



Adaptive entropy-based learning with dynamic artificial neural network

Pinto, Tiago; Morais, Hugo; Corchado, Juan Manuel

Published in:
Neurocomputing

Link to article, DOI:
[10.1016/j.neucom.2018.09.092](https://doi.org/10.1016/j.neucom.2018.09.092)

Publication date:
2019

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Pinto, T., Morais, H., & Corchado, J. M. (2019). Adaptive entropy-based learning with dynamic artificial neural network. *Neurocomputing*, 338, 432-440. <https://doi.org/10.1016/j.neucom.2018.09.092>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Adaptive Entropy-based Learning with Dynamic Artificial Neural Network

Tiago Pinto^{1,*}, Hugo Morais² and Juan Manuel Corchado¹

¹ BISITE Research Centre, University of Salamanca, Calle Espejo, 12, 37007 Salamanca, Spain; E-Mails: tpinto@usal.es (T.P.); corchado@usal.es (J.M.C.)

² AUTomation and Control Group – Department of Electrical Engineering / Technical University of Denmark (DTU), Elektrovej, Building 326, DK-2800 Kgs. Lyngby, Denmark; E-Mail: morais@elektro.dtu.dk (H.M.)

* Correspondence: tpinto@usal.es; Tel.: +351-913430665

Abstract: Entropy models the added information associated to data uncertainty, proving that stochasticity is not purely random. This paper explores the potential improvement of machine learning methodologies through the incorporation of entropy analysis in the learning process. A multi-layer perceptron is applied to identify patterns in previous forecasting errors achieved by a machine learning methodology. The proposed learning approach is adaptive to the training data through a re-training process that includes only the most recent and relevant data, thus excluding misleading information from the training process. The learnt error patterns are then combined with the original forecasting results in order to improve forecasting accuracy, using the Rényi entropy to determine the amount in which the original forecasted value should be adapted considering the learnt error patterns. The proposed approach is combined with eleven different machine learning methodologies, and applied to the forecasting of electricity market prices using real data from the Iberian electricity market operator – OMIE. Results show that through the identification of patterns in the forecasting error, the proposed methodology is able to improve the learning algorithms' forecasting accuracy and reduce the variability of their forecasting errors.

Keywords: Artificial Neural Networks, Electricity Market Prices, Entropy, Forecasting, Information Theory, Machine Learning

1. Introduction

Machine learning (and time series forecasting, in particular) is an attractive domain in the power and energy systems field, as it is essential to enable an adequate energy resources management [1]. With the increase of renewable generation, with an intermittent nature, and the consequent need for the increase in consumers' flexibility; forecasting energy generation, consumption and energy market prices, becomes crucial [2]. Regression approaches and Artificial Intelligence (AI) based methods, such as Artificial Neural Networks (ANN) are the most common approaches.

Several relevant advances have been accomplished in forecasting energy market prices. A novel learning algorithm based on improvement of conventional extended Kalman filter is proposed in [3], with the purpose of surpassing the limitations of feedforward ANN when training data contains a certain number of outliers. In [4] authors propose a model that estimates the speed of energy price adjustment to its target level by extracting energy price bubbles and showing that bubbles help explaining price adjustment. The work presented in [5] includes a study on the impact of intra-day price relationships and market fundamentals for short and mid-term forecasting of electricity prices. Market price forecasting models for India are proposed in [6], using Model Confidence Set (MCS) approach to test the utility of these models and picking up the "best" models. Other domain-specific market price forecasting approach is proposed in [7], with focus on the identification of market price spikes in the Ontario electricity market. The work presented in [8], on the other hand, studies the influence of specificities of the market and weather conditions on the electricity price. [9] addresses the perspective of market integration in the scope of day-ahead market price forecasting.

A review that reports the current state on the subject of extreme learning machine (ELM) is presented in [10]. ELM is used for classification, regression, clustering, feature selection, representational learning and many other learning tasks. Hardware implementation and parallel computation techniques have substantially sped up the training of ELM, making it feasible for big data processing and real-time reasoning. ELM are also used in [11] to improve wavelet neural networks in probabilistic forecasting of hourly electricity prices (please refer to [12] for a complete review on probabilistic forecasting of electricity market prices). ELM are also being used as core components of hybrid methods, such as [13], which combines ELM with Seasonal and Trend decomposition using Loess (STL). In fact, the trend for methods' hybridization is gaining significant strength in recent years; see e.g. [14], which proposes a hybrid multi-step ahead forecasting method for the energy system; [15], which proposes a novel method combining interval Holt's exponential smoothing method (HoltI) and multi-output support vector regression (MSVR) for interval-valued time series; [16], in which a hybrid forecasting approach is used for multi-objective optimization; or [17], which proposes a hybrid combination between wavelet-particle swarm optimization and adaptive neuro-fuzzy inference systems for short-term electricity prices forecasting. This type of hybridization generally shows good performance in dealing with multiple input variables. Also, with the objective of facilitating the combination of multiple variables, in [18], authors propose a clustering based methodology to identify similar days and time periods concerning solar intensity. Using this identification, the training process of forecasting methodologies is adapted in order to achieve better forecasting results.

In fact, identifying relevant data values and how these influence the variability of the data series, leads to the concept of entropy [19]. Entropy usually refers to stochasticity or uncertainty, and aims to measure the amount of information brought by each data point. The concept of information entropy was introduced by C. Shannon in [19]. Entropy has been often explored as a means of improving machine learning methodologies. The work proposed in [20] uses an entropy based method to process initial data of photovoltaic power generation and unify these data according to their expected information gain. The refined data is used to train an ELM for short-term forecasting. In [21] fast ensemble empirical mode decomposition and sample entropy are used to pre-process data. The prediction model is then built to forecast the sub-series, whose inputs and outputs are obtained in accordance to phase space reconstruction. This work uses real wind speed data for short-term wind speed forecasting. Also for wind speed forecasting, [22] proposes a novel hybrid model that uses wavelet decomposition based on entropy to create a set of data subseries and estimate their uncertainty. AdaBoost and wavelet filters are then used to forecast the wind speed values. Another hybrid methodology, this time for load forecasting, combining entropy based feature selection, fuzzy inductive reasoning, random forest and ANN is proposed in [23].

Entropy proves that stochasticity is not random; and the improvement of forecasting results when incorporating entropy analysis shows that uncertainty is associated to a degree of added information. Learning error is therefore also not random, and thereby it should be possible to identify error patterns and extract further relevant information to improve the learning process. However, as identified in this related work review, the existing forecasting methods focus on maximizing the forecasting performance by minimizing the prediction error according to the input data. There is no explicit work on analysing the tendencies of forecasting errors throughout the time, in order to identify patterns in these errors and incorporate these patterns as a way to reduce the original forecasting results. On the other hand, entropy is traditionally used in the feature extraction process, but this concept is not extended or applied to analysing the divergence/convergence between the forecasting errors and the actual forecasting values.

This work contributes to surpassing the identified gap by proposing an adaptive entropy-based learning methodology based on the analysis of the forecasting error, instead of using entropy to model and select training data, as performed by current entropy based forecasting models. A dynamic feedforward ANN with backpropagation is modelled to analyse and learn patterns in the historical set of forecasting errors achieved by a learning methodology. The proposed learning approach is adaptive to the training data through a re-training process that includes only the most recent and relevant data, thus excluding misleading information from the training process. Three

dynamic training strategies are proposed to model training data according to different perspectives when analysing the historical data patterns. The learnt error patterns are then combined with the original forecasting results in order to improve forecasting accuracy, using the Rényi entropy as measure to determine the amount in which the original forecast should be adapted taking into account the perceived error values. The proposed approach is combined with eleven different machine learning methodologies, and applied to the forecasting of electricity market prices using real data from the Iberian electricity market operator – OMIE [24]. Results show that through the identification of patterns in the forecasting error, the proposed methodology is able to improve the learning algorithms' forecasting accuracy.

After this introductory section, section 2 describes the proposed dynamic learning approach, including the proposed learning model using the considered entropy measures, and the dynamic training process. Section 3 presents the case study based on real electricity markets data, and section 4 presents the most relevant conclusions from this work.

2. Proposed methodology

1. Overview

The proposed learning model aims at identifying patterns in the history log of forecasting errors measured through the application of a given machine learning methodology. The objective is, therefore, to enable reaching predictions on the expected error of future forecasts provided by the supporting learning method, as depicted in Fig. 1.

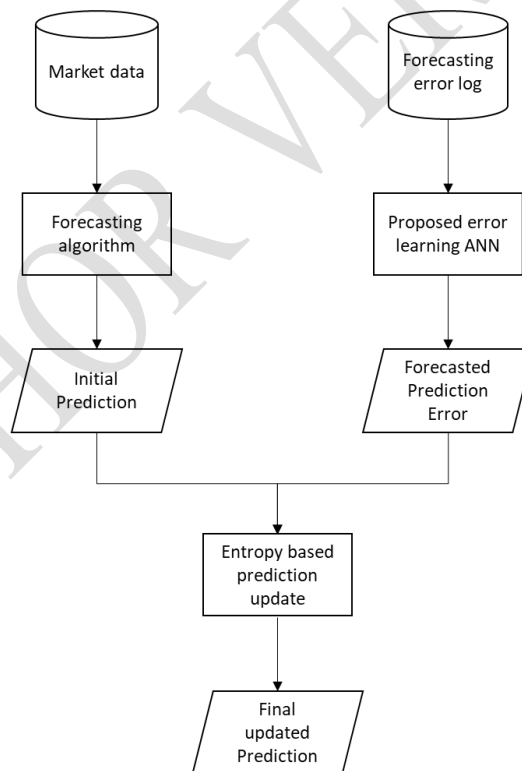


Figure 1. Flowchart of the proposed methodology

Fig. 1 shows that an initial prediction of the target output is provided by a supporting forecasting algorithm (which may be any forecasting method). In order to reach this prediction, the forecasting method analyses the historical data (in the case of the particular case study depicted in this paper, the historical electricity market data).

On the other hand, a historical log of the forecasting errors achieved by the same forecasting method is maintained, and updated every time the accuracy of a new forecast may be evaluated. The ANN described in section 2.2 is then applied in order to find patterns in the historical forecasting

errors, and trained according to the dynamic training process explained in section 2.3. This ANN provides as output the predicted/expected error associated to the most recent forecast that has been achieved.

Considering the expected error of the initial prediction, the initial forecasted value is updated according to the entropy based process explained in section 2.2, thus resulting in the final, updated, prediction that is the final output of the proposed learning approach.

2. Learning model

The forecasting errors may be positive if the predicted value is higher than the actual real value; or negative, if the predicted value is lower than the actual target value. The predicted/expected forecasting error y_t provided by the proposed learning model is used to adapt the initial forecasted value F_t , in order to make it converge towards the actual target value, at each time t . In order to determine the amount in which y_t should influence F_t , the Rényi entropy $H_\alpha(X)$ is used [25]. Rényi entropy is applied because it enables representing the amount of gained information by reflecting the difference between the reality and the forecast results. Among the several entropy and divergence methods available in the literature [26-31], Rényi entropy is the best solution for this problem, as it generalizes several of the most widely used entropy measures, namely the Hartley entropy [32], the Shannon entropy [19], the collision entropy and the min entropy. The Rényi entropy of order α is applied as in (1).

$$H_\alpha(y_t) = \frac{1}{1-\alpha} \log \sum p_{y_t}^\alpha \quad (1)$$

where $\alpha \geq 0$ and $\alpha \neq 1$. p_{y_t} is the probability associated to the predicted/expected forecasting error y_t in time t .

By using entropy as a measure for calculating the amount in which a new event should affect an already accomplished forecasted value, the proposed methodology enables adapting the current forecast and reaching a new forecasted value that is closer to the real value, as in (2).

$$F'_t = F_t - y_t H_\alpha(y_t) \quad (2)$$

The adapted forecasting value F'_t is therefore achieved by subtracting the expected error from the given original forecast, in a measure given the calculation of the Rényi entropy.

The learning process and prediction of the expected error y_t at each time is performed using a feedforward ANN, receiving as inputs the historic of forecasting errors achieved by the supporting forecasting methodology. The ANN considers two nodes in the intermediate layer (as result of extensive preliminary sensitivity analysis), and one output – the predicted forecasting error.

Backpropagation using the gradient descent method [33] has been used as training algorithm for the ANN. This requires calculating the derivative of the squared error function with respect to the weights of the network. The squared error function E for the single output neuron is defined as in (3).

$$E = \frac{1}{2}(t - y)^2 \quad (3)$$

where t is the target output for a training sample, and y is the actual output of the output neuron.

For each neuron j , its output o_j is defined by feedforward calculation, as in (4).

$$o_j = f\left(\sum_{k=1}^n w_{kj} x_k\right) \quad (4)$$

where n is the number of input units to neuron j , and w_{kj} is the weight between neurons k and j . Hence, the input for the activation function f of a neuron is the weighted sum of outputs o_k of the previous neurons. The used activation function f is the logistic function, a log-sigmoid function, which can be

defined as in (5).

$$f(z) = \frac{1}{1 + e^{-z}} \quad (5)$$

The backpropagation algorithm is used as the training method of the designed artificial neural network. The backpropagation algorithm includes the following steps [33]:

1. Initialize weights as small random numbers;
2. Introduce training data to the ANN and calculate the output by propagating the input forward through the network using (4);
3. Calculate the error using (3);
4. Propagate the sensitivities backward through the network by simply taking the derivative of the activation function (5) with respect to the network parameters;
5. Calculate w_{kj} updates;
6. Update the values of w_{kj} ;
7. Repeat steps 2 to 6 until all examples are classified correctly.

3. Dynamic training

In order to guarantee that the learning process is always updated throughout each simulation, the methodology is re-trained every time that new data is observed. The data used for training the ANN in the first iterations is the historic data from the electricity market. As the simulation results start appearing, these are used a continuation of the previous historic, *i.e.*, they will be used to complement the previous information, and will, as the time progresses, be the main subject of analysis for the training of the ANN. Fig. 2 provides a visualization of this process.

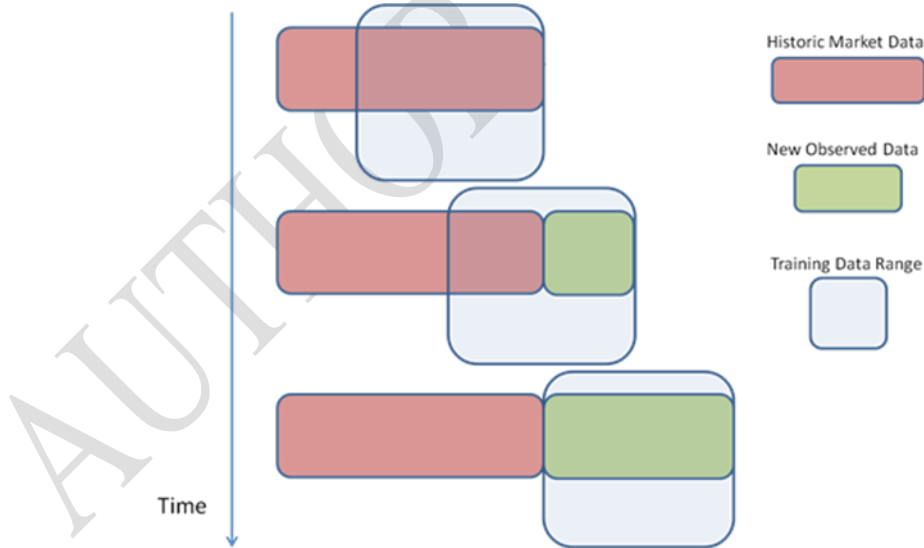


Figure 2. Data range for the training process of the proposed learning methodology

In addition to the re-training process that guarantees the constant update of the learning procedure, the training data is also selected carefully. Since the objective of the proposed approach is to find patterns in the forecasting error, the way historical error data should be approached is a crucial aspect. This process must be ad-hoc and dependent on the original training data itself. In the case addressed by this paper, the used training data refers to the historic electricity market prices. Since electricity market prices depend directly on generation of consumption in each moment, they have a strong seasonal variation (due to the variation in the use of heating and cooling devices). Also, there is a variation depending on the week day due to consumption patterns (especially from business days

to weekends). Also, there are differences from hour to hour due to the variability of generation (especially solar, which is only positive during the day) and consumption patterns. Due to this entropy, the training data must be selected carefully in order to improve the performance of the methodology. Three different training strategies are used, according to the entropy identified above.

These three approaches use the ANN with one value in the output layer - the value of the expected error, two intermediate nodes, and an input layer of four units. The input layer considers different values in the different approaches. These values depend on how the history of the error is considered:

- Strategy A - This strategy makes a prediction along the 24 hourly periods per day, using for the training of each period the error of the same hour for:
 - the previous day;
 - the previous 7 days;
 - the previous 14 days;
 - the previous 21 days.

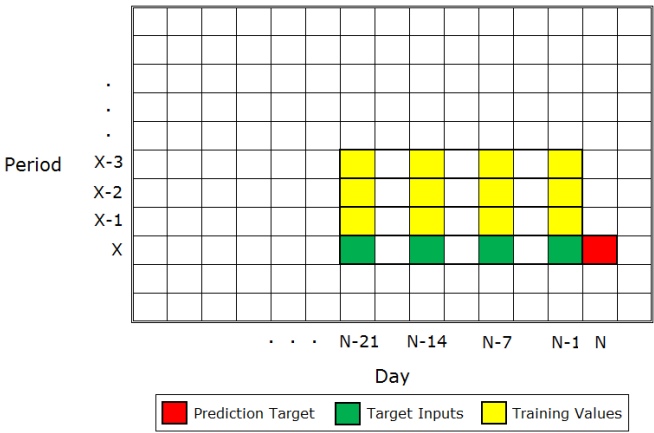


Figure 3. Strategy A training data structure

Having Fig. 3 as support, if an error prediction is required for day N and period X , the input values of the ANN are $N-1$, $N-7$, $N-14$ and $N-21$, all for hour X . The data of the previous hours is used to train the ANN, considering the same days.

- Strategy B - This strategy makes a prediction along the days, using the error of the following hourly periods:
 - previous period;
 - 2 previous periods;
 - 3 previous periods;
 - 4 previous periods.

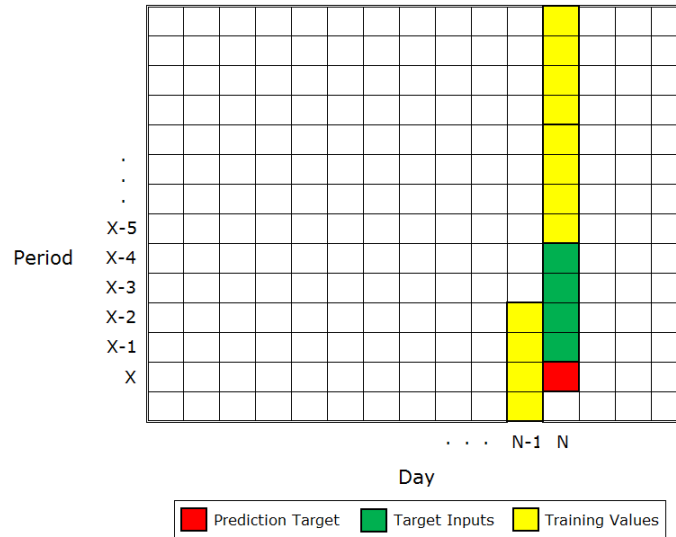


Figure 4. Strategy B training data structure

If an error prediction is required for day N and period X , the input values of the ANN are the previous four hourly periods, considering the same day. The previous periods are considered to train the ANN. When all the periods of the current day have contributed to the training, the periods of the prior day start being considered, as shown in Fig. 4.

- Strategy C - this strategy makes a prediction considering always the errors of the same period (the hourly period in question), using the error for:
 - the previous day;
 - the previous 7 days;
 - the previous 14 days;
 - the previous 21 days.

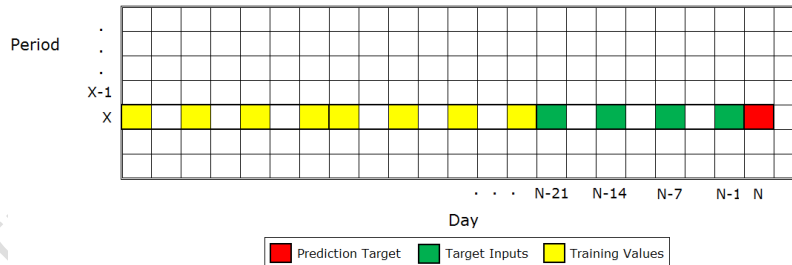


Figure 5. Strategy C training data structure

This approach always considers the same period. It ignores the data concerning all periods other than the required one, and uses the previous days to train the ANN, as depicted by Fig. 5.

3. Case study

1. Specifications

This section presents the case study used to assess the performance of the proposed methodology. The proposed approach is applied to the day-ahead forecasting of electricity market prices, using real data from the Iberian market operator – OMIE. Real data from 3 years is used, namely from 1, November 2014 to 31 October 2017. These data can be consulted and downloaded in [34]. From these data, the first two years are used as training data, and the last year as test and validation data. The used data has the characteristics presented in Table 1.

Table 1. *Experimental data characteristics overview*

Time interval	Max	Min	STD	Median	Mean
3 years (1, November 2014 to 31 October 2017)	12.32	0.00	1.31	6.84	6.32

Eleven different forecasting methods are used as basis approaches to provide the primary electricity market price forecasts. The proposed methodology is then applied to the analysis and learning of the forecasting errors originated by each of these approaches in order to reach the updated prediction values. The results from the application of the proposed methodology to the different methods is compared to the individual performance of the methods themselves. These methods have been chosen as they are the most widely applied methods to the electricity market forecasting problem in the literature. The considered basis forecasting methods are:

- AutoRegressive Integrated Moving Average (ARIMA) model;
- Feedforward ANN trained with backpropagation, as presented in [35];
- Support Vector Machine (SVM) for regression, as presented in [36];
- Wang and Mendel's (WM) Fuzzy Rule-Based System (FRBS) model [37];
- Hybrid neural Fuzzy Inference System (HyFIS) [38];
- Genetic fuzzy systems for fuzzy rule learning based on the MOGUL methodology (GFS.FR.MOGUL) [39];
- Genetic lateral tuning and rule selection of linguistic fuzzy systems (GFS.LT.RS) [40];
- The simplified TSK fuzzy rule generation method using heuristics and gradient descent method (FS.HGD) [41];
- Extreme Learning Machine (ELM) [42];
- Echo State Network (ESN) [43];
- Convolutional neural network [44].

In order to enable a fair comparison between these models, similar input and output structures have been used. The output of all methods is the predicted market price. The inputs are: (i) for ARIMA and SVM, the market price in the same hour of the previous four days, and the market price of the same hour, referring to one week, two weeks, three weeks and four weeks before the target day, (ii) for all the other methods, besides the inputs used for the ARIMA and SVM, inputs also include the volume of traded power in the same hour. This difference in training is due to the bigger difficulty of ARIMA and SVM in achieving quality regression results when associating different variables. The inclusion of extra data in this case only worsens the forecasting results, and it is, therefore, unnecessary, and thereby excluded.

2. Results

Fig. 6, 7 and 8 show a graphical comparison of the real market price, the price forecasted by the ANN and the adapted prediction price resulting from the proposed methodology, throughout the last 2 months of test data, using the three strategies considered for dynamic training, as explained in section 3.3.

Fig. 6 shows the results when using training strategy A, comparing its prediction with the value received by the auxiliary strategy (ANN), to which the expected error is applied. In this comparison there is presented the actual market price, so one can check the quality of the prediction strategy.

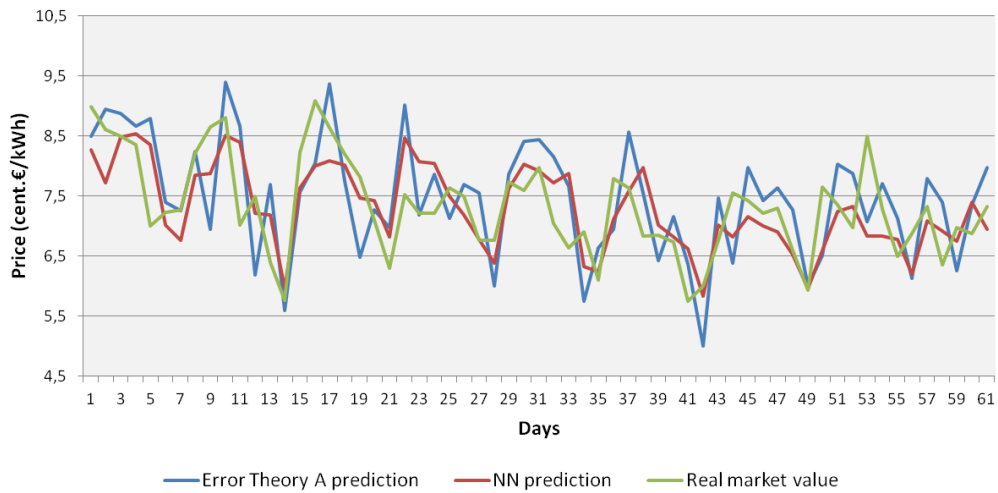


Figure 6. Results using training strategy A

As can be seen from Fig. 6, the calculation and application of the expected error, using strategy A, is very close to the actual market value. From Fig. 6 it is visible that the proposed methodology is able to improve the ANN forecasted values, as the green line is, in most cases, closer to the real market price values than the ANN forecast. The proposed approach is able to identify patterns on the ANN forecasting errors, learn from these errors history log, and use this learning process to adapt the prediction in a way that the predicted values become closer to the real market price values. In Fig. 7, the same information is presented, for the case in which strategy B is applied.

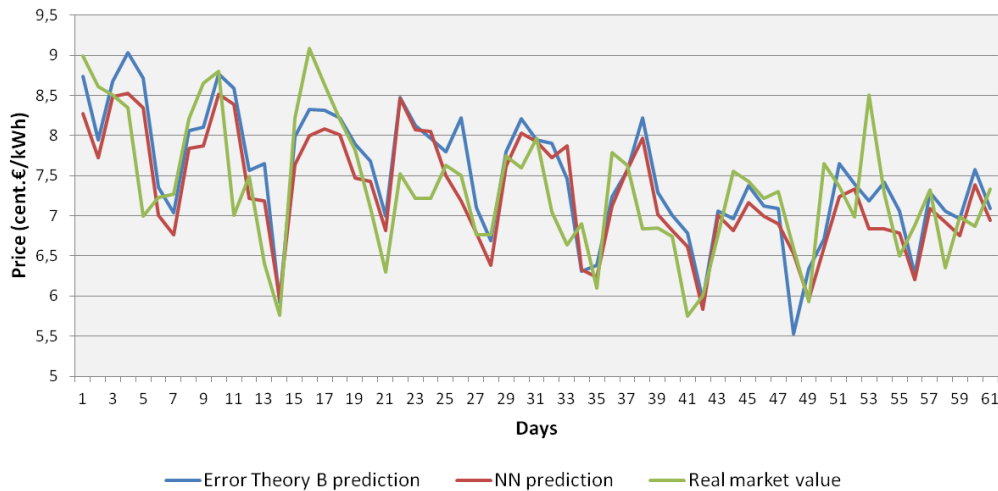


Figure 7. Results using training strategy B

From Fig. 7 it is visible that strategy B has had worst results when compared with strategy A, for the considered time period. Finally, Fig. 8 shows the results from the application of strategy C.

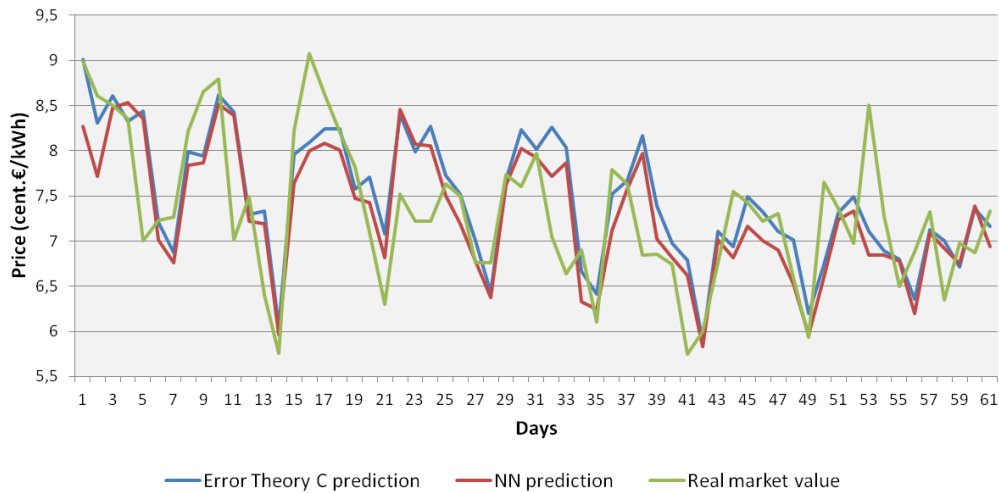


Figure 8. Results using training strategy C

Analysing the results when using strategy C, it can be seen that strategy C achieves better results than strategy B for the considered time period, regarding the level of proximity of the forecast with the actual values of the market, but still worse than strategy A.

From these results it can be concluded that the best strategy to be applied is training strategy A, which is the one applied in the experimental tests shown henceforward.

Table 2 presents the forecasting error achieved by the eleven forecasting methods, and the error achieved by the proposed methodology (using training strategy A) applied to each of these approaches. These results refer to the average results over the entire year of test data. The error is measured through the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The Standard Deviation (STD) of the forecast errors is also presented, as well as the value of Pearson Correlation (PC) [-1, 1] between the real data and the forecasted data [45].

Table 2. Results of the forecasting methods and of the proposed methodology

Method	Without the proposed methodology				With the proposed methodology			
	RMSE	MAPE (%)	STD	PC	RMSE	MAPE (%)	STD	PC
ARIMA	16.4	24.12	5.3	0.74	13.3	20.87	4.9	0.79
ANN	10.09	16.90	4.6	0.81	7.87	9.83	3.2	0.88
SVM	9.95	15.36	5.1	0.82	8.06	10.16	4.2	0.89
WM	12.31	20.06	6.3	0.78	11.45	17.63	4.8	0.81
HyFIS	7.25	10.32	3.6	0.88	5.36	7.87	2.9	0.93
GFS.FR.MOGUL	6.83	9.86	3.1	0.89	4.36	6.93	2.6	0.94
GFS.LT.RS	22.31	34.23	13.3	0.63	24.36	36.49	11.1	0.61
FS.HGD	18.14	27.88	8.4	0.68	14.80	24.42	7.2	0.70
ELM	9.45	14.98	4.9	0.83	7.21	9.37	3.1	0.90
ESN	8.89	13.26	4.2	0.86	6.92	8.89	3.0	0.91
CNN	17.23	26.21	8.6	0.69	13.12	23.87	8.2	0.71

From Table 2 it is visible that the proposed methodology has been able to reduce the forecasting error of nearly all the applied forecasting methodologies. The only exception is the GFS.LT.RS method, whose performance is bad from the start. There is no tendency found in the results, and thus the adaptation of the original value is not done properly. There is no found pattern in the errors achieved by this method. Nevertheless, the proposed approach is still able to decrease the errors STD, smoothing the variation of the errors outliers. In all the other methods the results are improved, regarding both the reduction of the forecasting error and of the STD. The method that achieves the best results is the GFS.FR.MOGUL, followed by the HyFIS, with both approaches being able to reach

a MAPE error below 8% when using the proposed methodology, which is a relevant result. Fig. 9 shows the MAPE errors of the different methodologies, to allow a visual comparison of their performance.

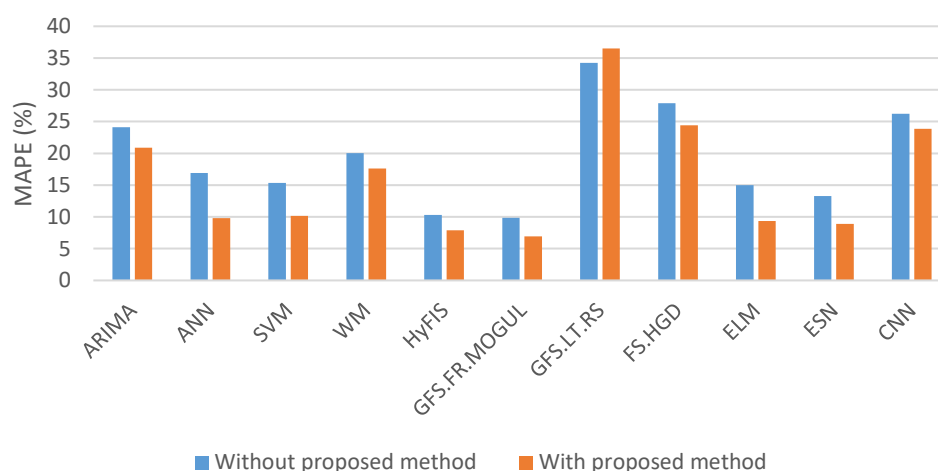


Figure 9. MAPE forecasting errors of the different methods and results when applying the proposed methodology

From Fig. 9 it is visible that the proposed method enables a reduction of the forecasting error of all methods but GFS.LT.RS, by capturing patterns in their forecasting errors, and adjusting the predicted value accordingly, in order to reduce the final prediction error. Table 3 shows the results of the forecasting error prediction, when analysing the error achieved by each of the forecasting methodologies.

Table 3. Results of the predictions' forecasting error prediction

Method	RMSE	MAPE (%)	STD
ARIMA	6.23	23.16	3.2
ANN	2.45	8.34	1.3
SVM	3.16	9.12	1.8
WM	6.32	24.17	2.3
HyFIS	4.26	10.34	2.2
GFS.FR.MOGUL	3.99	11.06	2.3
GFS.LT.RS	10.3	48.12	5.2
FS.HGD	6.06	22.34	3.6
ELM	2.95	9.01	1.7
ESN	2.89	8.86	1.6
CNN	5.98	21.06	3.8

Table 3 shows that the proposed model is able to predict the forecasting error of the different methodologies with an acceptable accuracy. By matching Table 3 with Table 2 it can be seen that the methods that have the better improvement when using the proposed methodology are those for which the proposed method is able to achieve a smaller error in predicting their forecasting errors. This is especially relevant in the case of the ANN (and also clearly visible for the SVM and GFS.FR.MOGUL), which is the method with the smaller forecasting error prediction, and the method with the biggest improvement when using the proposed approach. The contrary is visible in the case of the GFS.LT.RS method; the prediction of its forecasting error is not successful (as seen by the large values in Table 3), and this leads to no improvement when applying the proposed methodology (Table 2).

In order to validate the significance of the proposed method against the eleven considered benchmark methods, a statistical analysis is performed [46]. The Kruskal-Wallis test is the nonparametric test used to compare three or more independent samples. It indicates if there is a difference between at least two of them. This is used to test the null hypothesis that all populations have equal distribution functions against the alternative hypothesis that at least two of the populations have different distribution functions. In this way it is assumed that equality of averages when equality of equal distributions exists [47].

By the test Kruskal-Wallis it is possible to obtain the value of $p = 0$ that gives us indication of rejection of the null hypothesis that all data samples come from the same distribution at a 1% significance level. Given the result of the test that gives the indication of the null hypothesis, the comparison between the pairs of groups is made in order to verify which of the samples differ from each other.

The Bonferroni procedure is performed in order to make the comparison in pairs. Fig. 10 represents the 95% confidence interval for all sample groups (12 methods, in which group 1 is the proposed method using the best supporting forecasting approach: GFS.FR.MOGUL; and the other 11 groups are the 11 benchmark methods considered in this study). In this way, it is possible to see which groups differ in the value of the average, using the Bonferroni procedure.

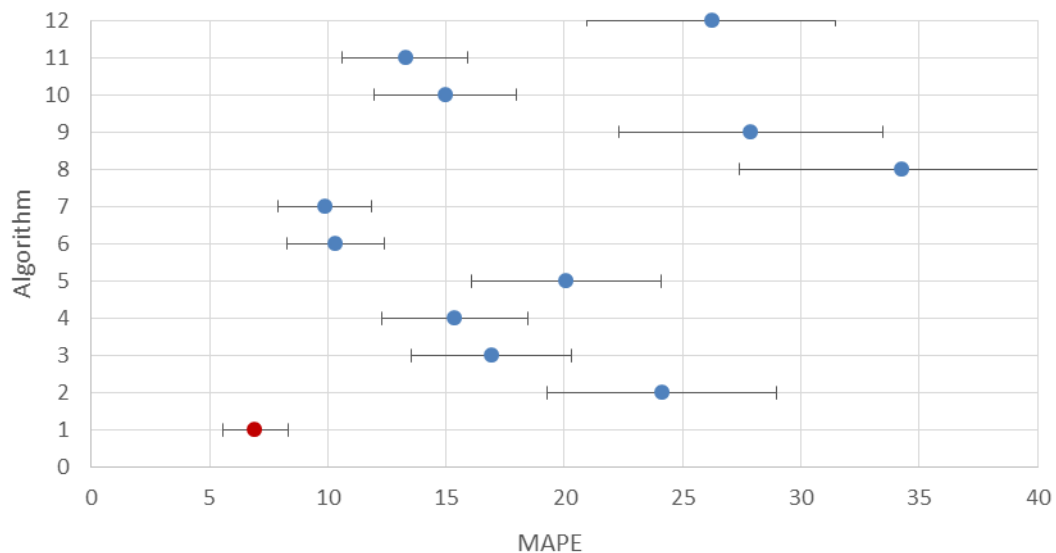


Figure 10. Bonferroni confidence interval by 95%

By analyzing the graph of Fig. 10, it is possible to observe that all methods have significantly different mean values. Table 4 shows the results of this analysis.

Group pairs			p-value
1	1	1	1
1	2	1	1
1	3	1	1
1	4	1	1
1	5	1	1
1	6	1	1
1	7	1	1
1	8	1	1
1	9	1	1
1	10	1	1
1	11	1	1
1	12	1	1

Since the p-value is equal to 1 in all these group tests, the null hypothesis where the groups are considered to have similar means with an error of 5% is accepted.

Taking into account this analysis, it is concluded that the applied benchmark methods achieve significantly different results, thus supporting the relevance of the proposed approach.

4. Conclusions

The study of how the uncertainty associated to different types of data can be modelled and how it influences the prediction of future events, is a relevant domain for several decades. Entropy has arisen as a recognized means to deal with this uncertainty and to extract the information that is given by the different data.

This paper incorporates learning of past data uncertainty into the forecasting process. This is accomplished by applying a multi-layer perceptron to analyse the historic of forecasting errors originated by a forecasting methodology, and reach predictions on the expected forecasting error. The expected error is then used to adapt the original forecasted value in order to reach a more accurate prediction, using the Rényi entropy as measure of the amount in which the adaption of the original forecasted value should be applied.

The proposed methodology has been applied and compared against eleven different state of the art forecasting methodologies, including ANN, SVM, ARIMA, ELM, ESN, and hybrid fuzzy based inference systems. The forecasting process has been experimented on a log of real electricity market data from the last three years. Results show that the proposed methodology is able to successfully find patterns in the history of forecasting errors originated by the different learning methodologies, and use these error predictions to improve the quality of the forecasts. The forecasting error has been decreased, as well as the standard deviation of the forecasting errors.

As a future work, different approaches will be experimented to learn the patterns from past error data, namely probabilistic approaches and hybrid forecasting methods. Also, the combination between the expected error and the original values will be enhanced by exploring different means of reaching intelligent and adaptive combinations of these values.

Acknowledgments: This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 703689 (project ADAPT). The authors also acknowledge the contribution to this manuscript provided by Tiago Pinto's Thesis: Adaptive Learning in Agents Behaviour: A framework for electricity markets simulation, from the School of Engineering, Polytechnic of Porto.

References

1. Hu, J., Harmsen, R., Crijns-Graus, W., Worrell, E., van den Broek, M., 2017. Identifying barriers to large-scale integration of variable renewable electricity into the electricity market: A literature review of market design, In *Renewable and Sustainable Energy Reviews*, 2017
2. Negrete-Pincetic, M., Wang, G., Arancibia, M., Kowli, A., Shafieepoofard, E., Meyn, S., 2017. The value of volatile resources in electricity markets, In *Sustainable Energy, Grids and Networks*, Volume 11, Pages 46-57, 2017
3. Najdan V., et.al. (2015). Robust sequential learning of feedforward neural networks in the presence of heavy-tailed noise. *Neural Networks*. 63: 31-47.
4. Seema Narayan, Paresh Kumar Narayan, Estimating the speed of adjustment to target levels: The case of energy prices, *Energy Economics*, Volume 62, February 2017, Pages 419-427
5. Maciejowska K. and Weron, R. (2016). Short- and Mid-Term Forecasting of Baseload Electricity Prices in the U.K.: The Impact of Intra-Day Price Relationships and Market Fundamentals. In *IEEE Transactions on Power Systems*. 31: 2:994-1005
6. Gupta, S. and Chitkara, P. (2017). Day ahead price forecasting models in thin electricity market. 2017 IEEE Power and Energy Conference at Illinois (PECI), Champaign, IL
7. Sandhu, H.S., Fang, L., Guan, L.: Forecasting day-ahead price spikes for the Ontario electricity market. *Electr. Power Syst. Res.* 141, 450–459 (2016)
8. Figueiredo, N.C., Silva, P.P.D., Bunn, D.: Weather and market specificities in the regional transmission of renewable energy price effects. *Energy*. 114, 188–200 (2016)
9. Lago, J., De Ridder, F., Vrancx, P., De Schutter, B.: Forecasting day-ahead electricity prices in Europe: The importance of considering market integration. *Appl. Energy*. 211, 890–903 (2018)

10. Gao H., et.al. (2015). Trends in extreme learning machines: A review. *Neural Networks*. 61:32–48.
11. Rafiei, M., Niknam, T. and Khooban, M. H. (2017). Probabilistic Forecasting of Hourly Electricity Price by Generalization of ELM for Usage in Improved Wavelet Neural Network. In *IEEE Transactions on Industrial Informatics*. 13:1:71–79.
12. Nowotarski, J., Weron, R.: Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renew. Sustain. Energy Rev.* 81, 1548–1568 (2018).
13. Xiong, T., Li, C., Bao, Y.: Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: Evidence from the vegetable market in China. *Neurocomputing*. 275, 2831–2844 (2018).
14. Du, P., Wang, J., Yang, W., Niu, T.: Multi-step ahead forecasting in electrical power system using a hybrid forecasting system. *Renew. Energy*. 122, 533–550 (2018).
15. Xiong, T., Li, C., Bao, Y.: Interval-valued time series forecasting using a novel hybrid HoltI and MSVR model. *Econ. Model.* 60, 11–23 (2017).
16. Wang, J., Yang, W., Du, P., Li, Y.: Research and application of a hybrid forecasting framework based on multi-objective optimization for electrical power system. *Energy*. 148, 59–78 (2018).
17. Catalão, J.P.S., Pousinho, H.M.I., Mendes, V.M.F.: Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting. *IEEE Trans. Power Syst.* 26, 137–144 (2011).
18. Pinto, T. et. al. (2015). Solar Intensity Characterization using Data-Mining to support Solar Forecasting. 12th International Conference in Distributed Computing and Artificial Intelligence, Advances in Intelligent Systems and Computing, 373:193–201. Springer International Publishing
19. Shannon, Claude E. (July–October 1948). "A Mathematical Theory of Communication". *Bell System Technical Journal*. 27 (3): 379–423.
20. Pingzhou Tang, Di Chen, Yushuo Hou, Entropy method combined with extreme learning machine method for the short-term photovoltaic power generation forecasting, In *Chaos, Solitons & Fractals*, Volume 89, 2016, Pages 243–248
21. Wei Sun, Yuwei Wang, "hort-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network, *Energy Conversion and Management*, Volume 157, 1 February 2018, Pages 1–12
22. Hui Liu, Zhu Duan, Feng-ze Han, Yan-fei Li, Big multi-step wind speed forecasting model based on secondary decomposition, ensemble method and error correction algorithm, *Energy Conversion and Management*, Volume 156, 15 January 2018, Pages 525–541
23. Sergio Jurado, Àngela Nebot, Fransisco Mugica, Narcís Avellana, Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques, In *Energy*, Volume 86, 2015, Pages 276–291
24. Operador del Mercado Iberico de Energia (OMIE) – Iberian market operator website, 2017. Available: <http://www.omie.es>
25. Rényi, A., "Probability Theory", American Elsevier Publishing Company, New York, 1970
26. Rao, S. and Principe, J., "Mean Shift: An Information Theoretic Perspective". *Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, 2008
27. Principe, J., "Information Theoretic Learning", Springer, Information Science and Statistics series, 2010
28. Fisher, R., "On the mathematical foundations of theoretical statistics", *Philosophical Transactions of the Royal Society*, vol. 222, pp. 309–368, 1922
29. Kullback, S. and Leibler, "On Information and Sufficiency", *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951
30. Liu, W. et al., "An Information Theoretic Approach of Designing Sparse Kernel Adaptive Filters", *Transactions on Neural Networks*, vol. 20, no. 12, pp. 1950–1961, 2008
31. Silverman, J. "The Arithmetic of Elliptic Curves", *Graduate Texts in Mathematics*, Springer, 1986
32. Hartley, H., "Transmission of Information", *Bell System Technical Journal*, vol. 7, no. 3, pp. 535 - 563, 1928
33. Rudd, K., Di Muro G. and Ferrari S. "A Constrained Backpropagation Approach for the Adaptive Solution of Partial Differential Equations," *IEEE Trans. on Neural Networks and Learning Systems*, vol. 25, no. 3, pp. 571,584, Mar. 2014
34. Operador del Mercado Iberico de Energia (OMIE) – Iberian market operator historic market prices, 2017. Available: <http://www.omie.es/reports/>
35. Pinto, T., et.al. (2012). Dynamic Artificial Neural Network for Electricity Market Prices Forecast. IEEE 16th International Conference on Intelligent Engineering Systems (INES 2012). Costa de Caparica, Portugal
36. Pinto, T., et al., Support Vector Machines for Decision Support in Electricity Markets' Strategic Bidding, *Neurocomputing*, Elsevier, vol. 172, p.p 438–445, January 2016
37. Wang, L. X. and Mendel, J. M. "Generating fuzzy rules by learning from examples", *IEEE Transactions on Systems, Man and Cybernetics*, vol.22, no.6, p.1414–1427, 1992

-
- 477 38. Kim, J. and Kasabov, N., HyFIS: adaptive neuro-fuzzy inference systems and their application to nonlinear
478 dynamical systems, *Neural Networks*, 12, 9, November 1999, 1301-1319
- 479 39. Herrera, F., Lozano, M. and Verdegay, J.L. "A learning process for fuzzy control rules using genetic algorithms",
480 *Fuzzy Sets and Systems*, vol. 100, pp. 143 - 158 (1998)
- 481 40. Alcalá, R., Alcalá-Fdez, J. and Herrera, F., "A proposal for the genetic lateral tuning of linguistic fuzzy systems and
482 its interaction with rule selection", *IEEE Trans. on Fuzzy Systems*, Vol. 15, No. 4, pp. 616 - 635 (2007).
- 483 41. Ishibuchi, H. et al, "Empirical study on learning in fuzzy systems by rice taste analysis", *Fuzzy Set and Systems*,
484 vol. 64, no. 2, pp. 129 - 144 (1994).
- 485 42. Zhang, Y., Yuan, Y., Wang, Y., Wang, G.: A novel multimodal retrieval model based on ELM. *Neurocomputing*.
486 277, 65–77 (2018).
- 487 43. Sotirios P. Chatzis, Yiannis Demiris, "Echo State Gaussian Process," *IEEE Transactions on Neural Networks*, vol.
488 22, no. 9, pp. 1435-1445, Sep. 2011
- 489 44. Dong, Z., Jing, C., Pei, M., Jia, Y.: Deep CNN based binary hash video representations for face retrieval. *Pattern*
490 *Recognit.* 81, 357–369 (2018).
- 491 45. Wang, J., Yang, W., Du, P., Niu, T.: A novel hybrid forecasting system of wind speed based on a newly developed
492 multi-objective sine cosine algorithm. *Energy Convers. Manag.* 163, 134–150 (2018).
- 493 46. Du P, Wang J Z, Guo Z H and Yang W D 2017 Research and application of a novel hybrid forecasting system
494 based on multi-objective optimization for wind speed forecasting *Energy Conversion and Management* 150 90-107
- 495 47. Theodorsson-Norheim, E. "Kruskal-Wallis test: BASIC computer program to perform nonparametric one-way
496 analysis of variance and multiple comparisons on ranks of several independent samples," *Comput. Methods*
497 *Programs Biomed.*, vol. 23, no. 1, pp. 57–62, 1986