Ocean Engineering 216 (2020) 108089

Contents lists available at ScienceDirect

Ocean Engineering

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

Short- and long-term energy flux prediction using Multi-Task Evolutionary Artificial Neural Networks

David Guijo-Rubio[1], Antonio M. Gómez-Orellana[1,*], Pedro A. Gutiérrez, César Hervás-Martínez

Department of Computer Science and Numerical Analysis, University of Córdoba, Córdoba, Spain

ARTICLE INFO

Keywords:
Wave energy flux prediction
Marine energy
Multi-task machine learning
Evolutionary artificial neural networks
Reanalysis data

ABSTRACT

This paper presents a novel approach to tackle simultaneously short- and long-term energy flux prediction (specifically, at $6h$, $12h$, $24h$ and $48h$ time horizons). The methodology proposed is based on the Multi-Task Learning paradigm in order to solve the four problems with a single model. We consider Multi-Task Evolutionary Artificial Neural Networks (MTEANN) with four outputs, one for each time prediction horizon. For this purpose, three buoys located at the Gulf of Alaska are considered. Measurements collected by these buoys are used to obtain the target values of energy flux, whereas, only reanalysis data are used as input values, allowing the applicability to other locations. The performance of three different basis functions (Sigmoidal Unit, Radial Basis Function and Product Unit) are compared against some popular state-of-the-art approaches such as Extreme Learning Machines and Support Vector Regressors. The results show that MTEANN methodology using Sigmoidal Units in the hidden layer and a linear output achieves the best performance. In this way, the multi-task methodology is an excellent and lower-complexity approach for energy flux prediction at both short- and long-term prediction time horizons. Furthermore, the results also confirm that reanalysis data is enough for describing well the problem tackled.

1. Introduction

In the last decades, there has been an increase in the carbon dioxide levels, and in consequence, the global mean temperatures have also risen. According to the special report regarding global warming done by the Intergovernmental Panel on Climate Change (IPCC) (Hoegh-Guldberg et al., 2018) the surface temperatures are increasing by about 0.2°C per decade. The ongoing rise of the average temperature of the Earth is leading to wildfires, the expansion of deserts or more intense storms, among others (Council et al., 2012). Therefore, the interest in renewable and eco-friendly energy sources such as solar, wind, tides and waves is in continuous growth (Ellabban et al., 2014). Although solar and wind are considered the most popular alternative sources of energy, they have the setback of not being constantly available, due to the intermittent nature of wind and sun. On the other hand, tides and waves not only benefit from being always in movement, but they also can be used in many parts of the world, including oceans or big seas. Furthermore, given that the energy storage systems are a major challenge (Palmer and Floyd, 2020) and that around 40% of the world's population live within 100 km of the coast, the wave and tidal renewable energies allow for a cost-effective transmission of electricity (Esteban and Leary, 2012).

Waves exhibit a stochastic nature, due to the influence of a great number of environmental elements. Therefore, they cannot be predicted straightforwardly as tides. Due to this stochastic behaviour, the reliability and confidence of wave energy generation need to be predicted beforehand in order to develop an stable source of energy. For the generation of electricity from wave energy, Wave Energy Converters (WECs) (Falcão, 2010; Aderinto and Li, 2018) are applied in order to transform the kinetic energy directly generated by the waves into electricity (Falnes and Kurniawan, 2020). To model the behaviour of wave energy, the two most important parameters are the significant wave height (H_s) and the wave energy flux (F_e), which have been widely studied in the literature from different perspectives (Nitsure et al., 2012): physical models (Ibarra-Berastegi et al., 2015), statistical models (Lin et al., 2020) or by the application of Machine Learning (ML) techniques (Cuadra et al., 2016), among others.

Focusing on ML approaches, some of the first works published were those of Deo and Naidu (1998) and Deo et al. (2001), which consisted in real-time forecasting of H_s by using Artificial Neural Network

* Corresponding author.
E-mail addresses: dguijo@uco.es (D. Guijo-Rubio), am.gomez@uco.es (A.M. Gómez-Orellana), pagutierrez@uco.es (P.A. Gutiérrez), chervas@uco.es (C. Hervás-Martínez).
[1] Contributed equally to this work.

https://doi.org/10.1016/j.oceaneng.2020.108089
Received 5 June 2020; Received in revised form 24 July 2020; Accepted 9 September 2020
Available online 16 September 2020

(ANN) models. Later, the ANN methodology has been validated by some other authors, such as Castro et al. (2014), who applied ANN models to predict the wave energy resource in the northern coast of Spain. Hadadpour et al. (2014) focused on forecasting the wave energy over horizons from 1 to 12 hours, in the southern part of the Caspian Sea. Support Vector Regressors (SVRs) have also been used for H_s forecasting leading to competitive results (Mahjoobi and Mosabbeb, 2009; Duan et al., 2016). Another approach was proposed by Cornejo-Bueno et al. (2016b), who developed a grouping genetic algorithm applied to Extreme Learning Machines (ELMs) to predict both H_s and F_e. In this work, a feature selection is carried out to select the most relevant features for the regressor machine. More recent works used an ensemble of ELMs for the prediction of H_s in 10 stations of varying terrains from Gulf of Mexico, Brazil and Korean region (Kumar et al., 2018) . Ali and Prasad (2019) also applied ELMs coupled with an improved complete ensemble empirical mode decomposition method, using adaptive noise, to predict H_s in the eastern coastal zones of Australia. Cornejo-Bueno et al. (2018) considered a regression genetic fuzzy system known as FRULER for the prediction of H_s and F_e in one buoy of the West Coast of the USA using measurements from other two neighbour buoys. With respect to the sort of data used, Fernández et al. (2015) proposed the use of meteorological data, obtained from NCEP/NCAR Reanalysis Project, and applied several ordinal classifiers for the prediction of both, H_s and F_e. On the other hand, Salcedo-Sanz et al. (2015) and Cornejo-Bueno et al. (2016a) used features extracted from X-band radar images as input variables, whereas Choi et al. (2020) used only raw ocean images to estimate real-time H_s using SVR and 2D and 3D deep learning approaches. ML techniques have also been used for forecast water-level in lakes at various horizons. In Kisi et al. (2015), the models were developed using SVM coupled with Firefly Algorithm (FA). The FA was applied to estimate the optimal SVM parameters. In Shiri et al. (2016), the authors used the popular ELM, whose results were compared with genetic programming and ANNs.

Despite the hard work done in H_s and F_e prediction, there is not, up to the knowledge of the authors, any previous study analysing the short- and long-term energy flux prediction simultaneously considering ML algorithms and reanalysis data for developing accurate flux of energy prediction models. This is the main motivation to carry out this study. In order to accomplish this goal, two architectures of multi-task ANNs are optimised by applying an evolutionary algorithm. Specifically, the ANN models have four outputs: the first two ones for short-term prediction ($6h$ and $12h$) and the second two for long-term prediction ($24h$ and $48h$). The purpose of the evolutionary algorithm is to optimise the structure of the ANNs proposed for the problem tackled. The methodology proposed performs flux of energy prediction using reanalysis variables as input data, that is, no observed data is required, what improves its applicability to other sites.

On the other hand, simultaneous energy flux prediction at different time horizons establishes a relationship between all of them, that can be exploited to obtain a single model to better solve all the prediction time horizons. In this sense, Multi-Task Learning (MTL) (Caruana, 1997) is a ML technique that exploits such relationship taking advantage of the latent information when performing related tasks simultaneously (Maurer et al., 2016; Dorado-Moreno et al., 2020). Hence, the model can infer such information to improve its performance.

Consequently, the methodology proposed in this study exhibits the following novelties: (1) A MTL approach is presented with the objective of short- and long-term energy flux prediction simultaneously. Four types of Multi-Task Evolutionary Artificial Neural Networks (MTEANN) are developed and studied. (2) Given that the methodology only uses reanalysis variables as input data, it could be extended to other locations. (3) A comprehensive comparison against some state-of-the-art regression techniques (ELM and SVR, among others) is presented. (4) A single MTEANN performs all the different time prediction horizons ($6h$, $12h$, $24h$ and $48h$), being less complex when compared against the literature techniques. (5) Moreover, MTEANN methodology exhibits the advantage of requiring just one model with less connections in comparison to the sum of the connections of four independent single-task models only performing one time prediction horizon.

The rest of the paper is organised as follows: in Section 2 a detailed description of the data used and the preprocessing steps applied is presented. Section 3 describes the methodology proposed for the energy flux prediction problem. The performance of the different MTEANN models and the comparison against the state-of-the-art techniques are studied and analysed in Section 4. Section 5 discusses the results achieved according to the best model. Finally, Section 6 gives some concluding remarks for closing the paper.

2. Data description and processing

This section introduces the data used in this study, which have been obtained from the following two well-known sources of information:

- *National Data Buoy Center* (NDBC) (National Data Buoy Center, 2020) maintains a network of buoys deployed in coastal and offshore waters around oceans and seas. Such buoys are equipped with assorted sensors to collect real-time marine meteorological and oceanographic observations, including significant wave height, average wave period or sea level pressure, among others.
- *National Center for Atmospheric Research* (NCAR) (Kalnay et al., 1996; Kistler et al., 2001) provides a global $2.5° \times 2.5°$ grid of reanalysis data of meteorological variables (e.g. pressure, air temperature, relative humidity, etc.), which are obtained by means of a climate model that assimilates weather observations collected from different sources.

Specifically, to carry out this study hourly collected measurements of three buoys and 4-times daily values of seven reanalysis variables have been selected during years 2015 to 2018. In that way, all input variables will be obtained from reanalysis data, whereas the output variables will be obtained from observed data collected by sensors installed in each buoy. The following subsections describe in detail such data.

2.1. Buoys measurements

The buoys subject to research are located in the Gulf of Alaska, as shown in Fig. 1. This region has a high frequency and variability regarding occurrence of storms, and, as a consequence, it is a zone of intense wave generation. In this sense, three buoys have been selected according to their proximity and depth of water:

- Buoy 46001 (LLNR 984) (NDBC - Station 46001, 2020) – Western Gulf of Alaska – 175NM SE of Kodiak, AK. Geographical location: 56.232N 147.949W
- Buoy 46066 (LLNR 984.1) (NDBC - Station 46066, 2020) – South Kodiak – 310NM SSW of Kodiak, AK. Geographical location: 52.765N 155.009W
- Buoy 46085 (LLNR 984.15) (NDBC - Station 46085, 2020) – Central Gulf of Alaska – 265NM West of Cape Ommaney, AK. Geographical location: 55.883N 142.482W

Although the flux of energy (F_e) is not directly measured by the buoys sensors, it can be calculated from two wave parameters measured by the buoys: the significant wave height (H_s) and the average wave period (T_e). In this way, for the three selected buoys, the F_e of each time instant has been obtained using the following equation:

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

where F_e is measured in kilowatts per metre, H_s is measured in metres and T_e is measured in seconds. Due to H_s is a kind of average wave height, F_e is defined in Eq. (1) as an average energy flux. Nevertheless, for simplicity it will be referred as flux of energy.

Given that the goal of this study is to predict the flux of energy generated by the waves at different time prediction horizons, only such buoy measurement will be considered when obtaining the datasets.

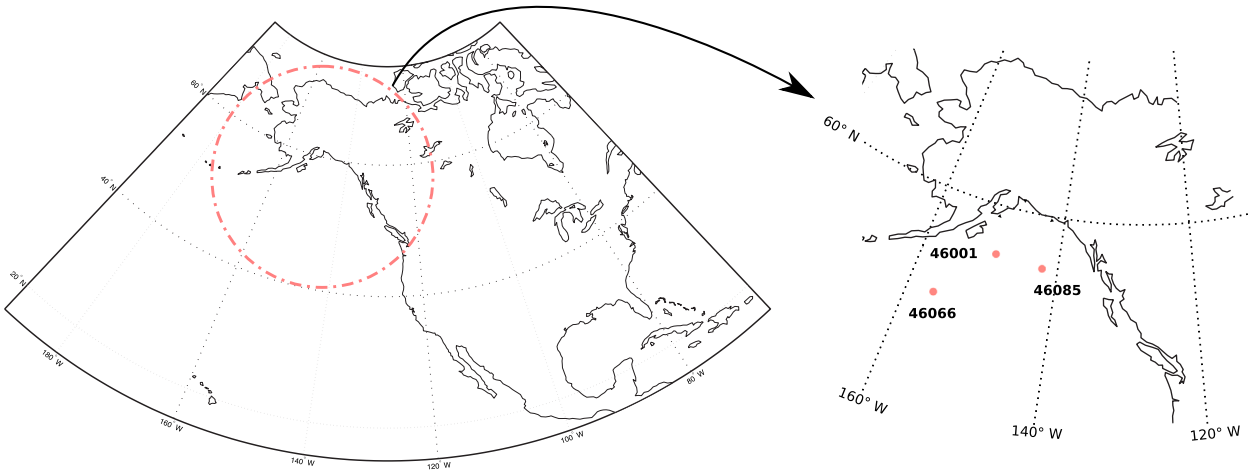


Fig. 1. Buoys location at Gulf of Alaska.

Table 1
Reanalysis variables.

Variable	Units
Air temperature	degK
Pressure	Pascals
Omega vertical velocity	Pascal/s
Precipitable water content	kg/m^2
Relative humidity	%
Component South–North of wind speed	m/s
Component West–East of wind speed	m/s

2.2. Reanalysis data grid

With respect to the reanalysis data used in this study, seven variables have been considered, shown in Table 1. The reasons behind the selection of these reanalysis variables are two: firstly, to have a good characterisation of wind and, secondly, to take advantage of the contribution of weather factors that exert on it.

On the one hand, waves, and consequently, flux of energy, are generated with the action of wind. In this sense, air temperature, pressure, omega vertical velocity and the South–North and West–East components of wind speed have been selected as appropriate descriptive variables of the wind in a given zone (Fernández et al., 2015; Dorado-Moreno et al., 2017; Ibarra-Berastegi et al., 2015).

On the other hand, atmospheric processes play an important role when tackling environmental modelling, because they provide useful information to better understand the formation and evolution of cyclones, storms and other related phenomena. In this way, and given that the three buoys under study are located at the Gulf of Alaska, precipitable water content is a key component to find patterns that could lead to severe weather events (Campmany et al., 2010; Wang et al., 2018; Choy et al., 2013). Furthermore, relative humidity provides additional information about the occurrence of deep convection (Siingh et al., 2014).

Given that the NNRP web page provides a global 2.5°×2.5° grid, for each reanalysis variable, the four closest geographical points (reanalysis nodes) to each buoy have been selected. In that way, each reanalysis file includes the data of one meteorological variable for each reanalysis node that surrounds the buoy, for the selected years (2015 to 2018). For a better understanding, in Fig. 2, an approximate representation of the four reanalysis nodes around the geographical location of the Buoy 46001 is shown.

As a result, three (one for each buoy) reanalysis data grids have been considered, as described in Table 2.

Finally, the reanalysis data have been obtained with the minimum temporal resolution available, that is, 4-times daily (00*h*, 06*h*, 12*h* and 18*h*).

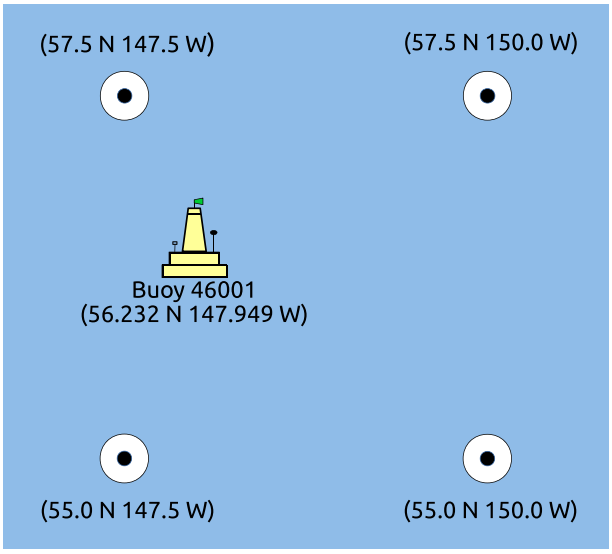


Fig. 2. Example of the four sub-grid reanalysis nodes around the *Buoy 46001*.

2.3. Obtaining the datasets

Firstly, a preprocessing task has been performed to recover the values of the measurements not recorded by the buoys. Note that such missing values were insignificant due to only the H_s and T_e variables have been used to calculate the flux of energy and because of the data integration process that is explained below.

After that, to create the dataset of each buoy including the reanalysis variables and the flux of energy, a data integration process or *matching* has been carried out to merge both sources of information. Taking into account that the buoys measurements were hourly collected (from 00: 50*h* to 23: 50*h*) and that the reanalysis data is available every 6 hours (00*h*, 06*h*, 12*h* and 18*h*), the *matching* has been performed every 6 hours with the nearest buoy measurement, discarding the remaining ones.

As a result of the *matching*, for each instance, both the value of each reanalysis variable and the value of the flux of energy correspond to the same time instant. Specifically, for the first instance such values refer to 2015/01/01 00: 00*h*, for the second one to 2015/01/01 06: 00*h* (+6*h*) and so on, until 2018/12/31 18: 00*h*. In this way, the values for each instance and time prediction horizon of the flux of energy (6*h*, 12*h*, 24*h* and 48*h*) are taken from one, two, four and eight instances in advance,

Table 2
Reanalysis data grid of each buoy.

Buoy	Geographical points	Reanalysis nodes (latitude N, longitude E)			
	(latitude N, longitude W)	Node 1	Node 2	Node 3	Node 4
46001 (LLNR 984)	(55.0N–57.5N , 147.5W–150.0W)	(55.0N, 210.0E)	(55.0N, 212.5E)	(57.5N, 210.0E)	(57.5N, 212.5E)
46066 (LLNR 984.1)	(52.5N–55.0N , 155.0W–157.5W)	(52.5N, 202.5E)	(52.5N, 205.0E)	(55.0N, 202.5E)	(55.0N, 205.0E)
46085 (LLNR 984.15)	(55.0N–57.5N , 140.0W–142.5W)	(55.0N, 217.5E)	(55.0N, 220.0E)	(57.5N, 217.5E)	(57.5N, 220.0E)

respectively. It is worth mentioning that instances from 2018/12/30 00: 00h to 2018/12/31 18: 00h required energy flow values calculated from 2019/01/01 00: 00h to 2019/01/02 18: 00h to fully accomplish the time prediction horizons.

Besides, given that the reanalysis data grid of each buoy is composed of 4 reanalysis nodes (see Table 2), the values for each reanalysis variable have been calculated as a weighted mean according to the distance from each reanalysis node to the location of the buoy. To calculate the distance, the *Haversine* equation (also known as the great circle distance) (de Smith et al., 2009) has been applied:

$$\begin{aligned} d(p_0, p_j) = \ & \arccos(\sin(lat_0) \cdot \sin(lat_j) \\ & \cdot \cos(lon_0 - lon_j) + \cos(lat_0) \\ & \cdot \cos(lat_j)), \end{aligned} \tag{2}$$

where p_0 is the geographical location of the buoy, p_j is the location of each reanalysis node, and lat and lon are the latitude and longitude of the points, respectively.

Once the distances are calculated, they are inverted and normalised as follows:

$$w_i = \frac{d(p_0, p_i)}{\sum_{j=1}^{N} d(p_0, p_j)}, \quad i = 1, \dots, N, \tag{3}$$

where w_i is the weight of the data from the ith reanalysis node. These weights will be used to obtain a final value for the input variable as a weighted average of the four values from the surrounding reanalysis nodes. In this way, the closest reanalysis nodes to the buoy will provide more information than the rest of the nodes, which are farther. Note that this weighting process takes advantage from the information provided by the 4 reanalysis nodes, without increasing neither the computational cost nor the complexity of the models trained on this data.

As a result of the whole process described, one dataset has been obtained for each buoy, containing 5844 instances (years 2015 to 2018). In turn, each dataset consists of 7 input variables (one for each weighted reanalysis variable) and 4 output variables (one for each time prediction horizon of the flux of energy: flux energy 6h ahead, 12h ahead, 24h ahead and 48h ahead).

3. Methodology: Evolutionary Artificial Neural Networks

In this section, the Artificial Neural Networks (ANN) models used in this paper are introduced along with the different type of basis functions considered. Besides, the evolutionary algorithm used to evolve the ANNs is also described in detail.

3.1. Artificial Neural Networks

ANNs (Bishop et al., 1995) exhibit features such as learning and adaptation, universal approximation abilities or feasibility for hardware implementation, among others, making them particularly attractive for modelling many non-linear real-world problems. In fact, they have become a widely used kind of non-linear models owing to the competitive results obtained in a huge variety of classification and regression problems. Although many different types of neural network architectures have been used, the most popular is the Multilayer Perceptron (MLP) (Bishop et al., 1995), which consists of one input layer, at least one hidden layer and one output layer.

As mentioned above, the goal of this study is to design ANNs able to predict flux of energy at four different time prediction horizons simultaneously, that is, 6h, 12h, 24h and 48h. Therefore, the problem is tackled following the multi-task regression paradigm that is formally defined as follows:

$$D = \left\{ \left(\mathbf{x}_i, y_{i6h}, y_{i12h}, y_{i24h}, y_{i48h} \right) ; i = 1, 2, \dots, N \right\}, \tag{4}$$

where D is the dataset used, $\mathbf{x}_i$ is a vector that contains the seven input weighted reanalysis variables, y_{i6h}, y_{i12h}, y_{i24h} and y_{i48h} are the flux of energy at each time prediction horizon, and N is the number of instances. In order to perform this task, a neural network model with four linear outputs is considered. Regardless the hidden layer type, the ANN output is given by the following expression:

$$f_q(\mathbf{x}, \mathbf{w}, \boldsymbol{\beta}) = \beta_{q0} + \sum_{j=1}^{m} \beta_{qj} B_j(\mathbf{x}, \mathbf{w}_j), \quad q = 1, \dots, Q, \tag{5}$$

where $B_j(\mathbf{x}, \mathbf{w}_j)$ represents the non-linear transformations performed by the basis function of each hidden neuron on the input vector $\mathbf{x}^{\mathbf{T}} = (x_1, x_2, \dots, x_d)$, being d its length, and considering the element $B_0(\mathbf{x}, \mathbf{w}_j) = 1$ as bias; $\boldsymbol{\beta}_q^T = (\beta_{q0}, \beta_{q1}, \beta_{q2}, \dots, \beta_{qm})$ are the weights from linear combination estimated from the data, including bias β_{q0}; $\mathbf{w}_j^T = (w_{j1}, w_{j2}, \dots, w_{jd})$ represents the parameters associated to each basis function; m is the number of basis functions used to minimise some defined error function; and Q is the number of outputs of the problem ($Q = 4$, in our case).

For the problem tackled, as it has been described in Section 2, the input layer of the neural network model is composed of 7 neurons, one for each input variable, whereas the output layer is composed of 4 neurons, one for each time prediction horizon of the flux of energy (6h, 12h, 24h and 48h). Concerning the neurons of the hidden layer, the following three basis functions will be analysed because of being the most widely used in the state-of-the-art:

- Sigmoidal Unit (SU) (Lippmann, 1989) can be considered as the most common basis function used. Sigmoidal neural networks or MLPs can approximate any continuous function accurately. However, they can get trapped in local minimum frequently during the training phase. According to the notation described above, SU basis function is expressed as follows:

$$B_j(\mathbf{x}, \mathbf{w}_j) = \frac{1}{1 + e^{-\left(w_{j0} + \sum_{i=1}^{d} w_{ji} x_i\right)}}, \qquad j = 1, \dots, m, \tag{6}$$

  w_{j0} being the bias.
- Product Unit (PU) (Martínez-Estudillo et al., 2008) basis function is an alternative to SU, and it provides multiplicative neurons instead of additives ones. PU based neural networks are able to represent strong interactions between input variables, though their training is more complicated because small changes in weights may cause large changes in the total error surface. According to the notation used, PU basis function is given by the following expression:

$$B_j(\mathbf{x}, \mathbf{w}_j) = \prod_{i=1}^{d} x_i^{w_{ji}}, \qquad j = 1, \dots, m, \tag{7}$$

  where $\mathbf{w}_j^T = (w_{j1}, w_{j2}, \dots, w_{jd})$, not considering bias ($w_{j0}$) in the inputs.

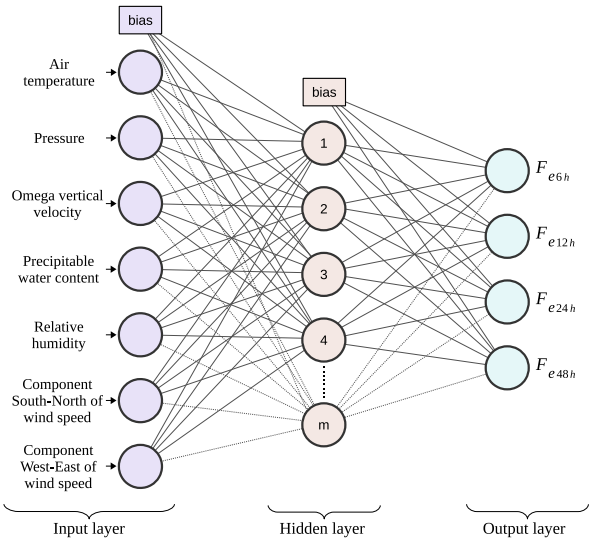


Fig. 3. ANN architecture. Bias is not considered for PUs and RBFs.

- Radial Basis Function (RBF) (Billings and Zheng, 1995) is a kernel function whereas PU an SU are projection functions. Therefore, RBF provides local behaviour for each neuron due to that each one is placed in distinct region of search space, reducing the number of local minima. According to the notation followed, RBF basis function is represented as follows:

$$B_j(\mathbf{x}, \mathbf{w}_j) = e^{-\frac{1}{2}\left(\frac{\sum_{i=1}^{d}(x_i - c_{ji})^2}{r_j}\right)}, \qquad j = 1, \ldots, m, \tag{8}$$

  where $\mathbf{w}_j = \{r_j, \mathbf{c}_j\}$, including the radium r_j and the centroid $\mathbf{c}_j$ of the Gaussian basis function of the jth hidden neuron.

The topology of the artificial neural networks considered is represented in Fig. 3, where each of this three types of basis functions can be used for the hidden layer.

Even though most ANNs use linear functions in the output layer as described in Eq. (5), for this study the performance of PUs in the output layer will be also evaluated. Specifically, an ANN composed of SUs in the hidden layer and PUs in the output layer is proposed. The motivation behind this architecture is that PUs are multiplicative functions able to infer the properties of an universal approximator (Schmitt, 2002). In this sense, this topology can exploit the interactions between the outputs of the hidden layer. Therefore, although the difficulty in the training step can be harder, they can result in better performance.

Up to now, and to the best of our knowledge, this is the first work considering such approach in flux of energy prediction. When using PU in the output layer, the ANN output is defined as follows:

$$f_q(\mathbf{x}, \mathbf{w}, \boldsymbol{\beta}) = \prod_{j=1}^{m} B_j(\mathbf{x}, \mathbf{w}_j)^{\beta_{qj}}, \quad q = 1, \ldots, Q. \tag{9}$$

3.2. Evolutionary Algorithms

Given the non-linearity of neural networks, the error surface exhibits high convolution and many local minima. As a consequence, the estimation of the parameters and structure of ANNs results in a very hard and complex task (Yao, 1999). The most widely used method to train neural networks is back-propagation, though it falls in local optima often. As an alternative, Evolutionary Algorithms (EAs) (Ser et al., 2019) were proposed as an efficient search method able to discover regions in the error surface where competitive solutions are located.

Therefore, the use of an EA to evolve ANNs is justified to determine an appropriate structure and training the weights. In this study, the algorithm proposed in Martínez-Estudillo et al. (2008) has been considered, and it is described in Algorithm 1.

Algorithm 1: Evolutionary Algorithm

generate a random population of size Np
evaluate and rank the networks according to MSE
repeat
  the best 10% of the population are cloned and then they replace the worst 10%
  parametric mutation is applied to the best 10% of the population
  structural mutation is applied to the remaining 90% of the population
  evaluate and rank the networks according to MSE
until *stopping criteria is fulfilled*
return *the best MSE neural network as final solution*

The algorithm begins the search generating an initial random population of neural networks. After that, the individuals are evaluated and ranked according to the Mean Squared Error (MSE, see Eq. (10)), and, in each iteration, the worst 10% of the population is replaced by the best 10% of the population. Next, all neural networks are evolved by means of parametric and structural mutations.

On the one hand, parametric mutation involves updating the weights of the neural network, but not the structure. For SU, PU and RBF neurons, it is done by adding a Gaussian noise (from a normal distribution with 0 mean and a variance which is decremented throughout the evolutionary process).

On the other hand, structural mutation allows exploring different structures with regards the connections between neurons and the number of nodes of the networks. In that sense, five types of structural mutations are applied: Add Link, Delete Link, Add Neuron, Delete Neuron and Neuron Fusion. Crossover is not used, because it is considered inefficient for evolving neural networks (Angeline et al., 1994).

The characteristics of the algorithm make it to fall into the category of Evolutionary Programming (EP). It has proved to be very efficient regardless of the initial search conditions, achieving competitive solutions without using gradients.

During evolutionary process, both structure and weights of the ANNs are optimised considering the MSE, defined as follows:

$$MSE(\mathbf{x}, \mathbf{w}, \boldsymbol{\beta}) = \frac{1}{Q}\sum_{q=1}^{Q}\left(\frac{1}{N}\sum_{i=1}^{N}\left(y_i^q - f_q(\mathbf{x}, \mathbf{w}, \boldsymbol{\beta})\right)^2\right), \tag{10}$$

where $f_q(\mathbf{x}, \mathbf{w}, \boldsymbol{\beta})$ represents the ANN output previously defined, y_i^q is the observed flux of energy for the instance i at the time prediction horizon q, Q is the number of outputs of the problem ($Q = 4$, in our case), and N is the number of instances.

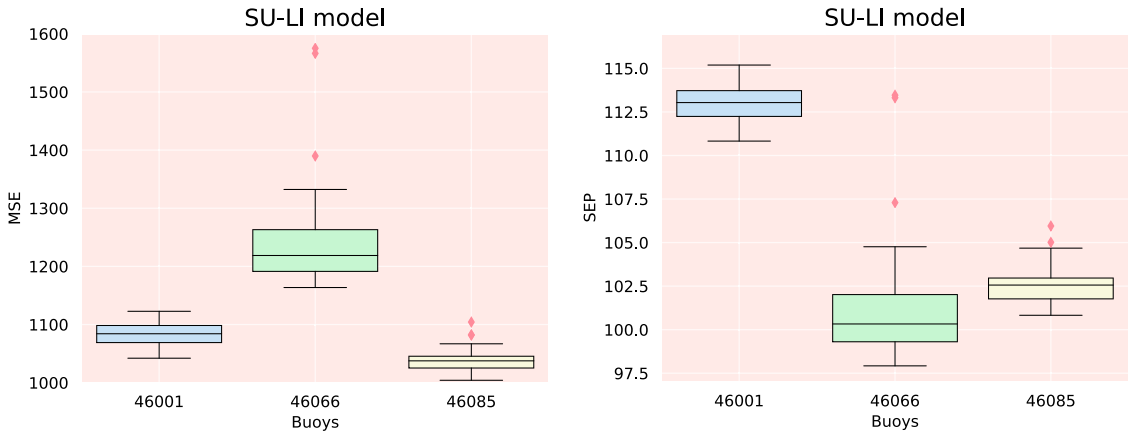
In this way, four different types of Multi-Task Evolutionary Artificial Neural Networks will be studied and they will be referred according to their architecture, that is, the basis function used in the hidden layer and in the output layer. In that sense, SU-LI, PU-LI and RBF-LI refer to the evolved models considering each basis function in the hidden layer (SU, PU and RBF, respectively) and linear (LI) output, whereas SU-PU refers to SU in the hidden layer and PU in the output layer. The Standard Error of Prediction (SEP) will also be considered in order to compare the performance of MTEANNs. This is a relative error that is expressed as a percentage, and it has the advantage of being dimensionless. It is described as follows:

$$SEP(\mathbf{x}, \mathbf{w}, \boldsymbol{\beta}) = \frac{100}{|\bar{y}|}\sqrt{MSE(\mathbf{x}, \mathbf{w}, \boldsymbol{\beta})}, \tag{11}$$

being $\bar{y}$ the average of all outputs of all instances in dataset.

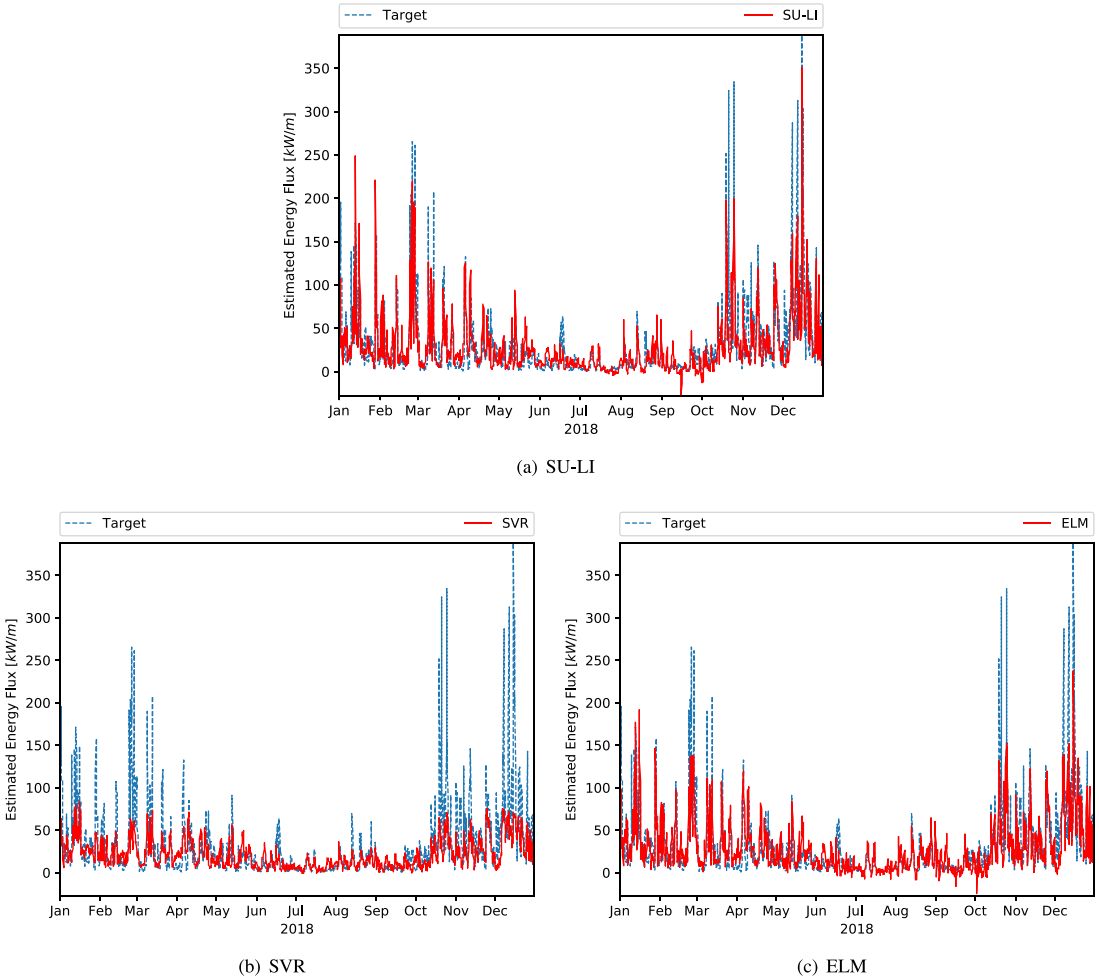
4. Experimental settings and results

In this section, we explain the experimental settings used for the different MTEANNs considered. Furthermore, the results achieved by these models and an extensive comparison against some state-of-the-art regression techniques are also presented.



(a) MSE (b) SEP

Fig. 4. Box-plot obtained from the 30 runs of SU-LI model for the three buoys considered.



(a) SU-LI

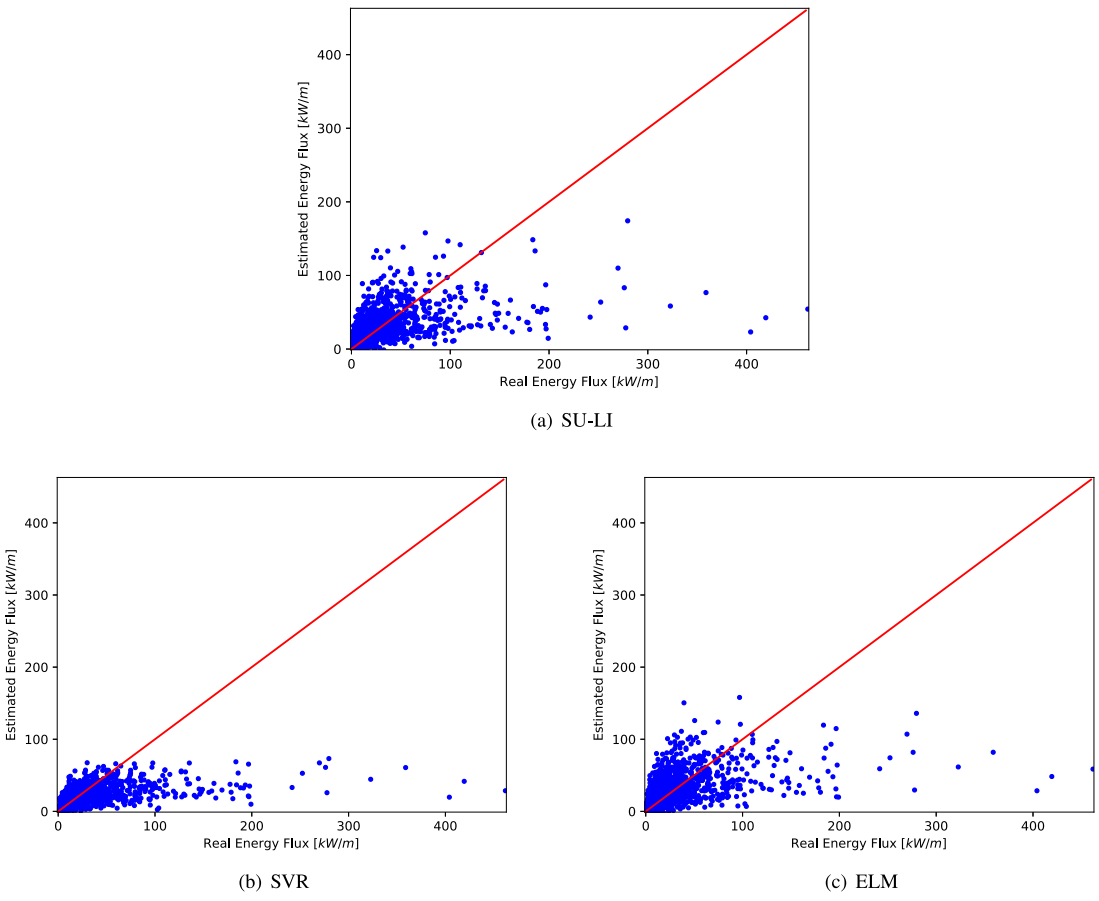
(b) SVR (c) ELM

Fig. 5. Predictions achieved by the best model and the two main state-of-the-art models for the buoy 46085 (prediction horizon $6h$).

4.1. Experimental settings

As previously described in Section 2, the original datasets are divided into two different sets: the first three years (2015–2017) for training the models (4384 instances), whereas the last year (2018) is used for testing (1460 instances).

The MTEANNs have been trained using the following parameters: the algorithm has been run 30 times using 5000 generations and a population size of 600. The number of nodes and links to be created or deleted falls in the ranges $[1, 5]$ and $[1, 7]$, respectively. Furthermore, the number of hidden nodes for the PU-LI is initialised in the range $[1, 3]$, due to its simplicity; in the case of SU-LI and SU-PU, as they need a higher number of hidden nodes, this number is initialised in the range $[1, 10]$; finally, in the case of RBF-LI, it varies in the range $[1, 2]$. Regarding the weights between input and hidden layer, they are initialised in the range $[-10, 10]$ for SU-LI, SU-PU and PU-LI, whereas



(a) SU-LI

(b) SVR

(c) ELM

Fig. 6. Scatter plots obtained by the best model and the two main state-of-the-art models for the buoy 46001.

Table 3
$Mean_{SD}$ values of MSE, SEP, and $\#conn$ obtained for the different buoys.

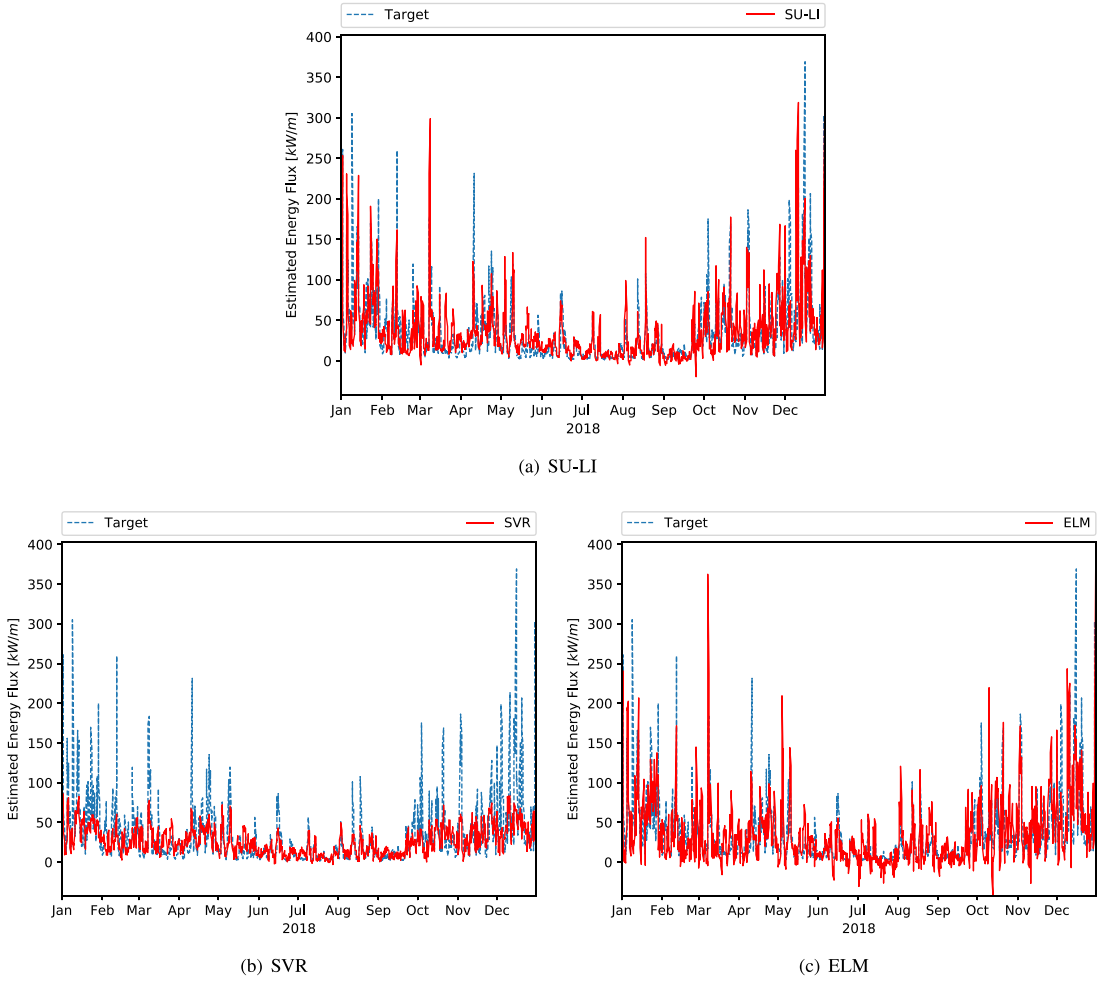
Buoy	Model	MSE					SEP (%)					#conn
		Average	6 h	12 h	24 h	48 h	Average	6 h	12 h	24 h	48 h	
46001	RBF-LI	$1157.32_{69.09}$	$859.99_{120.68}$	$981.15_{125.02}$	$1384.74_{107.21}$	$1403.41_{24.92}$	$117.45_{3.55}$	$101.70_{6.79}$	$108.63_{6.64}$	$128.64_{4.64}$	$130.83_{1.16}$	$\mathbf{78.47_{3.72}}$
	SU-LI	$\mathbf{1082.78_{21.31}}$	$\mathbf{649.31_{22.87}}$	$\mathbf{898.28_{46.57}}$	$\mathbf{1417.12_{45.08}}$	$\mathbf{1366.39_{23.43}}$	$\mathbf{112.98_{1.15}}$	$\mathbf{88.55_{1.56}}$	$\mathbf{104.09_{2.69}}$	$\mathbf{130.20_{2.06}}$	$\mathbf{129.09_{1.10}}$	$250.13_{7.73}$
	PU-LI	$1166.46_{55.57}$	$832.52_{31.10}$	$996.49_{49.53}$	$1463.30_{178.33}$	$1373.52_{45.92}$	$117.86_{2.43}$	$100.26_{1.88}$	$109.64_{2.71}$	$132.13_{7.28}$	$129.42_{2.13}$	$\mathit{120.43_{5.94}}$
	SU-PU	$1127.96_{45.00}$	$723.23_{126.89}$	$956.15_{68.45}$	$1459.44_{58.31}$	$1373.02_{28.24}$	$\mathit{115.51_{2.56}}$	$\mathit{93.17_{7.60}}$	$\mathit{107.36_{3.83}}$	$\mathit{132.12_{2.62}}$	$\mathit{129.40_{1.32}}$	$252.97_{8.83}$
	ELM	$\mathit{1083.89_{8.89}}$	$\mathit{736.38_{19.26}}$	$\mathit{875.13_{17.37}}$	$\mathit{1349.98_{14.97}}$	$\mathit{1374.09_{17.79}}$	$118.50_{1.26}$	$102.08_{1.71}$	$111.51_{1.48}$	$134.23_{2.21}$	$126.18_{2.08}$	$634.33_{54.51}$
46066	RBF-LI	$1477.87_{117.19}$	$1174.54_{103.37}$	$1314.61_{196.23}$	$1680.68_{305.14}$	$1741.63_{223.03}$	$110.61_{4.05}$	$99.75_{4.34}$	$105.06_{7.34}$	$117.99_{10.35}$	$119.64_{6.99}$	$\mathbf{81.37_{4.14}}$
	SU-LI	$\mathit{1250.52_{100.21}}$	$\mathit{880.80_{37.83}}$	$\mathit{1107.80_{168.44}}$	$\mathit{1439.68_{227.73}}$	$\mathit{1573.78_{45.69}}$	$\mathit{101.53_{3.82}}$	$\mathit{86.44_{1.86}}$	$\mathit{96.43_{6.90}}$	$\mathit{109.33_{8.01}}$	$\mathit{113.91_{1.63}}$	$250.10_{8.62}$
	PU-LI	$\mathbf{1250.37_{31.11}}$	$\mathbf{969.12_{47.33}}$	$\mathbf{1106.34_{68.76}}$	$\mathbf{1352.56_{45.66}}$	$\mathbf{1573.45_{35.37}}$	$101.84_{1.33}$	$90.66_{2.20}$	$96.56_{2.98}$	$106.23_{1.79}$	$113.90_{1.28}$	$\mathit{119.17_{6.33}}$
	SU-PU	$1288.09_{90.09}$	$910.77_{32.95}$	$1126.97_{95.06}$	$1434.50_{113.54}$	$1680.11_{298.47}$	$103.00_{3.17}$	$87.90_{1.59}$	$97.42_{4.07}$	$109.33_{4.27}$	$117.34_{9.46}$	$238.53_{8.00}$
	ELM	$1273.00_{27.87}$	$1037.48_{34.49}$	$1104.60_{40.79}$	$1320.89_{32.17}$	$1629.03_{37.06}$	$\mathbf{89.74_{0.85}}$	$\mathbf{84.57_{1.30}}$	$\mathbf{85.47_{1.28}}$	$\mathbf{90.55_{0.90}}$	$\mathbf{98.37_{1.20}}$	$619.67_{60.14}$
46085	RBF-LI	$1120.10_{63.71}$	$809.55_{107.40}$	$892.31_{123.50}$	$1290.32_{47.24}$	$1488.22_{46.91}$	$107.06_{3.18}$	$91.55_{5.81}$	$96.05_{6.27}$	$115.70_{2.10}$	$124.94_{1.95}$	$\mathbf{77.30_{4.24}}$
	SU-LI	$\mathbf{1038.77_{22.36}}$	$\mathbf{645.88_{30.72}}$	$\mathbf{788.78_{46.08}}$	$\mathbf{1237.22_{40.76}}$	$\mathbf{1483.20_{37.41}}$	$\mathbf{102.60_{1.18}}$	$\mathbf{81.91_{1.93}}$	$\mathbf{90.45_{2.59}}$	$\mathbf{113.30_{1.82}}$	$\mathbf{124.73_{1.57}}$	$250.87_{9.18}$
	PU-LI	$1183.67_{47.49}$	$836.63_{48.88}$	$960.60_{43.84}$	$1341.06_{39.64}$	$1596.39_{145.42}$	$110.08_{2.05}$	$93.21_{2.72}$	$99.84_{2.27}$	$117.96_{1.72}$	$129.30_{5.62}$	$\mathit{123.87_{4.77}}$
	SU-PU	$1110.04_{99.90}$	$720.29_{144.47}$	$903.18_{150.76}$	$1294.63_{82.02}$	$1522.06_{99.01}$	$\mathit{106.22_{4.85}}$	$\mathit{86.19_{7.77}}$	$\mathit{96.53_{7.69}}$	$\mathit{115.86_{3.57}}$	$\mathit{126.30_{4.05}}$	$256.67_{9.29}$
	ELM	$\mathit{1079.51_{15.09}}$	$\mathit{748.72_{20.93}}$	$\mathit{840.69_{22.73}}$	$\mathit{1250.95_{19.07}}$	$\mathit{1477.67_{23.27}}$	$113.09_{1.53}$	$97.55_{2.07}$	102.01_{2}	$122.95_{1.68}$	$129.87_{2.32}$	$634.33_{46.53}$

The best result is highlighted in **bold**; the second one best result is shown in *italics*.

for RBF-LI they are initialised in the range $[0.1, 0.9]$. Finally, the weights between the hidden and output layer are all initialised in the range $[-10, 10]$. Before the application of the MTEANN models, the input variables have been scaled to the range $[0.1, 0.9]$.[2]

Furthermore, we have carried out an extensive comparison against several state-of-the-art techniques, widely used for the energy flux prediction: Support Vector Regressor (SVR) (Vapnik, 2013), Linear Regression (LinearReg) (Bishop, 2006), Ridge Regression (RidgeReg) (Bishop, 2006), Lasso Regression (LassoReg) (Friedman et al., 2010), Elastic-Net (Zou and Hastie, 2005) and Extreme Learning Machines (ELM) (Huang et al., 2006). Note that the first three techniques are standard regressors (which can still be applied to multi-output problems), whereas the last three regressors can be adapted to Multi-Task learning, and we have considered the corresponding version. The main difference

[2] Further information of the parameters considered can be found in Martínez-Estudillo et al. (2006), Hervás et al. (2007) and Fernandez Caballero et al. (2010), whereas, more information regarding the ANN models can be obtained from Bishop et al. (1995).



(a) SU-LI

(b) SVR

(c) ELM

Fig. 7. Predictions achieved by the best model and the two main state-of-the-art models for the buoy 46066 (prediction horizon $6h$).

between these two groups of models is that standard regressors can be applied to multi-output problem, predicting two or more numerical values given an input pattern without sharing the inferred structure, i.e. a different model is built for each output. On the other hand, Multi-Task methods take advantage from inferring a common representation from the related tasks, in this case, the energy flux prediction at different time horizons, i.e. one model is able to tackle simultaneously all the tasks. Given their deterministic nature, all these techniques have been run once, except ELM, which has been run 30 times to provide a fair and robust comparison to the remaining techniques.

Moreover, in order to establish a fair and robust comparison of the methods, the parameters of such techniques have been selected using a 10-fold cross-validation over the training set. The best configuration is the one achieving the lowest MSE. Specifically, to adjust the kernel width, SVR uses the range $\{10^{-3}, 10^{-2}, \dots, 10^{3}\}$. RidgeReg, LassoReg and ElasticNet use the range $\{10^{-3}, 10^{-2}, \dots, 10^{3}\}$ to fit the regularisation parameter, and in the case of ElasticNet, the range $[0.10, 0.50, 0.70, 0.90, 0.95, 0.99, 1.00]$ is used for the ratio of the L1 penalisation weight. Finally, for ELM, RBF units are used in the hidden layer with a varying number of neurons chosen in the range $[10, 20, \dots, 60]$.

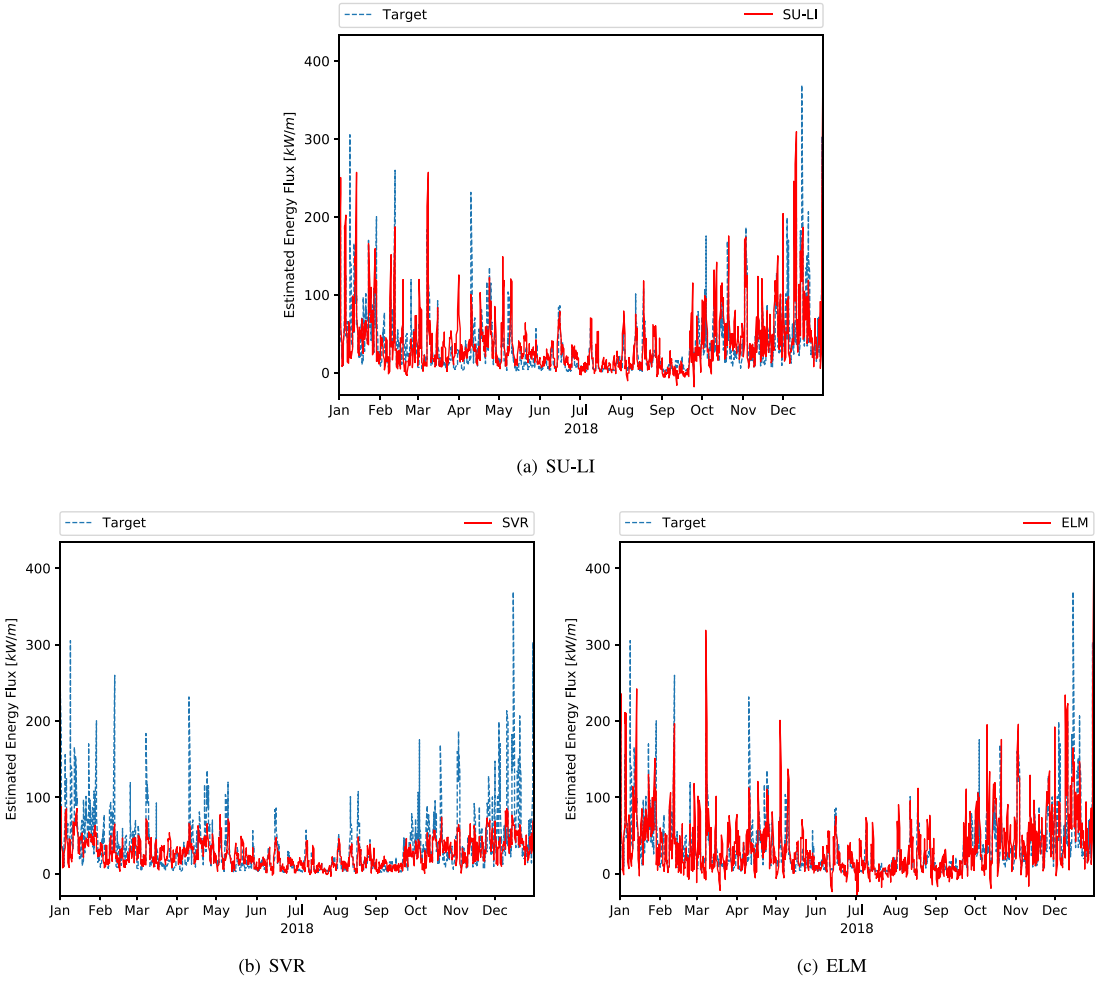
4.2. Results

The results achieved using the experimental settings in Section 4.1 are presented in Table 3 in terms of MSE, SEP and number of links (#*conn*). Given that all the MTEANN and ELM techniques have been run 30 times, the results are expressed as their mean and Standard Deviation (SD): $Mean_{SD}$. As can be seen, this Table 3 is divided into three parts, one for each buoy considered in the study: 46001, 46066 and 46085, respectively. Furthermore, in order to provide a fair comparison, the MSE and SEP values have been presented in two ways: for every output ($6h$, $12h$, $24h$ and $48h$) and in "Average" terms, which is the average of the values for every output.

From Table 3, we can see that MTEANN models achieve the best performance in terms of MSE. Specifically, for buoys 46001 and 46085, SU-LI is the one obtaining the best results in average, whereas for buoy 46066, PU-LI achieves the best results (although the differences with the results achieved by SU-LI are negligible). Regarding SEP, SU-LI also achieves the best values for the buoys 46001 and 46085, whereas for 46066, ELM obtains the best performance, being SU-LI the second one. Furthermore, it is worth mentioning that most of MTEANN models are stable, i.e. the standard deviation values are considerably small. Finally, regarding the number of connections, RBF-LI is the most simple model achieving competitive results, whereas PU-LI is the second one.

On the other hand, even though SU-LI achieves the best performance, it is worthy of mention that the results achieved by RBF-LI and PU-LI demonstrate that these models are also an excellent approach, especially due to the small number of connections used.

Moreover, Table 4 shows the results (in terms of MSE, SEP and #*conn*) for the best models, selected according to the best value of Average MSE, which is the metric used for the optimisation step. Table 4 includes the results obtained by the deterministic techniques considered, and it is also divided into three different parts, showing each part the results for the different buoys (46001, 46066 and 46085, respectively).



(a) SU-LI

(b) SVR

(c) ELM

Fig. 8. Predictions achieved by the best model and the two main state-of-the-art models for the buoy 46066 (prediction horizon $12h$).

As can be seen in Table 4, SU-LI is also achieving the best values in terms of MSE, with Average values of 1042.08, 1163.62 and 1004.12 for the buoys 46001, 46066 and 46085, respectively. Moreover, regarding SEP values, SU-LI also achieves the best values for buoys 46001 and 46085 (110.83% and 100.83%, respectively), whereas for buoy 46066, ELM obtains the best results (90.48%). Besides, regarding the number of connections, standard regressions models are the simplest ones, but, if we only consider the complex models (i.e. MTEANN, SVR and ELM), it can be seen that MTEANNs use fewer number of links in comparison against SVR and ELM. Although SU-PU do not stand out by its average results, it is worthy of mention that it achieves the second best results (for both, MSE and SEP) for the buoys 46001 and 46085, and competitive results for the buoy 46066, in terms of MSE. As can be seen, SU-PU best models have obtained a good performance although its average results are slightly worse. The reason of such behaviour is that small changes in the weights (exponents of the PU basis function) cause large changes in the outputs of the neural network, which makes SU-PU model unstable to train when compared to the other MTEANNs considered.

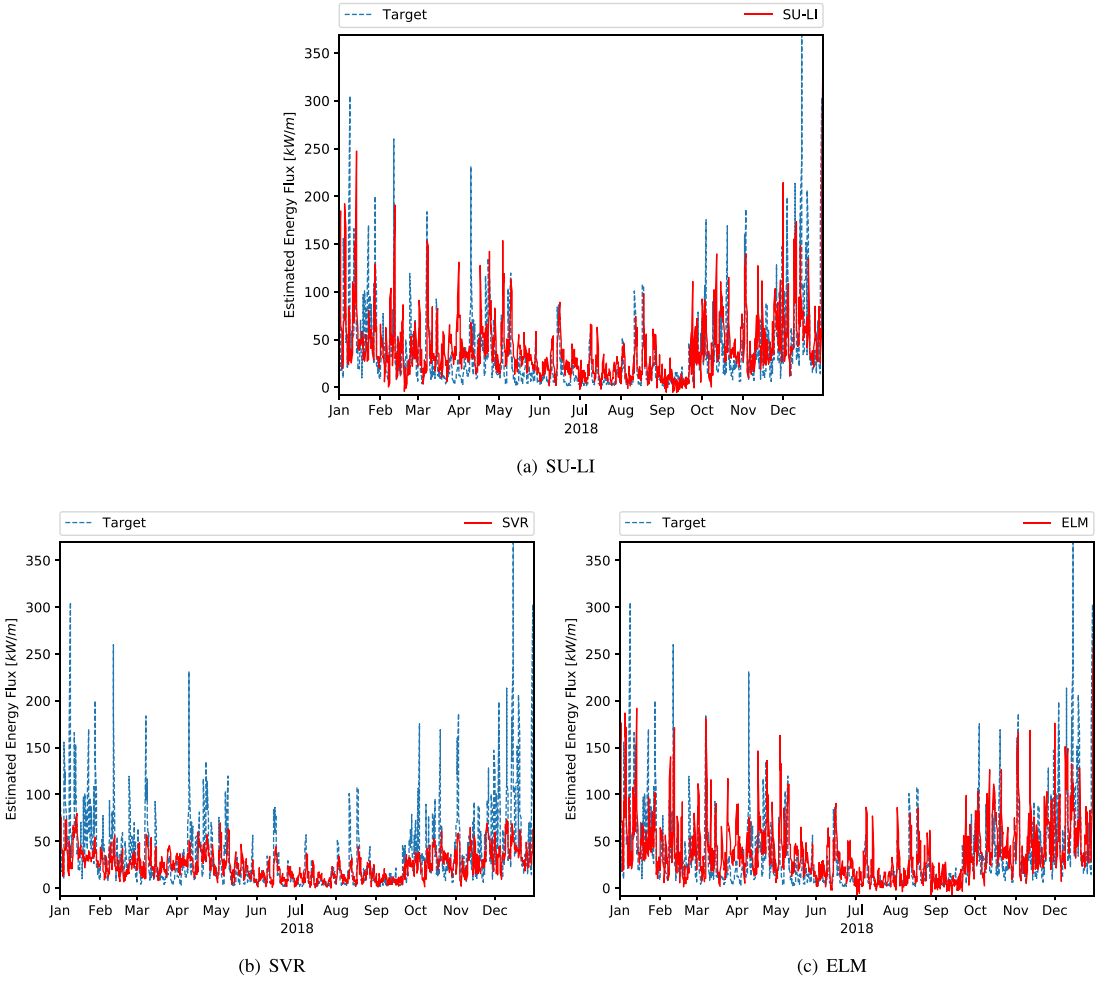
Therefore, the use of MTEANN models is justified due to they achieve the best and the second best values in terms of both, MSE and SEP for most of the buoys. Furthermore, the performance obtained by the MTEANN best models is considerably better than the one obtained for the state-of-the-art methods considered. Summarising, SU-LI structure does not only achieve competitive results in average terms but also obtains competitive values for the best models in terms of both, MSE and SEP, using a small number of connections.

In addition, the proposed MTEANN methodology has been compared against the same methodology but using a single-task scheme, i.e. EANNs predicting only one time prediction horizon. In this sense, the performance of SU-LI model has been compared to four single-task models that predict the flux of energy at $6h$, $12h$, $24h$ and $48h$, respectively, which are referred as SU-LI$_{ST}$. To train such models, the same parameter values described in Section 4.1 have been considered, though using 3000 generations because of single-task scheme is less demanding.

The results obtained are shown in Table 5. Concerning mean values, we can observe that the multi-task model (SU-LI) model achieves better Average MSE and SEP results in all buoys than the four single-task models (denoted as SU-LI$_{ST}$). Besides, SU-LI model also obtains the best MSE and SEP results in each time prediction horizon except for the $6h$ time prediction horizon of the buoy 46066. It can be checked that the differences in performance increase from $12h$ onward, being higher in long-term predictions, specially for buoys 46066 ($24h$ and $48h$) and 46085 ($48h$), demonstrating that MTEANN models benefit from the short-term prediction time horizons.

Regarding best models values, SU-LI model also achieves the best Average performance for each buoy. SU-LI$_{ST}$ models get best results for $6h$ time prediction horizon of all buoys. However, for $12h$, $24h$ (buoys 46001 and 46066) and $48h$ (all buoys) time prediction horizons, SU-LI model takes advantage of related-tasks, resulting in more accurate models.

These results confirm that considering a multi-task model for related energy flux prediction tasks reduces the average error and almost all individual errors. This is due to the fact that multi-task paradigm is



(a) SU-LI

(b) SVR

(c) ELM

Fig. 9. Predictions achieved by the best model and the two main state-of-the-art models for the buoy 46066 (prediction horizon $24h$).

able to benefit from what is learnt in a given task to improve the error in other related tasks. Moreover, given that MTEANN methodology requires only one model and less connections in comparison to the four single-task models, their implementation is easier and even achieves a better performance. Therefore, it can be considered more efficient for the problem tackled.

5. Discussion

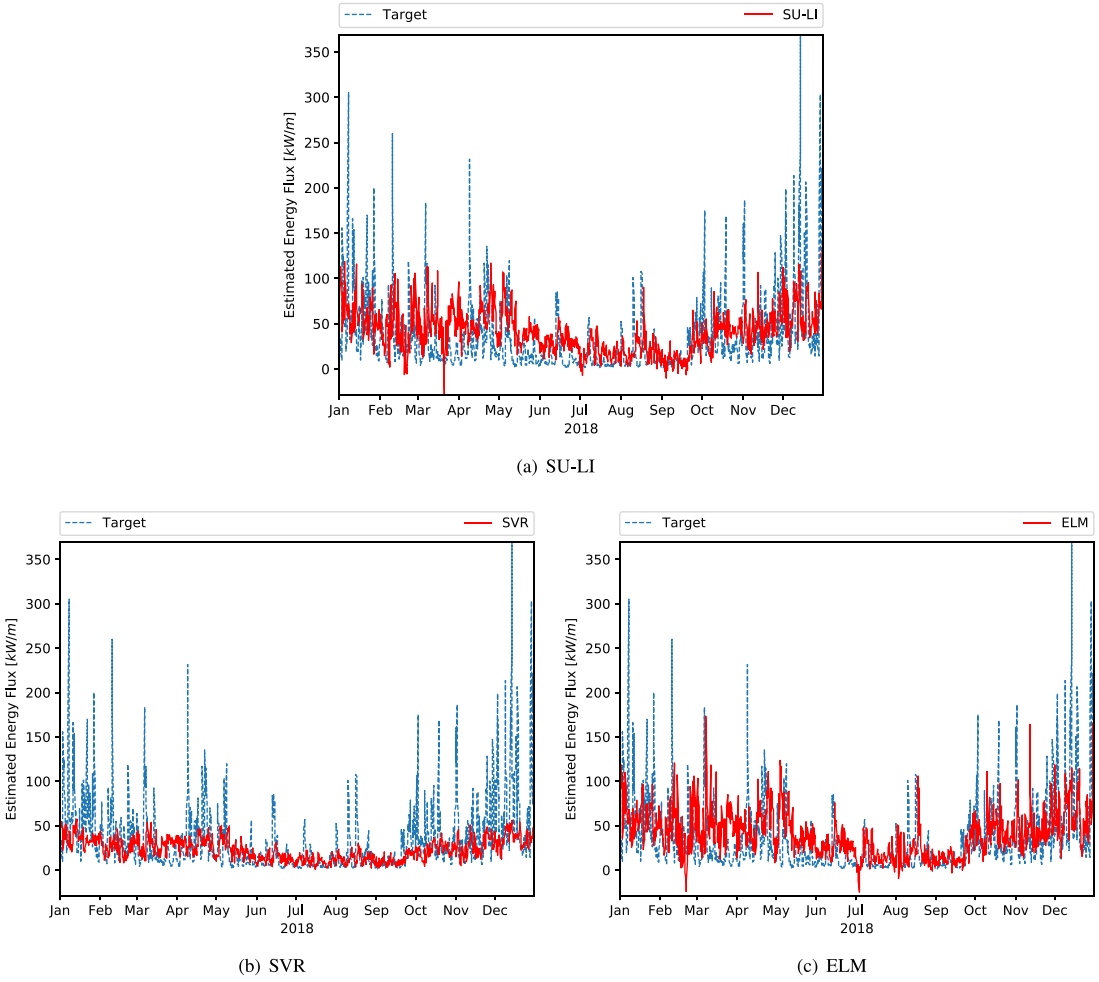
The main objective of this paper was to find a model able to accurately predict the energy flux for three buoys located in the Gulf of Alaska at four different time prediction horizons $6h$, $12h$, $24h$ and $48h$. In this way, we have found that SU-LI model is able to achieve competitive results in comparison against the most popular techniques in the state of the art. Furthermore, not only is this model able to achieve the best performance in terms of both MSE and SEP but also it is composed by a small number of links, in comparison to other models of similar complexity, such as SVR or ELM.

Regarding the results of the best models in Table 4, the SU-LI model achieves the best performance in terms of MSE for all the buoys, and the best results in SEP for two of the three buoys, with a small number of connections (less complexity). In order to demonstrate the robustness and stability of this model, the distribution of the results for the 30 executions shown in Table 3 is graphically presented in Fig. 4, considering both performance measures: MSE (a) and SEP (b). We can appreciate that the models developed are stable for all the buoys, having a small standard deviation and exhibiting only a few runs that could be considered as outliers, most of them for the buoy 46066.

Fig. 5 shows the performance of the SU-LI (a), SVR (b) and ELM (c) models for the buoy 46085 considering only one of the outputs (prediction horizon $6h$) and the first 1000 test samples for visualisation purposes. In these figures, the difference in performance can be appreciated between the best method (SU-LI) and two of the most popular techniques in the state-of-the-art (SVR and ELM). SVR achieves the lowest performance, not being able to predict the highest values of energy flux. Regarding ELM and SU-LI, both seem to correctly estimate most of the low values of energy flux. However, the main difference lies in the estimation of the high values, where SU-LI achieves a better performance, as can be seen between the 200th and 400th test samples.

Moreover, Fig. 6 shows the scatter plots for the prediction of energy flux at the time prediction horizon of $24h$ for the SU-LI (a), SVR (b) and ELM (c) models, considering the buoy 46001. We have chosen this time prediction horizon to demonstrate that MTEANNs are able to obtain a competitive performance also in the long-term prediction. Specifically, it can be seen that, as happened in Fig. 5, SVR is only able to correctly estimate low values of energy flux. Nevertheless, SU-LI and ELM are able to estimate more adequately also the high values of energy flux, being SU-LI specially better for those in the range from 100 to 200 kW/m.

So far, the analysis has been focused on the ability of the models to predict the energy flux at short-term and long-term. However, it is worthy to discuss the performance of the models from the time prediction horizon point-of-view. In this sense, Table 6 shows the average of each lead time for the three stations. Note that the "Average" column includes the mean and SD of the stochastic models, which have



(a) SU-LI

(b) SVR

(c) ELM

Fig. 10. Predictions achieved by the best model and the two main state-of-the-art models for the buoy 46066 (prediction horizon 48h).

been run 30 times, whereas the "Best" column includes errors of both stochastic and deterministic models of Table 4.

As can be concluded from Table 6, the best results are achieved for the 6h time prediction horizon, the error increasing for higher lead times (the lowest values for both MSE and SEP are achieved for the 6h time prediction horizon, whereas the highest values are obtained for the 48h time prediction horizon).

Furthermore, in order to provide a graphical comparison of the different time prediction horizons, Figs. 7, 8, 9 and 10 show the predictions achieved for the buoy 46066 by the best model (SU-LI) and the two main state-of-the-art models in the literature (SVR and ELM), for the prediction horizons 6h, 12h, 24h and 48h, respectively. For the three first prediction horizons (6h, 12h and 24h), it can be seen that, in general, ELM and SU-LI models achieve satisfactory results, although those from SU-LI are slightly better. As expected, the results for the long-term prediction horizon (48h) are less accurate, but the models are able to follow the general behaviour of the energy flux.

6. Conclusions

In this paper, we have analysed the performance of four different types of Multi-Task Evolutionary Artificial Neural Networks (MTEANN) in the field of marine energy flux prediction. Specifically, observed data from three buoys located in the Gulf of Alaska have been considered in order to measure the energy flux, whereas seven variables from reanalysis data have been used as input variables. Therefore, the prediction of the energy flux is performed only considering reanalysis data, avoiding missing data problems and allowing the applicability to other locations. Moreover, the target values have been taken at 4 different time prediction horizons (6h, 12h, 24h and 48h) with the main goal of predicting not only short-term values but also long-term ones. The principal objective of this paper was to obtain a multi-task model achieving competitive results with lower-complexity, i.e. simple models with a small number of connections. In this sense, MTEANNs using different sort of basis functions (Product Units—PUs, Sigmoidal Units—SUs and Radial Basis Functions—RBFs) have been tested, resulting in the combination of SUs in the hidden layer with linear units in the output layer (SU-LI), the one that achieves the best results for the three buoys considered. In addition, RBF-LI and PU-LI approaches also achieve competitive results, benefiting also from its low-complexity. Concerning the SU-PU approach, the best models have obtained a good performance although the average results of the 30 runs are slightly worse (less stability). Furthermore, these MTEANN models have been compared against several state-of-the-art approaches, including the popular Support Vector Regressors (SVR) and Extreme Learning Machines (ELM), which have been proved to be competitive for the problem tackled, yet obtaining worse results than SU-LI. Therefore, we have shown that the proposed approaches are able to solve the problem with a high level of accuracy, demonstrating that MTEANN models are an excellent approach for the prediction of energy flux at both short- and long-term time prediction horizons. Finally, the results obtained also demonstrated that the proposed methodology achieve accurate energy flux predictions using reanalysis data as input, without requiring additional observed data during the test phase. As future research, it will be interesting to broaden the study including buoys from other locations and to test hybrid models able to take advantage of the combination of different basis functions.

Table 4
Best models values of MSE, SEP, and *#conn* obtained for the different buoys.

Buoy	Model	MSE					SEP (%)					*#conn*
		Average	6 h	12 h	24 h	48 h	Average	6 h	12 h	24 h	48 h	
46001	RBF-LI	1057.60	713.59	797.47	1322.94	1396.41	111.82	92.84	98.11	125.82	130.51	79
	SU-LI	**1042.08**	**643.20**	**821.37**	**1344.99**	**1358.78**	**110.83**	**88.14**	**99.57**	**126.86**	**128.74**	260
	PU-LI	1088.74	755.09	899.32	1346.30	1354.25	113.78	95.50	104.19	126.92	128.52	129
	SU-PU	*1053.67*	*620.45*	*873.76*	*1380.23*	*1340.23*	*111.41*	*86.57*	*102.69*	*128.51*	*127.86*	235
	SVR	1308.68	1146.96	1156.44	1433.80	1497.56	188.92	167.24	166.28	198.96	223.24	138944
	LinearReg	1383.32	1339.32	1335.36	1443.48	1415.04	129.28	128.76	129.44	133.36	125.52	*32*
	RidgeReg	1383.32	1339.80	1335.80	1442.84	1414.76	129.08	128.56	129.16	133.04	125.52	*32*
	LassoReg	1383.71	1342.37	1338.10	1441.04	1413.35	128.55	128.71	128.74	131.70	125.06	**29**
	ElasticNet	1385.22	1344.91	1340.23	1441.55	1414.20	128.21	128.33	128.24	131.18	125.11	**29**
	ELM	1066.19	719.66	859.62	1325.20	1360.27	116.22	100.79	111.03	130.32	122.74	660
46066	RBF-LI	1342.10	1097.89	1234.79	1401.65	1634.07	105.71	96.52	102.06	108.15	116.08	85
	SU-LI	**1163.62**	**815.80**	**975.37**	**1322.09**	**1541.21**	97.92	83.20	90.71	105.04	112.73	261
	PU-LI	1190.28	905.96	1074.82	1281.84	1498.50	99.37	87.68	95.22	103.43	111.16	110
	SU-PU	1182.01	856.35	1012.39	1321.99	1537.30	98.82	85.25	92.42	105.04	112.59	229
	SVR	*1175.24*	*973.60*	*982.16*	*1200.40*	*1544.72*	135.68	117.20	117.52	138.56	169.44	139136
	LinearReg	1304.44	1215.60	1218.88	1274.28	1508.96	93.92	93.36	93.28	92.36	96.68	*32*
	RidgeReg	1293.64	1202.20	1206.08	1264.96	1501.28	93.44	92.72	92.60	91.96	96.48	*32*
	LassoReg	1302.18	1212.82	1216.50	1272.44	1506.97	93.86	93.30	93.19	92.32	96.64	**29**
	ElasticNet	1289.16	1196.74	1200.88	1261.16	1497.85	*93.28*	*92.49*	*92.40*	*91.82*	*96.42*	**29**
	ELM	1234.49	997.03	1071.48	1280.42	1589.05	**90.48**	**85.52**	**87.30**	**90.88**	**98.22**	550
46085	RBF-LI	1055.96	689.03	787.19	1237.91	1509.73	103.56	84.63	90.40	113.35	125.85	86
	SU-LI	**1004.12**	**613.56**	**757.77**	**1205.67**	**1439.47**	**100.83**	**79.86**	**88.69**	**111.86**	**122.89**	253
	PU-LI	1127.92	775.42	900.50	1321.45	1514.32	107.40	89.78	96.69	117.11	126.04	121
	SU-PU	*1020.07*	*622.46*	*742.52*	*1210.39*	*1504.93*	*101.49*	*80.44*	*87.80*	*112.08*	*125.65*	257
	SVR	1360.72	1156.64	1190.08	1456.40	1639.68	180.08	158.36	160.08	187.08	214.72	139056
	LinearReg	1358.20	1251.20	1281.20	1412.48	1488.00	126.80	124.52	125.60	129.00	128.12	*32*
	RidgeReg	1358.12	1252.12	1281.12	1410.60	1488.64	126.76	124.44	125.48	128.84	128.16	*32*
	LassoReg	1357.93	1252.67	1279.92	1406.38	1492.75	126.27	124.16	125.07	128.17	127.68	**29**
	ElasticNet	1357.93	1252.67	1279.92	1406.38	1492.75	126.27	124.16	125.07	128.17	127.68	**29**
	ELM	1045.63	701.61	801.41	1212.19	1467.33	108.79	92.20	96.95	119.55	126.44	660

The best result is highlighted in **bold**; the second one best result is shown in *italics*.

Table 5
Comparison of SU-LI and SU-LI$_{ST}$ models in terms of MSE, SEP, and *#conn* obtained for the different buoys.

Buoy	Model	$Mean_{SD}$ values										
		MSE					SEP (%)					*#conn*
		Average	6 h	12 h	24 h	48 h	Average	6 h	12 h	24 h	48 h	
46001	SU-LI	$\mathbf{1082.78_{21.31}}$	$\mathbf{649.31_{22.87}}$	$\mathbf{898.28_{46.57}}$	$\mathbf{1417.12_{45.08}}$	$\mathbf{1366.39_{23.43}}$	$\mathbf{112.98_{1.15}}$	$\mathbf{88.55_{1.56}}$	$\mathbf{104.09_{2.69}}$	$\mathbf{130.20_{2.06}}$	$\mathbf{129.09_{1.10}}$	$\mathbf{250.13_{7.73}}$
	SU-LI$_{ST}$	$1116.86_{24.30}$	$661.45_{46.09}$	$937.33_{53.63}$	$1440.16_{36.85}$	$1428.48_{53.07}$	$114.72_{1.28}$	$89.33_{3.02}$	$106.32_{2.98}$	$131.26_{1.68}$	$131.98_{2.40}$	$744.00_{16.76}$
46066	SU-LI	$\mathbf{1250.52_{100.21}}$	$880.80_{37.83}$	$\mathbf{1107.80_{168.44}}$	$\mathbf{1439.68_{227.73}}$	$\mathbf{1573.78_{45.69}}$	$\mathbf{101.53_{3.82}}$	$86.4_{1.86}$	$\mathbf{96.43_{6.90}}$	$\mathbf{109.33_{8.01}}$	$\mathbf{113.91_{1.63}}$	$\mathbf{250.10_{8.62}}$
	SU-LI$_{ST}$	$1457.60_{174.90}$	$\mathbf{869.97_{26.34}}$	$1253.46_{227.45}$	$1744.86_{635.43}$	$1962.12_{210.10}$	$108.69_{5.28}$	$\mathbf{85.91_{1.30}}$	$102.50_{8.33}$	$119.32_{18.33}$	$127.03_{6.71}$	$729.07_{29.42}$
46085	SU-LI	$\mathbf{1038.77_{22.36}}$	$\mathbf{645.88_{30.72}}$	$\mathbf{788.78_{46.08}}$	$\mathbf{1237.22_{40.76}}$	$\mathbf{1483.20_{37.41}}$	$\mathbf{102.60_{1.18}}$	$\mathbf{81.91_{1.93}}$	$\mathbf{90.45_{2.59}}$	$\mathbf{113.30_{1.82}}$	$\mathbf{124.73_{1.57}}$	$\mathbf{250.87_{9.18}}$
	SU-LI$_{ST}$	$1101.94_{43.81}$	$661.74_{42.13}$	$848.39_{81.36}$	$1270.02_{51.61}$	$1627.61_{98.87}$	$105.51_{2.04}$	$82.89_{2.59}$	$93.75_{4.34}$	$114.78_{2.30}$	$130.62_{3.90}$	$751.23_{19.17}$
Buoy	Model	Best models values										
		MSE					SEP (%)					*#conn*
		Average	6 h	12 h	24 h	48 h	Average	6 h	12 h	24 h	48 h	
46001	SU-LI	**1042.08**	643.20	**821.37**	**1344.99**	**1358.78**	**110.83**	88.14	**99.57**	**126.86**	**128.74**	**260**
	SU-LI$_{ST}$	1051.62	**595.39**	856.59	1379.00	1375.50	111.12	**84.80**	101.68	128.45	129.53	710
46066	SU-LI	**1163.62**	815.80	**975.37**	**1322.09**	**1541.21**	**97.92**	83.20	**90.71**	**105.04**	**112.73**	**261**
	SU-LI$_{ST}$	1228.20	**805.35**	1067.32	1394.61	1645.53	100.48	**82.67**	94.89	107.88	116.49	675
46085	SU-LI	**1004.12**	613.56	757.77	1205.67	**1439.47**	**100.83**	79.86	88.69	111.86	**122.89**	**253**
	SU-LI$_{ST}$	1008.68	**609.79**	**737.09**	**1203.31**	1484.53	100.91	**79.61**	**87.48**	**111.75**	124.80	708

The best result is highlighted in **bold**.

Table 6
Average of each lead time for the three stations.

Lead time	Average		Best	
	MSE	SEP (%)	MSE	SEP (%)
6 h	835.68 ± 61.95	92.37 ± 3.41	1000.22	105.17
12 h	990.21 ± 84.21	100.47 ± 4.05	1075.08	109.20
24 h	1381.14 ± 90.46	118.44 ± 3.67	1336.12	122.05
48 h	1510.40 ± 75.46	122.88 ± 2.92	1478.75	127.42

CRediT authorship contribution statement

David Guijo-Rubio: Methodology, Validation, Formal analysis, Investigation, Writing - original draft, Visualization. **Antonio M. Gómez-Orellana:** Methodology, Software, Validation, Investigation, Data curation, Writing - original draft. **Pedro A. Gutiérrez:** Conceptualization, Software, Formal analysis, Writing - review & editing. **César Hervás-Martínez:** Conceptualization, Resources, Supervision, Project administration, Funding acquisition.

Acknowledgements

This research has been partially supported by the "Ministerio de Economía, Industria y Competitividad" of Spain (Ref. TIN2017-85887-C2-1-P) and the "Fondo Europeo de Desarrollo Regional (FEDER) y de la Consejería de Economía, Conocimiento, Empresas y Universidad de la Junta de Andalucía" (Ref. UCO-1261651). D. Guijo-Rubio's research has been supported by the FPU Predoctoral Program from Spanish Ministry of Science, Innovations and Universities (Grant Ref. FPU16/02128).

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.oceaneng.2020.108089.

References

Aderinto, T., Li, H., 2018. Ocean wave energy converters: Status and challenges. Energies 11 (5), 1250. http://dx.doi.org/10.3390/en11051250.

Ali, M., Prasad, R., 2019. Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition. Renew. Sustain. Energy Rev. 104, 281–295. http://dx.doi.org/10.1016/j.rser.2019.01.014.

Angeline, P.J., Saunders, G.M., Pollack, J.B., 1994. An evolutionary algorithm that constructs recurrent neural networks. Trans. Neural Netw. 5 (1), 54–65. http://dx.doi.org/10.1109/72.265960.

Billings, S.A., Zheng, G.L., 1995. Radial basis function network configuration using genetic algorithms. Neural Netw. 8 (6), 877–890. http://dx.doi.org/10.1016/0893-6080(95)00029-Y.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer.

Bishop, C.M., et al., 1995. Neural Networks for Pattern Recognition. Oxford University Press.

Campmany, E., Bech, J., Rodríguez-Marcos, J., Sola, Y., Lorente, J., 2010. A comparison of total precipitable water measurements from radiosonde and sunphotometers. Atmos. Res. 97 (3), 385–392. http://dx.doi.org/10.1016/j.atmosres.2010.04.016.

Caruana, R., 1997. Multitask Learning. Mach. Learn. 28 (1), 41–75. http://dx.doi.org/10.1023/A:1007379606734.

Castro, A., Carballo, R., Iglesias, G., Rabuñal, J., 2014. Performance of artificial neural networks in nearshore wave power prediction. Appl. Soft Comput. 23, 194–201. http://dx.doi.org/10.1016/j.asoc.2014.06.031.

Choi, H., Park, M., Son, G., Jeong, J., Park, J., Mo, K., Kang, P., 2020. Real-time significant wave height estimation from raw ocean images based on 2D and 3D deep neural networks. Ocean Eng. 201, 107129. http://dx.doi.org/10.1016/j.oceaneng.2020.107129.

Choy, S., Wang, C., Zhang, K., Kuleshov, Y., 2013. GPS Sensing of precipitable water vapour during the March 2010 Melbourne storm. Adv. Space Res. 52 (9), 1688–1699. http://dx.doi.org/10.1016/j.asr.2013.08.004.

Cornejo-Bueno, L., Borge, J.N., Alexandre, E., Hessner, K., Salcedo-Sanz, S., 2016a. Accurate estimation of significant wave height with support vector regression algorithms and marine radar images. Coast. Eng. 114, 233–243. http://dx.doi.org/10.1016/j.coastaleng.2016.04.007.

Cornejo-Bueno, L., Nieto-Borge, J., García-Díaz, P., Rodríguez, G., Salcedo-Sanz, S., 2016b. Significant wave height and energy flux prediction for marine energy applications: A grouping genetic algorithm – Extreme Learning Machine approach. Renew. Energy 97, 380–389. http://dx.doi.org/10.1016/j.renene.2016.05.094.

Cornejo-Bueno, L., Rodríguez-Mier, P., Mucientes, M., Nieto-Borge, J., Salcedo-Sanz, S., 2018. Significant wave height and energy flux estimation with a Genetic Fuzzy System for regression. Ocean Eng. 160, 33–44. http://dx.doi.org/10.1016/j.oceaneng.2018.04.063.

Council, N.R., et al., 2012. Climate Change: Evidence, Impacts, and Choices: Set of 2 Booklets, with DVD. National Academies Press.

Cuadra, L., Salcedo-Sanz, S., Nieto-Borge, J., Alexandre, E., Rodríguez, G., 2016. Computational intelligence in wave energy: Comprehensive review and case study. Renew. Sustain. Energy Rev. 58, 1223–1246. http://dx.doi.org/10.1016/j.rser.2015.12.253.

de Smith, M.J., Goodchild, M.F., Longley, P.A., 2009. Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools, third revised ed. Matador, p. 516.

Deo, M., Jha, A., Chaphekar, A., Ravikant, K., 2001. Neural networks for wave forecasting. Ocean Eng. 28 (7), 889–898. http://dx.doi.org/10.1016/S0029-8018(00)00027-5.

Deo, M., Naidu, C.S., 1998. Real time wave forecasting using neural networks. Ocean Eng. 26 (3), 191–203. http://dx.doi.org/10.1016/S0029-8018(97)10025-7.

Dorado-Moreno, M., Cornejo-Bueno, L., Gutiérrez, P., Prieto, L., Hervás-Martínez, C., Salcedo-Sanz, S., 2017. Robust estimation of wind power ramp events with reservoir computing. Renew. Energy 111, 428–437. http://dx.doi.org/10.1016/j.renene.2017.04.016.

Dorado-Moreno, M., Navarin, N., Gutiérrez, P., Prieto, L., Sperduti, A., Salcedo-Sanz, S., Hervás-Martínez, C., 2020. Multi-task learning for the prediction of wind power ramp events with deep neural networks. Neural Netw. 123, 401–411. http://dx.doi.org/10.1016/j.neunet.2019.12.017.

Duan, W., Han, Y., Huang, L., Zhao, B., Wang, M., 2016. A hybrid EMD-SVR model for the short-term prediction of significant wave height. Ocean Eng. 124, 54–73. http://dx.doi.org/10.1016/j.oceaneng.2016.05.049.

Ellabban, O., Abu-Rub, H., Blaabjerg, F., 2014. Renewable energy resources: Current status, future prospects and their enabling technology. Renew. Sustain. Energy Rev. 39, 748–764. http://dx.doi.org/10.1016/j.rser.2014.07.113.

Esteban, M., Leary, D., 2012. Current developments and future prospects of offshore wind and ocean energy. Appl. Energy 90 (1), 128–136. http://dx.doi.org/10.1016/j.apenergy.2011.06.011, Energy Solutions for a Sustainable World, Special Issue of International Conference of Applied Energy, ICA2010, April 21-23, 2010, Singapore.

Falcão, A.F.d.O., 2010. Wave energy utilization: A review of the technologies. Renew. Sustain. Energy Rev. 14 (3), 899–918. http://dx.doi.org/10.1016/j.rser.2009.11.003.

Falnes, J., Kurniawan, A., 2020. Ocean Waves and Oscillating Systems: Linear Interactions Including Wave-Energy Extraction, Vol. 8. Cambridge University Press.

Fernández, J., Salcedo-Sanz, S., Gutiérrez, P., Alexandre, E., Hervás-Martínez, C., 2015. Significant wave height and energy flux range forecast with machine learning classifiers. Eng. Appl. Artif. Intell. 43, 44–53. http://dx.doi.org/10.1016/j.engappai.2015.03.012.

Fernandez Caballero, J.C., Martinez, F.J., Hervas, C., Gutierrez, P.A., 2010. Sensitivity versus accuracy in multiclass problems using memetic Pareto evolutionary neural networks. IEEE Trans. Neural Netw. 21 (5), 750–770. http://dx.doi.org/10.1109/TNN.2010.2041468.

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33 (1), 1–22, Articles.

Hadadpour, S., Etemad-Shahidi, A., Kamranzad, B., 2014. Wave energy forecasting using artificial neural networks in the Caspian sea. Proc. Inst. Civil Eng. - Maritime Eng. 167 (1), 42–52. http://dx.doi.org/10.1680/maen.13.00004.

Hervás, C., Gutierrez, P.A., Silva, M., Serrano, J.M., 2007. Combining classification and regression approaches for the quantification of highly overlapping capillary electrophoresis peaks by using evolutionary sigmoidal and product unit neural networks. J. Chemometr. 21 (12), 567–577. http://dx.doi.org/10.1002/cem.1082.

Hoegh-Guldberg, O., Jacob, D., Bindi, M., Brown, S., Camilloni, I., Diedhiou, A., Djalante, R., Ebi, K., Engelbrecht, F., Guiot, J., et al., 2018. Impacts of 1.5 C Global Warming on Natural and Human Systems. Global warming of 1.5° C. An IPCC Special Report, IPCC Secretariat.

Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., 2006. Extreme learning machine: Theory and applications. Neurocomputing 70 (1), 489–501. http://dx.doi.org/10.1016/j.neucom.2005.12.126.

Ibarra-Berastegi, G., Saénz, J., Esnaola, G., Ezcurra, A., Ulazia, A., 2015. Short-term forecasting of the wave energy flux: Analogues, random forests, and physics-based models. Ocean Eng. 104, 530–539. http://dx.doi.org/10.1016/j.oceaneng.2015.05.038.

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., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Jenne, R., Joseph, D., 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Amer. Meteorol. Soc. 77 (3), 437–471. http://dx.doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kisi, O., Shiri, J., Karimi, S., Shamshirband, S., Motamedi, S., Petković, D., Hashim, R., 2015. A survey of water level fluctuation predicting in Urmia Lake using support vector machine with firefly algorithm. Appl. Math. Comput. 270, 731–743. http://dx.doi.org/10.1016/j.amc.2015.08.085.

Kistler, R., Collins, W., Saha, S., White, G., Woollen, J., Kalnay, E., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., Fiorino, M., 2001. The NCEP–NCAR 50–year reanalysis: Monthly means CD–ROM and documentation. Bull. Amer. Meteorol. Soc. 82 (2), 247–267. http://dx.doi.org/10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2.

Kumar, N.K., Savitha, R., Mamun, A.A., 2018. Ocean wave height prediction using ensemble of Extreme Learning Machine. Neurocomputing 277, 12–20. http://dx.doi.org/10.1016/j.neucom.2017.03.092.

Lin, Y., Dong, S., Tao, S., 2020. Modelling long-term joint distribution of significant wave height and mean zero-crossing wave period using a copula mixture. Ocean Eng. 197, 106856. http://dx.doi.org/10.1016/j.oceaneng.2019.106856.

Lippmann, R.P., 1989. Pattern classification using neural networks. IEEE Commun. Mag. 27 (11), 47–50. http://dx.doi.org/10.1109/35.41401.

Mahjoobi, J., Mosabbeb, E.A., 2009. Prediction of significant wave height using regressive support vector machines. Ocean Eng. 36 (5), 339–347. http://dx.doi.org/10.1016/j.oceaneng.2009.01.001.

Martínez-Estudillo, F., Hervás-Martínez, C., Gutiérrez, P., Martínez-Estudillo, A., 2008. Evolutionary product-unit neural networks classifiers. Neurocomputing 72 (1), 548–561. http://dx.doi.org/10.1016/j.neucom.2007.11.019, Machine Learning for Signal Processing (MLSP 2006) / Life System Modelling, Simulation, and Bio-inspired Computing (LSMS 2007).

Martínez-Estudillo, A., Martínez-Estudillo, F., Hervás-Martínez, C., García-Pedrajas, N., 2006. Evolutionary product unit based neural networks for regression. Neural Netw. 19 (4), 477–486. http://dx.doi.org/10.1016/j.neunet.2005.11.001.

Maurer, A., Pontil, M., Romera-Paredes, B., 2016. The benefit of multitask representation learning. J. Mach. Learn. Res. 17 (1), 2853–2884. http://dx.doi.org/10.5555/2946645.3007034.

National Data Buoy Center, 2020. National oceanic and atmospheric administration of the USA (NOAA). http://www.ndbc.noaa.gov/. (Accessed 1 June 2020).

NDBC - Station 46001, 2020. NDBC - Station 46001 Recent Data. https://www.ndbc.noaa.gov/station_page.php?station=46001. (Accessed 1 June 2020).

NDBC - Station 46066, 2020. NDBC - Station 46066 Recent Data. https://www.ndbc.noaa.gov/station_page.php?station=46066. (Accessed 1 June 2020).

NDBC - Station 46085, 2020. NDBC - Station 46085 Recent Data. https://www.ndbc.noaa.gov/station_page.php?station=46085.(Accessed 1 June 2020).

Nitsure, S., Londhe, S., Khare, K., 2012. Wave forecasts using wind information and genetic programming. Ocean Eng. 54, 61–69. http://dx.doi.org/10.1016/j.oceaneng.2012.07.017.

Palmer, G., Floyd, J., 2020. Electricity: A new challenge for storage. In: Energy Storage and Civilization: A Systems Approach. Springer International Publishing, Cham, pp. 71–88. http://dx.doi.org/10.1007/978-3-030-33093-4_5.

Salcedo-Sanz, S., Borge, J.N., Carro-Calvo, L., Cuadra, L., Hessner, K., Alexandre, E., 2015. Significant wave height estimation using SVR algorithms and shadowing information from simulated and real measured X-band radar images of the sea surface. Ocean Eng. 101, 244–253. http://dx.doi.org/10.1016/j.oceaneng.2015.04.041.

Schmitt, M., 2002. On the complexity of computing and learning with multiplicative neural networks. Neural Comput. 14 (2), 241–301. http://dx.doi.org/10.1162/089976602741121.

Ser, J.D., Osaba, E., Molina, D., Yang, X.-S., Salcedo-Sanz, S., Camacho, D., Das, S., Suganthan, P.N., Coello, C.A.C., Herrera, F., 2019. Bio-inspired computation: Where we stand and what's next. Swarm Evol. Comput. 48, 220–250. http://dx.doi.org/10.1016/j.swevo.2019.04.008.

Shiri, J., Shamshirband, S., Kisi, O., Karimi, S., Bateni, S.M., Nezhad, S.H.H., Hashemi, A., 2016. Prediction of water-level in the Urmia lake using the extreme learning machine approach. Water Resour. Manag. 30 (14), 5217–5229. http://dx.doi.org/10.1007/s11269-016-1480-x.

Siingh, D., Buchunde, P., Singh, R., Nath, A., Kumar, S., Ghodpage, R., 2014. Lightning and convective rain study in different parts of India. Atmos. Res. 137, 35–48. http://dx.doi.org/10.1016/j.atmosres.2013.09.018.

Vapnik, V., 2013. The Nature of Statistical Learning Theory. Springer science & business media, http://dx.doi.org/10.1007/978-1-4757-2440-0.

Wang, X., Zhang, K., Wu, S., Li, Z., Cheng, Y., Li, L., Yuan, H., 2018. The correlation between GNSS-derived precipitable water vapor and sea surface temperature and its responses to El Niño–Southern Oscillation. Remote Sens. Environ. 216, 1–12. http://dx.doi.org/10.1016/j.rse.2018.06.029.

Yao, X., 1999. Evolving artificial neural networks. Proc. IEEE 87 (9), 1423–1447. http://dx.doi.org/10.1109/5.784219.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B Stat. Methodol. 67 (2), 301–320. http://dx.doi.org/10.1111/j.1467-9868.2005.00503.x.