Significant Wave Height and Energy Flux Estimation with a Genetic Fuzzy System for Regression

L. Cornejo-Bueno [a], P. Rodríguez-Mier [b], M. Mucientes [b], J. C. Nieto-Borge [a], S. Salcedo-Sanz* [a],

[a] *Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares, Spain.*

[b] *Centro Singular de Investigación en Tecnoloxías da Información (CiTIUS) Universidade de Santiago de Compostela. Santiago de Compostela, Spain.*

Abstract

This paper proposes a regression Genetic Fuzzy System (GFS, FRULER) for a problem of sea wave parameters estimation from neighbor buoys, with application on wave energy systems. FRULER is a recently proposed, three-staged algorithm that combines an instance selection method for regression, a multigranularity fuzzy discretization of the input variables and an evolutionary algorithm to generate accurate and simple Takagi-Sugeno-Kant (TSK) fuzzy rules. We have applied FRULER to a real problem of significant wave height and energy flux prediction in one buoy of the West Coast of the USA (California), from values of other two neighbor buoys. In the case of the significant wave height, FRULER is able to obtain a robust prediction with only three rules, which in addition are fully interpretable, since they clearly separate swell situations from wind-sea in the prediction. In both cases, the variables used in the significant wave prediction are completely different and can be identified as relevant for the specific case (swell or wind-sea). In the case of the energy flux prediction, the interpretation of the rules provided by FRULER is more difficult, since eight rules are necessary to obtain the prediction. Even in this case, several rules can be clearly classified as swell predictors, and the rest of the rules describe local wind situation of waves. This study shows that the GFSs are useful tools to obtain robust and interpretable predictions in ocean wave parameter estimation problems.

Key words: Significant Wave Height; Wave Energy Flux; Marine Systems; Genetic Fuzzy Systems.

1 Introduction

Sea waves parameters prediction problems play an important role in many different ocean engineering tasks, such as the design of marine structures like oil platforms or harbours [1,2] or in the design and management of marine energy systems [3,4], like the proper operation of wave energy converters [5], among others. Among the different sea wave parameters, the significant wave height (H_{m_0}) and the wave energy flux (P) are the two most studied in the literature. In the case of H_{m_0}, it is usually estimated using in-situ sensors, such as buoys, recording time series of wave elevation information. The significant wave height prediction is useful for a number of ocean engineering systems, including facilities safety in situations of severe weather [6,7]. Regarding the sea wave energy flux P, it is more related to marine energy [8,9], which is currently one of the most promising sources of renewable energy, still minor at a global level, but playing a major role in several offshore islands [10,11]. In this case, the accurate estimation of the wave energy flux P is relevant to characterize the wave energy production from Wave Energy Converters (WECs) facilities [12].

The research work on wave parameters' prediction systems has been intense in the last years, with special incidence in machine learning or soft-computing approaches. One of the first works on this topic was the direct prediction of H_{m_0} using artificial neural networks in [13]. Improvements on this prediction system were further presented in [14]. Neural networks have also been applied to other problems of H_{m_0} and P prediction, such as [15], where H_{m_0} and P are inferred from observed wave records using time series neural networks. In [16] a neural network is applied to estimate the wave energy resource in the northern coast of Spain. In [18] and [19] different hybrid algorithms mixed with an Extreme Learning Machine neural network were proposed for the estimation of H_{m_0} and P, in the context of marine energy applications. Alternative methods based on different computational approaches have been recently proposed. For example, in [20] different soft-computing techniques are tested for H_{m_0} prediction. Support Vector regression (SVR) has also been applied to marine energy related problems such as in [21]. Similarly, [22] and [23] proposed to feed SVR approaches with information from radar sources in order to obtain an accurate prediction of H_{m_0} and P parameters. Classification approaches have been applied in [24] to analyze and predict H_{m_0} and P ranges in buoys for marine energy applications. In [25], use of genetic programming for H_{m_0} reconstruction problems was proposed.

* Corresponding author: Sancho Salcedo-Sanz. Department of Signal Processing and Communications, Universidad de Alcalá. 28871 Alcalá de Henares. Madrid. Spain. Ph: +34 91 885 6731 Fax: +34 91 885 6699

Email address: sancho.salcedo@uah.es (S. Salcedo-Sanz*).

In spite of this important research work on the application of different Soft-Computing approaches in wave parameters prediction, the use of fuzzy-based systems in this field is still under exploration. One of the first works on the application of fuzzy systems to wave parameters estimation is [34], where the performance of an Adaptive-Network-Based Fuzzy Inference System (ANFIS) in a problem of wave parameters estimation is investigated on real data from lake Ontario. In [31] a Takagi-Sugeno-rule-based Fuzzy Inference System (FIS) was developed for forecasting sea wave parameters. Input data such as wind speed and direction, and lagged-wave characteristics were used in this case. The model was successfully applied to data from an oceanographic buoy deployed in the Aegean Sea (Greece). In [32] a forecasting system based on the combination of wavelet and fuzzy logic approaches was proposed, for wave parameters prediction up to 48 hours. The idea of the system is to use the wavelet technique to separate time series into its spectral bands, and then these spectral bands are estimated individually by the fuzzy logic approach. Results in different deep sea and coastal buoys of the Pacific coast of the USA are shown. In [17] a hybrid genetic algorithm-adaptive network-based fuzzy inference system model was developed to forecast H_{m_0} and the peak spectral period at lake Michigan. In [33] a neuro-fuzzy approach is used to define spatial variability of the significant wave height assumed as a regionalized variable, in such a way that it is possible to estimate the significant wave height values of a selected station from neighboring stations which exhibit similar features. Experiments with buoys in the Gulf of Mexico show the performance of this approach. In [27] Fuzzy Inference Systems (FIS) are combined with Adaptive Network-based Fuzzy Inference Systems (ANFIS) for the modelling of non-stationary time series for an improved prediction of wind and wave parameters. This system was tested with data from weather models in the North Atlantic.

In this paper we propose the application of FRULER [39] (a Genetic Fuzzy System (GFS) for regression [40,41]), in a problem of sea wave parameters estimation from neighbor buoys. FRULER is based on Takagi-Sugeno-Kant (TSK) rules, and it is composed of an instance selection method for regression, a multi-granularity fuzzy discretization of the input variables, and an evolutionary algorithm that uses a fast and scalable method with Elastic Net regularization to generate accurate and simple TSK-1 fuzzy rules. We show the performance of FRULER in a problem of H_{m_0} and P estimation using three buoys in the West Coast of the USA (California). Results are discussed in terms of the good interpretability of the obtained rules, which allow an intuitive explanation of the wave parameters reconstruction using a reduced number of rules and predictive variables.

The rest of the paper is organized as follows: next section details the calculation of the parameters of interest in ocean wave characterization, H_{m_0} and P in this case. Section 3 presents the TSK rules, which are learned through

FRULER. Section 4 briefly describes the main components of FRULER [39]: the instance selection method, the multi-granularity fuzzy discretization, and the evolutionary algorithm. Section 5 presents the experimental part of the paper, where the performance of FRULER is tested for H_{m_0} and P predictions in the Western coast of the USA. Finally, Section 6 closes the paper with some final conclusions and remarks on this research.

2 Wave parameters of interest: calculation of H_{m_0} and P

The accurate characterization of wave parameters is essential in the evaluation of marine systems. Specifically, in the case of wave energy plants or wave energy converters, it is necessary to characterize the amount of wave energy available at a particular location, which is given by parameters such as H_{m_0} and P. In order to obtain these parameters, it is necessary to focus on the water surface, and take into account the framework of the linear wave theory, in which the vertical wave elevation ($\eta(\mathbf{r},t)$) at a point $\mathbf{r} = (x,y)$ on the sea surface at time t can be assumed as a superposition of different monochromatic wave components [28,29]. This model is appropriate when the free wave components do not vary appreciably in space and time (i.e., statistical temporal stationarity and spatial homogeneity can be assumed [29]).

In this model, the concept of "sea state" refers to the sea area and the time interval in which the statistical and spectral characteristics of the wave do not change considerably. The parameters of a given sea state are then the combined contribution of all parameters from different sources. For example, the "wind sea" occurs when the waves are caused by the energy transferred between the local wind and the free surface of the sea. The "swell" is the situation in which the waves have been generated by winds blowing on another far area (for instance, by storms), and propagate towards the region of observation. Usually, sea states are the composition of these two pure states, forming multi-modal or mixed seas. In a given sea state, the wave elevation $\eta(\mathbf{r},t)$ with respect to the mean ocean level can be assumed as a zero-mean Gaussian *stochastic process*, with statistical symmetry between wave maxima and minima. A buoy deployed at point $\mathbf{r}_B$ takes samples from this process, $\eta(\mathbf{r}_B,t_j)$ $j = 1,2,\cdots,t_{\mathrm{MAX}}$, generating thus a time series of empirical vertical wave elevations. The Discrete Fourier Transform (DFT) of this sequence, using the Fast Fourier Transform (FFT) algorithm, allows estimating the *energy spectral density* $S(f)$. Its spectral moments of order n can be computed as follows:

$$m_n = \int_0^\infty f^n S(f) df. \tag{1}$$

The Significant Wave Height (SWH) is then defined as the average (in meters) of the highest one-third of all the wave heights during a 20-minute sampling period, and it has been widely researched. It can be calculated from moment of order 0 in Equation (1), as follows:

$$H_{m_0} = 4 \cdot (m_0)^{1/2} \tag{2}$$

On the other hand, the wave energy flux is a first indicator of the amount of wave energy available in a given area of the ocean. Wave energy flux P, or power density per meter of wave crest [30], can be computed as:

$$P = \frac{\rho g^2}{4\pi} \int_0^\infty \frac{S(f)}{f} df = \frac{\rho g^2}{4\pi} m_{-1} = \frac{\rho g^2}{64\pi} H_{m_0}^2 \cdot T_e, \tag{3}$$

where ρ is the sea water density (1025 kg/m^3), g the acceleration due to gravity, $H_{m_0} = 4\sqrt{m_0}$ is the spectral estimation of the significant wave height, and $T_e \equiv T_{-1,0} = m_{-1}/m_0$ is an estimation of the mean wave period, normally known as the period of energy, which is used in the design of turbines for wave energy conversion. Expression (3) (with H_{m_0} in meters and T_e in seconds) leads to:

$$P = 0.49 \cdot H_{m_0}^2 \cdot T_e, \tag{4}$$

measured in kW/m, which helps engineers estimate the amount of wave energy available when planning the deployment of WECs at a given location.

3 TSK Fuzzy Rules

The TSK fuzzy rule model was developed by Takagi, Sugeno and Kang [36,37] as a mathematical tool to fuzzily describe systems using rules, where the antecedent is defined with linguistic variables and the consequent is represented as a polynomial function of the input variables. These rules are usually denoted as TSK-0 for polynomials of order 0, TSK-1 for polynomials of order 1, and so on. The most common function for the consequent of a TSK rule is a linear combination of the $\{X_1, \ldots, X_p\}$ input variables (TSK-1):

$$\begin{aligned} &\text{If } X_1 \text{ is } A_1 \text{ and } X_2 \text{ is } A_2 \text{ and } \ldots \text{ and } X_p \text{ is } A_p \text{ then} \\ &Y = \beta_0 + X_1 \cdot \beta_1 + X_2 \cdot \beta_2 + \cdots + X_p \cdot \beta_p \end{aligned} \tag{5}$$

where X_j represents the j-th input variable; A_j is the linguistic fuzzy term for X_j; β_j is the coefficient associated with X_j in the consequent part of the rule,

representing the weight of the importance of each variable in the consequent part; and Y is the output variable.

The final output of a TSK fuzzy rule base system with m TSK fuzzy rules can be computed as the weighted average of each outputs Y_k produced by each rule, where the weight corresponds with the matching degree:

$$\hat{y} = \frac{\sum_{k=1}^{m} h_k \cdot Y_k}{\sum_{k=1}^{m} h_k} \tag{6}$$

In Equation (6), the h_k represents the weight applied to each output, and it is computed as the matching degree between the antecedent of each rule r_k and the inputs $(x_1, x_2, \ldots, x_p)$:

$$h_k = T(A_1^k(x_1), A_2^k(x_2), \ldots, A_p^k(x_p)) \tag{7}$$

where A_j^k is the linguistic fuzzy term for the j-th input variable in the k-th rule and T is the t-norm conjunctive operator, usually the minimum function.

One of the main advantages of the TSK fuzzy rule systems is that they represent a good trade-off between accuracy and interpretability due to:

- Full description of the input space using linguistic terms in the antecedent of the rules.
- Ability to learn accurate solutions using different well-studied statistical methods to optimize the coefficients of the polynomials.
- Easy understanding of the relationship between the inputs and outputs using a linear combination of the input variables in the consequent of the rules.

Although the interpretability of TSK fuzzy rules is not as easy a task as in the case of the Mamdani approach [38], the use of TSK rules can still bring useful information to many domains.

4 FRULER: Fuzzy RUle Learning through Evolution for Regression

FRULER (Fuzzy RUle Learning through Evolution for Regression) [39] is a Genetic Fuzzy System that automatically generates accurate TSK-1 fuzzy rules for regression problems[1]. The goal of FRULER is to generate fuzzy knowledge bases with a high accuracy while keeping a low complexity in the

[1] For a detailed description of FRULER, please refer to [39].

learned rules. There is also a scalable distributed version of FRULER, which is able to solve regression problems for Big Data [42]. The main three components of FRULER consist of an instance selection method for regression, a multi-granularity fuzzy discretization of the input variables, and an evolutionary algorithm that uses Elastic Net regularization to generate simple and accurate TSK-1 fuzzy rule bases. The first two components of FRULER are part of a two-stage preprocessing step that was included in the algorithm in order to improve the accuracy and simplicity of the fuzzy rules obtained by the evolutionary algorithm.

4.1 Instance selection

The aim of the instance selection method is to reduce the variance of the models by generating rules using the most representative examples. The technique is an improvement of the CCISR (Class Conditional Instance Selection for Regression) algorithm [43], which is an adaptation for regression of the instance selection method for classification CCIS (Class Conditional Instance Selection) [44].

The selection process starts calculating a score for each of the instances based on the distances to similar and different examples. Then, an initial core of instances is selected, sorted by the score. After this, the instance selection method iteratively selects instances and adds them to the set until the error increases. Finally, in order to further improve the number of selected instances, the method applies a post-processing step to select examples close to the decision boundary.

4.2 Multi-granularity fuzzy discretization.

In this step, the fuzzy linguistic labels are automatically generated from data through fuzzy discretization (Figure 1), using a multi-granularity approach where each variable var is divided into i fuzzy labels, i.e., $g^i_{var} = \{A^{i,1}_{var}, \ldots, A^{i,i}_{var}\}$.

The generation of the fuzzy linguistic labels consists of a two-step process. First, the variable is discretized to obtain a set of split points C^g for each granularity g. These split points are generated in an interative way, adding a new split point at each step to generate two new intervals, which preserves the interpretability between contiguous granularities. Then, the fuzzy labels are defined for each granularity using the selected split points. In FRULER, the discretization process searches for the split point that minimizes the mean squared error when a linear model is applied to each of the resulting intervals.

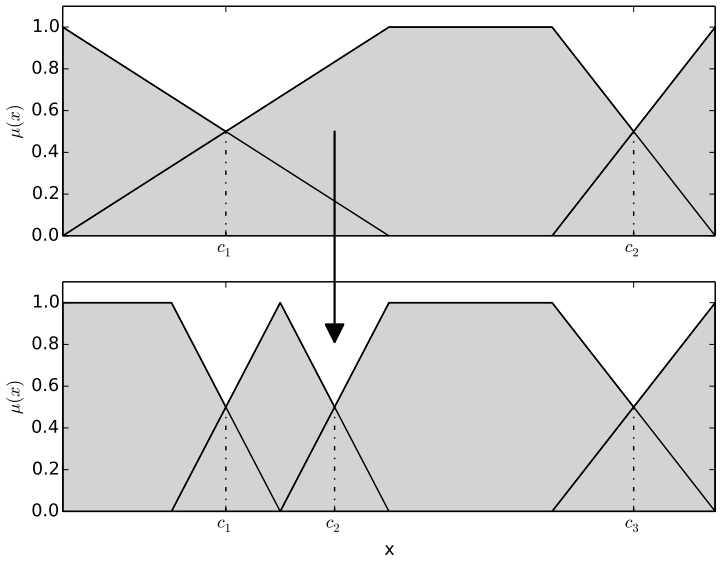

Fig. 1. Example of multi-granularity discretization.

4.3 Evolutionary algorithm

Finally, the objective of the evolutionary algorithm is to learn accurate and simple TSK-1 rule bases for a regression problem. The integration of the evolutionary algorithm with the preprocessing stage is as follows:

(1) The instance selection process is executed over the training examples E_{tra} in order to obtain a subset of representative examples E_S.
(2) The multi-granularity fuzzy discretization algorithm obtains the fuzzy partitions for each input variable.
(3) The evolutionary algorithm searches for the best data base configuration using the obtained fuzzy partitions, generates the entire linguistic TSK rule base using E_S and evaluates the different rule bases using E_{tra}.

The chromosome in the evolutionary algorithm is codified with a double coding scheme ($C = C_1 + C_2$). In this encoding C_1 represents the granularity of each input variable. In turn, C_2 represents the lateral displacements of the split points of the input variables fuzzy partitions.

FRULER uses the Wang & Mendel algorithm to create the antecedent part of the rule base for each individual. The coefficients of the consequent part of the rules, i.e., the importance weight of each variable in the consequent, is automatically learned using the Elastic Net method [45]. Elastic Net linearly combines the ℓ_1 (Lasso regularization) and ℓ_2 (Ridge regularization) penalties

of the Lasso and Ridge methods, minimizing the following equation:

$$\hat{\beta} = \arg\min_{\beta} \|Y - X \cdot \beta\|_2^2 + \lambda \cdot \alpha \cdot \|\beta\|_2^2 + \lambda \cdot (1 - \alpha) \cdot \|\beta\|_1 \tag{8}$$

where β is the coefficients vector, Y is the outputs vector, X is the inputs matrix, λ is the regularization parameter and α represents the trade-off between ℓ_1 and ℓ_2 penalization. In order to solve the minimization problem of Elastic Net (Equation (8)), we used Stochastic Gradient Descent (SGD).

The rule base is generated using only those examples in E_s. This way, those examples that are not representative are not taken into account and, therefore, the method avoids the generation of too specific rules and reduces the time needed to create the rule base.

The fitness function is:

$$\mathit{fitness} = MSE(E_{tra}) = \frac{1}{2 \cdot |E|} \sum_{i=1}^{|E|} (F(x^i) - y^i)^2, \tag{9}$$

where E_{tra} is the full training dataset and $F(x^i)$ is the output obtained by the knowledge base for input x^i. Using all the examples for evaluation can be seen, in some way, as a validation process, as the rule base was constructed with a subset of them (E_S).

5 Experiments and results

To evaluate the performance of the proposed approach, we present some experiments in which we tackle the estimation of the significant wave height and energy flux with FRULER. We focus on two aspects of the system: its interpretability and accuracy. Furthermore, in order to validate that FRULER is able to learn interpretable knowledge bases without loss of accuracy, we also compare the results against two other accurate but non-interpretable state-of-the-art predictive models.

5.1 Problem description

Figure 2 and Table 1 show the three buoys considered in this study at the Western coast of the USA, and their main characteristics [46]. In this case we consider the reconstruction of buoy 46069 from a number of predictive variables from the other two buoys. For this, 10 predictive variables measured at each neighbor buoy are considered (a total of 20 predictive variables to carry out the reconstruction). Table 2 shows details on the predictive variables for

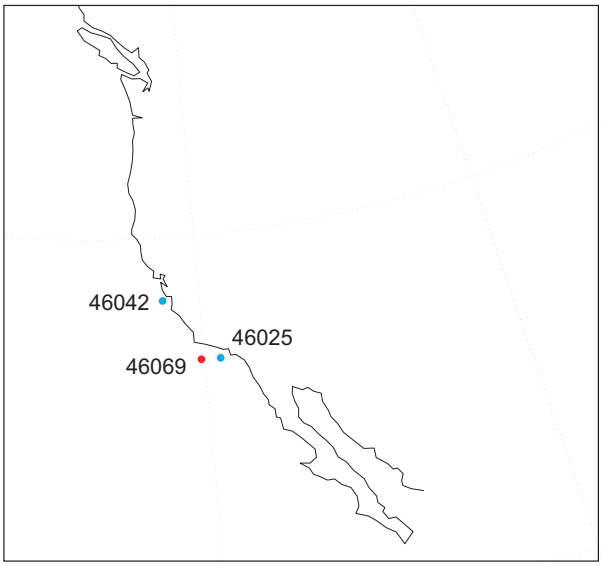

Fig. 2. Western USA buoys considered in this study. In red the buoy where the H_{m_0} and P predictions are carried out from data of the blue buoys.

Table 1
Geographic coordinates and buoy's description.

Characteristics	Station 46025	Station 46042	Station 46069
	33°44'58"N 119°3'10"W	36°47'29"N 122°27'6"W	33°40'28"N 120°12'42"W
Site elevation	sea level	sea level	sea level
Air temp. height	4 m above site elevation	4 m above site elevation	4 m above site elevation
Anemometer height	5 m above site elevation	5 m above site elevation	5 m above site elevation
Barometer elevation	sea level	sea level	sea level
Sea temp. depth	0.6 m below water line	0.6 m below water line	0.6 m below water line
Water depth	905.3 m	2098 m	1020.2 m
Watch circle radius	1327 yards	2108 yards	1799 yards

this problem. Data for two complete years (1st January 2009 to 31st December 2010) with a hourly time resolution were used. Note that complete data (without missing values in both predictive and objective variables) are available for the considered period in the three buoys.

5.2 Methodology

Data were divided into training and test sets in such a way that data from the complete year 2009 were used for the training, and data from the complete year 2010 were used for the test, to estimate how well the proposed approach

Table 2
Predictive variables used in the experiments.

Acronym	Predictive variable	units
WDIR	Wind direction	[degrees]
WSPD	Wind speed	[m/s]
GST	Gust speed	[m/s]
WVHT	Significant wave height	[m]
DPD	Dominant wave period	[sec]
APD	Average period	[sec]
MWD	Direction DPD	[degrees]
PRES	Atmospheric pressure	[hPa]
ATMP	Air temperature	[Celsius]
WTMP	Water temperature	[Celsius]

performs on unseen data. We trained each model using the training data set and we measured the performance of the models on the test set in terms of the Root Mean Square Error (RMSE) and the Nash-Sutcliffe Efficiency Coefficient (CE). The coefficient of efficiency CE proposed by Nash and Sutcliffe is a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance [47]. It is defined as:

$$CE = 1 - \frac{\sum_{i=1}^{n}(y^i - F(x^i))^2}{\sum_{i=1}^{n}(y^i - \bar{y})^2} \tag{10}$$

where y^i are the true observations and $F(x^i)$ are the predicted values. The range of CE lies between $(-\inf, 1]$, where a $CE = 1$ indicates a perfect fit, and negative values of the coefficient indicates that the model performs arbitrarily worse.

Finally, to validate the interpretability of the rules obtained with FRULER, we analyzed the resulting knowledge bases, and we provide an explanation of the different rules that are consistent with the underlying physical processes of the observed phenomena.

5.3 Results

FRULER has been applied to the prediction of H_{m_0} and P from neighbor buoys (Figure 2), obtaining accurate estimations but, also, highly interpretable

Table 3
Comparison of the results obtained by the proposed FRULER with state of the art algorithms for this problem.

Algorithm	RMSE (H_{m_0}) (m)	CE (H_{m_0})	RMSE (P) (kW/m)	CE (P)
FRULER	0.408	0.710	2.734	0.729
GGA-ELM	0.435	0.563	2.874	0.760
GGA-SVR	0.392	0.740	2.663	0.756

results supported by the underlying physical processes. First, we compare the Root Mean Square (RMSE) error and the CE coefficient of FRULER with the results of the state of the art hybrid algorithms in the same problem. Specifically, we compare the performance of the FRULER approach with a hybrid Grouping Genetic Algorithm – Extreme Learning Machine (GGA-ELM) and a Support Vector Regression algorithm (GGA-SVR), reported in [19]. Table 3 shows the obtained results. As can be seen, FRULER performance is competitive in terms of the quality of results obtained, with slightly better performance than GGA-ELM and slightly worse results than GGA-SVR, both for H_{m_0} and P predictions. Although the performance of all the algorithms is similar, GGA-ELM and GGA-SVR are black-boxes approaches, whereas FRULER provides fully interpretable results, accordingly to the physical characteristics of the prediction problem. This is the main advantage of the prediction system proposed in this paper. Moreover, the scatter plots of the predictions are shown in Figures 3 and 4, where the reader can appreciate the similarity in the results among the three methods compared.

Regarding the interpretation of the specific solutions found, first we analyze the results for the H_{m_0} prediction case. Figures 5, 6 and 7 show the three rules obtained by FRULER for the prediction of H_{m_0}, and the most important variables involved in the final prediction. The variables on the left side of the figures represent the multi-granularity dicretization of the variables in the antecedent part of the rule, as automatically learned by FRULER. In Figure 5, variables MWD_46042 and GST_46025 were selected for the antecedent part, and were discretized into two labels. We can interpret these labels as "Low" and "High". In the case of MWD_46042, the definition of "Low" goes from 0 to 342, whereas "High" is interpreted as values above 221. Note that values between 221.6 and 342 have different degrees of membership for "Low" and "High" labels. As for the case of the GST_46025, "Low" means those values between 0 and 21.2, and "High" above 5.7.

Panel on the right shows the weight of each variable in the consequent part, i.e., the coefficient β_j of each variable X_j. These are the values selected by FRULER that minimize the expression represented in Equation (8), during the learning process.

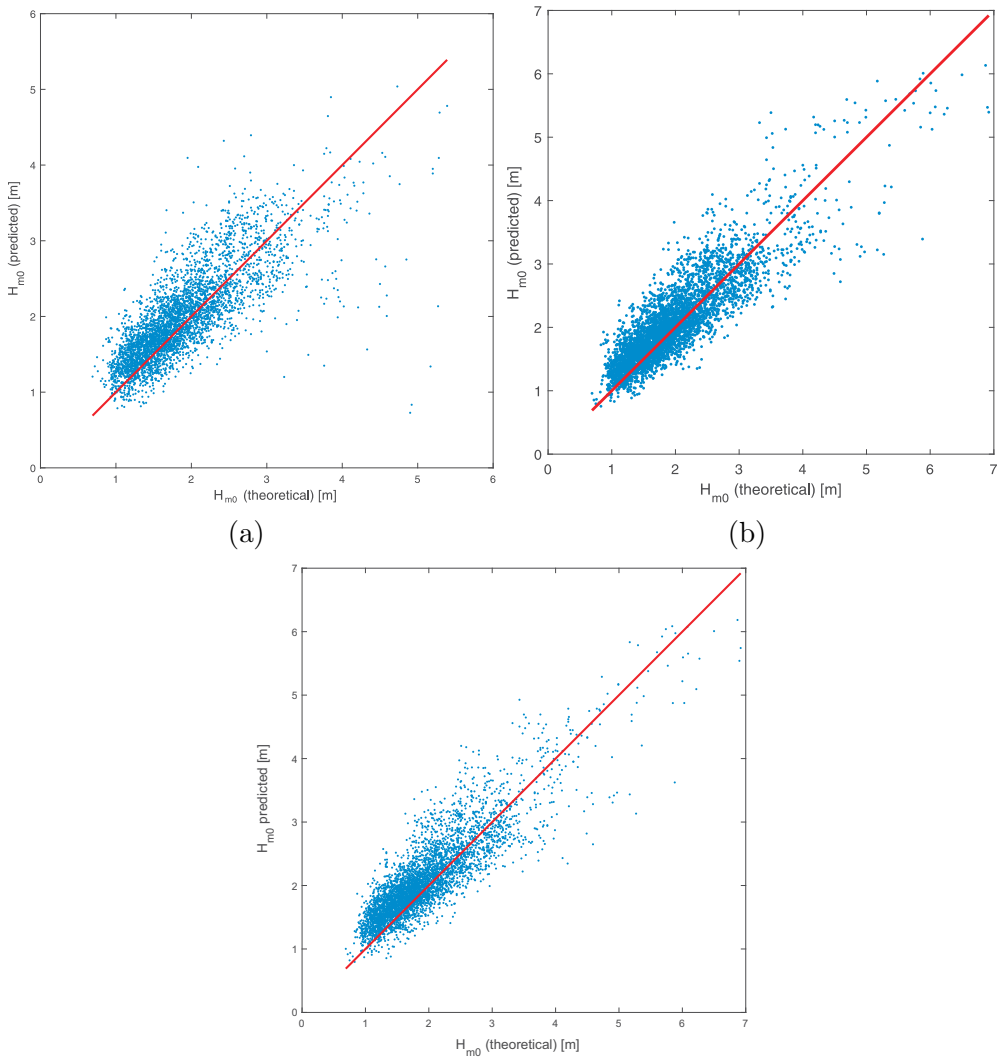

Fig. 3. Scatter plots in the problem of H_s prediction by the ELM, SVR and FRULER approaches after the feature selection process with the GGA-ELM approach; (a) ELM; (b) SVR; (c) FRULER

Following the definition of a TSK rule, as described in Equation (5), the TSK rule represented in Figure 5 can be defined as:

$$\begin{aligned}&\text{IF } MWD\_46042 \text{ is “High” and } GST\_46025 \text{ is “Low” then}\\ &Y = 0.07 \cdot WSPD\_46042 + 0.08 \cdot GST\_46042 + \cdots + 0.12 \cdot WTMP\_46025 \quad (11)\end{aligned}$$

As can be seen, all the rules are triggered based only on two antecedent variables (MWD_46042 and GST_46025, i.e mean ocean waves direction and wind gust), and they mainly correspond to two different physical situations of *swell* (Rule 1 and Rule 3) and wind-sea (Rule 2). Recall that swell is defined as

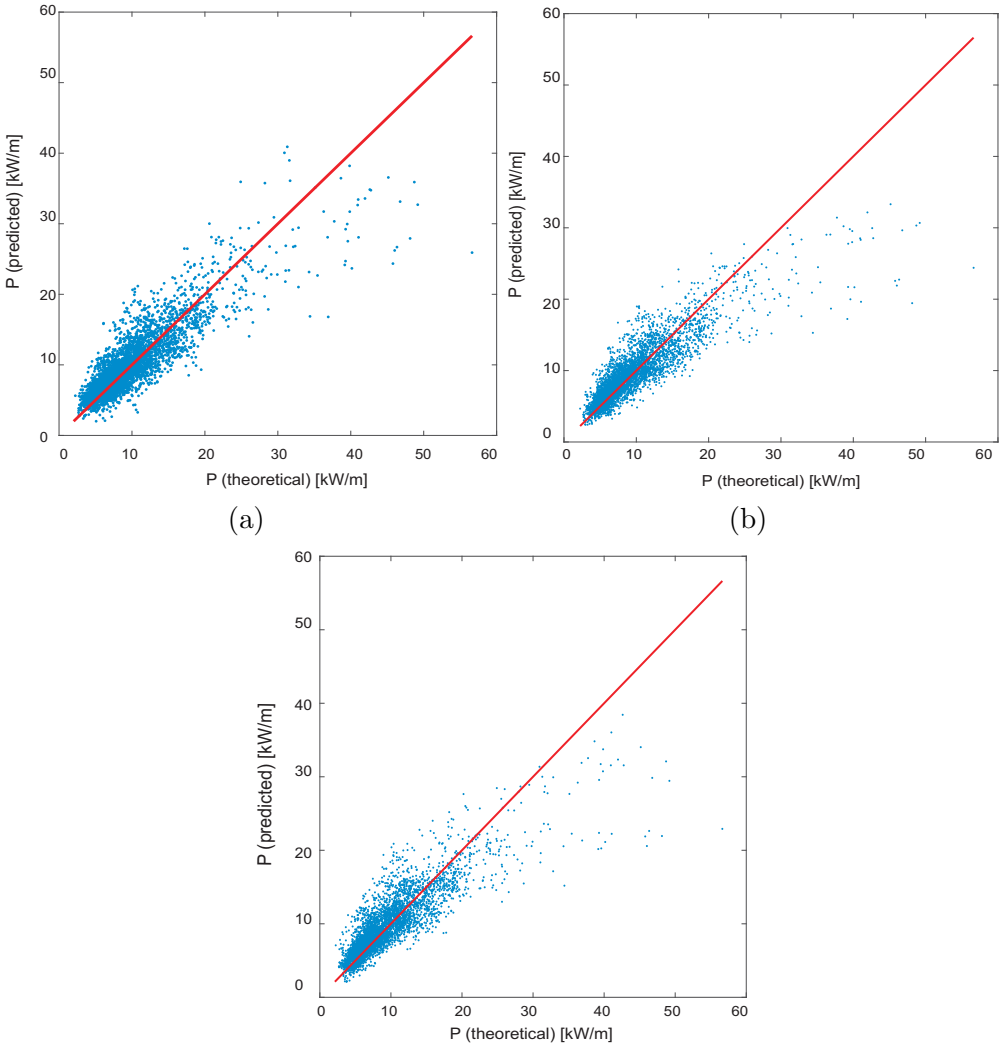

Fig. 4. Scatter plots in the problem of P prediction by the ELM, SVR and FRULER approaches after the feature selection process with the GGA-ELM approach; (a) ELM; (b) SVR; (c) FRULER

those ocean surface waves which are not generated by the immediate local wind, but instead by distant weather systems, such as storms etc. Note that swell is opposed to a locally generated wind wave (wind-sea), which is mainly produced by local mechanisms such as wind blowing. In particular, for Rules 1 and 3, we are in a situation of swell in both cases, but the difference between them due to variable MWD (mean wave direction). Note that in Rule 1, the antecedent MWD_46042 goes from values of 221.6° to 360°, whereas in Rule 3, MWD_46042 covers a different range of directions from 0° to 342° approximately.

This affects the influence that the consequent variables have in the prediction of H_{m_0}. For Rule 1, we can appreciate the great impact that the following

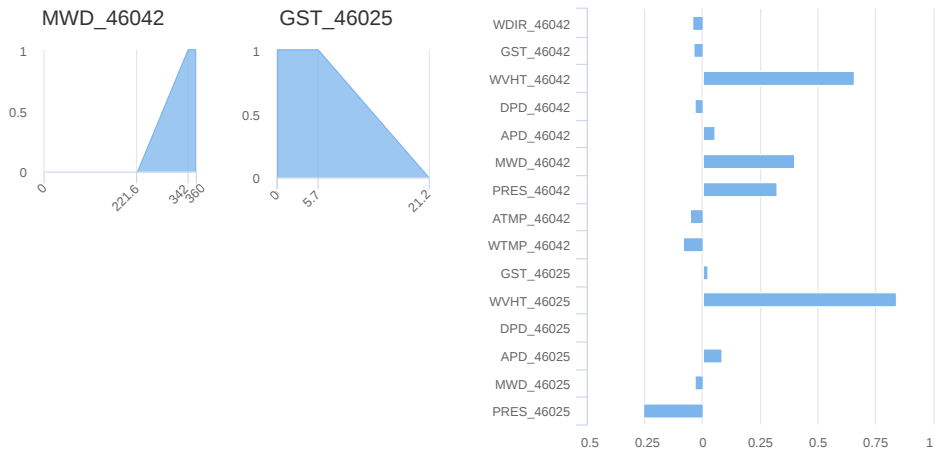

Fig. 5. Rule 1 for the Significant Wave Height.

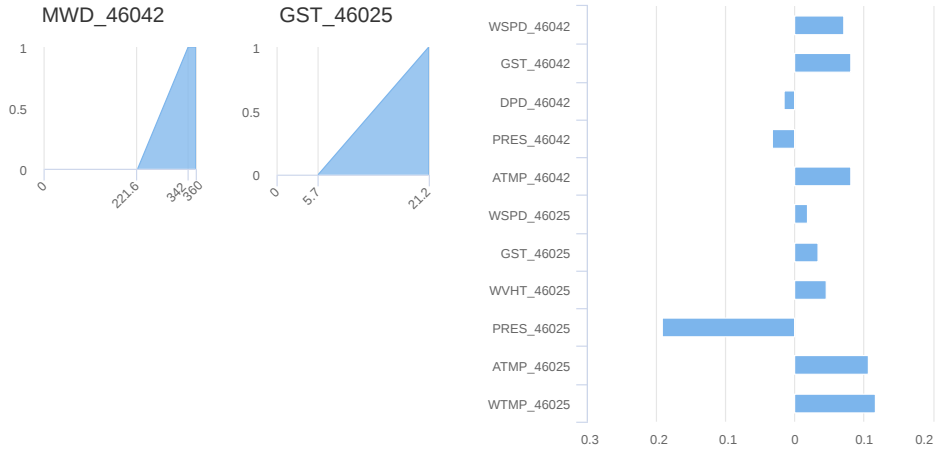

Fig. 6. Rule 2 for the Significant Wave Height.

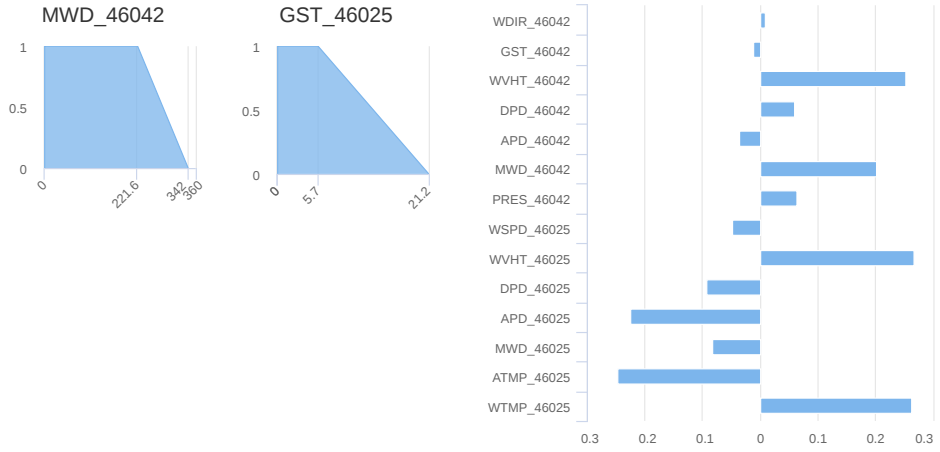

Fig. 7. Rule 3 for the Significant Wave Height.

consequent variables have in the prediction: $WVHT\_46042$, MWD_46042,

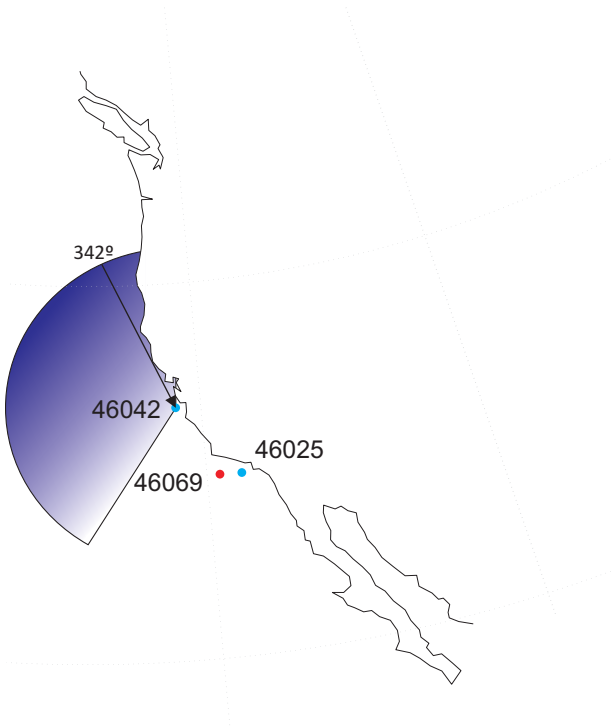

Fig. 8. Predominant wave direction in Rule 1 for the Significant Wave Height.

$PRES\_46042$, $WVHT\_46025$, and $PRES\_46025$. This is corroborated in Figure 5. On the other hand, in Rule 3, the consequent variables that affect the most in the prediction of H_{m_0} are: $WVHT\_46042$, MWD_46042, $WVHT\_46025$, DPD_46025, APD_46025, MWD_46025, $ATMP\_46025$, and $WTMP\_46025$. Note how the value of H_{m_0} in both neighbor buoys is included in the prediction, as expected, plus a number of alternative variables to make the prediction more accurate. Figure 7 shows that, in Rule 3, the overall influence of the consequent variables belonging to buoy 46025 is much higher than in Rule 1 — all the variables from buoy 46025 but $WSPD\_46025$ are important for the prediction. This seems reasonable, since in this case the swell mainly comes from the South, where buoy 46025 is located. Figures 8 and 9 show these two cases of swell in relation with the MWD in buoy 46042. Note the clear difference in the swell situation captured by each rule (swell from the North-west and swell from the South).

Rule 2 shows a completely different scenario. This rule takes into account wind-sea situations, which are characterized by an irregular sea, where local meteorological values have more relevance than in the previous case of swell. In this case, we must highlight the antecedent GST_46025 because it is decisive to determine if we are in a swell situation or wind sea condition. In Rule 2, we can see that the values of GST_46025 go from 5.7 m/s to 21.2 m/s, which are larger values than those in Rules 1 and 3 — label GST_46025 has the maximum membership values in the range 0-5.7 m/s. This fact causes

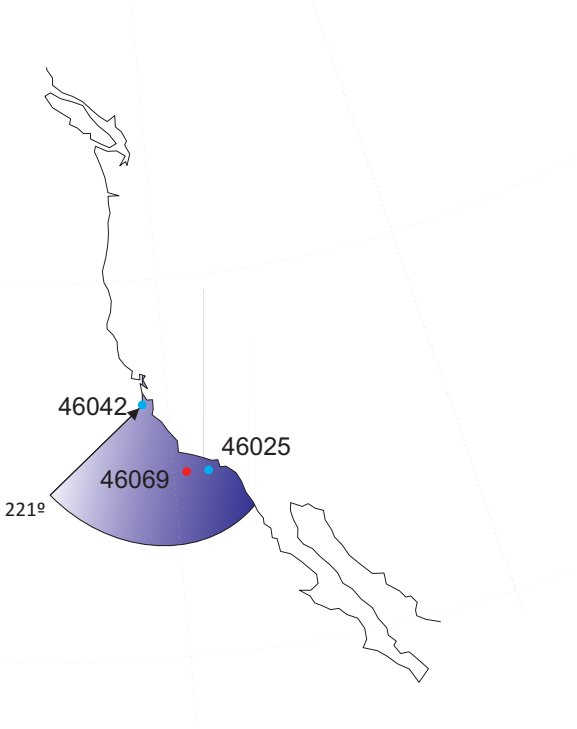

Fig. 9. Predominant wave direction in Rule 3 for the Significant Wave Height.

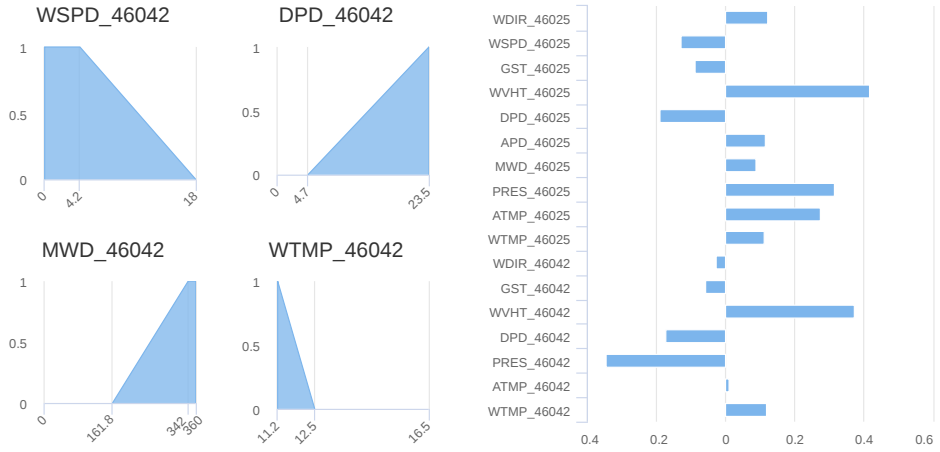

Fig. 10. Rule 1 for the Energy Flux.

higher wind gusts, that produce the physical situation of wind-sea, captured in Rule 2. The variables with a higher influence in the prediction of Rule 2 are: $WSPD\_46042$, GST_46042, $ATMP\_46042$, $PRES\_46025$, $ATMP\_46025$, and $WTMP\_46025$. It is very interesting to check out that the local meteorological variables have more importance in this case, as shown in Figure 6.

Regarding the prediction of the wave energy flux P, FRULER provides 8 rules (Figures 10 to 17), triggered based on 4 antecedent variables ($WSPD\_46042$, DPD_46042, MWD_46042 and $WTMP\_46042$). Note that in this case, the

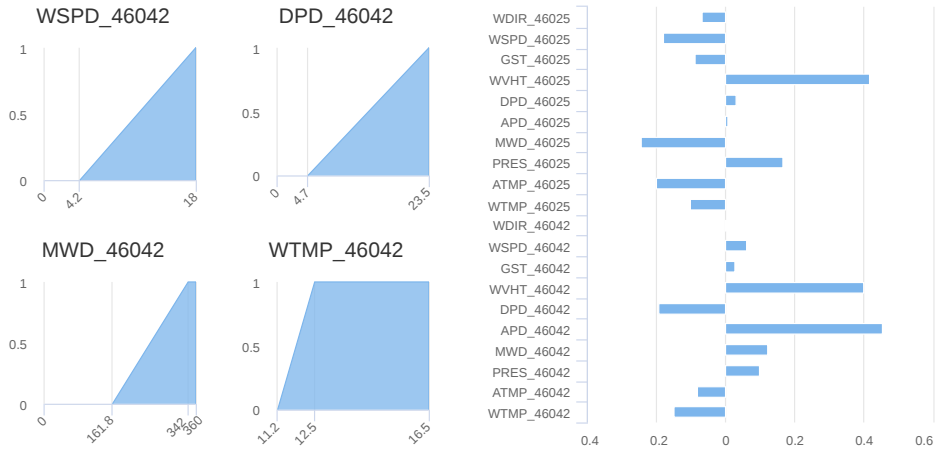

Fig. 11. Rule 2 for the Energy Flux.

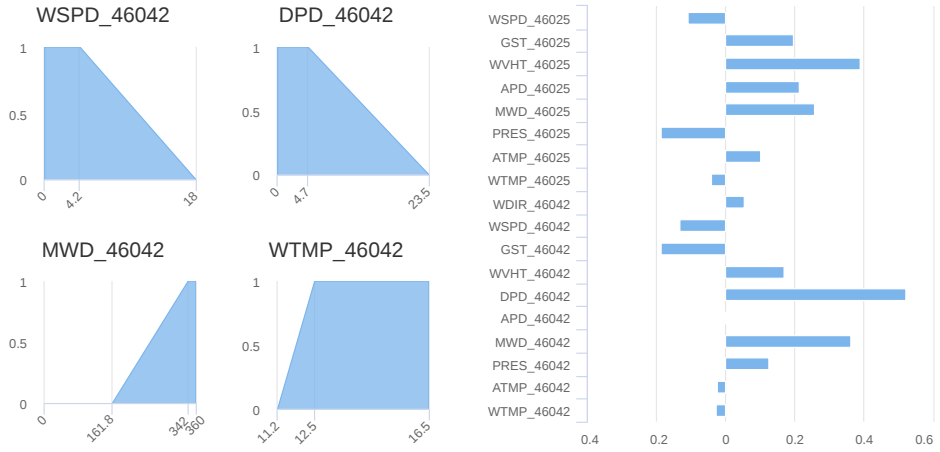

Fig. 12. Rule 3 for the Energy Flux.

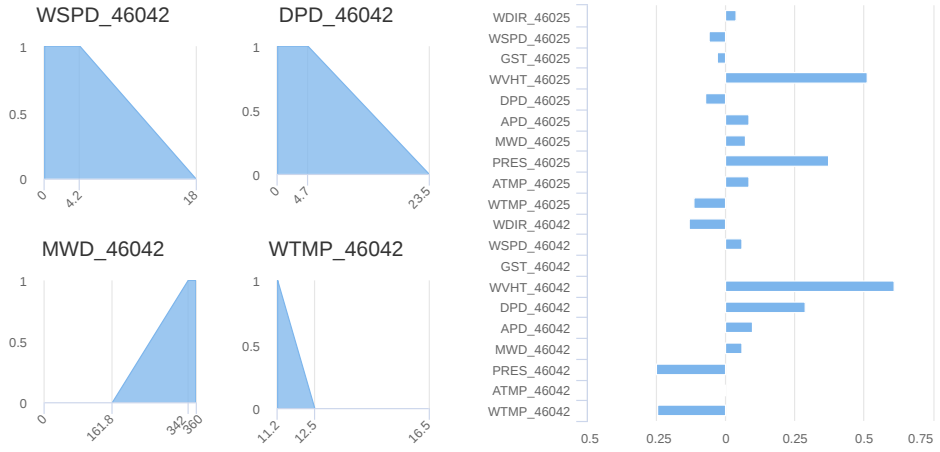

Fig. 13. Rule 4 for the Energy Flux.

antecedent variables only refer to buoy 46042. A more in depth analysis of

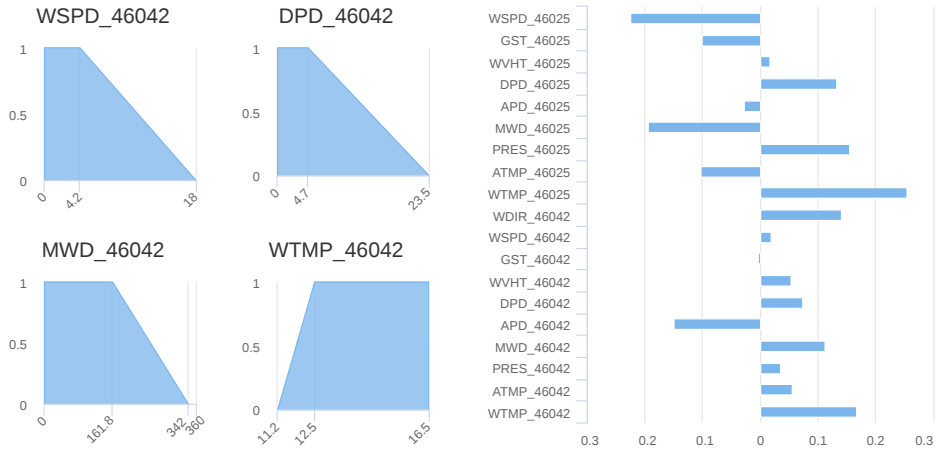

Fig. 14. Rule 5 for the Energy Flux.

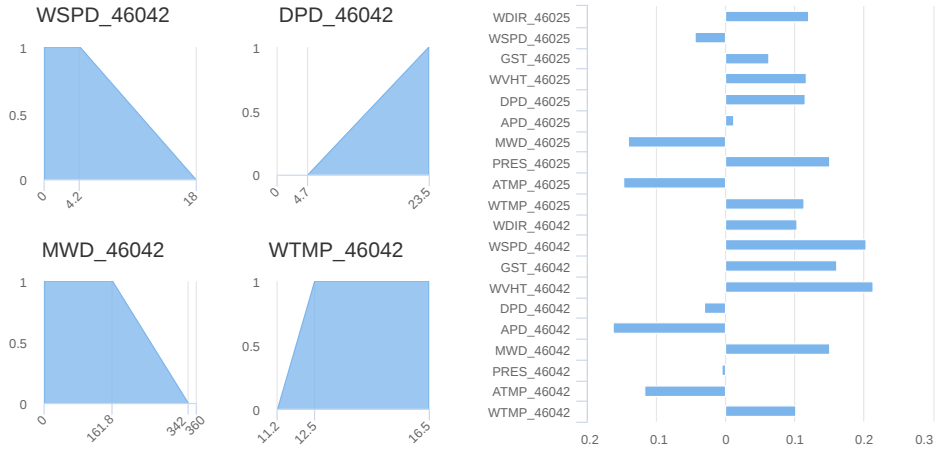

Fig. 15. Rule 6 for the Energy Flux.

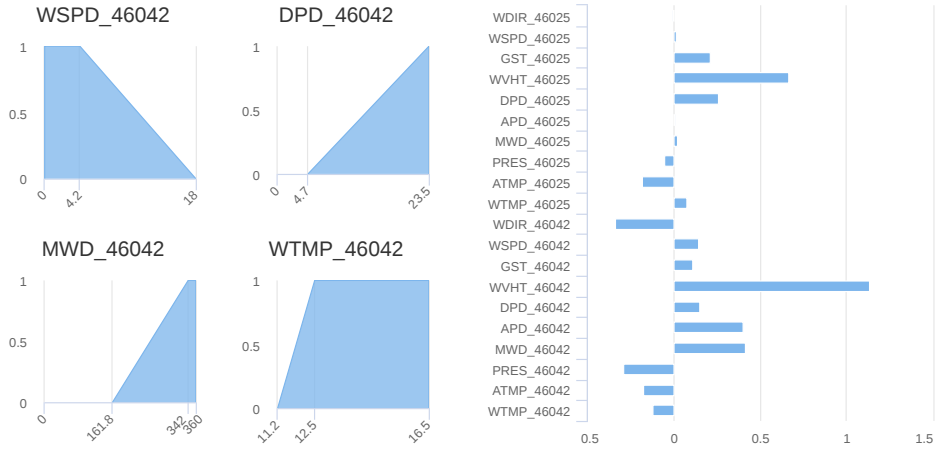

Fig. 16. Rule 7 for the Energy Flux.

these antecedent variables shows that wind speed and the dominant wave

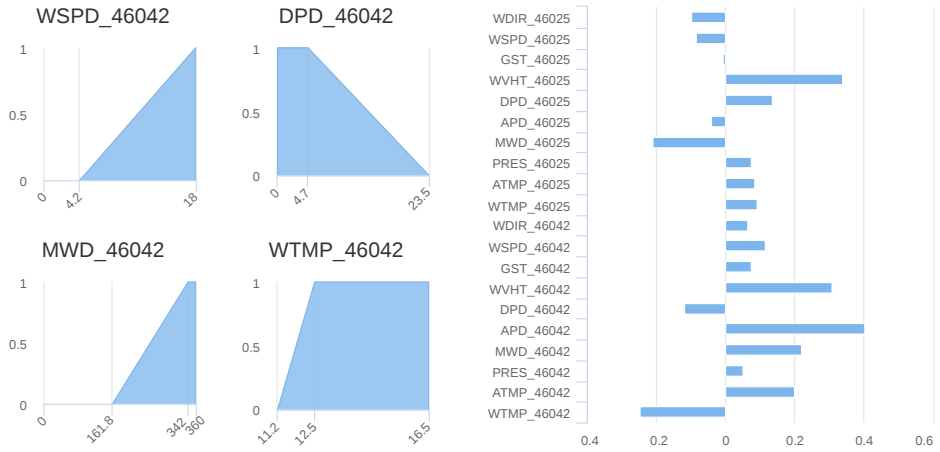

Fig. 17. Rule 8 for the Energy Flux.

period (value and direction) together with the water temperature are the main variables taken into account by the rules learned with FRULER. The inclusion of the dominant wave period and its direction in this case is significant, as it is fully related to the wave energy flux (see Equation (4)). A direct interpretation of the rules in this case is harder, since there are 8 rules. However, it is possible to locate interpretable swell cases in Rules 1, 6 and 7 (i.e. high values of period, low value of local wind). The differences are given by the dominant period direction (north in the case of Rules 1 and 7, and mainly south in the case of Rule 6). Rules 1 and 7 are different due to the values of temperature. In the latter case, the higher the temperature, the more important are the variables of buoy 46042 in the prediction, mainly the significant wave height and the variables related to the wave period, as expected. The rest of the rules provided by FRULER describe situations of wind-sea wave, and other local events related to the prediction of wave energy flux.

In summary, the rules learned by FRULER are able to obtain very accurate and interpretable predictions, both for significant wave height and wave energy flux estimations. The comparison with the state of the art approaches reflects that there are no significant differences in accuracy between the best method (GGA-SVR) and FRULER — +4% in RMSE for H_{m_0} and +3% for P. Moreover, GGA-SVR is a black-box approach, while the linguistic rules generated by FRULER are fully interpretable, and the generated knowledge bases have a low number of rules with few antecedent variables.

6 Conclusions

In this paper we have applied FRULER in two different problems related to ocean wave energy — significant wave height and wave energy flux predictions

— using data from three buoys in the California coast. The FRULER approach is based on an evolutionary algorithm to obtain Takagi-Sugeno-Kant (TSK) fuzzy rules, together with an instance selection method and a multi-granularity fuzzy discretization of the input variables. We have shown how the proposed system is able to obtain robust and very accurate predictions for both objective variables, with a precision very similar to the best state of the art approach. Moreover, the predictions obtained by FRULER are fully interpretable in the case of the significant wave height, and partially interpretable for the harder case of the energy flux problem.

Acknowledgement

This work has been partially supported by projects TIN2014-54583-C2-2-R and TIN2014-56633-C3-1-R — co-funded by FEDER program — of the Spanish Ministerial Commission of Science and Technology (MICYT), by the Comunidad Autónoma de Madrid under project S2013ICE-2933_02, and the Consellería de Cultura, Educación e Ordenación Universitaria — grant GRC2014/030, accreditation 2016-2019, ED431G/08 and the European Regional Development Fund (ERDF). We acknowledge support by DAMA network TIN2015-70308-REDT.

References

[1] F. Comola, T. Lykke Andersen, L. Martinelli, H.F. Burcharth and P. Ruol, "Damage pattern and damage progression on breakwater roundheads under multidirectional waves," Coastal Engineering, vol. 83, pp. 24-35, 2014.

[2] S.W. Kim and K.D. Suh, "Determining the stability of vertical breakwaters against sliding based on individual sliding distances during a storm," Coastal Engineering, vol. 94, pp. 90-101, 2014.

[3] R.A. Arinaga, K.F. Cheung, "Atlas of global wave energy from 10 years of reanalysis and hindcast data," Renewable Energy, vol. 39, pp. 49-64, 2012.

[4] M. Esteban, D. Leary, "Current developments and future prospects of offshore wind and ocean energy," Applied Energy, vol. 90, pp. 128-136, 2012.

[5] I. López, J. Andreu, S. Ceballos, I. Martínez de Alegría and I. Kortabarria, "Review of wave energy technologies and the necessary power-equipment," Renewable and Sustainable Energy Reviews, vol. 27, pp. 413-434, 2013.

[6] S. Rao and S.Mandal,2005. "Hindcasting of storm waves using neural networks," Ocean Engineering, vol. 32, pp.667-684, 2005.

[7] H. K. Chang, J. C. Liou, S. J. Liu and S. R. Liaw, "Simulated wave-driven ANN model for typhoon waves," Advances in Engineering Software, vol. 42, no. 1-2, pp. 25-34, 2011.

[8] A.S. Bahaj, "Generating electricity from the oceans," Renewable and Sustainable Energy Reviews, vol. 15, pp. 3399-3416, 2011.

[9] A.F. Falcão, "Wave energy utilization: A review of the technologies," Renewable and Sustainable Energy Reviews, vol. 14, pp. 899-918, 2010.

[10] M. Fadaeenejad, R. Shamsipour, S.D. Rokni, C. Gomes, "New approaches in harnessing wave energy: With special attention to small islands," Renewable and Sustainable Energy Reviews, vol. 29, pp. 345-354, 2014.

[11] L. Rusu, C. Guedes-Soares, "Wave energy assessments in the Azores islands," Renewable Energy, vol. 45, pp. 183-196, 2012.

[12] L. Cuadra, S. Salcedo-Sanz, J. C. Nieto-Borge, E. Alexandre and G. Rodríguez, "Computational Intelligence in wave energy: comprehensive review and case study," Renewable and Sustainable Energy Reviews, vol. 58, pp. 1223-1246, 2016.

[13] M.C. Deo and C.S. Naidu, "Real time wave prediction using neural networks," Ocean Engineering, vol. 26(3), pp. 191-203, 1998.

[14] J.D. Agrawal and M.C. Deo, "Wave parameter estimation using neural networks," Marine Structures, vol. 17, pp. 536-550, 2004.

[15] C.P. Tsai, C. Lin and J.N. Shen, "Neural network for wave forecasting among multi-stations," Ocean Engineering, vol. 29(13), pp. 1683-1695, 2002.

[16] A. Castro, R. Carballo, G. Iglesias and J.R. Rabuñal, "Performance of artificial neural networks in nearshore wave power prediction," Applied Soft Computing, vol. 23, pp. 194-201, 2014.

[17] M. Zanaganeh, S. Jamshid-Mousavi and A.F. Etemad-Shahidi, "A hybrid genetic algorithm-adaptive network-based fuzzy inference system in prediction of wave parameters," Engineering Applications of Artificial Intelligence, vol. 22(8), pp. 1194-1202, 2009.

[18] E. Alexandre, L. Cuadra, JC. Nieto-Borge, G. Candil-García, M. del Pino and S. Salcedo-Sanz, "A Hybrid Genetic Algorithm – Extreme Learning Machine approach for accurate significant wave height reconstruction," Ocean Modelling, vol. 92, pp. 115-123, 2015.

[19] L. Cornejo-Bueno, J. C. Nieto-Borge, P. García-Díaz, G. Rodríguez and S. Salcedo-Sanz, "Significant Wave Height and Energy Flux Prediction for Marine Energy Applications: A Grouping Genetic Algorithm – Extreme Learning Machine Approach," Renewable Energy, vol. 97, pp. 380-389, 2016.

[20] J.Mahjoobi, A. Etemad-Shahidi and M.H. Kazeminezhad, "Hindcasting of wave parameters using different soft computing methods," Applied Ocean Research, vol. 30(1), pp. 28-36, 2008.

[21] J. Mahjoobi and E.A. Mosabbeb, "Prediction of significant wave height using regressive support vector machines," Ocean Engineering, vol. 36(5), pp. 339-347, 2009.

[22] S. Salcedo-Sanz, J. C. Nieto-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.

[23] L. Cornejo-Bueno, J. C. Nieto-Borge, E. Alexandre, K. Hessner and S. Salcedo-Sanz, "Accurate estimation of significant wave height with support vector regression algorithms and marine radar images," Coastal Engineering, vol. 114, pp. 233-243, 2016.

[24] J.C. Fernández, S. Salcedo-Sanz, P.A. Gutiérrez, E. Alexandre and C. Hervás-Martínez, "Significant wave height and energy flux range forecast with machine learning classifiers," Engineering Applications of Artificial Intelligence, vol. 43, pp. 44-53, 2015.

[25] S.P. Nitsure, S.N. Londhe and K.C. Khare, "Wave forecasts using wind information and genetic programming," Ocean Engineering, vol. 54, pp. 61-69, 2012.

[26] M.Özger, "Prediction of ocean wave energy from meteorological variables by fuzzy logic modeling," Expert Systems with Applications, vol. 38, pp. 6269-6274, 2011.

[27] C. Stefanakos, "Fuzzy time series forecasting of nonstationary wind and wave data," Ocean Engineering, vol. 121, pp. 1-12, 2016.

[28] J.C. Nieto-Borge, K. Reichert, K. Hessner, "Detection of spatio-temporal wave grouping properties by using temporal sequences of X-band radar images of the sea surface," Ocean Modelling, vol. 61, pp. 21-37, 2013.

[29] Y. Goda, Random seas and design of maritime structures, World Scientific, 2010.

[30] B.G. Cahill and T. Lewis, "Wave energy resource characterization of the Atlantic marine energy test site," International Journal of Marine Energy, 2013, vol. 1, pp. 3-15.

[31] G. Sylaios, F. Bouchette, V. A. Tsihrintzis and C. Denamiel, "A fuzzy inference system for wind-wave modeling," Ocean Engineering, vol. 36, pp. 1358-1365, 2009.

[32] M.Özger, "Significant wave height forecasting using wavelet fuzzy logic approach," Ocean Engineering, vol. 37, pp. 1443-1451, 2010.

[33] M.Özger, "Neuro-fuzzy approach for the spatial estimation of ocean wave characteristics," Advances in Engineering Software, vol. 40, no. 9, pp. 759-765, 2009.

[34] M.H. Kazeminezhad, A. Etemad-Shahidi, S.J. Mousavi, "Application of fuzzy inference system in the prediction of wave parameters," Ocean Engineering, vol. 32, pp. 1709-1725, 2005.

[35] R. Hashim, C. Roy, S. Motamedi, S. Shamshirband and D. Petkovic, "Selection of climatic parameters affecting wave height prediction using an enhanced Takagi-Sugeno-based fuzzy methodology," Renewable and Sustainable Energy Reviews, vol. 60, pp. 246-257, 2016.

[36] T. Takagi, M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," IEEE Transactions on Systems, Man and Cybernetics, vol. 1, pp. 116-132, 1985.

[37] M. Sugeno, G. Kang, Structure identification of fuzzy model, Fuzzy sets and systems 28 (1) (1988) 15–33.

[38] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," International Journal of Man-Machine Studies, vol. 7, no. 1, pp. 1-13, 1975.

[39] I. Rodríguez-Fernández, M. Mucientes, A. Bugarín, "FRULER: Fuzzy rule learning through evolution for regression," Information Sciences, vol. 354, pp. 1-18, 2016. 1–18.

[40] O. Cordón, F. Herrera, F. Hoffmann and L. Magdalena, "Genetic fuzzy systems: evolutionary tuning and learning of fuzzy knowledge bases," World Scientific press, 2001.

[41] V. C. Finotto, W.R.L. da Silva, M. Valasek and P. Stemberk, "Hybrid fuzzy-genetic system for optimising cabled-truss structures," Advances in Engineering Software, vol. 62-63, pp. 85-96, 2013.

[42] I. Rodríguez-Fernández, M. Mucientes, A. Bugarín, "S-FRULER: Scalable fuzzy rule learning through evolution for regression, Knowledge-Based Systems, vol. 110, pp.255-266, 2016.

[43] I. Rodríguez-Fernández, M. Mucientes, A. Bugarín, "An instance selection algorithm for regression and its application in variance reduction," in: Proceedings of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1-8, 2013.

[44] E. Marchiori, "Class conditional nearest neighbor for large margin instance selection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 2, pp. 364-370, 2010.

[45] H. Zou, T. Hastie, "Regularization and variable selection via the elastic net," Journal of the Royal Statistical Society, vol. 67, no. 2, pp. 301-320, 2005.

[46] NOAA, National Data Buoy Center: http://www.ndbc.noaa.gov/.

[47] J.E. Nash and J.V. Sutcliffe, "River flow forecasting through conceptual models part I-A discussion of principles," Journal of hydrology, vol. 10, no. 3, pp. 282-290, 1970.