Engineering Applications of Artificial Intelligence 43 (2015) 44–53

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Significant wave height and energy flux range forecast with machine learning classifiers

J.C. Fernández [a], S. Salcedo-Sanz [b,*], P.A. Gutiérrez [a], E. Alexandre [b], C. Hervás-Martínez [a]

[a] Department of Computer Science and Numerical Analysis, University of Córdoba, Rabanales Campus, Albert Einstein Building, 3rd Floor, 14071 Córdoba, Spain

[b] Department of Signal Processing and Communications, University of Alcalá, 28871 Alcalá de Henares, Madrid, Spain

ARTICLE INFO

Article history:
Received 23 September 2014
Received in revised form
12 March 2015
Accepted 20 March 2015
Available online 29 April 2015

Keywords:
Wave energy prediction
Ordinal classification
Multi-class classification
Significant wave height
Flux of energy
Wave energy converters

ABSTRACT

In this paper, the performance of different ordinal and nominal multi-class classifiers is evaluated, in a problem of wave energy range prediction using meteorological variables from numerical models. This prediction could be used in problems of wave energy conversion in renewable and sustainable systems for energy supply. Specifically, the work is focused on *ordinal classifiers*, that have provided excellent performance in previous applications. The proposed techniques are novel with respect to alternative classification and regression techniques used up to date, the former not considering the order relation between classes in a multi-class problem and the latter needing, in general, more complex models. Another important novelty of the paper is to consider meteorological variables from numerical models as inputs of the classifiers, which has not been done before, to our knowledge, in this context. For this, a data matching is carried out between meteorological data, obtained from NCEP/NCAR Reanalysis Project in four points around the two buoys subjected to study (a buoy in the Gulf of Alaska and another one in the Southeast of United States), and the wave height or wave period collected by sensors in each buoy. Using this matching, the problem is tackled as an ordinal multi-class classification problem and the objective is to predict the range of height of the wave produced in each buoy and the range of energy flux generated. The classifiers to be compared and the model proposed are fully evaluated in both buoys. The results obtained are promising, showing an acceptable reconstruction by ordinal methods with respect to nominal ones in terms of wave height and energy flux.

1. Introduction

Marine energy is currently a hot topic in renewable and sustainable systems for energy supply (López et al., 2013; Heras-Saizarbitoria et al., 2013). It refers to a set of different technologies that exhibit a clear potential for sustainable growth, do not generate greenhouse gases, and that are potentially able to convert part of the huge energy of oceans into electricity. Some of the marine energy technologies that are currently under exploitation are off-shore wind energy, ocean thermal and tidal and wave energy conversion. Among these technologies, off-shore wind energy is currently the most exploited, but wave energy conversion is pushing hard as a new and promising energy source many countries are betting for. In fact, a number of public institutions and companies in British islands (Lawrence et al., 2013; The Offshore Renewable Energy in Scotland Website; The Pelamis Wave Power Website) are aiming at exploiting the huge wave energy resource (Arinaga and Cheung, 2012; Esteban and Leary, 2012), which, according to The Pelamis Wave Power Website, is equivalent to three times the current UK electricity demand. A review of the major issues involved in the generation of electricity from ocean energy, including wave energy conversion and its related economical aspects, can be found in Bahaj (2011).

Ocean wave energy is the energy source that converts potential and kinetic energy of waves into electricity, using devices called Wave Energy Converters (WECs) (Falcão, 2010; Lindroth and Leijon, 2011; Alamian et al., 2014). WECs transform the energy of waves into electricity by means of either the vertical oscillation of waves or the linear motion of waves, and exhibit some important advantages when compared to the other sources of ocean energy sources: (1) WECs have a much lower impact on ecosystems than tidal devices; (2) there are "ideal" areas for marine wave energy in Europe (Norway, UK, Ireland, Portugal), North coast of US and Canada, southern coast of Australia, northern coast of New Zealand, and Japan (in the sense of having great wave power density located near populated regions demanding energy); (3) wave energy is the most concentrated form of

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

http://dx.doi.org/10.1016/j.engappai.2015.03.012
0952-1976/

renewable energy (about 1000 times denser than wind) and is less variable on an hourly basis than wind energy; (4) the wave energy resource in a given area can be predicted within hours in advance, though this prediction is difficult because of their stochastic nature and the large amount of factors which influence wave height (meteorology, sea depth, closeness to the coast, etc.). A comprehensive review of these and other beneficial properties of WEC-based technologies can be found in Falcão (2010).

Even though wave energy exhibits a number of advantages when compared to tidal energy conversion, however, waves are more difficult to characterise and predict because of their stochastic nature. Even more, wave energy flux can exhibit nonlinear variability, with irregular extreme events (Reikard, 2013). As a consequence of this complexity, wave resource prediction become a crucial topic for the design, deployment, and control of WECs (Richter et al., 2013; Fusco and Ringwood, 2010a,b), that require a proper characterisation of waves. Alternative applications of wave height prediction are decision making in operational works at sea, risk evaluation of marine energy facilities, etc. Note that the wave height characterisation can be based on either physical models or data-driven ones. Data for characterising waves can be basically obtained from radars and buoys arrays, which generate time series. Using these time series, the corresponding wave spectrum, $S(f)$, can be computed. In turn, based on its spectral moments, a number of wave parameters, which are particularly useful to estimate the *wave power density* at a given location, can be defined. The most important wave parameters in this regard are the *significant wave height* (H_s) and the *wave energy period* (T_e). Note that the wave's *flux of energy* (F_e), which will be used by the WEC in order to generate electricity, can be obtained from H_s and T_e as $F_e = 0.49 \cdot H_s^2 \cdot T_e$. As mentioned, waves stochastic nature makes very difficult the prediction of wave energy resource, so the research work on this topic has been intense in the last few years. Focusing on machine learning approaches (Bishop, 2006) to wave energy prediction, most of them use artificial neural networks. One of the first approaches to predict H_s is due to Deo and Naidu (1998), who proposed the use of artificial neural networks in this problem. Improvements on this proposed system were presented in a posterior work (Agrawal and Deo, 2004). In Tsai et al. (2002), H_s and T_e are predicted from the observed wave records using time series neural networks. In Zanaganeh et al. (2009) (an improvement of Kazeminezhad et al., 2005), a hybrid genetic algorithm-adaptive network-based fuzzy inference system model was developed to forecast H_s and the peak spectral period in Lake Michigan from wind speed, fetch length and wind duration. In this methodology, both clustering and rule base parameters are simultaneously optimised using genetic algorithms and artificial neural networks. Recently, in Castro et al. (2015), a neural network is applied to estimate the wave energy resource in the northern coast of Spain. There have been other works that apply more efficient regression methods, based on machine learning and soft computing (Mahjoobi et al., 2008), such as support vector regression (Mahjoobi and Mosabbeb, 2009), genetic programming (Nitsure et al., 2012) or fuzzy logic (Özger, 2011). Alternative approaches using numerical models of atmosphere and ocean, usually hybridised with time series prediction, can also be found in the literature (Reikard et al., 2011; Akpinar and Kömürcü, 2013).

This paper deals with a problem of marine wave energy prediction (specifically H_s and F_e prediction), including different novelties in its resolution. Firstly, it is proposed the application of machine learning classifier algorithms for this problem. This approach is completely novel, since the previous works had tackled the problem, to the best of our knowledge, by applying regression techniques. Secondly, the problem is tackled as an ordinal multi-class classification problem, from the point of view that any real variable can be discretised in different categories, reducing the amount of information involved and simplifying this way the problem: a predefined number of ordered categories or ranges in terms of H_s or F_e is enough for obtaining practical information. The categories obtained through discretisation of a real value exhibit an order, and this order can be taken into account both in classifier construction and evaluation, choosing appropriate performance measures for that (Verwaeren et al., 2012; Cruz-Ramírez et al., 2014; Baccianella et al., 2009). Ordinal techniques studied take this consideration, achieving an error in the predictor that usually improves that of nominal classifiers. Modelling this prediction problem as a multi-class classification challenge allows the application of novel and powerful approaches existing in the literature, that have not been discussed before in a problem of marine wave energy prediction. This kind of multi-class classification problems are sometimes referred to as *ordinal regression* (Chu and Keerthi, 2007; Sun et al., 2010; Gutiérrez et al., 2012) and this work will compare the performance of standard multi-class classifiers and ordinal regression ones in this problem of H_s and F_e prediction. Another novelty of the problem is that the different classification techniques discussed in this paper are applied by using meteorological predictive variables from global numerical models. Specifically, meteorological data from The NCEP/NCAR Reanalysis Project are used in four different grid points around each buoy subjected to study in order to conform the independent variables for H_s and F_e prediction. For this purpose, a matching procedure is carried out every 6 h between the meteorological variables obtained from NCEP/NCAR Reanalysis Project and between H_s and F_e hourly collected by each buoy, which are obtained from the National Data Buoy Center (NDBC) (National Oceanic and Atmospheric Administration), belonging to the National Data Buoy Service of the United States. The performance of the different classifiers discussed in this problem has been evaluated in the prediction of H_s and F_e from real data of a buoy located in the Gulf of Alaska and another one located in the Southeast of United States.

The rest of this paper is structured as follows: Firstly, the next section briefly presents the formulation for an ordinal classification problem. Secondly, the more important ordinal classification methods in the literature for tackling this problem are described. In Section 4, the datasets used in this paper, including the buoys and reanalysis data are detailed. The specific problem modelling considered, which includes the estimation/prediction of H_s and F_e in a 6 h time-horizon, are also stated in this section. Section 5 presents the experimental part of the paper and the results. Finally, Section 6 gives some concluding remarks for closing the paper.

2. Ordinal classification

The classification of items into naturally ordered classes is commonly found in many supervised learning problems (Gutiérrez et al., 2013; Sánchez-Monedero et al., 2014), which are referred to as ordinal classification (also known ordinal regression) problems, where an ordinal scale (Lippmann, 1989) is used to label the patterns – for instance, a teacher who rates his/her students using labels (A, B, C, and D) that have a natural order among them ($A \succ B \succ C \succ D$). However, they are usually tackled by using methods intended for the classification of nominal classes, where the order relation is not considered.

Taking into account the nature of the problem studied in this paper, different class labels can be defined based on different ranges of H_s and F_e, in such way that an ordered relationship exists (the smallest to the largest, in increasing order) between the labels ($\mathcal{C}_1 \prec \mathcal{C}_2 \prec \cdots \prec \mathcal{C}_Q$, being Q the number of classes or labels). A class with lower order provides less energy than a class with higher order. This ordered rank implies an additional restriction for the classification problem. As a result, the nominal multi-class classification problem turns into an ordinal classification one.

When the ordinal nature of the target variable is not obvious or has been defined a posteriori, nominal classifiers can also be applied to ordinal problems (Bender and Grouven, 2006) and they can even yield better performance. However, they completely ignore the ordering of the labels by considering them as independent classes. This makes them require larger training sets with respect to ordinal approaches (Kramer et al., 2001). Ordinal classifiers are specifically built to exploit the existing order among classes of the target variable. For classifier evaluation, the rank of the label is usually considered, i.e., the position of the label on an ordinal scale, which is usually expressed by the following function $\mathcal{O}(\mathcal{C}_q) = q, q \in \{1, 2, \ldots, Q\}$.

Note that H_s can be treated in an ordinal manner by assigning labels $H_1, H_2, \ldots, H_Q$ to the different (increasing) H_s intervals. Regarding F_e, it can also be considered as an ordinal variable, with labels $F_1, F_2, \ldots, F_Q$ for increasing values of F_e. In this work, both the H_s and the corresponding F_e are discretised in 4, 5 and 6 different categories, trying to cover the most important kinds of waves from the point of view of their energy production. The ranges for each label in H_s and F_e are detailed in Section 4. Then, the more important ordinal classification methods are used for making predictions in H_s and F_e in a 6 h time-horizon. These methods are described in the next section.

3. Ordinal classifiers

One of the main objectives of this paper is to test several methods to address the prediction of H_s and F_e as a multi-class classification problem. To perform a prediction in a 6 h time-horizon, some of the best performing and more commonly used machine learning methods in the literature have been selected (Gutiérrez et al., 2012). All of them provide models to classify ordinal data, generally taking advantage of the order in the labels.

Threshold models are the most common approaches for ordinal classification (Verwaeren et al., 2012; Gutiérrez et al., 2012). They try to model ordinal regression problems from a thresholded regression perspective, assuming that some underlying real-valued outcomes exist, but they are unobservable. They involve estimating: (1) a projection function $f(x)$ that intends to uncover the nature of the assumed underlying outcome; and (2) a threshold vector, $\mathbf{b} = (b_1, b_2, \ldots, b_{Q-1}) \in \mathbb{R}^{Q-1}$, to represent the intervals in the range of $f(x)$, fulfilling $b_1 \le b_2 \le \cdots \le b_{Q-1}$. The ordinal response is modelled as real intervals in one dimension, and the algorithms try to find a direction to project the patterns and the set of thresholds to divide the real line into consecutive intervals (representing ordinal categories).

From a statistical background, one of the most popular threshold methods is the Proportional Odds Model (POM). POM is a logit model that can be viewed quite simply as an extension of logistic regression. The hallmark of the POM is that the odds ratio for a predictor can be interpreted as a summary of the odds ratios obtained from separate binary logistic regressions using all possible cut points ($\mathbf{b}$) of the ordinal outcome. This technique allows ordered data to be modelled by analysing it as a number of dichotomies. A binary logistic regression model compares one dichotomy (for example, waves exceeding 4 m or not, energy flux exceeding 10 kW/m or not, etc.), whereas POM compares a number of dichotomies by arranging the ordered categories into a series of binary comparisons (each binary comparison can be viewed as a separate logistic regression model resulting in $Q-1$ models). A detailed formulation of this model is available at McCullagh (1980).

Support Vector Machines (SVMs) are the most common kernel classifier for statistical pattern recognition (Cortes and Vapnik, 1995). They originally are linear classification models which select a set of patterns, named as support vectors. By the application of the kernel trick, they can obtain nonlinear decision boundaries. Although they were initially formulated for binary classification problems (finding a maximum margin separating hyperplane that allows to classify two classes), they have been extended for multi-class environments (Hsu and Lin, 2002), ordinal classification problems (Chu and Keerthi, 2007) and for standard regression (Smola and Schlkopf, 2004). SVMs have also been adapted to fit into this generic form of threshold models. Two support vector approaches for Ordinal Regression (SVOR) are proposed in Chu and Keerthi (2007). These methods use the threshold vector $\mathbf{b}$, together with parallel hyper-planes which are specified by a normal vector, $\mathbf{w}$. In order to define parallel discriminant hyperplanes for the ordinal scales, the SVM approaches share the common objective of seeking $Q-1$ parallel discriminant hyperplanes represented by a common vector $\mathbf{w}$ and the scalars biases, $\mathbf{b}$, to properly separate training data into ordered classes: (1) SVOR with Explicit constraints (SVOR-EX), based on defining a quadratic programming problem where the last set of constraints assuring the order between the thresholds explicitly appears in the optimisation problem and where the slacks for the q-th parallel hyperplane are defined for all patterns of class q and $q+1$; and (2) SVOR with Implicit constraints (SVOR-IM), based on redefining again the quadratic programming problem, following this principle: instead of considering only the errors from the samples of adjacent categories, samples in all the categories are allowed to contribute errors for each hyperplane. In this way, the ordinal inequalities on the thresholds are implicitly satisfied at the optimal solution.

Another method reformulated to tackle ordinal regression is the discriminant analysis (Sun et al., 2010). Discriminant analysis is usually not considered as a classification technique by itself, but rather as a supervised dimensionality reduction. Nonetheless, it is widely used for that purpose, since, as a projection method, the definition of thresholds can be used to discriminate the classes. In general, to allow the computation of the optimal projection for the data, this algorithm analyses two main objectives: the maximisation of the between-class distance, and the minimisation of the within-class distance, by using variance-covariance matrices and the Rayleigh coefficient. To export this idea to ordinal regression, only one projection vector is optimised and a minimum separation constraint is imposed on the closest projections of points of contiguous pairs of classes, leading to a proper ordering of the patterns. This preserves the ordinal information and avoid some serious ordinal misclassification errors. The methodology (in its kernel version) is known as Kernel Discriminant Learning for Ordinal Regression (KDLOR) (Sun et al., 2010).

In machine learning, a recent and very popular algorithm to obtain single-layer feedforward neural networks is the Extreme Learning Machines (ELMs) (Huang et al., 2012). In this algorithm input weights of an Single-Hidden Layer Feedforward Network (SLFN) can be randomly chosen (according to any continuous sampling distribution), and the output weights of the SLFN can be analytically determined by the minimum norm least-squares solutions. Since there are no iterative steps, the training speed of ELM can be hundreds of times faster than traditional gradient-based learning algorithms. ELMs have demonstrated good scalability and generalisation performance. ELMs have been adapted to ordinal regression (ELMOR) (Deng et al., 2010), considering the ordinal relationship among categories. The key of ELMOR is applying the output coding strategy of Frank and Hall (2001), that imposes the class ordering restriction. In the ELMOR algorithm with a single ELM classifier, the encoding for ordinal regression is implemented with Q output nodes where Q is the number of ranks or intervals. For each training example $\mathbf{x}$, the target output $\mathbf{t}$ is encoded into Q bits $[t_1, t_2, \ldots, t_Q]$. Considering the order of the categories, if a data point $\mathbf{x}$ belongs to the q-th rank, it is automatically classified into lower-order categories $(1, 2, \ldots, q-1)$. So the target output of $\mathbf{x}$ is encoded as $\mathbf{t} = [1, 1, \ldots, 1, -1, -1, \ldots, -1]$, where $\mathbf{t}_i$ $(1 \le i \le q)$ is set to 1 and other elements -1. The output coding matrix is a $Q \times Q$ matrix where the number of coding bits is equal to the number of ranks Q. The q-th row represents the output coding vector of q-th rank, and every column represents one output coding bit corresponding to an output node.

Finally, ordinal regression can be reduced to binary classification by data replication methods in what is known as augmented binary classification (Lin and Li, 2012), using the output coding strategy in Frank and Hall (2001), and an extension, replication and specific weighting of the obtained patterns. In Cardoso and da Costa (2007), a technique known as Ordinal Neural Networks (ONN), based in applying this replication method to neural networks, is used for ordinal multi-classification tasks. The activation function of the output node of a neural network is set to the *logsig* function and the model is trained to predict the posterior probabilities when fed with the original input variables and the variables generated by the data replication method.

4. Obtaining the datasets

The datasets used in this paper for each buoy includes values from two sources:

1. H_s and T_e as standard meteorological data collected by the sensors in the buoys. One of them is located in the Gulf of Alaska and the other one in the Southeast of the United States. The data have been directly obtained from the National Oceanic and Atmospheric Administration (NOAA), National Data Buoy Center (NDBC) (National Oceanic and Atmospheric Administration), a part of the National Weather Service (NWS), that operates and maintains a network of data collecting buoys and coastal stations. The next subsection details the location for each buoy.
2. Meteorological variables from The NCEP/NCAR Reanalysis Project. Reanalysis data includes four sea level variables in a grid of four points around each buoy (North, South, East and West): air temperature, sea level pressure, the zonal component of the wind and the meridional component of the wind. These four variables have been chosen because they are able to characterise quite well the local wind in a given zone. Of course, there may be contributions to wave height and flux from phenomena that occur at long distances from the considered buoy (groundswell, etc). However, the local contribution of the wind (and related variables, such as temperatures to take into account breezes, mainly in WECs near the shore, or pressure) is still the most important factor to predict significant heights and flux. On the other hand, note that four points are considered surrounding the buoy, what may be useful to detect possible groundswell or related phenomena with a certain direction, and also it takes into account the pressure wind and temperature effect (possible breezes). The selection of the four grid points is detailed in Section 4.2.

Therefore, for each buoy, 16 input variables are considered, four meteorological variables for each of the four points around each buoy. H_s and F_e are the objective variables. F_e is calculated from H_s and T_e, and it is detailed in the next subsection. With these two data sources a matching procedure is carried out: the meteorological variables on the one hand, and H_s and F_e on the other. The following subsections explain this in detail.

Both, the reanalysis meteorological data and standard meteorological data from the buoys are extracted from January 1st (00:00) to December 31st (23:00), for the years 2012 and 2013.

4.1. Selecting the buoys

The standard meteorological data for each buoy are collected hourly in the NDBC where they are stored in downloadable annual text files. Two buoys from the NDBC with a sufficient range of values for H_s and F_e have been selected to validate the performance of machine learning classifiers:

- Station 46001 (LLNR 984) (National Data Buoy Center) – Western Gulf of Alaska – 175NM SE of Kodiak, AK. Geographical location: 56.304N 147.920W (56°18′16″N 147°55′13″W). This location lies in the northern extra-tropical storm belt, and, as a result, it is a region of active wave generation. It is a deep-water buoy in which the average height of waves is very high.
- Station 41013 (LLNR 815) (National Data Buoy Center) - Frying Pan Shoals, NC Buoy. Geographical location: 33.436N 77.743W (33°26′11″N 77°44′35″W). This location lies in the East Coast of the USA, and it is a near-shore buoy, with a smaller average height of waves than in the previous buoy considered.

For each selected buoy, it is possible to obtain the value of energy flux considering H_s (measured in *meters*) and T_e (measured in *seconds*) as follows:

$$F_e = 0.49 \cdot H_s^2 \cdot T_e, \tag{1}$$

where F_e is the energy flux generated by the waves measured in *kilowatts per meter*. Note that H_s and F_e are the two physical phenomena to be predicted in this study. Note also that F_e is defined in Eq. (1) as an average energy flux (H_s is a kind of average wave height), though for simplicity it will be referred just as *energy flux*.

The objective variables (H_s and F_e) must be discretised in order to tackle the problem as a classification one. The minimum and maximum values of H_s and F_e are 0.43–11.14 m and 0.31–665.85 kW/m, respectively, for the buoy in the Gulf of Alaska, and 0.34–6.21 m and 0.25–158.16 kW/m, for the buoy in the Southeast of United States. To check out how the change in the discretisation of H_s and F_e can affect the prediction task, datasets with different number of classes according to the thresholds used for each one have been built. Considering the expert knowledge about the problem, the threshold values used were selected trying to obtain the most representative classes of the behaviour of the waves for each station and geographical location. Note that this discretisation may be adjustable depending on each specific buoy situation, i.e., depending on the normal sea conditions where the WEC is installed, and even the WEC type. Note also that the inclusion of many different classes may compromise the performance of the classification task. Tables 1–3 show the class thresholds for 4, 5 and 6 classes respectively for each buoy and for H_s and F_e.

4.2. Meteorological data grid

The well-known NCEP/NCAR Reanalysis Project (Kalnay et al., 1996) is a joint project between the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR), at the National Oceanic and Atmospheric Administration (NOAA)/Earth System Research Laboratory (ESRL) Physical Sciences Division (The NCEP/NCAR Reanalysis Project). The goal

Table 1
Discretisation in 4 classes for H_s and F_e.

Buoy data	Class A	Class B	Class C	Class D
Station 46001 – Western Gulf of Alaska				
H_s (m)	[0, 1.5)	[1.5, 2.5)	[2.5, 4)	[4, ∞)
F_e (kW/m)	[0, 7)	[7, 20)	[20, 50)	[50, ∞)
Station 41013 – Southeast of United States				
H_s (m)	[0, 1)	[1, 1.5)	[1.5, 3)	[3, ∞)
F_e (kW/m)	[0, 2)	[2, 4)	[4, 10)	[10, ∞)

Table 2
Discretisation in 5 classes for H_s and F_e.

Buoy data	Class A	Class B	Class C	Class D	Class E
Station 46001 – Western Gulf of Alaska					
H_s (m)	$[0, 1.5)$	$[1.5, 2.5)$	$[2.5, 3.5)$	$[3.5, 5)$	$[5, \infty)$
F_e (kW/m)	$[0, 7)$	$[7, 20)$	$[20, 45)$	$[45, 80)$	$[80, \infty)$
Station 41013 – Southeast of United States					
H_s (m)	$[0, 1)$	$[1, 1.5)$	$[1.5, 2)$	$[2, 3)$	$[3, \infty)$
F_e (kW/m)	$[0, 1)$	$[1, 2)$	$[2, 4)$	$[4, 10)$	$[10, \infty)$

Table 3
Discretisation in 6 classes for H_s and F_e.

Buoy Data	Class A	Class B	Class C	Class D	Class E	Class F
Station 46001 – Western Gulf of Alaska						
H_s (m)	$[0, 1.5)$	$[1.5, 2.2)$	$[2.2, 3)$	$[3, 4.2)$	$[4.2, 5.2)$	$[5.2, \infty)$
F_e (kW/m)	$[0, 5)$	$[5, 12)$	$[12, 25)$	$[25, 45)$	$[45, 80)$	$[80, \infty)$
Station 41013 – Southeast of United States						
H_s (m)	$[0, 0.7)$	$[0.7, 1)$	$[1, 1.5)$	$[1.5, 2)$	$[2, 3)$	$[3, \infty)$
F_e (kW/m)	$[0, 1)$	$[1, 2)$	$[2, 3)$	$[3, 4)$	$[4, 10)$	$[10, \infty)$

of this joint effort was to produce new atmospheric analyses using historical data (1948 onwards) and as well to produce analyses of the current atmospheric state (Climate Data Assimilation System, CDAS). Reanalysis data have been used in surface for tackling this problem of H_s and F_e prediction. For each meteorological data listed above (air temperature, sea level pressure, the zonal component of the wind and the meridional component of the wind), the four closest points around of each buoy have been selected, obtaining a meteorological data sub-grid within the following ranges:

- Station 46001 (LLNR 984): 55.0–57.5 grades North for latitude and 145.5–150.0 grades West for longitude. Therefore, the observations for each of the physical variables were taken from the following points (latitude N, longitude E): (55.0N, 210.0E), (55.0N, 212.5E), (57.5N, 210.0E) and (57.5N, 212.5E). Fig. 1 shows an approximate representation of the four grid points around the buoy. For the next buoy, Station 41013, the representation is similar.
- Station 41013 (LLNR 815): 32.5–35.0 grades North for latitude and 77.5–80.0 grades West for longitude. Therefore, for this buoy, the physical observations selected are in each of the following geographic points (latitude N, longitude E): (32.5N, 280.0E), (32.5N, 282.5E), (35.0N, 280.0E) and (35.0N, 282.5E).

Reanalysis Meteorological data are stored in The NCEP/NCAR Reanalysis Project web page 6-hourly, daily and monthly, and they can be extracted in files with Network Common Data Form (NetCDF) (Unidata) format (`.nc` extension), a special type of file for representing scientific data. The data have been extracted with a time horizon of 6 h because it is the minimum temporal resolution of the Reanalysis data used. In addition, a time horizon of 6 h seems a more reasonable time for detecting changes in sea conditions. Note, however, that the system can be trained with different prediction time horizons, depending on the final objective of this prediction.

The NCEP/NCAR Reanalysis Project web page allows you to select a range of years for a meteorological variable within a geographic grid arranged every 2.5°. Hence, as the four closest geographic points to each buoy were needed, it has been necessary to extract a NetCDF file for each the four meteorological variables. In other words, each file extracted contains the values for one meteorological variable in the four closest geographical

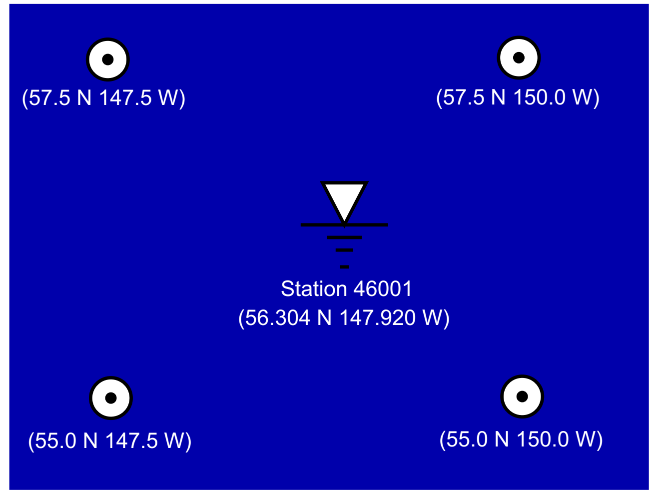

Fig. 1. Closest geographical points around the Station 46001 (56.304N 147.920W).

points around one buoy for the years 2012–2013. Once obtained the NetCDF files, a Java software using a NetCDF-Java library (implemented by us) was used to obtain, for each buoy, the input variables of our problem, so that a pattern is formed by the values of air temperature, sea level pressure, the zonal component of the wind and the meridional component of the wind in each one of the four closest to the buoy at a specific time, taken every 6 h. Then, a matching every 6 h is done with respect to H_s and F_e (which are taken hourly) for each buoy. This is also done automatically by the Java software implemented. Consequently, for an specific discretisation (number of classes), for each buoy, two datasets are considered, one for H_s and another for F_e.

4.3. Final composition of the datasets

In this subsection, some problems found during the composition of the datasets are discussed.

It has been necessary to carry out a preprocessing for cleaning the data (H_s and T_e) extracted for each buoy from the NDBC web page, removing missing hours and invalid data. The data extracted from the buoys are stored hourly, thereby, the total number of hourly measurements, taking into account the number of days in each month and if the year is leap, should be 8784 for 2012 and 8760 for 2013. On the other hand, although a buoy has measurements at a specific time, the measured values can be null. If it happens, null values are marked in the downloadable annual text files with the value 99.00. That said, at the Station 46001 – Western Gulf of Alaska, the null values plus complete missing hours for 2012 and 2013 are a total of 361 hourly measurements. For the Station 41013 – Southeast of United States there are a total of 57 hourly measurements less. Compared with the 17544 (8784+8760) possible measurements for the years 2012–2013 the number of missing patterns is less than 2.5%. Note that these missing values may be even less because the matching with meteorological variables is 6-hourly. In this case, if there is no data for a certain hourly date and time in the buoy, then the pattern is discarded.

Regarding the meteorological variables, as the measurements are taken 6-hourly, the total number of measurements for 2012 are 1464 (8784/6) and for 2013 are 1460 (8760/6), so that, the maximum number of patters for each buoy in the years 2012–2013 could be 1464+1460=2924. Considering all these factors, the total number of patterns obtained for prediction, joining 2012 and 2013, are 2759 for the Station 46001 – Western Gulf of Alaska and 2920 for the Station 41013 – Southeast of United States. The characteristics of the datasets, taking into account the number of classes in the discretisation procedure, are shown in Table 4.

Fig. 2 shows a small example of the dataset obtained for H_s (Station 46001), using 4 classes. The structure for F_e is similar, changing the objective variable.

5. Experiments

Regarding the experimental design, due to the fact that the majority of methods used in the comparison procedure are deterministic, a stratified holdout with 30 random splits is performed, where approximately 75% of the instances were used for the training set and the remaining 25% for the test or generalisation set, maintaining the original distribution of classes. The objective of this design is to evaluate the results as a function of the different training/test splits.

Apart from all ordinal methods presented in Section 3, two important and relevant nominal algorithms based on Support Vector Machine methodology are considered (which do not take the class order into account) to compare the problem under study. Furthermore, a well-known regression method is also added to the comparison in order to study the performance of regression techniques in predicting H_s and F_e. In this way, all possible approach types (regression, classification and ordinal classification) have been included in the comparison.

The objective of the nominal methodologies is to check the difference with ordinal methods. SVM methodologies can be said to be the best performing algorithms in machine learning to handle binary classification. The *1 versus 1* (SVC1V1) and *1 versus all* (SVC1VA) approaches (Hsu and Lin, 2002) are considered, because they are the two main approaches when applying SVM to multiclass problems. In the SVC1V1 method, a binary SVM classifier is built for each pair of classes, resulting in $Q(Q-1)/2$ SVMs trained to differentiate the patterns of one class from those of another class. The classification of unseen patterns is carried out according to the majority voting strategy, each SVM is considered to be a voting and the pattern is designated to be in a class with the maximum number of votes. The SVC1VA method constructs Q classifiers, where the ith SVM is trained with all the samples in ith class with positive labels, and all other examples with negative ones. As having a total of Q decision functions, a pattern is assigned to the class which has the largest value of the decision function. SVC1VA tries to maximise the margin between two groups of data and when date are not linear separable, a penalty term for reducing the number of training error is used. The basic concept behind SVM is to search for a balance between a regularisation term and the training errors.

With respect to the regression technique, a Support Vector Machine algorithm (the Support Vector Regression, SVR, method) is used in the comparison. Concretely, the ϵ-SVR model considered in Smola and Schlkopf (2004) is considered, one of the best performing implementations in this context. The actual values of H_s and F_e are used for training the regressor. During the test phase, the predictions are discretised into 4, 5 or 6 classes to compare the accuracy of the method with respect to the rest of algorithms. For this, the thresholds defined in Tables 1–3 are considered.

The hyperparameters for each method were selected by using a nested fivefold cross-validation over the training set and taking into account the lowest cross-validation error obtained with respect to the Mean Absolute Error performance (see Section 5.1), *MAE*, in order to favour better ordering classifier. The parameter configurations are the following:

- The kernel function considered for all kernel methods (SVOREX, SVORIM, SVR, KDLOR, SVC1V1 and SVC1VA) was the Gaussian function. For every tested kernel method, the kernel width was selected within these values $\sigma \in \left\{10^{-3}, 10^{-2}, \ldots, 10^{3}\right\}$.
- The POM algorithm is used with the *logit* link function, the most extended one.
- The range of ϵ (width of the tube within which no penalty is associated in the training loss function) for ϵ-SVR was $\epsilon \in \left\{10^{0}, 10^{1}, \ldots, 10^{3}\right\}$.

Table 4
Characteristics of the datasets for each buoy. A total of 12 datasets: 6 datasets for each buoy, 3 for H_s (4, 5 and 6 classes) and 3 for F_e.

Buoy	#Patterns	#Inputs	#Classes	#Patterns H_s[a]	#Patterns F_e[a]
46001	2847	16	4	619-823-895-422	744-756-753-506
41013	2920			1077-965-806-72	858-863-762-437
46001	2847	16	5	619-823-697-448-172	744-756-682-325-252
41013	2920			1077-965-469-337-72	233-625-863-762-437
46001	2847	16	6	619-603-625-556-204-152	519-587-617-459-325-252
41013	2920			318-759-965-469-337-72	233-625-521-342-762-437

[a] The thresholds for each dataset are described in Section 4.1.

				57.5x210 (latxlon)				57.5x212.5 (latxlon)				55x210 (latxlon)				55x212.5 (latxlon)				
#YY	MM	DD	hh	air1	uwnd1	vwnd1	psl1	air2	uwnd2	vwnd2	psl2	air3	uwnd3	vwnd3	psl3	air4	uwnd4	vwnd4	psl4	Class-Hs (wave height)
#yr	mo	dy	hr	c	m/s	m/s	bar	c	m/s	m/s	bar	c	m/s	m/s	bar	c	m/s	m/s	bar	
2012	1	1	0	-5.60	0.90	4.60	1.01	-1.50	1.90	5.70	1.01	-3.10	0.10	14.60	1.01	0.00	5.50	16.50	1.00	B
2012	1	1	6	-4.00	4.00	11.20	1.01	1.30	0.80	12.20	1.00	-2.20	1.20	18.00	1.00	1.30	2.90	17.50	1.00	B
2012	1	1	12	-3.30	6.10	14.80	1.00	2.10	2.50	16.30	1.00	-3.30	4.80	21.00	1.00	0.30	4.20	21.90	0.99	C
2012	1	1	18	-0.90	7.60	12.70	1.00	4.00	8.50	15.60	0.99	-3.30	7.80	16.00	1.00	-0.10	10.00	19.50	0.99	D
2012	1	2	0	-1.00	12.90	10.60	0.99	2.50	18.20	14.00	0.98	-3.20	10.40	7.10	1.00	-1.80	15.20	11.90	0.99	D
2012	1	2	6	-0.10	12.30	1.80	0.99	1.50	17.40	5.30	0.99	-0.50	11.00	1.20	1.00	-0.70	13.20	1.80	0.99	C
2012	1	2	12	1.40	9.40	1.90	0.99	2.90	12.30	2.00	0.99	0.50	10.80	3.20	1.00	0.60	10.10	2.50	1.00	C
2012	1	2	18	1.00	10.50	2.20	1.00	4.20	9.60	3.80	1.00	0.30	7.20	0.80	1.00	1.00	5.00	3.60	1.00	C
2012	1	3	0	0.00	4.80	2.10	1.00	3.40	5.00	0.50	1.00	-0.50	0.20	2.20	1.00	0.60	0.80	1.20	1.00	C
2012	1	3	12	0.40	3.30	5.70	1.00	3.90	5.40	1.60	1.00	1.50	10.20	9.90	0.99	2.90	14.60	7.30	0.99	A
2012	1	3	18	1.90	6.20	8.90	0.99	5.00	7.70	7.60	0.99	2.30	10.80	12.60	0.98	3.10	12.80	11.40	0.98	C
2012	1	4	0	2.50	2.40	11.50	0.99	5.60	3.80	12.80	0.99	2.10	9.10	16.00	0.98	2.80	10.10	11.70	0.98	C
2012	1	4	6	1.00	4.20	12.40	0.99	4.90	2.60	14.50	0.98	1.00	3.10	18.00	0.98	2.50	5.80	13.90	0.98	C
2012	1	4	12	-0.50	13.70	10.90	0.99	3.70	11.40	11.80	0.98	-1.40	7.00	18.90	0.99	0.70	2.90	17.90	0.98	D
2012	1	4	18	-1.70	14.50	7.10	0.99	2.90	15.10	8.30	0.98	-3.40	12.20	10.20	0.99	-1.00	12.20	14.20	0.99	D

Fig. 2. Part of the final dataset matching the buoy data and the reanalysis data with 6-hourly measurements, for H_s, Station 46001 and 4 classes.

- The cost parameter C of all SVM methods was selected within the range $C \in \{10^{-3}, 10^{-2}, \dots, 10^{3}\}$, and for the KDLOR methodology, within the range $C \in \{10^{-1}, 10^{0}, 10^{1}\}$. The u parameter to avoid singularities in the covariance matrices for KDLOR was considered within the range $u \in \{10^{-6}, 10^{-3}, 10^{-2}\}$.
- For the ONN methodology, the sigmoidal activation function is used and the number of nodes in the hidden layer was allowed values in the range $\{5, 10, 15, \dots, 40\}$.
- For ELMOR, a higher numbers of hidden neurons is considered, $H \in \{5, 10, 20, \dots, 100\}$, given that it relies on sufficiently informative random projections (Huang et al., 2012).

5.1. Performance evaluation

Metrics specifically designed to measure the degree of discrepancy (in the ordinal scale) between the predicted and real categories are needed to evaluate classifiers for ordinal regression problems. There are different alternatives (Cardoso and da Costa, 2007; Cruz-Ramírez et al., 2014; Baccianella et al., 2009), and there is an agreement that, in an ordinal regression problem, the aim is to evaluate two factors: whether the patterns are generally well classified, and whether the classifier tends to predict a class as close to the real class as possible.

Let us suppose an ordinal classification problem with Q classes and n patterns with a classifier g obtaining a $Q \times Q$ confusion matrix. n_{jk} represents the number of times the patterns are predicted by classifier g to be in class k when they really belong to class j and n is the total number of patterns. From this assumptions, the Correct Classification Rate (CCR) or accuracy (the most common metric for measuring the performance of a nominal classifier) can be obtained as

$$CCR = \frac{1}{n} \sum_{j=1}^{Q} n_{jj},$$

where CCR values range from 0 to 1.

On the other hand, MAE metric is defined as the average absolute deviation of the predicted class from the true class (i.e., deviation in number of categories of the ordinal scale). This measure allows to check out whether the classifier tends to predict a class as close to the real class as possible:

$$MAE = \frac{1}{n} \sum_{j,k=1}^{Q} |j-k| \, n_{jk} = \frac{1}{n} \sum_{i=1}^{n} e(\mathbf{x}_i),$$

where $e(\mathbf{x}_i) = |\mathcal{O}(y_i) - \mathcal{O}(y_i^*)|$ is the distance between the true (y_i) and the predicted (y_i^*) ranks, and $\mathcal{O}(C_j) = j$ is the position of a label or class in the ordinal rank. Then, MAE values range from 0 to $Q-1$.

5.2. Results

Table 5 shows the averages (*Mean*) and the standard deviations (*SD*), for H_s and F_e at the Station 46001 – Gulf of Alaska, of the test errors for the 30 models generated (one for each stratified hold-out) from the different machine learning classifiers compared. This table shows the results for 4, 5 and 6 classes. As a reference, the accuracy obtained by random guessing are $CCR = 33.3\%$ for 3 classes, $CCR = 25.0\%$ for 4 classes and $CCR = 16.6\%$ for 6 classes. Because of this, and given that there are several external factors not included in the dataset (sea depth, closeness to the coast, etc.), which can increase the uncertainty and the noise in the variable to be predicted, the results obtained can be considered as satisfactory. It must be noted that the models obtained are trained in data-driven fashion, leading to very good accuracy results. However, they do not allow a direct physical interpretation, because they are highly nonlinear and the relations between the input variables include several interactions.

Table 5
Average (*Mean*) and standard deviation (*SD*) obtained for the test set in terms of *CCR* and *MAE* for H_s and F_e at the Station 46001 – Gulf of Alaska. The best result is marked in boldface and the second one in italics.

Metric	Method	4 classes (*Mean* ± *SD*)	5 classes (*Mean* ± *SD*)	6 classes (*Mean* ± *SD*)
Significant wave height, H_s				
CCR	ELMOR	58.580 ± 1.381	52.845 ± 2.288	46.159 ± 1.322
	KDLOR	59.285 ± 1.358	53.638 ± 1.991	46.618 ± 1.392
	ONN	57.671 ± 1.727	51.643 ± 1.524	45.430 ± 2.034
	POM	56.324 ± 1.544	50.981 ± 1.659	44.324 ± 1.428
	SVC1V1	*60.758 ± 1.499*	54.464 ± 1.593	47.870 ± 2.454
	SVC1VA	60.295 ± 1.226	53.792 ± 1.851	*48.222 ± 1.875*
	SVOREX	**60.836 ± 1.212**	**55.140 ± 2.030**	**48.357 ± 1.565**
	SVORIM	60.271 ± 1.151	*54.551 ± 2.228*	47.783 ± 1.697
	SVR	48.531 ± 1.667	41.188 ± 1.953	32.285 ± 2.803
MAE	ELMOR	0.437 ± 0.015	0.524 ± 0.024	0.653 ± 0.019
	KDLOR	0.442 ± 0.016	0.527 ± 0.026	0.656 ± 0.023
	ONN	0.451 ± 0.020	0.540 ± 0.018	0.672 ± 0.034
	POM	0.461 ± 0.016	0.544 ± 0.020	0.682 ± 0.021
	SVC1V1	0.422 ± 0.015	0.515 ± 0.018	0.664 ± 0.031
	SVC1VA	0.435 ± 0.013	0.540 ± 0.022	0.682 ± 0.029
	SVOREX	**0.414 ± 0.012**	**0.497 ± 0.021**	*0.620 ± 0.023*
	SVORIM	*0.416 ± 0.012*	*0.497 ± 0.022*	**0.619 ± 0.025**
	SVR	0.550 ± 0.021	0.646 ± 0.028	0.832 ± 0.048
Flux of energy, F_e				
CCR	ELMOR	55.304 ± 1.885	50.208 ± 1.753	41.300 ± 1.502
	KDLOR	56.671 ± 2.085	50.377 ± 1.901	42.314 ± 1.636
	ONN	53.222 ± 2.488	48.556 ± 1.788	39.918 ± 2.458
	POM	53.043 ± 1.660	48.966 ± 1.477	38.599 ± 1.470
	SVC1V1	*57.821 ± 1.612*	*52.676 ± 1.891*	42.855 ± 1.614
	SVC1VA	57.251 ± 2.195	52.043 ± 1.498	*43.319 ± 1.795*
	SVOREX	**58.014 ± 1.813**	**52.870 ± 1.588**	**43.449 ± 1.906**
	SVORIM	57.386 ± 1.660	51.599 ± 1.529	42.894 ± 1.818
	SVR	45.527 ± 2.084	41.053 ± 2.241	33.029 ± 1.527
MAE	ELMOR	0.499 ± 0.023	0.585 ± 0.024	0.740 ± 0.023
	KDLOR	0.491 ± 0.024	0.593 ± 0.027	0.747 ± 0.025
	ONN	0.528 ± 0.033	0.613 ± 0.026	0.776 ± 0.048
	POM	0.523 ± 0.021	0.607 ± 0.021	0.783 ± 0.021
	SVC1V1	0.490 ± 0.022	0.586 ± 0.023	0.766 ± 0.022
	SVC1VA	0.504 ± 0.026	0.610 ± 0.022	0.794 ± 0.034
	SVOREX	*0.467 ± 0.022*	**0.550 ± 0.019**	*0.721 ± 0.022*
	SVORIM	**0.467 ± 0.019**	*0.557 ± 0.019*	**0.712 ± 0.022**
	SVR	0.681 ± 0.029	0.781 ± 0.036	1.045 ± 0.029

From a descriptive point of view, the best result in CCR for predicting H_s considering a discretisation of 4 classes is obtained by the SVOREX method, which obtains a value of 60.836 ± 1.212; and the second best result is obtained by the method SVC1V1, with a value of 60.758 ± 1.499. With respect to the MAE metric, the best result is obtained again by the SVOREX methodology, with a value of 0.414 ± 0.012, and the second best result changes, being obtained by the SVORIM methodology with a value of 0.416 ± 0.012. In the case of F_e, the best result in CCR is also obtained by SVOREX method, with a value of 58.014 ± 1.813, and the second best result is also obtained by the SVC1V1 method, with a value of 57.821 ± 1.612. Concerning the MAE metric, the best result is for the SVORIM method, with a value of 0.467 ± 0.019, and the second best result is obtained by the SVOREX method, with a value of 0.467 ± 0.022. Note that these results are very similar, except that the variance of the SVORIM method is slightly lower than the SVOREX method. Summarising, the best results in CCR for H_s and F_e are obtained by the SVOREX method and the best results in MAE for H_s and F_e are obtained by the SVOREX and SVORIM methods.

For 5 and 6 classes, the prediction of H_s and F_e is worse in terms of CCR and MAE. Thus, for 5 classes, the best result in H_s and F_e is obtained by the SVOREX method both in CCR and MAE. For

Table 6
Average (*Mean*) and standard deviation (*SD*) obtained for the test set in terms of *CCR* and *MAE* for H_s and F_e at the Station 41013 – Southeast of United States. The best result is marked in boldface and the second one in italics.

Metric	Method	4 classes (*Mean* ± *SD*)	5 classes (*Mean* ± *SD*)	6 classes (*Mean* ± *SD*)
Significant wave height, H_s				
CCR	ELMOR	62.829 ± 1.736	55.717 ± 1.369	47.137 ± 1.512
	KDLOR	61.530 ± 1.251	55.808 ± 1.287	43.840 ± 1.702
	ONN	62.968 ± 1.943	50.708 ± 2.383	46.584 ± 2.545
	POM	62.274 ± 1.612	55.630 ± 1.552	47.110 ± 1.459
	SVC1V1	65.100 ± 1.704	59.388 ± 1.495	49.776 ± 1.594
	SVC1VA	64.178 ± 1.573	57.575 ± 1.522	48.274 ± 1.447
	SVOREX	**65.169 ± 1.703**	**60.155 ± 1.318**	**50.539 ± 1.775**
	SVORIM	*65.128 ± 1.794*	*59.904 ± 1.338*	*50.434 ± 1.329*
	SVR	30.913 ± 0.626	19.708 ± 0.843	19.311 ± 0.717
MAE	ELMOR	0.390 ± 0.018	0.483 ± 0.017	0.592 ± 0.018
	KDLOR	0.405 ± 0.014	0.488 ± 0.015	0.657 ± 0.019
	ONN	0.389 ± 0.021	0.568 ± 0.034	0.608 ± 0.039
	POM	0.392 ± 0.017	0.479 ± 0.019	0.594 ± 0.018
	SVC1V1	0.370 ± 0.018	0.454 ± 0.017	0.571 ± 0.020
	SVC1VA	0.388 ± 0.018	0.488 ± 0.018	0.610 ± 0.022
	SVOREX	*0.360 ± 0.017*	*0.429 ± 0.014*	*0.546 ± 0.018*
	SVORIM	**0.359 ± 0.018**	**0.427 ± 0.015**	**0.543 ± 0.015**
	SVR	0.932 ± 0.018	1.068 ± 0.016	1.185 ± 0.017
Flux of energy, F_e				
CCR	ELMOR	49.548 ± 1.845	45.622 ± 1.528	38.483 ± 1.867
	KDLOR	50.498 ± 1.581	40.694 ± 1.468	35.388 ± 1.705
	ONN	50.699 ± 1.718	45.447 ± 2.255	36.429 ± 4.787
	POM	49.251 ± 1.736	45.018 ± 1.683	38.265 ± 1.664
	SVC1V1	**53.904 ± 1.816**	47.571 ± 1.779	**43.900 ± 1.194**
	SVC1VA	52.329 ± 1.477	46.539 ± 1.546	42.027 ± 1.667
	SVOREX	52.416 ± 1.437	*48.059 ± 1.769*	*42.868 ± 1.248*
	SVORIM	*52.890 ± 1.291*	**48.498 ± 1.712**	40.909 ± 1.273
	SVR	28.498 ± 4.431	29.689 ± 2.250	28.050 ± 1.134
MAE	ELMOR	0.553 ± 0.023	0.625 ± 0.020	0.878 ± 0.032
	KDLOR	0.562 ± 0.018	0.697 ± 0.018	0.944 ± 0.032
	ONN	0.548 ± 0.021	0.637 ± 0.029	0.966 ± 0.179
	POM	0.559 ± 0.020	0.638 ± 0.022	0.890 ± 0.031
	SVC1V1	0.535 ± 0.022	0.616 ± 0.020	0.908 ± 0.032
	SVC1VA	0.575 ± 0.021	0.672 ± 0.028	0.965 ± 0.031
	SVOREX	*0.515 ± 0.015*	*0.594 ± 0.021*	*0.866 ± 0.027*
	SVORIM	**0.505 ± 0.015**	**0.579 ± 0.017**	**0.818 ± 0.025**
	SVR	0.994 ± 0.202	1.008 ± 0.060	1.429 ± 0.082

6 classes, the best result for H_s and F_e is obtained by the SVOREX and SVORIM method in *CCR* and *MAE* respectively.

Table 6 shows the averages (*Mean*) and the standard deviations (*SD*), for H_s and F_e at the Station 41013 – Southeast of United States, of the test errors for the 30 models generated (one for each stratified hold-out) from the different machine learning classifiers compared. This table shows the results for 4, 5 and 6 classes.

Again, from a descriptive point of view, the best result in *CCR* for predicting H_s considering a discretisation of 4 classes is obtained by the SVOREX method, which obtains a value of 65.169 ± 1.703; and the second best result is obtained by the method SVORIM, with a very similar value of 65.128 ± 1.794. With respect to the *MAE* metric, this fact is reversed, the best result is obtained by the SVORIM methodology, with a value of 0.359 ± 0.018, and the second best result is obtained by the SVOREX methodology with a value of 0.360 ± 0.017. In the case of F_e, the best result in *CCR* is obtained by the SVC1V1 method, with a value of 53.904 ± 1.816, and the second best result is also obtained by the SVORIM method, with a value of 52.890 ± 1.291. Concerning the *MAE* metric, the best result is for the SVORIM method, with a value of 0.505 ± 0.015, and the second best result is obtained by the SVOREX method, with a value of 0.515 ± 0.015. Summarising, the best results in *CCR* for H_s and F_e are obtained by the methods SVOREX and SVC1V1, respectively, and the best results in *MAE* for H_s and F_e are always obtained by the SVORIM method.

Table 7
Mean ranking results in *CCR* and *MAE* obtained by all the methods tested and the 12 datasets used. The best result is marked in boldface and the second one in italics.

Method	Mean ranking	
	CCR	*MAE*
ELMOR	5.83	4.17
KDLOR	6.08	5.96
ONN	6.75	6.46
POM	7.33	6.29
SVC1V1	*2.50*	3.58
SVC1VA	3.42	6.54
SVOREX	**1.33**	*1.75*
SVORIM	2.75	**1.25**
SVR	9.00	9.00

For 5 and 6 classes, the prediction of H_s and F_e is also worse in terms of *CCR* and *MAE*. Thus, for 5 classes and H_e, the best result is obtained by the SVOREX and SVORIM methods in *CCR* and *MAE*, respectively. For the prediction of F_e, the best result is obtained by the SVORIM method both in *CCR* and *MAE*. For 6 classes and H_e, the best result is obtained by the SVOREX and SVORIM methods in *CCR* and *MAE*, and for F_e, the SVC1V1 method obtains the best result in *CCR* and the SVORIM method obtains the best result in *MAE*.

The performance for F_e prediction is worse than the one obtained for H_s. The reason is that this problem does not have such a clear ordinal component as H_s, since in F_e the contribution of T_e is significant, which makes this variable harder to be predicted. On the other hand, note that the range of F_e values is much wider than that of H_s, resulting in a higher variance, more difficult to be represented by the model.

In order to better summarise these results, Table 7 shows the test mean rankings in terms of *CCR* and *MAE*, for the 9 methods considered in these experiments along all the 12 datasets: 6 datasets for each buoy, 3 for H_s (4, 5 and 6 classes) and 3 for F_e. For each dataset, a ranking of 1 is given to the best method, in average, and a 9 is given to the worst one. From this table, it can be seen that most support vector methods (except SVR) achieve the best performance. However, ordinal methods, SVOREX and SVORIM, achieve better performance respect to SVC1V1, SVC1VA and SVR, specially in *MAE*. For *CCR*, the decomposition of SVC1V1 also achieves a quite good performance, although one should take into account that this measure do no penalise mistakes according to the ordinal scale.

To quantify whether a statistical difference exists among the algorithms compared, a procedure is employed to compare multiple classifiers in multiple datasets (Demsar, 2006), using the average ranking (for *CCR* and *MAE*) of each method shown in Table 7. This procedure begins with a non-parametric Friedman's test, establishing the significance level at $\alpha = 0.05$. The test rejects the null-hypothesis that all of the algorithms perform similarly in mean ranking for all the metrics. Confidence interval, C_0, for Friedman's test is $C_0 = (0, F_{(\alpha=0.05)} = 2.05)$ and the corresponding F-values, F, for *CCR* and *MAE* are $F_{CCR} = 92.75 \notin C_0$ and $F_{MAE} = 57.87 \notin C_0$ respectively. Note that for *CCR* the differences are larger.

On the basis of this rejection and following the guidelines in Demsar (2006), the best performing methods (i.e., SVOREX and SVORIM) are considered as control methods for the following tests. These two methods are compared to the rest according to their rankings. Holm's test is considered, which involves computing the following statistics for comparing the average ranks (R_i and R_j) of the

Table 8
Results of the Holm procedure in terms of *CCR* and *MAE* using SVOREX and SVORIM as control methods, corrected α values, compared method and p-values, ordered by the number of comparison (i).

Control alg.: SVOREX			*CCR*		*MAE*	
i	$\alpha^*_{0.05}$	$\alpha^*_{0.10}$	Method	p_i	Method	p_i
1	0.0063	0.0125	SVR	0.0000_{++}	SVR	0.0000_{++}
2	0.0071	0.0143	POM	0.0000_{++}	SVC1VA	0.0000_{++}
3	0.0083	0.0167	ONN	0.0000_{++}	ONN	0.0000_{++}
4	0.0100	0.0200	KDLOR	0.0001_{++}	POM	0.0001_{++}
5	0.0125	0.0250	ELMOR	0.0001_{++}	KDLOR	0.0002_{++}
6	0.0167	0.0333	SVC1VA	0.0624	ELMOR	0.0307_{+}
7	0.0250	0.0500	SVORIM	0.2051	SVC1V1	0.1010
8	0.0500	0.1000	SVC1V1	0.2967	SVORIM	0.6547
Control alg.: SVORIM						
1	0.0063	0.0125	SVR	0.0000_{++}	SVR	0.0000_{++}
2	0.0071	0.0143	POM	0.0000_{++}	SVC1VA	0.0000_{++}
3	0.0083	0.0167	ONN	0.0004_{++}	ONN	0.0000_{++}
4	0.0100	0.0200	KDLOR	0.0029_{++}	POM	0.0000_{++}
5	0.0125	0.0250	ELMOR	0.0058_{++}	KDLOR	0.0000_{++}
6	0.0167	0.0333	SVOREX	0.2051	ELMOR	0.0091_{++}
7	0.0250	0.0500	SVC1VA	0.5510	SVC1V1	0.0369_{+}
8	0.0500	0.1000	SVC1V1	0.8231	SVOREX	0.6547

Statistically significant win for $\alpha = 0.05$ $(++)$ and $\alpha = 0.10$ $(+)$.

i-th and j-th method:

$$z = \frac{R_i - R_j}{\sqrt{\frac{J(J+1)}{6T}}},$$

where J is the number of algorithms and T is the number of datasets. The z value is used to find the corresponding p-value, which is then compared with an appropriate level of significance α. The test uses an adjusted value for α in order to compensate for multiple comparisons. The hypotheses are ordered by their significance, according to the p-values. Holm's test compares p_i with $\alpha^*_{\text{Holm}} = \alpha/(J-i)$, starting from the lowest p value. If the corresponding hypothesis is rejected, p_{i+1} is compared with $\alpha/(J-i+1)$. If not, the test ends. The results of the test are included in Table 8. Several conclusions can be drawn. The results of SVOREX are promising. Concretely, in *CCR*, it obtains significant differences with respect to all methods except the nominal support vector approaches, SVC1V1 and SVC1VA, and the ordinal support vector approach SVORIM. For the *MAE* metric, SVOREX also obtains significant differences with respect to all methods except SVC1V1 and SVORIM methodologies, but, in this case, SVOREX also obtains significant differences on SVC1VA. If the SVORIM method is considered as a control method, it significantly outperforms all other algorithms for *MAE* (except for SVOREX where non-significant differences were found), but for *CCR*, SVORIM does not statistically outperform the SVOREX method and the nominal support vector approaches, SVC1V1 and SVC1VA. Note again that accuracy is not an appropriate measure for ordinal regression dataset.

That said, generally, the ordinal methods SVOREX and SVORIM are the best election using *CCR* and *MAE* as comparison metrics for predicting H_s and F_e. In the case of *CCR*, the nominal methods SVC1V1 and SVC1VA also obtain competitive performance. Although regression is the most common approach to tackle this problem in previous works, note that the SVR technique has obtained significant worse results, indicating that the additional complexity derived from considering the problem as a regression task limits the performance in terms of *CCR* and *MAE*. Therefore, the ordinal methodologies are a good option for obtaining a first coarse prediction of the wave energy produced by a given WEC, they can be applied to systems at any location and also they can be tuned by the practitioner specialist by just adjusting the labels of the classes, depending on the WEC system managed and the wave zone under study.

6. Conclusions

In this paper eight different ordinal and nominal classifiers and one support vector regression algorithm have been evaluated in a problem of marine energy prediction. Specifically, significant wave height (H_s) and energy flux (F_e) prediction in a 6 h time-horizon is considered. Up until now, this problem had only been tackled with regression techniques, and, to our knowledge, this is the first work that introduces classification techniques in marine energy prediction. Moreover, the prediction of H_s and F_e is carried out by using meteorological input variables, from weather numerical models, including four grid points surrounding the considered WEC, what is also a novelty in these type of forecasting problems. The results obtained have shown that ordinal classifiers such as SVOREX and SVORIM obtain a good reconstruction in terms of H_s with respect to nominal classification and regression methods, taking into account the difficulty of the problem. These approaches are less accurate when F_e is considered as the objective variable, because of the contribution of the T_e component, but SVOREX and SVORIM also get the best results compared to other methods. Nominal methods, SVC1VA and SVC1V1, obtain in some cases similar results in terms of prediction accuracy (although this measure is not appropriate for range prediction problems), but they obtain worse results than ordinal methods for metrics which consider the severity of the error committed. The methodologies proposed in this paper based on multi-class classification are novel for marine energy prediction, and allows to tune the algorithms and even thresholds classes depending on the WEC system or zone under study, in order to improve the performance of the prediction system.

Acknowledgements

This work has been partially supported by the projects TIN2014-54583-C2-1-R and TIN2014-54583-C2-2-R, of the Spanish Ministerial Commission of Science and Technology (MICYT), FEDER funds and the P11-TIC-7508 project of the Junta de Andalucía (Spain).

References

Agrawal, J.D., Deo, M.C., 2004. Wave parameter estimation using neural networks. Mar. Struct. 17, 536–550.
Akpinar, A., Kömürcü, M.I., 2013. Assessment of wave energy resource of the Black Sea based on 15-year numerical hindcast data. Appl. Energy 101, 502–512.
Alamian, R., Shafaghat, R., Miri, S.J., Yazdanshenas, N., Shakeri, M., 2014. Evaluation of technologies for harvesting wave energy in Caspian sea. Renew. Sustain. Energy Rev. 32, 468–476.
Arinaga, R.A., Cheung, K.F., 2012. Atlas of global wave energy from 10 years of reanalysis and hindcast data. Renew. Energy 39, 49–64.
Baccianella, S., Esuli, A., Sebastiani, F., 2009. Evaluation measures for ordinal regression. In: Proceedings of the Ninth International Conference on Intelligent Systems Design and Applications, ISDA'09, pp. 283–287.
Bahaj, A.S., 2011. Generating electricity from the oceans. Renew. Sustain. Energy Rev. 15, 3399–3416.
Bender, R., Grouven, U., 2006. Ordinal logistic regression in medical research. J. R. Coll. Physicians Lond. 31 (5), 546–551.
Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer, New York.
Cardoso, J.S., da Costa, J.F.P., 2007. Learning to classify ordinal data: the data replication method. J. Mach. Learn. Res. 8, 1393–1429.
Castro, A., Carballo, R., Iglesias, G., Rabuñal, J.R., 2014. Performance of artificial neural networks in nearshore wave power prediction. Applied Soft Computing 23, 194–201. http://dx.doi.org/10.1016/j.asoc.2014.06.031.
Chu, W., Keerthi, S.S., 2007. Support vector ordinal regression. Neural Comput. 19 (3), 792–815.
Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297.

Cruz-Ramírez, M., Hervás-Martínez, C., Sánchez-Monedero, J., Gutiérrez, P.A., 2014. Metrics to guide a multi-objective evolutionary algorithm for ordinal classification. Neurocomputing 135, 21–31.
Demsar, J., 2006. Statistical comparisons of classifiers over multiple data sets. J. Mach. Learn. Res. 7, 1–30.
Deng, W.-Y., Zheng, Q.-H., Lian, S., Chen, L., Wang, X., 2010. Ordinal extreme learning machine. Neurocomputing 74 (1), 447–456.
Deo, M.C., Naidu, C.S., 1998. Real time wave prediction using neural networks. Ocean Eng. 26 (3), 191–203.
Esteban, M., Leary, D., 2012. Current developments and future prospects of offshore wind and ocean energy. Appl. Energy 90, 128–136.
Falcão, A.F.de O., 2010. Wave energy utilization: a review of the technologies. Renew. Sustain. Energy Rev. 14, 899–918.
Frank, E., Hall, M., 2001. A simple approach to ordinal classification. In: Proceedings of the 12th European Conference on Machine Learning, EMCL'01, pp. 145–156.
Fusco, F., Ringwood, J.V., 2010. Short-term wave prediction with AR models in real-time optimal control of wave energy converters. In: Proceedings of IEEE International Symposium on Industrial Electronics, ISIE 2010, pp. 2475–2480.
Fusco, F., Ringwood, J.V., 2010b. Short-term wave prediction for real-time control of wave energy converters. IEEE Trans. Sustain. Energy 1 (2), 99–106.
Gutiérrez, P.A., Pérez-Ortiz, M., Fernández-Navarro, F., Sánchez-Monedero, J., Hervás-Martínez, C., 2012. An experimental study of different ordinal regression methods and measures. In: Hybrid Artificial Intelligent Systems. Lecture Notes in Computer Science, vol. 7209. Springer Berlin, Heidelberg, pp. 296–307.
Gutiérrez, P.A., Salcedo-Sanz, S., Hervás-Martínez, C., Carro-Calvo, L., Sánchez-Monedero, J., Prieto, L., 2013. Ordinal and nominal classification of wind speed from synoptic pressure patterns. Eng. Appl. Artif. Intell. 26 (3), 1008–1015.
Heras-Saizarbitoria, I., Zamanillo, I., Laskurain, I., 2013. Social acceptance of ocean wave energy: a case study of an OWC shoreline plant. Renew. Sustain. Energy Rev. 27, 515–524.
Hsu, C.W., Lin, C.J., 2002. A comparison of methods for multi-class support vector machines. IEEE Trans. Neural Netw. 13 (2), 415–425.
Huang, G.-B., Zhou, H., Ding, X., Zhang, R., 2012. Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst. Man Cybern. B 42 (2), 513–529.
Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Am. Meteorol. Soc. 77, 437–470.
Kazeminezhad, M.H., Etemad-Shahidi, A., Mousavi, S.J., 2005. Application of fuzzy inference system in the prediction of wave parameters. Ocean Eng. 32 (14–15), 1709–1725.
Kramer, S., Widmer, G., Pfahringer, B., De Groeve, M., 2001. Prediction of ordinal classes using regression trees. Fundam. Inform. 47 (1), 1–13.
Lawrence, J., Sedgwick, J., Jeffrey, H., Bryden, I., 2013. An overview of the U.K. marine energy sector. Proc. IEEE 101(4), 876–890.
Lin, H.-T., Li, L., 2012. Reduction from cost-sensitive ordinal ranking to weighted binary classification. Neural Comput. 24 (5), 1329–1367.
Lindroth, S., Leijon, M., 2011. Offshore wave power measurements: a review. Renew. Sustain. Energy Rev. 15, 4274–4285.
Lippmann, R.P., 1989. Pattern classification using neural networks. IEEE Trans. Neural Netw. 27, 47–64.
López, I., Andreu, J., Ceballos, S., Martínez de Alegría, I., Kortabarria, I., 2013. Review of wave energy technologies and the necessary power-equipment. Renew. Sustain. Energy Rev. 27, 413–434.
Mahjoobi, J., Mosabbeb, E.A., 2009. Prediction of significant wave height using regressive support vector machines. Ocean Eng. 36, 339–347.
Mahjoobi, J., Etemad-Shahidi, A., Kazeminezhad, M.H., 2008. Hindcasting of wave parameters using different soft computing methods. Appl. Ocean Res. 30 (1), 28–36.
McCullagh, P., 1980. Regression models for ordinal data. J. R. Stat. Soc. Ser. B: Methodol. 42 (2), 109–142.
National Data Buoy Center (NDBC). Station 46001 (LLNR 984) – Wester Gulf of Alaska – 175NM SE of Kodiak, AK. [Online]. Available at: ⟨http://www.ndbc.noaa.gov/station_history.php?station=46001⟩.
National Data Buoy Center (NDBC). Station 41013 (LLNR 815) – Frying Pan Shoals, NC Buoy. [Online]. Available at: ⟨http://www.ndbc.noaa.gov/station_history.php?station=41013⟩.
National Oceanic and Atmospheric Administration (NOAA), National Data Buoy Center (NDBC). [Online]. Available at: ⟨http://www.ndbc.noaa.gov⟩.
Nitsure, S.P., Londhe, S.N., Khare, K.C., 2012. Wave forecasts using wind information and genetic programming. Ocean Eng. 54, 61–69.
Özger, M., 2011. Prediction of ocean wave energy from meteorological variables by fuzzy logic modeling. Exp. Syst. Appl. 38 (5), 6269–6274.
Reikard, G., 2013. Integrating wave energy into the power grid: simulation and prediction. Ocean Eng. 73, 168–178.
Reikard, G., Pinson, P., Bidlot, J.R., 2011. Forecasting ocean wave energy: the ECMWF wave model and time series methods. Ocean Eng. 38, 1089–1099.
Richter, M., Magaña, M.E., Sawodny, O., Brekken, T.K., 2013. Nonlinear model predictive control of a point absorber wave energy converter. IEEE Trans. Sustain. Energy 4 (1), 118–126.
Sánchez-Monedero, J., Salcedo-Sanz, S., Gutiérrez, P.A., Casanova-Mateo, C., Hervás-Martínez, C., 2014. Simultaneous modelling of rainfall occurrence and amount using a hierarchical nominal-ordinal support vector classifier. Eng. Appl. Artif. Intell. 34, 199–207.
Smola, A., Schlkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14 (3), 199–222.
Sun, B.-Y., Li, J., D Wu, D., Zhang, X.-M., Li, W.-B., 2010. Kernel discriminant learning for ordinal regression. IEEE Trans. Knowl. Data Eng. 22 (6), 906–910.
The NCEP/NCAR Reanalysis Project. NOAA/ESRL Physical Sciences Division. [Online]. Available at: ⟨http://www.esrl.noaa.gov/psd/data/reanalysis/reanalysis.shtml⟩.
The Offshore Renewable Energy in Scotland Website. [Online]. Available at: ⟨http://www.scotland.gov.uk/Topics/marine/marineenergy⟩.
The Pelamis Wave Power Website. [Online]. Available at: ⟨http://www.pelamiswave.com/⟩.
Tsai, C-P., Lin, C., Shen, J-N., 2002. Neural network for wave forecasting among multi-stations. Ocean Eng. 29 (13), 1683–1695.
Unidata. Network Common Data Form (NetCDF). [Online]. Available at: ⟨http://www.unidata.ucar.edu/software/netcdf/⟩.
Verwaeren, J., Waegeman, W., De Baets, B., 2012. Learning partial ordinal class memberships with kernel-based proportional odds models. Comput. Stat. Data Anal. 56 (4), 928–942.
Zanaganeh, M., Jamshid-Mousavi, S., Etemad-Shahidi, A.F., 2009. A hybrid genetic algorithm-adaptive network-based fuzzy inference system in prediction of wave parameters. Eng. Appl. Artif. Intell. 22 (8), 1194–1202.