSP where the number of features involved is small. In problems with a large amount of features, the search space size is huge (it grows at least as 2^N, where N is the number of features considered), so exhaustive search is not a computational viable option. This is why meta-heuristic approaches are mainly considered for wrapper feature selection systems. As Table 1 reveals, the number and characteristics of meta-heuristics approaches

Table 1
Summary of the most important articles applying feature selection methods in energy applications.

Reference	FSP type	Predictor	FSP algorithm
Wind energy			
[21]	wrapper	NN	PSO
[22]	wrapper	NN, K-NN	PSO,DE
[23]	wrapper	NN	GA
[24]	wrapper	GP	Greedy
[25]	wrapper	ELM	CRO
[27]	wrapper	ELM	CRO (HS)
[28]	filter	NN	Linear Correlation Coefficient
[28]	wrapper	NN	Exhaustive search
[29]	feature-extraction	SVR, RSVR	PCA
[30]	filter	NARX	ReliefF
[32]	wrapper	NN, SVR	Linear Regression
[33]	filter	SVR	Kullback-Leibler divergence
[34]	wrapper	ELM	PSO, GS
[35]	wrapper	ELM	BSA
[36]	feature-extraction	NN,SVR,RF	PCA, recursive elimination
Solar energy			
[37]	feature-extraction	Linear Regression	Image Processing techniques
[38]	wrapper	NN	Exhaustive search
[39]	filter	SVM	correlation-based
[41]	filter	NNs, SVR	correlation-based
[42]	Intrinsic	ANFIS	ANFIS
[43]	wrapper	Linear regression	niching GA
[44]	wrapper	ELM	GGA
[45]	wrapper	ELM	CRO
[46]	Filter	DNN, SVR	correlation-based
Marine energy			
[8]	wrapper	ELM	CRO
[47]	Intrinsic	ANFIS	ANFIS
[48]	wrapper	ELM	GGA
Energy-related applications			
[49]	wrapper	ELM	PSO
[50]	filter	Autocorrelation, MI, RReliefF	NN, Linear regression
[51]	filter	Entropy-based	NN, RF
[52]	wrapper	MSVR	Firefly, Local search
[53]	feature-extraction	wavelets	–

used in FSPs in large, and, moreover, it is difficult to decide what technique is better, since there is not an unified framework for FSP, and each problem/application have different properties, so it fully depends on the specific type of FSP tackled.

The choosing of a given meta-heuristic as global search approach in wrapper FSPs is therefore a problem itself, and usually researchers try specific approaches without a clear idea whether it is the best option. In order to alleviate this situation, in this paper we explore the idea of using several global search mechanisms at the same time, in such a way that the combination of search procedures leads to better solutions than applying a single, sometimes wrongly chosen, search pattern. We do this by using a co-evolution approach with special characteristics: the CRO-SL algorithm [54], which enables the combination of different search procedures in a single population, with a controlled computational burden. In the next section we show a case study of the application of the CRO-SL algorithm with several global search approaches to a wrapper feature selection problem in wind energy, and will show the improvement in performance of this co-evolution approach.

4. Case study: FSP in wind speed prediction systems with the CRO-SL algorithm

This section presents a case study where we discuss a novel prediction approach for a FSP in wind energy. Specifically, the problem consists of forecasting wind speed for a wind farm, considering input variables from a numerical meso-scale weather model. The new system (Fig. 2) consists of a hybrid approach, mixing a meta-heuristic algorithm for feature selection (CRO-SL) with a ELM for prediction. It is, therefore, a wrapper approach for the FSP. After the FSP, the wind speed prediction is obtained by means of another ELM, trained only with the selected variables after the FSP. We structure this case study as follows: first we describe the CRO-SL algorithm and the ELM main characteristics. The specific FSP in data from a real wind farm in Spain is then presented. The computational experiments and results obtained are finally discussed.

4.1. The Coral reefs optimization algorithm with substrate layer

The original CRO algorithm [55,56] is a recently proposed class of evolutionary algorithm, based on the main processes of coral reproduction and reef formation that occur in nature. The algorithm maintains a population of possible solutions of a given optimization problem (corals), which will be evolved by applying different evolution mechanisms (operators), based on the real processes occurring in a coral reef. The basic CRO algorithm can be improved by considering very specific interactions in the reef. For example, different studies have shown that successful recruitment in coral reefs (i.e., successful settlement and subsequent survival of larvae) depends on the type of substrate on which they fall after the reproduction process [57]. This specific characteristic of the coral reefs was first included in the CRO in [58], in order to solve different instances of the Model Type Selection Problem for energy applications. The CRO-SL is, however, a much more general approach, and it can be defined as an algorithm for competitive co-evolution, where each substrate layer represents different searching or exploration processes, such as defined in [54]. This approach has been successfully applied to solve a number of difficult optimization problems recently [59–61]. Since we use here the CRO-SL as searching mechanism for the optimal set of variables, we first describe in detail this evolutionary-type algorithm.

The basic CRO process for searching follows the steps given in Algorithm 1.

Algorithm 1. Pseudo-code for the CRO algorithm.

Require Valid values for the parameters controlling the CRO algorithm
ENSURE A single feasible individual with optimal value of its *fitness*
Initialize the algorithm
for each iteration of the simulation
Update values of influential variables: predation probability, etc.
Sexual reproduction processes (broadcast spawning and brooding)
Settlement of new corals
Predation process
Evaluate the new population in the coral reef
end for
Return the best individual (final solution) from the reef

We detail here the specific definition of the different searching operators that form the classical CRO algorithm:

1. **Sexual reproduction**: The CRO model implements two different kinds of sexual reproduction: external and internal.
   (a) **External sexual reproduction** or *broadcast spawning*: the corals eject their gametes to the water, from which male-female couples meet and combine together to produce a new larva by sexual crossover. In Nature, some species are able to combine their gametes to generate mixed polyps even though they are different from each other. In the CRO algorithm, external sexual reproduction is applied to a usually high fraction F_b of the corals. The couple selection can be done uniformly at random or by resorting to any fitness proportionate selection approach (e.g. roulette wheel). In the original version of the CRO, standard

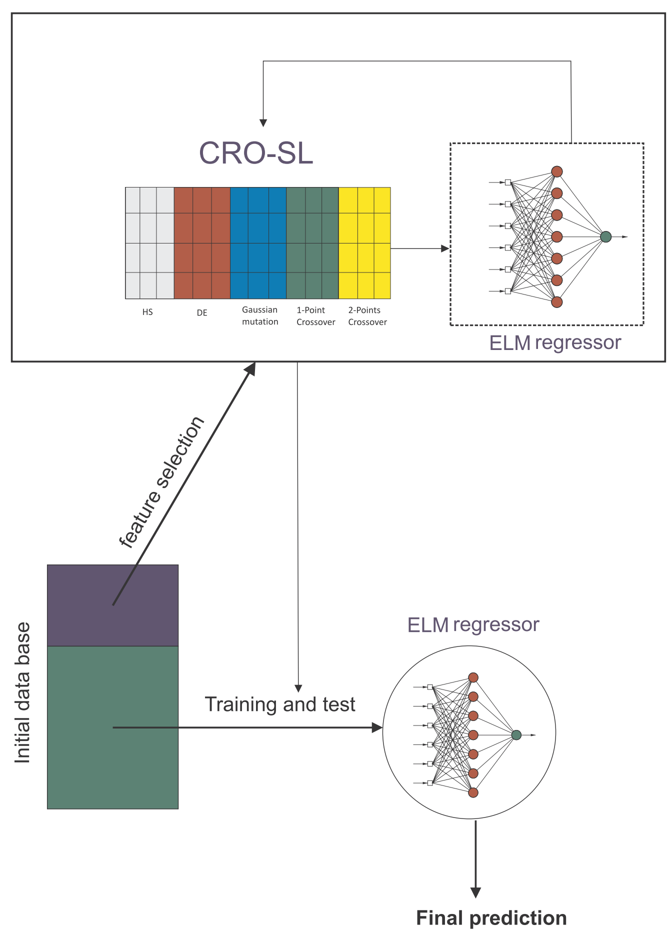

Fig. 2. Hybrid CRO-SL-ELM system for wind speed prediction with feature selection.

crossover (one point or two-points) are applied in the broadcast spawning process.

(b) **Internal sexual reproduction** or *brooding*: CRO applies this method to a fraction $(1 - F_b)$ of the corals in the reef. The brooding process consists of the formation of a coral larva by means of a random mutation of the brooding-reproductive coral (self-fertilization considering hermaphrodite corals). The produced larvae are then released out to the water in a similar fashion than that of the larvae generated through broadcast spawning.

2. **Larvae settlement**: once all larvae are formed at iteration k through reproduction, they try to settle down and grow in the reef. Each larva will randomly attempt at setting in a square (i, j) of the reef. If the location is empty (free space in the reef), the coral grows therein no matter the value of its health function. By contrast, if another coral is already occupying the square at hand, the new larva will set only if its health function is better than the fitness of the existing coral. We define a number of attempts $\mathcal{N}_{att}$ for a larva to set in the reef: after $\mathcal{N}_{att}$ unsuccessful tries, it will not survive to following iteration.
3. **Depredation**: corals may die during the reef formation phase of the reef. At the end of each iteration, a small number of corals can be preyed, thus liberating space in the reef for the next iteration. The depredation operator is applied under a very small probability P_d, and exclusively to a fraction F_d of the worse health corals.

In order to extend this basic CRO to form the CRO-SL approach, we consider different *substrate layers*, which in this case are different searching mechanisms, depending on the reef zone. The inclusion of the substrate layers in the CRO can be done, in a general way, in a straightforward manner: we redefine the artificial reef considered in the CRO in such a way that each cell of the square grid Ψ representing the reef is now defined by 3 indexes (i, j, t), where i and j stand for the cell location in the grid, and $t \in T$ defines the substrate layer, by indicating which structure (searching operator in this case) is associated with the cell (i, j). Each coral in the reef is then processed in a different way depending on the specific substrate layer where it falls after the reproduction process. Specifically, we have considered four well-known search operators in the CRO-SL, well adapted to the FSP in wind speed prediction tackled: a two-points crossover, a multi-point crossover, a Harmony Search and finally a Differential Evolution-based search operator:

- **Two points crossover** (2PX): 2PX [62] is one of the most used recombination/crossover operators in evolutionary algorithms. In the standard version of the operator, two parents from the reef population, randomly chosen, are provided as input. A recombination operation is then carried out by randomly choosing two crossover points, and interchanging each part of the corals (solutions) between those points.
- **Multi-points crossover** (MPX): Similar to the 2PX, but in this case the recombination between the parents is carried out considering a high number of crossover points (M), and a binary template which indicates whether each part of one parent is interchanged with the corresponding of the other parent.
- **Differential Evolution-based operator** (DE): This operator is based on the evolutionary algorithm in [63], a method with powerful global search capabilities. DE introduces a differential mechanism to explore the search space, which may be exploited in the CRO-SL as part of the different operators applied to promote the evolution of the solutions reef.
- **Harmony Search-based operator** (HS): Harmony Search [64] is a population based meta-heuristic that mimics the improvisation of a music orchestra while it is composing a melody. This method integrates concepts such as harmony aesthetics or note pitch as an analogy for the optimization process, resulting in a good exploratory algorithm, specially well-suited for integer encodings, such as the feature selection tackled in this paper.

A prediction procedure is necessary in the adaptation of the CRO-SL to solve a FSP for wind speed prediction, in order to evaluate the quality of the feature selected for every coral (solution) in the CRO-SL. In this case, the prediction procedure considered is a neural network trained by means of an ELM algorithm, described in the next subsection.

4.1.1. The extreme learning machine

An ELM [26] is a fast learning method based on the structure of MLPs, with a novel way of training feed-forward neural networks. One of the most important characteristics of the ELM training is the randomness in the process where the network weights are set, obtaining, in this way, a pseudo-inverse of the hidden-layer output matrix. The simplicity of this technique makes the training algorithm extremely fast. Moreover, it has an outstanding performance when compared to other learning methods, including classical MLPs training [65]. Furthermore, the ELM presents universal approximation capability, as well as classification capability, as proven in [66].

The ELM algorithm can be explained as follows: given a training set

$$\mathbb{T} = (\mathbf{x}_i, \upsilon_i) | \mathbf{x}_i \in \mathbb{R}^n, \upsilon_i \in \mathbb{R}, i = 1, \ldots, l,$$

an activation function $g(x)$ and number of hidden nodes $(\widetilde{N})$,

1. Randomly assign inputs weights $\mathbf{w}_i$ and bias b_i, $i = 1, \ldots, \widetilde{N}$.
2. Calculate the hidden layer output matrix $\mathbf{H}$, defined as

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1\mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_{\widetilde{N}}\mathbf{x}_1 + b_{\widetilde{N}}) \\ \vdots & \cdots & \vdots \\ g(\mathbf{w}_1\mathbf{x}_l + b_1) & \cdots & g(\mathbf{w}_{\widetilde{N}}\mathbf{x}_N + b_{\widetilde{N}}) \end{bmatrix}_{l \times \widetilde{N}} \tag{2}$$

3. Calculate the output weight vector β as

$$\beta = \mathbf{H}^{\dagger}\mathbf{T}, \tag{3}$$

where $\mathbf{H}^{\dagger}$ stands for the Moore-Penrose inverse of matrix $\mathbf{H}$ [26], and $\mathbf{T}$ is the training output vector, $\mathbf{T} = [\upsilon_1, \ldots, \upsilon_l]^T$.

The number of hidden nodes $(\widetilde{N})$ is a free parameter of the ELM training, and it can be fixed initially, or in a best convenient way, it must be estimated for obtaining good results as a part of a validation set in the learning process. Hence, scanning a range of $\widetilde{N}$ values is the solution for this problem. The Matlab ELM implementation by G. B. Huang, freely available in the Internet [67] has been used.

4.2. FSP tackled: wind data and predictive variables

As previously mentioned, here the FSP deals with the wind speed prediction in a wind farm, using predictive variables coming from a numerical weather model. Therefore, we consider real wind data from a wind farm located in northern Spain, Fig. 3. Ten years of data are considered, from 11/01/2002 to 29/10/2012, 80% of them will be used to train the prediction system, whereas the remainder 20% will be used for test purposes. Regarding predictive variables, a total of 98 meteorological variables serve as initial input parameters. All the predictive variables have been obtained from a meso-scale model (WRF) [68], in a node which corresponds to the wind farm situation. The WRF is a powerful meso-scale numerical weather prediction system designed for atmospheric research and also for operational forecasting needs. It was developed in collaboration by the NCAR, NCEP, FSL, AFWA, the Naval Research Laboratory, the University of Oklahoma, and the FAA of the USA. Note that there is a large number of variables coming from the WRF model, which are expected to provide detailed information to obtain a good prediction of the wind speed. The CRO-SL-ELM is then applied as a system to obtain a wind speed prediction from these

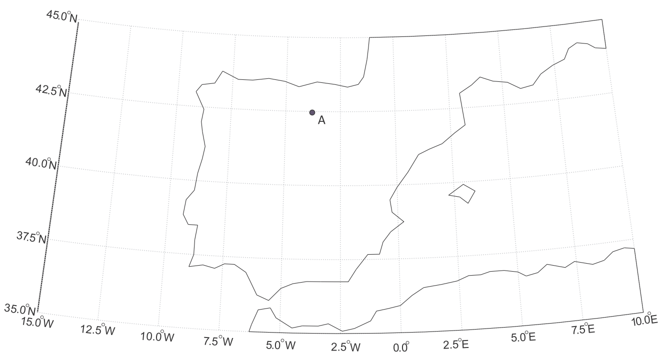

Fig. 3. Wind farm considered for the experiments.

variables. Note that the system tries to obtain the relationship between the predictors and the objective real wind speed in the wind farm. Table 2 summarizes the complete set of variables from the WRF mesoscale model considered. Note that there are 49 predictive variables in the table, and we also consider the moving average in a period of 24 h (*ma*) of all the variables, so a final set of 98 predictors is considered. Among the variables considered, we have chosen those which may have direct relationship with the wind speed in the objective site, for example wind speed and direction in the WRF nodes at different levels, pressure, temperatures at different levels, etc. We also consider some indirect measurements which are displayed in the table as $g(x)$, where g stands for a function such as logarithmic or exponential, and x stands for the corresponding direct predictive variable.

We carry out different experiments and discussions, including the experimental evaluation of different objective functions to test the quality of the CRO-SL solutions, and also evaluate different prediction time-horizons (hourly and daily wind speed prediction), in terms of the features selected by the algorithm. Regarding this latter point, the ELM is trained with an hourly prediction time horizon, and then a daily prediction from the average of hourly results is also obtained. Thus, we use different objective functions to evaluate the features selected by the CRO-SL:

$$f_1(\mathbf{x}) = r_h^2, \tag{4}$$

This first objective function only considers the Pearson coefficient r^2 for the wind speed hourly prediction.

$$f_2(\mathbf{x}) = 0.4 \cdot r_h^2 + 0.6 \cdot r_d^2, \tag{5}$$

The second objective function considers both the hourly and daily prediction time-horizons, using some weights to combine both predictions. The last objective function does the same, but with a small weight in the wind speed hourly component:

$$f_3(\mathbf{x}) = 0.1 \cdot r_h^2 + 0.9 \cdot r_d^2, \tag{6}$$

4.3. Results

As initial reference values we show in Table 3 the results obtained by an ELM and a MLR algorithm [69] (used as baseline approach), with the 98 input variables, i.e. without any feature selection mechanism. As can be seen, the hourly prediction of the wind speed with these methods from numerical weather variables is good, obtaining correlation coefficient of 0.69 and 0.68, respectively. Table 4 shows the performance of the hybrid CRO-SL-ELM algorithm. A comparison with a CRO-ELM approach similar to that in [25] and with a CRO-MLR and CRO-SL-MLR is also shown in this table. It is possible to see that the CRO-SL-ELM is able to improve the accuracy of the ELM on its own. This indicates that a process of feature selection is positive for improving the prediction capability of the system. Regarding the comparison with the CRO, Table 4 shows that the system with the CRO-SL obtains better results than the system with the basic CRO. It is an indication that the CRO-SL is able to better explore the search space. The results obtained also show that the hybrid CRO-SL-ELM outperforms CRO-MLR and CRO-SL-MLR, i.e. the ELM works in general better than the MLR as final prediction mechanism. The feature selection process is also positive when the MLR was used as final regression, improving the performance of the MLR on its own. Note that the solution provided by the basic CRO included 52 features out of the initial 98, but the CRO-SL provided a better solution with 25 features. Table 5 shows the best set of features according to the CRO-SL with f_1 objective function.

The effect of including several objective functions with different weights for hourly and daily prediction (see Equations ((4) to (6))), can be evaluated by analyzing Table 6. Note that in daily prediction the inclusion of a feature selection mechanism also improves the performance of the prediction system (see also Table 3 to compare the results

Table 2
Complete set of predictive features from the WRF model for wind speed prediction.

Variable	Name
wind speed(10 m)	ϑ_{10}
wind direction(10 m)	θ_{10}
wind speed(20 m)	ϑ_{20}
wind direction(20 m)	θ_{20}
wind speed(50 m)	ϑ_{50}
wind direction(50 m)	θ_{50}
wind speed(100 m)	ϑ_{100}
wind direction(100 m)	θ_{100}
wind speed(200 m)	ϑ_{200}
wind direction(200 m)	θ_{200}
wind speed(500hPa)	ϑ_{500hPa}
wind direction(500hPa)	θ_{500hPa}
temperature(0 m)	T_0
temperature(2 m)	T_2
temperature(20 m)	T_{20}
temperature(50 m)	T_{50}
specific humidity (2 m)	q
pressure(0 m)	P
long wave down(0 m)	Iw_d
short wave down(0 m)	Sw_d
precipitation (0 m)	p
wind speed(80 m)	ϑ_{80}
wind direction(80 m)	θ_{80}
power(80 m)	P_w
ϑ_{10}^2	ϑ_{10}^2
ϑ_{10}^3	ϑ_{10}^3
$g(T_{20}/T_{50})$	$g_{T_{20/50}}$
$g(T_2/T_{50})$	$g_{T_{2/50}}$
$g(T_0/T_{50})$	$g_{T_{0/50}}$
$g(\vartheta_{500}/\vartheta_{50})$	$g_{\vartheta_{500/50}}$
$g(\vartheta_{500}/\vartheta_{20})$	$g_{\vartheta_{500/20}}$
$g(\vartheta_{500}/\vartheta_{10})$	$g_{\vartheta_{500/10}}$
$g(\vartheta_{50}/\vartheta_{20})$	$g_{\vartheta_{50/20}}$
$g(\vartheta_{50}/\vartheta_{10})$	$g_{\vartheta_{50/10}}$
$g(\vartheta_{20}/\vartheta_{10})$	$g_{\vartheta_{20/10}}$
u_{10}	u_{10}
v_{10}	v_{10}
u_{20}	u_{20}
v_{20}	v_{20}
u_{50}	u_{50}
v_{50}	v_{50}
u_{100}	u_{100}
v_{100}	v_{100}
u_{200}	u_{200}
v_{200}	v_{200}
u_{500hPa}	u_{500hPa}
v_{500hPa}	v_{500hPa}
u_{80}	u_{80}
v_{80}	v_{80}

Table 3
Results of the hourly and daily wind speed estimation by the ELM and MLR with all features considered (98).

	RMSE [m]	MAE [m]	r^2
hourly			
ELM	1.7523	1.3309	0.6915
MLR	2.1102	1.6652	0.6831
daily			
ELM	0.4867	0.3588	0.9033
MLR	1.3083	1.1873	0.8888

Table 4
Comparative best results of the hourly wind speed prediction by the ELM and MLR, with different fitness functions in the CRO and CRO-SL algorithms.

	CRO			CRO-SL		
FITNESS	RMSE [m]	MAE [m]	r^2	RMSE [m]	MAE [m]	r^2
ELM						
$f_1(\mathbf{x})$	1.7245	1.3192	0.6990	1.6714	1.2686	**0.7199**
$f_2(\mathbf{x})$	1.7316	1.3246	0.6938	1.6729	1.2813	0.7167
$f_3(\mathbf{x})$	1.7181	1.3225	0.6976	1.6604	1.2707	0.7178
MLR						
$f_1(\mathbf{x})$	1.7970	1.3729	0.6880	1.7352	1.3265	0.6951
$f_2(\mathbf{x})$	1.7778	1.3716	0.6890	1.7301	1.3322	0.6987
$f_3(\mathbf{x})$	1.7695	1.3555	0.6926	1.7229	1.3207	0.7016

Table 5
Best set of features selected by the CRO-SL ($f_1(\mathbf{x})$, 25 features).

Feature	Variable
1	ϑ_{10}
2	θ_{10}
5	ϑ_{50}
7	ϑ_{100}
12	θ_{500hPa}
13	T_0
14	T_2
28	$g_{T_{2/50}}$
29	$g_{T_{0/50}}$
31	$g_{v_{500/20}}$
39	v_{10}
40	u_{50}
41	v_{50}
42	u_{100}
43	v_{100}
45	v_{200}
48	u_{80}
49	v_{80}
54	$ma(\vartheta_{50})$
56	$ma(\vartheta_{100})$
62	$ma(T_0)$
86	$ma(v_{10})$
88	$ma(v_{20})$
90	$ma(v_{50})$
98	$ma(v_{80})$

without feature selection). In this case the proposed CRO-SL-ELM system is able to obtain an accuracy of 0.92, improving the 0.90 of the ELM without feature selection procedure. In spite of this improvement in prediction performance, note that f_1 is again the objective function which produces the best results in terms of daily wind speed prediction, which indicates that objective functions f_2 and f_3 (which included specific daily prediction terms), are not able to improve the prediction in this specific problem.

Table 6
Comparative best results of the daily wind speed estimation by the ELM and MLR, with different fitness in the CRO-ELM and CRO-SL-ELM algorithm.

	CRO			CRO-SL		
FITNESS	RMSE [m/s]	MAE [m/s]	r^2	RMSE [m/s]	MAE [m/s]	r^2
ELM						
$f_1(\mathbf{x})$	0.4806	0.3649	0.9170	0.4675	0.3658	**0.9271**
$f_2(\mathbf{x})$	0.5018	0.3826	0.9138	0.4505	0.3241	0.9203
$f_3(\mathbf{x})$	0.5175	0.3751	0.9027	0.4740	0.3627	0.9236
MLR						
$f_1(\mathbf{x})$	0.6679	0.4744	0.8867	0.5585	0.3769	0.8913
$f_2(\mathbf{x})$	0.6289	0.4566	0.8930	0.5676	0.3939	0.8980
$f_3(\mathbf{x})$	0.6523	0.4553	0.8768	0.5648	0.4012	0.9006

Another comparison is carried out by considering the best solutions found by the CRO-ELM and CRO-SL-ELM algorithms, and use them in alternative regression mechanisms. In this case, we consider the SVR [70] and the MLP neural network, trained with the Levenberg-Marquardt algorithm [65], as final regression mechanisms. We have chosen these regressors since they have also provided good results in alternative problems. Table 7 shows the performance of both approaches using the solutions from the CRO-ELM and CRO-SL-ELM. As can be seen, the SVR is not able to improve the performance of the ELM in the final prediction. Moreover, it seems that the features selected by the CRO-ELM degrade the SVR performance significantly. The SVR using as inputs the solution by the CRO-SL-ELM performs better, but it is still not able to improve the prediction of the ELM as final regressor. This behaviour of the SVR can be determined by the number of features, since the CRO-ELM obtains a solution with 52, whereas the best solution by the CRO-SL-ELM has 25. The MLP has a better behaviour in this problem, producing better results than the SVR. The application of the feature selection mechanism seems to be positive in this case, and the MLP results improve when the CRO or the CRO-SL algorithms for FSP are previously applied. In this case, the system with the MLP as final regressor obtains better results than using the MLR, but it is worse than the CRO-SL-ELM previously analyzed.

Fig. 4 shows the scatter plot with and without feature selection (ELM and CRO-SL-ELM). This figure can be complemented with Fig. 5, which shows the temporal performance of the predictions with and without feature selection mechanism. In general both predictions are quite accurate, which shows the good performance of the ELM as regressor/predictor. Note that it is possible to detect differences due to the feature selection process, such as better wind speed peaks and valleys reconstruction, which produce the improvement in the prediction accuracy obtained with the CRO-SL-ELM algorithm.

Table 7
Comparative best results of the hourly wind speed estimation by a SVR and MLP regressors, with all features, CRO-ELM and CRO-SL-ELM, considering fitness function f_1.

	RMSE [m/s]	MAE [m/s]	r^2
SVR (All features)	2.8174	2.1308	0.1954
SVR (FSP CRO)	2.6290	1.9730	0.3966
SVR (FSP CRO-SL)	1.6908	1.2845	0.7112
MLP (All features)	1.8261	1.4070	0.6680
MLP (FSP CRO)	1.7436	1.3295	0.6913
MLP (FSP CRO-SL)	1.7084	1.3034	0.7013

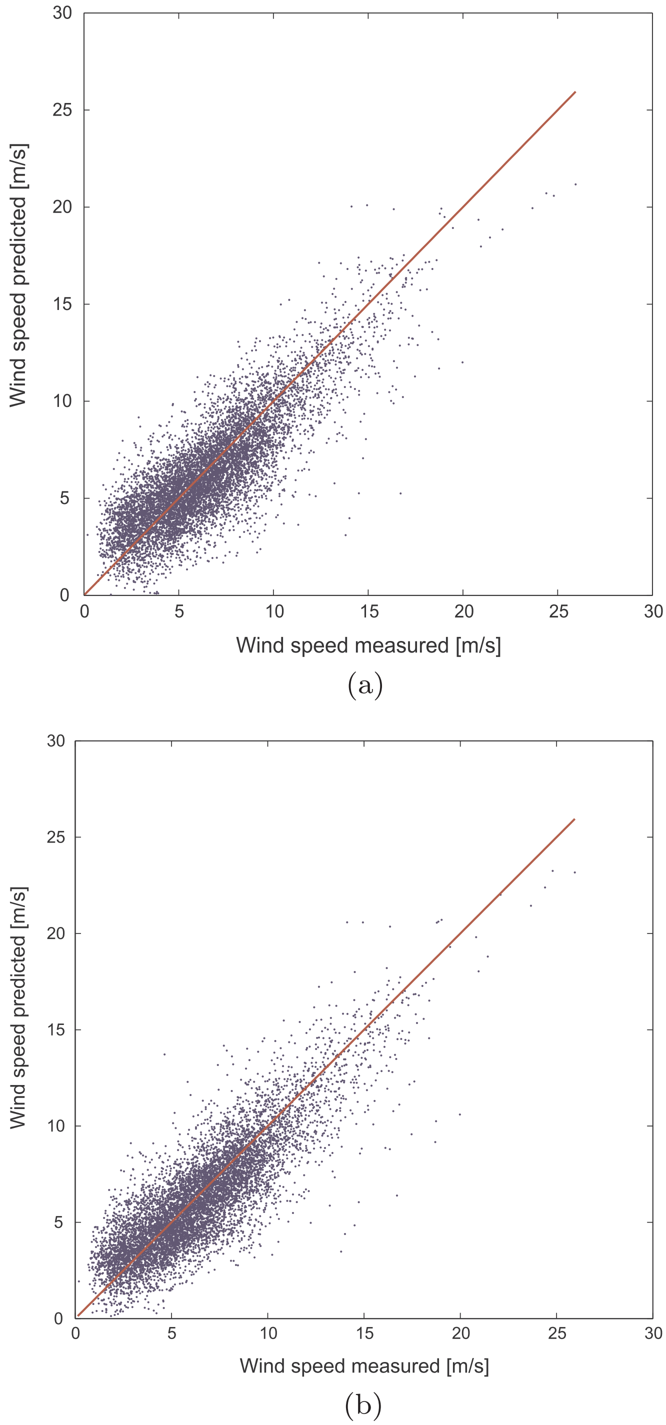

Fig. 4. Scatter plots of the wind speed hourly estimation by the ELM method for the test data set: (a) without feature selection; (b) with CRO-SL for the selection.

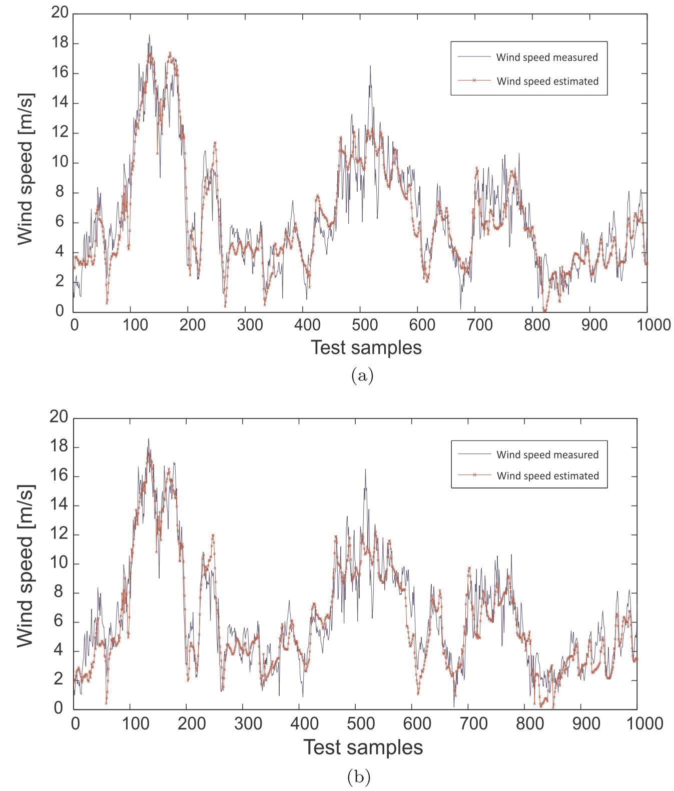

Fig. 5. Temporal evolution of the wind speed hourly estimation by the ELM method for the test data set: (a) without feature selection; (b) with CRO-SL for the selection.

5. Discussion and recommendations

We have analyzed a novel wrapper approach based on a multi-operator evolutionary-type algorithm (the CRO-SL) hybridized with a fast training neural network (the ELM). The robustness of the proposed approach allows including different objective functions in the CRO-SL, in order to guide the FSP. With this, we have evaluated the possibility of improving the ELM prediction performance in different time-horizons (hourly and daily in this case). We have shown that is it possible to do so, with better improvement in the hourly time-horizon prediction. The comparison with alternative final regression has shown that the feature selection mechanism may affect their performance differently: we have analyzed the case of a SVR where the features obtained by the CRO-SL-ELM degrade the performance. On the contrary, the performance of an MLP seems to be enhanced by the feature selection pre-processing step. Thus, the wrapper approach performance may be very dependent on the final prediction mechanism applied. A very important point to be noted is that a fast-training prediction algorithm is needed for the feature selection step, when the wrapper methodology is selected. Since a global search procedure is involved in this process, the faster the prediction algorithm training, the better the search carried out, since more generations or search cycles could be considered. In this sense, the ELM is really well fitted for this task, since it is extremely fast in its training sequence, and in addition its prediction performance is very competitive. Alternative approaches related to filter feature selection mechanisms are known to offer a poorer performance than their wrapper FSP counterparts in prediction problems. In particular cases with a huge number of very correlated features, hybrid filter-wrapper approaches may work, using a first step with a filter mechanism to remove the most related features. Then a wrapper similar to the one discussed in this paper could be applied.

6. Conclusions

This paper reviews the most important characteristics and applications of FSPs in renewable energy prediction problems. First, the definition of the FSP is given, and its main applications in different prediction problems (associated with the most important renewable energy sources such as wind, solar and marine) are discussed. The analysis carried out have shown a large number of different algorithms used in FSP as predictors, and also in terms of the search techniques to specify the best set of features selected. In this paper, we also propose a novel algorithm for the FSP which combines the searching potential of different approaches. It is the Coral Reefs Optimization algorithm with

Substrate Layer, which is combined with an ELM for regression. We discuss the performance of this algorithm in a FSP related to wind energy, where the input variables come from a numerical weather model, whereas the objective variables is the wind speed measured in a real wind farm in Spain. Another novelty of this approach is that we have included three different objective functions to guide the search for the best set of features, which combine different prediction time-horizons, from hourly to daily. Another novelty of the proposed approach is that we have included three different objective functions to guide the search for the best set of features, which combine different prediction time-horizons, from hourly to daily. The results obtained have shown that the proposed approach is able to obtain an excellent prediction of the wind speed at different prediction-time horizons, and the feature selection part of the algorithm works well, by improving the performance of the final regressor. Future lines of research in feature selection aim to explore novel techniques related to deep learning in large FSP problems. The use of techniques such as *auto-encoders* could be of help in FSP-related problems. Also, the exploration of novel hybrid approaches with fast regressors and accurate search approaches is a promising field to be explored, mainly in the design of high-quality wrapper FSP algorithms.

Acknowledgements

This work has been partially supported by Comunidad de Madrid, under project number S2013/ICE-2933, and by project TIN2014-54583-C2-2-R of the Spanish Ministerial Commission of Science and Technology (MICYT).

References

[1] Ren 21. Renewable 2016 Global Status Report, http://www.ren21.net/; 2016 [Accessed 21 March 2018].
[2] Dvorák P, Martinát S, Van der Horst D, Frantál B, Turecková K,. Renewable energy investment and job creation; a cross-sectoral assessment for the Czech Republic with reference to EU benchmarks. Renew Sustain Energy Rev 2017;69:360–8.
[3] Rafindadi AA, Ozturk I. Impacts of renewable energy consumption on the German economic growth: evidence from combined cointegration test. Renew Sustain Energy Rev 2017;75:1130–41.
[4] Thapar S, Sharma S, Verma A. Economic and environmental effectiveness of renewable energy policy instruments: best practices from India. Renew Sustain Energy Rev 2016;66:487–98.
[5] Jung J, Villaran M. Optimal planning and design of hybrid renewable energy systems for microgrids. Renew Sustain Energy Rev 2017;75:180–91.
[6] Costa A, Crespo A, Navarro J, Lizcano G, Madsen H, Feitosa E. A review on the young history of the wind power short-term prediction. Renew Sustain Energy Rev 2008;12:1725–44.
[7] Yadav AK, Chandel SS. Solar radiation prediction using Artificial Neural Network techniques: a review. Renew Sustain Energy Rev 2014;33:772–81.
[8] Cuadra L, Salcedo-Sanz S, Nieto-Borge JC, Alexandre E, Rodríguez G. Computational intelligence in wave energy: comprehensive review and case study. Renew Sustain Energy Rev 2016;58:1223–46.
[9] Blum A, Langley P. Selection of relevant features and examples in Machine Learning. Artif Intell 1997;97:245–71.
[10] Weston J, Mukherjee S, Chapelle O, Pontil M, Poggio T, Vapnik V. Feature selection for SVMs. NIPS'00 In: Proceedings of the 13th International Conference on Neural Information Processing Systems; 2000; Denver, CO. MIT Press Cambridge; 2000.
[11] John G, Kohavi R, Pfleger K. Irrelevant features and the subset selection problem. In: Proceedings of the 11th International Conference on Machine Learning; 1994 July 10-13; New Brunswick, NJ; 1994.
[12] Kohavi R, John G. Wrappers for features subset selection. Int J Dig Libr 1997;1:108–21.
[13] Yang J, Honavar V. Feature subset selection using a genetic algorithm. IEEE Intell Syst 1998;13:44–9.
[14] Salcedo-Sanz S, Prado-Cumplido M, Pérez-Cruz F, Bousoño-Calzón C. Feature selection via genetic optimization. In: Proceedings of the International Conference on Artificial Neural Networks, ICANN2002; Aug 2002; Madrid, Spain. Lecture Notes in Computer Science. Springer-Verlag. p. 547–52.
[15] Torkkola K, Campbell WM. Mutual information in learning feature transformations. In: Proceedings of the 17th International Conference on Machine Learning; 2000; San Francisco, CA. p. 1015—22.
[16] Torkkola K. On feature extraction by mutual information maximization. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing; 2002; p. 821–4.
[17] Solorio-Fernández S, Carrasco-Ochoa JA, Martínez-Trinidad JF. A new hybrid filter-wrapper feature selection method for clustering based on ranking. Neurocomputing 2016;214:866–80.
[18] Huda S, Abdollahian M, Mammadov M, Yearwood J, Ahmed S, Sultan I. A hybrid wrapper-filter approach to detect the source(s) of out-of-control signals in multivariate manufacturing process. Eur J Oper Res 2014:857–70.
[19] Huda S, Abawajy J, Alazab M, Abdollalihian M, Islam R, Yearwood J. Hybrids of support vector machine wrapper and filter based framework for malware detection. Fut Gen Comp Syst 2016;55:376–90.
[20] Ferreira AJ, Figueiredo MT. Incremental filter and wrapper approaches for feature discretization. Neurocomputing 2014;123:60–74.
[21] Jursa R. Variable selection for wind power prediction using particle swarm optimization. In: Proceedings of the 9th Genetic and Evolutionary Computation Conference (GECCO); 2007; London, England. p. 2059–65.
[22] 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.
[23] Gupta RA, Kumar R, Bansal AK. Selection of input variables for the prediction of wind speed in wind farms based on genetic algorithms. Wind Eng 2011;35(6):649–60.
[24] Kou P, Liang D, Gao F, Gao L. Probabilistic wind power forecasting with online model selection and warped gaussian process. Energy Convers Manag 2014;84:649–63.
[25] Salcedo-Sanz S, Pastor-Sánchez A, Prieto L, Blanco-Aguilera A, García-Herrera R. Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization - Extreme learning machine approach. Energy Convers Manag 2014;87:10–8.
[26] Huang GB, Zhu QY. Extreme learning machine: theory and applications. Neurocomputing 2006;70:489–501.
[27] Salcedo-Sanz S, Pastor-Sánchez A, Del Ser J, Prieto L, Geem ZW. A coral reefs optimization algorithm with harmony search operators for accurate wind speed prediction. Ren Energy 2015;75:93–101.
[28] Carta JA, Cabrera P, Matías JM, Castellano F. Comparison of feature selection methods using ANNs in MCP-wind speed methods. A case study. Appl Energy 2015;158:490–507.
[29] Kong X, Liu X, Shi R, Lee KY. Wind speed prediction using reduced support vector machines with feature selection. Neurocomputing 2015;169:449–56.
[30] Kumar S, López D. Feature selection used for wind speed forecasting with data driven approaches. J Eng Sci Technol Rev 2015;8:124–7.
[31] Lin T, Horne BG, Tino P, Giles CL. Learning long-term dependencies in NARX recurrent neural networks. IEEE Trans Neural Netw 1996;7:1329–51.
[32] Zhang C, Wei H, Zhao J, Liu T, Zhu T, Zhang K. Short-term wind speed forecasting using empirical mode decomposition and feature selection. Renew Energy 2016;96:727–37.
[33] Jiang Y, Huang G. Short-term wind speed prediction: hybrid of ensemble empirical mode decomposition, feature selection and error correction. Energy Convers Manag 2017;144:340–50.
[34] Zheng W, Peng X, Lu D, Zhang D, Liu Y, Lin Z, Lin L. Composite quantile regression extreme learning machine with feature selection for short-term wind speed forecasting: a new approach. Energy Convers Manag 2017;151:737–52.
[35] Zhang C, Zhou J, Li C, Fu W, Peng T. A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting. Energy Convers Manag 2017;143:360–76.
[36] Feng C, Cui M, Hodge BM, Zhang J. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. Appl Energy 2017;190:1245–57.
[37] Fu CL, Cheng HY. Predicting solar irradiance with all-sky image features via regression. Sol Energy 2013;97:537–50.
[38] Yadav AK, Malik H, Chandel SS. Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models. Renew Sustain Energy Rev 2014;31:509–19.
[39] Wang F, Zhen Z, Mi Z, Sun H, Su S, Yang G. Solar irradiance feature extraction and support vector machines based weather status pattern recognition model for short-term photovoltaic power forecasting. Energy Build 2015;86:427–38.
[40] Pérez-Ortiz M, Jiménez-Fernández S, Gutiérrez PA, Alexandre E, Hervás-Mart´ínez C, Salcedo-Sanz S. A review of classification problems and algorithms in renewable energy applications. Energies 2016;9:1–27.
[41] Rana M, Koprinska I, Agelidis VG. Univariate and multivariate methods for very short-term solar photovoltaic power forecasting. Energy Convers Manag 2016;121:380–90.
[42] Mohammadi K, Shamshirband S, Petkovic D, Khorasanizadeh H. Determining the most important variables for diffuse solar radiation prediction using adaptive neuro-fuzzy methodology; case study: city of Kerman. Iran Renew Sustain Energy Rev 2016;53:1570–9.
[43] Will A, Bustos J, Bocco M, Gotaya J, Lamelas C. On the use of niching genetic algorithms for variable selection in solar radiation estimation. Renew Energy 2011;50:168–76.
[44] Aybar-Ruiz A, Jiménez-Fernández S, Cornejo-Bueno L, Casanova-Mateo C, Sanz-Justo J, Salvador-González P, Salcedo-Sanz S. A novel Grouping Genetic Algorithm-Extreme Learning Machine approach for global solar radiation prediction from numerical weather models inputs. Sol Energy 2016;132:129–42.
[45] Salcedo-Sanz S, Jiménez-Fernández S, Aybar-Ruiz A, Casanova-Mateo C, Sanz-Justo J, García-Herrera R. A CRO-species optimization scheme for robust global solar radiation statistical downscaling. Renew Energy 2017;111:63–76.
[46] García-Hinde O, Terrén-Serrano G, Hombrados-Herrera MA, Gómez-Verdejo V, Jiménez-Fernández S, Casanova-Mateo C, Sanz-Justo J, Martínez-Ramón M, Salcedo-Sanz S. Evaluation of dimensionality reduction methods applied to numerical weather models for solar radiation forecasting. Eng Appl Artif Intell

2018;69:157–67.
[47] Hashim R, Roy C, Motamedi S, Shamshirband S, Petkovic D. Selection of climatic parameters affecting wave height prediction using an enhanced Takagi-Sugeno-based fuzzy methodology. Renew Sustain Energy Rev 2016;60:246–57.
[48] Cornejo-Bueno L, Nieto-Borge JC, García-Dáz P, Rodríguez G, Salcedo-Sanz S. Significant wave height and energy flux prediction for marine energy applications: a grouping genetic algorithm - Extreme Learning Machine approach. Renew Energy 2016;97:380–9.
[49] Ahila R, Sadasivam V, Manimala K. An integrated PSO for parameter determination and feature selection of ELM and its application in classification of power system disturbances. Appl Soft Comput 2015;32:33–7.
[50] Koprinska I, Rana M, Agelidis VG. Correlation and instance based feature selection for electricity load forecasting. Knowl-Based Syst 2015;82:29–40.
[51] Jurado S, Nebot A, Múgica F, Avellana N. Hybrid methodologies for electricity load forecasting: entropy-based feature selection with machine learning and soft computing techniques. Energy 2015;86:276–91.
[52] Hu Z, Bao Y, Chiong R, Xiong T. Mid-term interval load forecasting using multi-output support vector regression with a memetic algorithm for feature selection. Energy 2015;84:419–31.
[53] Jiang Y, Tang B, Qin Y, Liu W. Feature extraction method of wind turbine based on adaptive Morlet wavelet and SVD. Renew Energy 2011;36:2146–53.
[54] Salcedo-Sanz S, Camacho-Gómez C, Molina D, Herrera F. A Coral Reefs Optimization algorithm with substrate layers and local search for large scale global optimization. In: Proceedings of the IEEE Conference on Evolutionary Algorithms; 2016; Vancouver, Canada. p. 1–8.
[55] Salcedo-Sanz S, Del Ser J, Landa-Torres I, Gil-López S, Portilla-Figueras JA. The Coral Reefs Optimization algorithm: a novel metaheuristic for efficiently solving optimization problems. Sci World J 2014;739768:1–15.
[56] Salcedo-Sanz S. A review on the coral reefs optimization algorithm: new development lines and current applications. Prog Artif Intell 2017;6:1–15.
[57] Vermeij MJ. Substrate composition and adult distribution determine recruitment patterns in a Caribbean brooding coral. Mar Ecol Prog Ser 2005;295:123–33.
[58] Salcedo-Sanz S, Muñoz-Bulnes J, Vermeij M. New Coral Reefs-based Approaches for the Model Type Selection Problem: a Novel Method to Predict a Nation's Future Energy Demand. Int J Bio-inspir Comput 2016;10:3.
[59] Mallol-Poyato R, Jiménez-Fernández, del Ser J. A novel Coral Reefs Optimization algorithm with substrate layers for optimal battery scheduling optimization in micro-grids. Soft Comput J 2016;20:4287–300.
[60] Salcedo-Sanz S, Camacho-Gómez C, Magdaleno A, Pereira E, Lorenzana A. Structures vibration control via tuned mass dampers using a co-evolution coral reefs optimization algorithm. J Sound Vib 2017;393:62–75.
[61] Camacho-Gómez C, Wang X, Pereira E, Díaz IM, Salcedo-Sanz S. Active vibration control design using the Coral Reefs Optimization with Substrate Layer algorithm. Comput Struct 2018;157:14–26.
[62] Eiben AE, Smith JE. Introduction to evolutionary computing. Springer-Verlag, Natural Computing Series 1st edition, 2003.
[63] Storn R, Price K. Differential evolution - A simple and efficient heuristic for global optimization over continuous spaces. J Glob Optim 1997;11:341–59.
[64] Geem ZW, Kim JH, Loganathan GV. A new heuristic optimization algorithm: harmony Search. Simulation 2001;76(2):60–8.
[65] Hagan MT, Menhaj MB. Training feed forward network with the Marquardt algorithm. IEEE Trans Neural Netw 1994;5(6):989–93.
[66] Huang GB, Zhou H, Ding X, Zhang R. Extreme Learning Machine for Regression and Multiclass Classification. IEEE Trans Syst Man Cybern Part B Cybern 2012;42(2):513–29.
[67] Huang GB. ELM matlab code. 〈http://www.ntu.edu.sg/home/egbhuang/elm_codes.html〉; [Accessed 15 December 2016].
[68] Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Wang W, Powers JG. A Description of the Advanced Research WRF Version 2. National Center for Atmospheric Reserach, Mesoscale and Microscale Meteorology Division, Technical Note, 2005.
[69] Brown SH. Multiple linear regression analysis: a matrix approach with MATLAB. Alabama Journal of Mathematics, Spring/Fall, 2009.
[70] Smola AJ, Schölkopf B. A tutorial on support vector regression. Stat Comput 2004;14:199–222.