Hybridization of neural network models for the prediction of extreme significant wave height segments

Antonio M. Durán-Rosal*, Juan C. Fernández*, Pedro A. Gutiérrez* and César Hervás-Martínez*
* Department of Computer Science and Numerical Analysis, University of Córdoba,
Rabanales Campus, Albert Einstein Building, 3rd Floor, 14071, Córdoba, Spain.
Email: {i92duroa, jfcaballero, pagutierrez, chervas}@uco.es

Abstract—**This work proposes a hybrid methodology for the detection and prediction of Extreme Significant Wave Height (ESWH) periods in oceans. In a first step, wave height time series is approximated by a labeled sequence of segments, which is obtained using a genetic algorithm in combination with a likelihood-based segmentation (GA+LS). Then, an artificial neural network classifier with hybrid basis functions is trained with a multiobjetive evolutionary algorithm (MOEA) in order to predict the occurrence of future ESWH segments based on past values. The methodology is applied to a buoy in the Gulf of Alaska and another one in Puerto Rico. The results show that the GA+LS is able to segment and group the ESWH values, and the neural network models, obtained by the MOEA, make good predictions maintaining a balance between global accuracy and minimum sensitivity for the detection of ESWH events. Moreover, hybrid neural networks are shown to lead to better results than pure models.**

I. INTRODUCTION

Prediction of Significant Wave Height (SWH) values is an important issue for the design of coastal and offshore structures [1], for oil and gas extraction, for estimation of renewable energy production and for other activities. In recent years, Machine Learning (ML) techniques have been used in order to improve the prediction of SWH treated as a regression problem [2].

Specifically, artificial neural network (ANN) models have been successfully used in reconstruction and prediction of waves, improving the accuracy of numerical models [3], [4]. When applying ANNs, there are many different options in the literature, such as sigmoidal unit neural networks (SUNN) [5], radial basis function (RBF) neural networks [5], product unit neural networks (PUNN) [6], or some related methods, such as projection pursuit learning [7], generalized additive models [8], or hybridization of multivariate adaptive splines (MARS) [9]. All of them can be seen as different basis functions models. The main problem of these approaches is to adjust the optimal number and the typology of the corresponding basis functions.

Hybrid models of ANNs can be found in the literature, where different activation/transfer functions are used for the nodes in the hidden layer. The hybridization of different basis functions has been made using either one single hybrid hidden layer or several connected pure layers. According to Duch and Jankowski [10], mixed transfer functions within one network may be introduced in two ways. In the first way, a constructive method selects the most promising functions from a pool of candidates in RBF-like architecture, and it is added to the network [11]. In the second approach, starting from a network that already contains several types of functions (such as Gaussian and sigmoidal functions), pruning or regularization techniques are used to reduce the number of functions [10]. Iulian proposed a methodology including three distinct modules implementing a hybrid feed-forward ANN, namely a Gaussian type RBF network, a principal component analysis (PCA) process, and a multi-layer perceptron ANN [12]. In this context, it is worth emphasizing the paper by Cohen and Intrator [13], which is based on the duality and complementary properties of projection-based functions (SU and PU) and kernel typology (RBF). This hybridization of models has been justified theoretically in [14], who demonstrated that any continuous function can be decomposed into two mutually exclusive functions, such as radial (kernel functions) and crest ones (based on the projection). Although theoretically this decomposition is justified, in practice, it is difficult to apply gradient methods to separate the different locations of a function (in order to adjust them by means of a combination of RBF) and then to estimate the residual functions by means of a functional approach based on projections, all without getting trapped in local optima in the procedure of error minimizations [9]. In order to avoid these problems, evolutionary algorithms (EAs) can be used for automatically designing and adjusting the structure and weights of hybrid ANNs [15], and this is methodology applied in this paper.

Coming back to SWH prediction, the interest of developing novel methodologies for tackling the prediction of Extreme SWHs (ESWH) has increased during last years [16], [17]. In general, these methodologies use probabilistic distributions of extreme values [18], where the main assumption is that a global change in time causes the occurrence of greater and longer SWHs, affecting ports, beaches, and fishing and navigation activities.

ESWH prediction models also assume that yearly samples of weather are i.i.d. However, large inter-year variability is frequently observed, and this hypothesis should not be assumed to be true [19]. Several evidences suggest that storm intensity over the ocean is changing due to climate change, and, because of this reason, the parameters of the probabilistic distribution of extreme values should change too [20].

In this paper, we propose a novel methodology where the i.i.d. assumption is not made. As a first step, a time series segmentation genetic algorithm is combined with a likelihood-based segmentation (GA+LS) and applied to discover four types of SWH segments, where one of them is related to extreme waves (ESWHs). Then, from this segmentation, prediction is tackled by using a binary classifier with the aim of deciding if, at a given moment, the next segment is going to be a ESWH segment or not, using the characteristics of previous ones. The classifier is trained by a multi-objective evolutionary algorithm (MOEA), especially adapted to imbalanced datasets, given that the number of ESWH segments is much lower than the number of non ESWH segments. The population of individuals in the MOEA is formed by feed-forward hybrid ANNs with a hidden layer of SUs or PUs in combination with RBFs. The ANNs are fitted to perform a balanced classification, optimizing both their typology, structure and weights.

This paper is organized as follows: Section II includes the proposed methodology, Section III presents the description of the time series, the performed experiments and the discussion of the results. Finally, Section IV concludes with some tips.

II. Methodology

The proposed methodology can be divided into two stages (detection and prediction). Firstly, to obtain the segments of the SWH time series, a GA is used. In this stage, the algorithm tries to divide the time series into a set segments and a clustering of these segments is done to group them. Once the segmentation is finished, the cluster grouping ESWH segments is decided, and the time series is converted into a sequence of labels (corresponding to the clusters assigned to each segment). Finally, this sequence is used for prediction (second stage), where, based on the characteristics of past segments, we try to predict whether the next segment corresponds to ESWH or not. The summary of this methodology can be see in Fig. 1.

A. Stage 1: Detection of ESWH segments

The formal definition of the segmentation stage is the following: given a time series $Y = \{y_n\}_{n=1}^{N}$, our goal is to find a set of m segments defined by $\mathbf{t} = \{t_i\}_{i=1}^{m-1}$ cut points. So, the segments $\mathbf{s} = \{s_1, s_2, \dots, s_m\}$ are defined as $s_1 = \{y_1, \dots, y_{t_1}\}, s_2 = \{y_{t_1}, \dots, y_{t_2}\}, \dots, s_m = \{y_{t_{m-1}}, \dots, y_N\}$. After this step, it is necessary the clustering of these segments in k groups, where $(k << m)$ and one of them is related to ESWH. A label $l \in \{1, \dots, k\}$ is assigned to each segment. The segmentation is finally improved with a likelihood-based method (assuming a beta probabilistic distribution of extreme values for all the segments), where some new cut points are discovered.

The segmentation GA is similar to the one the proposed in [21]. However, in this paper, two main changes have been made:

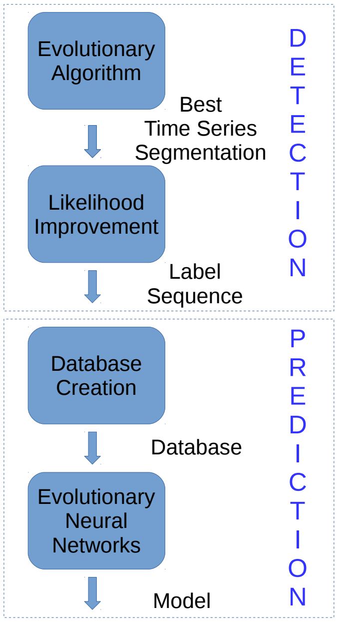


Fig. 1. Summary of the proposed methodology.

1) The first one corresponds to the fitness function. The clustering index used in [21] (COP index) has been substituted by the Caliński and Harabasz index (CH), which allows us to determine well separated and compact clusters. In addition, CH leads to very good results in the experimental study presented in [22]. CH is defined by:

$$CH = \frac{(\sum_{i=1}^{k} n_i ||\bar{\mathbf{c}}_\mathbf{i} - \bar{\mathbf{s}}||^2) \cdot (m-k)}{(\sum_{i=1}^{k} \sum_{\mathbf{s} \in c_i} ||\mathbf{s} - \bar{\mathbf{c}}_\mathbf{i}||^2) \cdot (k-1)}, \quad (1)$$

where m is the number of patterns (mapped segments in our case), n_i is the number of segments of cluster i, $\bar{\mathbf{c}}_\mathbf{i}$ is the centroid of cluster i, $\bar{\mathbf{s}}$ is the overall mean of all mapped segments, and $||\bar{\mathbf{c}}_\mathbf{i} - \bar{\mathbf{s}}||^2$ is the Euclidean distance between two vectors. According to [21], a mapped segment $\mathbf{s}_i$ is a 5-dimensional array, including the variance (S_i^2), the kurtosis (γ_{1i}), the skewness (γ_{2i}), the slope of a linear regression (a_i), and the autocorrelation (AC_i), of the values of the i-th segment. In this way, $\mathbf{s}_i = (S_i^2, \gamma_{1i}, \gamma_{2i}, a_i, AC_i)$.

2) The second one is that the likelihood-based segmentation procedure assumes a beta distribution instead a normal distribution. The beta distribution is specifically designed for correctly representing extreme values.

Further details of the hybrid segmentation GA can be checked in [21].

B. Stage 2: Prediction of ESWH segments

Once the segmentation is finished, the best time series segmentation given by the GA+LS algorithm is used, and the ESWH cluster is identified. The database for the prediction stage is constructed as follows. The three mapped segments previous to the segment to be predicted are used as independent variables and the dependent variable is a binary label indicating whether the next segment corresponds to extreme waves (i.e. if it is a ESWH segment or not). Some other details about the database creation are given in Section III.

The main inconvenient of the derived dataset is that it is extremely imbalanced with respect to the ESWH class, highly outnumbered by non extreme segments. For this reason, the MOEA described in [23] has been considered, which tackles this kind of problems by using two non-cooperative metrics or fitness functions: the Correctly Classified Rate, CCR, and the sensitivity of the worst classified class, named as Minimum Sensitivity, MS. These metrics obtain good results in the maximization of the overall precision, maintaining the accuracy per class as balanced as possible, as shown in [23].

The proposed MOEA is based on the original NSGA2 algorithm [24], and it has been modified and adapted for the evolution of hybrid ANN models with one feed forward hidden layer, given their ability as classification methods [25], [26]. The algorithm, named $NSGA2-ANN$ in this work and whose scheme is shown in Figure 2, estimates the coefficients and the structure of each ANN in the evolutionary population. The following subsections include further details about the algorithm.

1) Hybrid basis functions: Given that the algorithm in [23] evolved pure ANNs, some specific features have to be included to consider hybrid ANNs. From a statistical point of view, feed forward ANN models with one hidden-layer are generalized linear regression models, considering a linear combination of non-linear projections of the input variables, $f_j(\mathbf{x}, \mathbf{w}_j)$ in the following way:

$$f(\mathbf{x}, \boldsymbol{\theta}) = \beta_0 + \sum_{j=1}^{M} \beta_j f_j(\mathbf{x}, \mathbf{w}_j), \tag{2}$$

where M is the number of non-linear combinations, $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \mathbf{w}_1, \dots, \mathbf{w}_M\}$ are the parameters of the model, $f_j(\mathbf{x}, \mathbf{w}_j)$ are each of the basis functions, $\mathbf{w}_j$ are the set of parameters associated to each basis functions, and $\mathbf{x} = \{x_1, \dots, x_k\}$ are the input variables associated to the problem. This kind of models, which include ANN, are called linear models of basis functions [27].

In this paper, non-parametric adaptive methods are used to estimate the binary label, ESWH segment or not (nESWH). We try to determine the f function so that the true relationship between the response variable $y \in \{\text{ESWH}, \text{nESWH}\}$ and the dependent variables $\mathbf{x}$ can be uncovered, where $\hat{y} = f(\mathbf{x})$. To do so, the training dataset, D, consists of n observed responses at some known predictor locations, so $D = \{\mathbf{x}_l, y_l\}$ for $l = 1, 2, \dots, n$.

More concretely, let $f : A \subset \mathbb{R}^k \rightarrow \{\text{ESWH}, \text{nESWH}\}$ and the dataset $D = \{(\mathbf{x}_l, f(\mathbf{x}_l)) : \mathbf{x}_l \in A\}$. Supposing $\varepsilon > 0$, we define two functions $g_1 = f_{\text{RBF}} + f_{\text{PU}}$ and $g_2 = f_{\text{RBF}} + f_{\text{SU}}$, such that $|f(\mathbf{x}) - g_1(\mathbf{x})| < \varepsilon$ and $|f(\mathbf{x}) - g_2(\mathbf{x})| < \varepsilon$, $\forall \mathbf{x} \in A$. The functions are given by:

$$f_{\text{RBF}}(\mathbf{x}) = \sum_{j=1}^{m_1} \alpha_j \exp\left(\frac{||\mathbf{x} - \mathbf{c}_j||}{{r_j}^2}\right), \tag{3}$$

$$f_{\text{PU}}(\mathbf{x}) = \sum_{j=1}^{m_2} \beta_j \prod_{i=1}^{k} x_i^{w_{ji}}, \tag{4}$$

$$f_{\text{SU}}(\mathbf{x}) = \sum_{j=1}^{m_3} \gamma_j \frac{1}{1 + e^{-\sum_{i=1}^{k} w_{ji} x_i}}, \tag{5}$$

where $\mathbf{c}_j$ and r_j are the centroid and the radius of the j-th RBF unit, respectively; α_j, β_j and γ_j are the weights of the output neuron with respect to the j-th hidden neuron; and w_{ji} are the weights of the j-th hidden neuron with respect to the i-th input neuron.

Given a combination of basis functions, the algorithm has to find the number of neurons of each type (m_1, m_2 and m_3) and the parameters of the ANN ($\mathbf{c}_j$, r_j, α_j, β_j, γ_j, w_{ji}) trying to minimize the classification error of the training dataset.

For evaluating the error, the output layer is interpreted from the point of view of probability through the use of the *Softmax* activation function [27]. Therefore, a function $g(\mathbf{x}, \boldsymbol{\theta})$ is used to predict the membership degree of an input pattern with respect to one of the two classes. Taking this consideration into account, it can be seen that the class predicted by the ANN corresponds to the neuron on the output layer whose g value is the greatest:

$$g_l(\mathbf{x}, \boldsymbol{\theta}) = \frac{\exp f_l(\mathbf{x}, \boldsymbol{\theta})}{\sum_{j=1}^{2} \exp f_j(\mathbf{x}, \boldsymbol{\theta})}, l = 1, 2. \tag{6}$$

2) NSGA2-ANN: Following the scheme shown in Figure 2, $NSGA2-ANN$ begins with the random generation of a population of ANNs. Once the individuals have been evaluated using the CCR and MS objective functions[1], the population is sorted depending on the Pareto dominance concept [28]. After that, new individuals (for the offspring) are generated by binary tournament techniques. Five mutation operators are applied to the individuals created by the selection, four structural mutations (add neurons, delete neurons, add links and delete links) and one parametric mutation, which consists in adding a Gaussian noise to the weights of the links of ANNs throughout the evolution. This Gaussian noise is a random value of a normal distribution with zero mean and $1 - A(c)$ variance, where $A(c)$ is the fitness value of the mutated individual c (Note that the variance is a decreasing value along the generations, for more details see [29]). Once the mutation process is carried out, the mutated population (offspring) is evaluated for both objective functions, and it is combined with the parent population, the size of the population being $2N$. The resulting population is sorted according to the Pareto dominance concept, and the N best individuals are selected as parents for the next generation. The process starts again if the maximum number of generations have not been reached.

---

[1] The definition of these functions can be found in [23]

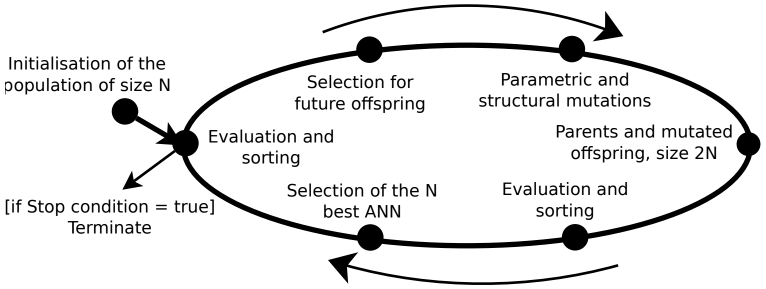


Fig. 2. General scheme for $NSGA2-ANN$.

Once the Pareto front is built into the last generation of training, two strategies are used to choose the best individuals for the addressed problem. The first strategy selects the best model in CCR, which is the upper individual of the Pareto front, named $NSGA2-ANN-CCR$. The second strategy selects the best model in MS, which is the bottom individual from the Pareto front, named $NSGA2-ANN-MS$. The MS individual is very important to predict when ESWH occur, which, in this problem, corresponds with the minority class.

The basis functions used in the population of $NSGA2-ANN$ can be one of these five options: 1) pure sigmoidal units (SU) [30], 2) pure product units (PU) [31], 3) pure RBF units (RBF) [5], 4) hybrid units combination of SU+RBF, and 5) hybrid units combination of PU+RBF [15]. The models with the highest values in CCR and MS metrics, which are also interpretable (i.e. with a lower number of neurons and weights) will be used to decided which combination is better (SU, PU, RBF, SU+RBF or PU+RBF). Therefore, the resulting methodologies for the experimental section are named $NSGA2-ANN-CCR-XX$ or $NSGA2-ANN-MS-XX$, being XX one of the five combinations of basis functions.

III. EXPERIMENTS AND RESULTS

The time series for the first detection stage are built from the SWH of two buoys belonging to the National Data Buoy Center of the USA [32] and are located in different places to check the robustness of the proposed methodology: 1) a buoy in the Gulf of Alaska, with identification number 46001, and 2) a buoy at northeast Puerto Rico, with identification number 41043. Data from 1st January 2008 to 31st December 2013 are considered for the buoy 46001, and data from 1st January 2011 to 31st December 2015 are considered for the buoy 41043, using a 6-hour time horizon approximation, with a total of 8767 and 7303 observations, respectively. In TABLE I, the main characteristics of these buoys are shown: minimum, maximum, mean and standard deviation of SWH, Kurtosis coefficient, skewness of SWH, water depth and location. The complete time series used in the experiments can be seen in Fig. 3.

For the second prediction stage, a database is built by using the associated labels of the detection phase. The results of the detection stage are a sequence of segments, summarized by the 5 characteristics discussed in Section II-A together with a label indicating the assigned cluster. The dataset used to determine if a future segment will be considered as ESWH event considers the characteristics of the three previous segments (with a total of 15 inputs). Each pattern of the dataset is formed by the input variables $\mathbf{s}_{i-3}$, $\mathbf{s}_{i-2}$, $\mathbf{s}_{i-1}$ and the binary output at the time i.

TABLE I
INFORMATION OF THE CONSIDERED BUOYS, 46001 AND 41043.

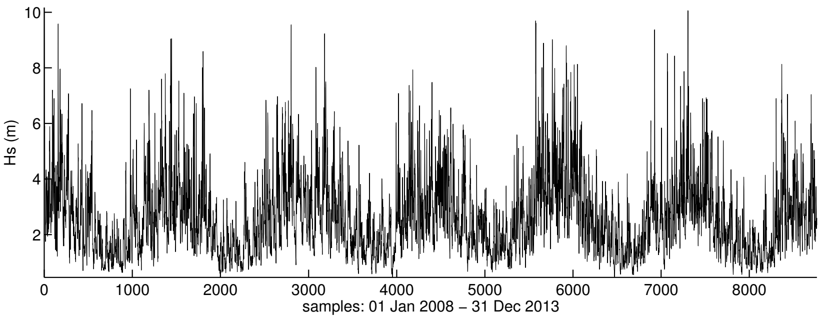
Buoy Id	Min(m)	Max(m)	Mean(m)	SD(m)
46001	0.46	10.17	2.653	1.404
41043	0.03	6.12	1.763	0.591
Buoy Id	Kurtosis	Skewness(m)	Depth(m)	Location
46001	4.714	1.162	4206	56.304 N 147.920 W
41043	5.451	1.06	5289	21.132 N 64.856 W
Buoy Id	#Patterns	#Inputs	#Classes	#Patterns per class
46001	379	15	2	347-32
41043	298	15	2	264-34

The total number of patterns obtained for the prediction phase (see TABLE I) for the buoy 46001 is 379 (corresponding to the number of segments obtained in the detection stage minus three). For the buoy 41043, the total number of patterns is 298. Note that the number of nESWH patterns in each buoy (347 and 264) is very with respect to the prediction of ESWH (32 and 34). This proportion clearly leads to an unbalanced classification problem.

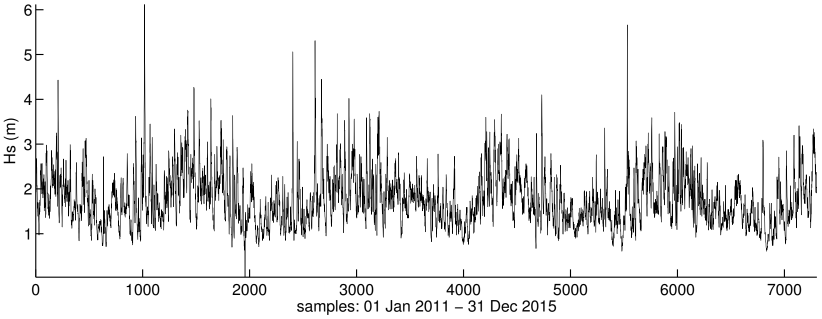
A. Experimental design

In the detection stage, the following parameters has to be adjusted: population size (P), crossover and mutation probabilities (p_c and p_m), percentage of mutated cut point ($\%m_{points}$), number of generations (g), initial minimum and maximum segment size (s_{min} and s_{max}), number of iterations of k-means ($\#k$-mean) and number of clusters (k). TABLE II shows the values of the parameters obtained by a trial and error procedure.

On the prediction stage, and for each buoy, a stratified holdout cross-validation procedure is applied for training and test the ANN models. Approximately, 75% of the patterns were randomly selected for the training set and the remaining 25% for the test set. Given that the $NSGA2-ANN$ algorithm is stochastic, it was run 30 times, using the parameter values



(a) Buoy Id 46001



(b) Buoy Id 41043

Fig. 3. Time series of the buoy id 46001 in the Gulf of Alaska and the buoy id 41043 in Puerto Rico.

TABLE II
PARAMETERS USED IN THE DETECTION STAGE.

P	p_c	p_m	$\%m_{points}$	g
100	0.8	0.2	20%	100
s_{min}	s_{max}	$\#k$-means	$k*$	
20	120	20	4 - 5	

* $k = 5$ for buoy id 46001, and $k = 4$ for buoy id 41043

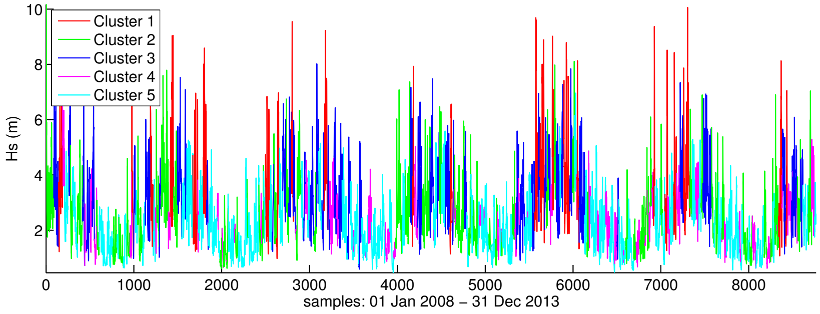
established and discussed in [23]. The only difference is that, in this work, local search (or lifetime learning) is not applied. The five different options concerning the transfer functions of the models are tested. The algorithm obtains a Pareto front where the best model in CCR and the best model in MS are selected. The support vector machine classifier, SVM, a decision tree algorithm, $C4.5$, and their cost sensitive versions ($SVM{-}CS$ and $C4.5 - CS$, respectively) have also been run using the Weka machine learning framework [33]. These algorithms are used for comparison, because they can be considered some of the most representative methods in the literature for classification tasks, and also because their results show how the MS metric in $NSGA2 - ANN$ is essential to obtain a good classification for both classes. To evaluate the performance of all methods, the most common metrics for binary classification have been selected [34]: CCR, MS and AUC or Area under Roc Curve.

B. Discussion

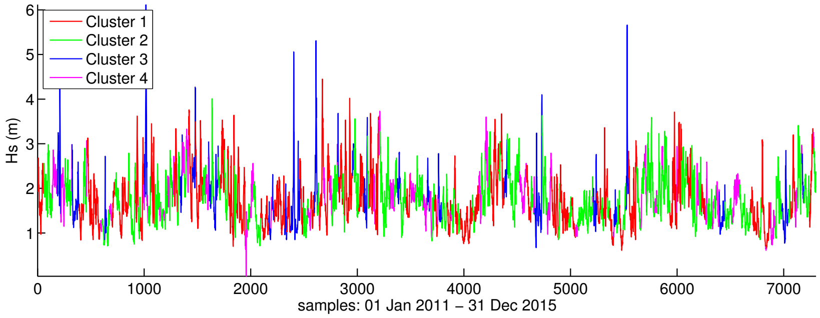
Fig. 4 shows the best segmentations produced by the GA in combination with the likelihood-based segmentation, in terms of the fitness function. As we can see, in buoy id 46001, the results show that the red cluster groups segments with extreme wave height. In the case of buoy id 41043, the cluster which groups ESWH segments is the dark blue one. In addition, the mean results of the 30 runs of our algorithm allow us observe that applying the likelihood-based segmentation (assuming a beta distribution) to the final solution of the GA improved fitness function from 4629.27 to 6804.93 in the case of buoy 46001, and from 4494.99 to 4775.17 in the case of buoy id 41043. For this reason, the application of this type of distribution to the segmentation/detection of ESWHs can be confirmed to be suitable.

After selecting considering the best fitness segmentation, we tackle the prediction stage. For this stage, Table III shows the mean and standard deviation (SD) of the results extracted from 30 runs for each configuration and for each buoy.

As can be observed, SVM and $C4.5$ methodologies lead to the best results in CCR. The problem of these algorithms is that they do not classify patterns of the minority class, that is, they do not predict correctly ESWH segments, obtaining values of MS in the range $0 - 5\%$ for both buoys (not being useful for these problems). These results are far worse than those obtained in MS by the $NSGA2{-}ANN$ methodologies. With respect to AUC, the results of SVM and $C4.5$ are worse than the values provided by the other methods for buoy id 46001, and very similar for buoy id 41043 (in this buoy, the



(a) Segmentation of Buoy Id 46001



(b) Segmentation of Buoy Id 41043

Fig. 4. Segmentations of Buoy id 46001 and Buoy id 41043 time series.

prediction is more difficult for all methodologies). The cost sensitive versions of these algorithms get better results in terms of MS and AUC than the same versions without costs, but $NSGA2-ANN$ methodologies continue to be the best ones.

After observing that SVM, $C4.5$, $SVM-CS$ and $C4.5-CS$ methodologies do not achieve good results in terms of MS and AUC compared to the proposed methodology, the following results obtained with the $NSGA2-ANN$ methodologies can be discussed:

- **Buoy id 46001:** The $NSGA2-ANN-CCR-SU+RBF$ and $NSGA2-ANN-CCR-PU+RBF$ hybrid ANNs obtain the best results in CCR, 0.905 ± 0.012 and 0.900±0.016, respectively, and low values (although better than SVM and $C4.5$) in MS. $NSGA2-ANN-MS-SU+RBF$ and $NSGA2-ANN-MS-PU+RBF$ obtain the best results in MS, 0.605±0.089 and 0.634±0.078 respectively. The results in CCR for these methodologies are not as high as those obtained for the methodologies that use the CCR extreme of the Pareto front, but they are acceptable, 0.743 ± 0.039 and 0.748± 0.039, respectively. Therefore, for this buoy, hybrid ANN models are found to be more useful than pure ANN models, because the combinations $SU+RBF$ and $PU+RBF$ get good results in CCR and MS.
- **Buoy id 41043:** The $NSGA2-ANN-CCR-PU+RBF$ methodology obtains the best results in CCR, 0.864±0.022, and a low value (although better than SVM and $C4.5$) in MS. The hybrid $NSGA2-ANN-MS-PU+RBF$ method obtains the best result in MS, 0.341±0.140, while the CCR result, 0.586±0.069, is not so good. $NSGA2-ANN-MS-RBF$ leads to a good compromise between CCR and MS, with values of 0.703±0.052 and 0.326±0.170, respectively. Therefore, the methodology recommended for this buoy is $NSGA2-ANN-MS-RBF$, providing very acceptable values for the three metrics.

IV. CONCLUSIONS

This paper presents a two-stage methodology to detect and predict segments where there are one or more extreme height waves. In the first stage, a detection of ESWH events with the combination of a genetic algorithm and a likelihood-based segmentation is carried out. The second stage uses the segmentation obtained to perform binary predictions, deciding if ESWH segments will occur or not.

On the one hand, in the detection stage, results agree that the application of the likelihood-based segmentation assuming a beta distribution is suitable for this type of time series with extreme values. On the other hand, for the prediction stage, the use of a MOEA for designing ANN models improves the prediction of ESWH events, because traditional methods tends to ignore the minority class.

Among the ANN architectures proposed, those using hybridization of the basis functions should be considered, because these hybrid methodologies lead to better results than pure models.

TABLE III

PREDICTION RESULTS: CCR, MS AND AUC RESULTS FOR THE TEST SETS. MEAN AND STANDARD DEVIATION (SD) ARE EXTRACTED FROM 30 RUNS AND ARE EXPRESSED IN THE RANGE $[0, 1]$.

Buoy 46001			
Method	CCR	MS	AUC
	Mean±SD	Mean±SD	Mean±SD
NSGA2-ANN-CCR-**SU**	0.880 ± 0.025	0.075 ± 0.117	0.660±0.094
NSGA2-ANN-MS-**SU**	0.780 ± 0.040	0.584 ± 0.089	**0.731±0.062**
NSGA2-ANN-CCR-**PU**	0.898 ± 0.021	0.038 ± 0.058	0.591±0.100
NSGA2-ANN-MS-**PU**	0.755 ± 0.034	0.598 ± 0.092	0.697±0.040
NSGA2-ANN-CCR-**RBF**	0.879 ± 0.020	0.133 ± 0.143	0.663±0.109
NSGA2-ANN-MS-**RBF**	0.814 ± 0.033	0.525 ± 0.133	0.711±0.077
NSGA2-ANN-CCR-**SU+RBF**	0.905 ± 0.012	0.046 ± 0.116	0.700±0.077
NSGA2-ANN-MS-**SU+RBF**	0.743 ± 0.039	*0.605 ± 0.089*	*0.714±0.064*
NSGA2-ANN-CCR-**PU+RBF**	0.900 ± 0.016	0.042 ± 0.068	0.680±0.078
NSGA2-ANN-MS-**PU+RBF**	0.748 ± 0.039	**0.634 ± 0.078**	0.709±0.049
SVM	**0.916±0.000**	0.000±0.000	0.500±0.000
C4.5	*0.906±0.017*	0.017±0.043	0.502±0.016
SVM-CS	0.737±0.048	0.473±0.137	0.620±0.068
C4.5-CS	0.819 ± 0.037	0.188 ± 0.138	0.532±0.063

Buoy 41043			
Method	CCR	MS	AUC
	Mean±SD	Mean±SD	Mean±SD
NSGA2-ANN-CCR-**SU**	0.842 ± 0.027	0.041 ± 0.074	0.453±0.1154
NSGA2-ANN-MS-**SU**	0.674 ± 0.058	0.270 ± 0.162	0.480±0.118
NSGA2-ANN-CCR-**PU**	0.856 ± 0.017	0.041 ± 0.062	0.467±0.121
NSGA2-ANN-MS-**PU**	0.641 ± 0.050	0.219 ± 0.138	0.454±0.121
NSGA2-ANN-CCR-**RBF**	0.848 ± 0.024	0.048 ± 0.070	*0.546±0.079*
NSGA2-ANN-MS-**RBF**	0.703 ± 0.052	*0.326 ± 0.170*	**0.562±0.104**
NSGA2-ANN-CCR-**SU+RBF**	0.834 ± 0.026	0.037 ± 0.053	0.410±0.114
NSGA2-ANN-MS-**SU+RBF**	0.665 ± 0.055	0.244 ± 0.122	0.449±0.113
NSGA2-ANN-CCR-**PU+RBF**	0.864 ± 0.022	0.037 ± 0.061	0.494±0.098
NSGA2-ANN-MS-**PU+RBF**	0.586 ± 0.069	**0.341 ± 0.140**	0.452±0.119
SVM	**0.880±0.000**	0.000±0.000	0.500±0.000
C4.5	*0.875±0.019*	0.000±0.000	0.497±0.011
SVM-CS	0.625±0.086	0.290±0.172	0.486±0.083
C4.5-CS	0.726 ± 0.077	0.300 ± 0.160	0.542±0.070

The best result for each metric are in boldface and the second best result in italics.

ACKNOWLEDGMENT

This work has been subsidized by the project TIN2014-54583-C2-1-R of the Spanish Ministry of Economy and Competitiveness (MINECO), FEDER funds and the P11-TIC-7508 project of the Junta de Andalucía (Spain). Antonio M. Durán-Rosal's research has been subsidized by the FPU Predoctoral Program of the Spanish Ministry of Education, Culture and Sport (MECD), grant reference FPU14/03039.

REFERENCES

[1] A. Altunkaynak, "Prediction of significant wave height using spatial function," *Ocean Engineering*, vol. 106, pp. 220–226, 2015.

[2] S. Salcedo-Sanz, J. N. Borge, L. Carro-Calvo, L. Cuadra, K. Hessner, and E. Alexandre, "Significant wave height estimation using SVR algorithms and shadowing information from simulated and real measured X-band radar images of the sea surface," *Ocean Engineering*, vol. 101, pp. 244–253, 2015.

[3] O. Makarynskyy and D. Makarynska, "Wave prediction and data supplementation with artificial neural networks," *Journal of Coastal Research*, pp. 951–960, 2007.

[4] D. I. Gopinath and G. Dwarakish, "Wave prediction using neural networks at New Mangalore Port along west coast of India," *Aquatic Procedia*, vol. 4, pp. 143–150, 2015.

[5] C. M. Bishop, *Neural networks for pattern recognition*. Oxford university press, 1995.

[6] R. Durbin and D. Rumelhart, "Products Units. A computationally powerful and biologically plausible extension to backpropagation networks," *Neural Computation*, vol. 1, pp. 133–142, 1989.

[7] J. H. Friedman and W. Stuetzle, "Projection pursuit regression," *Journal of the American statistical Association*, vol. 76, no. 376, pp. 817–823, 1981.

[8] T. J. Hastie and R. J. Tibshirani, *Generalized additive models*. CRC Press, 1990, vol. 43.

[9] J. H. Friedman, "Multivariate adaptive regression splines," *The annals of statistics*, pp. 1–67, 1991.

[10] W. Duch and N. Jankowski, "Transfer functions: hidden possibilities for better neural networks." in *ESANN*. Citeseer, 2001, pp. 81–94.

[11] W. Duch, R. Adamczak, and G. H. Diercksen, "Constructive density estimation network based on several different separable transfer functions." in *ESANN*, 2001, pp. 107–112.

[12] I. B. Ciocoiu, "Hybrid feedforward neural networks for solving classification problems," *Neural Processing Letters*, vol. 16, no. 1, pp. 81–91, 2002.

[13] S. Cohen and N. Intrator, "A hybrid projection-based and radial basis function architecture: initial values and global optimization," *Pattern Analisys and Applications*, vol. 5, pp. 113–120, 2002.

[14] D. L. Donoho and I. M. Johnstone, "Projection-based approximation and a duality with kernel methods," *The Annals of Statistics*, pp. 58–106, 1989.

[15] P. A. Gutiérrez, C. Hervás, M. Carbonero, and J. C. Fernández, "Combined projection and kernel basis functions for classification in evolutionary neural networks," *Neurocomputing*, vol. 72, no. 13, pp. 2731–2742, 2009.

[16] X. Feng, M. Tsimplis, M. Yelland, and G. Quartly, "Changes in significant and maximum wave heights in the Norwegian Sea," *Global and Planetary Change*, vol. 113, pp. 68–76, 2014.

[17] G. Muraleedharan, C. Lucas, and C. G. Soares, "Regression quantile models for estimating trends in extreme significant wave heights," *Ocean Engineering*, vol. 118, pp. 204–215, 2016.

[18] J. Q. H. Nørgaard and T. L. Andersen, “Can the Rayleigh distribution be used to determine extreme wave heights in non-breaking swell conditions?” *Coastal Engineering*, vol. 111, pp. 50–59, 2016.

[19] T. Kitano, S. Jayaprasad, and W. Kioka, “An Extended Poisson test for Detecting the Difference Between the Past and Future Rates of Extremes of Sea wave Heights,” *Procedia Engineering*, vol. 116, pp. 583–591, 2015.

[20] Y. Fan, I. M. Held, S.-J. Lin, and X. L. Wang, “Ocean warming effect on surface gravity wave climate change for the end of the twenty-first century,” *Journal of Climate*, vol. 26, no. 16, pp. 6046–6066, 2013.

[21] A. M. Durán-Rosal, M. de la Paz-Marín, P. A. Gutiérrez, and C. Hervás-Martínez, “Applying a Hybrid Algorithm to the Segmentation of the Spanish Stock Market Index Time Series,” in *International Work-Conference on Artificial Neural Networks*. Springer, 2015, pp. 69–79.

[22] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, “An extensive comparative study of cluster validity indices,” *Pattern Recognition*, vol. 46, no. 1, pp. 243–256, 2013.

[23] J. C. Fernández, F. J. Martínez, C. Hervás, and P. A. Gutiérrez, “Sensitivity versus Accuracy in Multi-class Problems using Memetic Pareto Evolutionary Neural Networks,” *IEEE Transactions on Neural Networks*, vol. 21, no. 5, pp. 750–770, 2010.

[24] K. Deb, A. Pratab, S. Agarwal, and T. Meyarivan, “A fast and elitist multiobjective genetic algorithm: NSGA2,” *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[25] C. Bishop, *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[26] G. Zhang, “Neural Networks for Classification: A Survey,” *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, vol. 30, no. 4, pp. 451–462, 2000.

[27] C. M. Bishop, “Pattern recognition,” *Machine Learning*, vol. 128, 2006.

[28] P. Vilfredo, *Cours D’Economie Politique*. Lausanne: F. Rouge, 1896, vol. I and II.

[29] A. C. Martínez-Estudillo, C. Hervás-Martínez, F. J. Martínez-Estudillo, and N. García-Pedrajas, “Hybridization of evolutionary algorithms and local search by means of a clustering method,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 36, no. 3, pp. 534–545, 2005.

[30] R. P. Lippmann, “Pattern classification using neural networks,” *IEEE communications magazine*, vol. 27, no. 11, pp. 47–50, 1989.

[31] W. Duch and N. Jankowski, “Survey of neural transfer functions,” *Neural Computing Surveys*, vol. 2, no. 1, pp. 163–212, 1999.

[32] http://www.ndbc.noaa.gov/, “National Data Buoy Center.” National Oceanic and Atmospheric Administration of the USA (NOAA), accessed June 29, 2016.

[33] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA data mining software: an update,” *Special Interest Group on Knowledge Discovery and Data Mining Explorer Newsletter*, vol. 11, no. 1, pp. 10–18, November 2009.

[34] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Information Processing and Management*, vol. 45, pp. 427–437, 2009.