Energy Conversion and Management 99 (2015) 62–71

Contents lists available at ScienceDirect

Energy Conversion and Management

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

# One-year-ahead energy demand estimation from macroeconomic variables using computational intelligence algorithms

S. Salcedo-Sanz [a,*], J. Muñoz-Bulnes [a], J.A. Portilla-Figueras [a], J. Del Ser [b]

[a] Department of Signal Processing and Communications, Universidad de Alcalá, Madrid, Spain
[b] OPTIMA Area, Tecnalia Research & Innovation, Zamudio, Spain

ARTICLE INFO

*Article history:*
Received 30 July 2014
Accepted 29 March 2015
Available online 24 April 2015

*Keywords:*
Energy demand estimation
Computational intelligence
Extreme Learning Machines
Harmony Search

ABSTRACT

This paper elaborates on a problem of one-year ahead estimation of energy demand based on macroeconomic variables. To this end, two different Computational Intelligence approaches are herein evaluated: (1) a modified Harmony Search (HS) optimization algorithm with an exponential prediction model and (2) an Extreme Learning Machine (ELM). In the case of the HS, a feature selection of the best set of features for the prediction is carried out jointly with the optimization of the model's parameters. On the other hand, the ELM will be tested with and without the feature selection carried out by the HS approach. We describe several modifications on the proposed HS, which include a hybrid encoding with a binary part for the feature selection, and a real part to tune the parameters of the prediction model. Other adaptations focused on the HS operators are also introduced. The performance of both approaches has been assessed in a real application scenario, corresponding to the total energy demand estimation in Spain, in which we have 14 macroeconomic variables with history values for the last 30 years, including the recent crisis period starting in 2008. The performance of the proposed HS and ELM models incorporating feature selection is shown to provide an accurate one-year-ahead forecast at a higher prediction's accuracy when compared to previous proposals in the literature. Specifically, the HS and ELM approaches are able to improve the results of a previous approach (based on a genetic algorithm), obtaining an improvement over 15% in this problem of energy demand estimation. As a final experimental evaluation of the proposed algorithm, a similar problem of one-year ahead $CO_2$ emissions estimation from macroeconomic variables is also tackled, and also in this case the HS and ELM are able to obtain significant improvements over a previous approach based on evolutionary computation, over 10% of improvement in this problem.



## 1. Introduction

In the last decades energy demand has increased sharply at a worldwide scale, pushed up by the globalization phenomenon, a rapid population growth, an aggressive industrialization of developing countries and the high standard of life in developed nations [1]. In this context it is well known that, as the economy grows, the energy demand increases exponentially, what brings along important environmental issues that may compromise the future of next generations. Currently, 80% of the energy demand in the world is covered by non-renewable energy sources such as coal or petroleum, with more dramatic values of this indicator foreseen at developing countries. Another point to be taken into account lies on the fact that industry is the responsible for more than 50% of the energy demand in the world. Consequently, countries with a growing industrial activity happen to be more energy demanding than other with economies based on alternative sectors. Managing medium and long-term energy demand has become a key problem with impact in all countries' economies and nations' development.

Some years ago different studies predicted an increase of the overall energy demand of more than 50% in the next 20 years, in what seemed an unstoppable process [2–6,9]. However, all these forecasts and projections failed after the deep world crisis that started in 2008. The main issue with energy demand estimation problems at a national level is that they depend on macroeconomic variables, which are annually calculated in the majority of cases. Thus, very few data are usually available for constructing the prediction models for energy demand estimation. Furthermore,

* Corresponding author at: Department of Signal Theory and Communications, Universidad de Alcalá, 28871 Alcalá de Henares, Madrid, Spain. Tel.: +34 91 885 6731; fax: +34 91 885 6699.

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

http://dx.doi.org/10.1016/j.enconman.2015.03.109
0196-8904/

economies of the world 30 years ago were completely different than the current ones, what restricts the variety of historical macroeconomic indicators that can be considered for energy demand estimation.

Having said this, the first approach to tackle this problem was proposed in [3], where a genetic algorithm (GA) was used to obtain the parameters of an exponential prediction model. Specifically, the model proposed in [3] is based on four input macroeconomic variables (Gross Domestic Product or GDP, population, import size and export size) for Turkey, with data recorded from the early 80s to the first years of 2000. The prediction of the energy demand was done for the same year than the input variables (i.e. features' importance is studied for the same year than that of the energy demand, instead of considering the prediction over a given time horizon). Linear and exponential models were considered, whereas the GA was proposed to be a basic binary algorithm, with standard crossover, flip mutation, and a tournament selection. The objective function to be optimized was a measure of mean quadratic error between the real data and the result given by the model, computed over a training set, i.e. a fraction of the available data. With the obtained models it was proven that energy demand in the future could be estimated by projecting variations in the affecting factors (input variables). In this case, future projections predicted a continuous increase of the energy demand in Turkey for the forthcoming 20 years.

The majority of the subsequent literature has since then focused on testing the performance of different evolutionary-type algorithms when applied to this problem, such as Particle Swarm Optimization (PSO) [4,5]) or hybrid approaches based on PSO and Ant Colony Optimization (ACO) [7]. Another hybrid approach blending together PSO and GA has been recently reported in [6,8,10] for energy demand estimation in China. Other approaches have instead elaborated on prediction models from a different approach than the exponential ones used in [3]. Thus, in [11] several new models based on logarithmic and alternative exponential functions are used, optimized by a real-encoding genetic algorithm. All these previous approaches consider a reduced number of affecting factors (input variables or features) from which the obtained projections show a sustained increase of the energy demand in next years. In all cases the training years do not include data beyond 2005, i.e. all years thereafter are missing important events expected to impact on the quality of the performed prediction (e.g. the 2008 global crisis).

In this paper energy demand estimation is tackled from a novel perspective, which combines evolutionary solvers and neural computation algorithms towards an efficient solving methodology. First, we focus on a one-year-ahead energy demand prediction problem: this must be regarded as a major difference with previous approaches where energy estimation is analyzed (i.e. they relate input variables and energy demand, all taken at the same year). In addition, we consider a higher number of predictive (input) variables than previous approaches, with a feature selection procedure to yield the best set of input variables that must be considered for the predictive model. On this purpose we propose to use the Harmony Search (HS) algorithm [12] – a recent evolutionary optimization approach based on mimicking the music generation and improvisation processes – which has obtained very good results in a number of applications [13]. The manuscript describes the proposed approach thoroughly and analyzes its performance when applied with an exponential prediction model to the one-year-ahead energy demand forecast in Spain. The study is further extended by considering a novel neural computation approach – Extreme Learning Machine (ELM) – as the predictive model, which is applied to the complete spectrum of available input variables, as well as to the best set of features obtained by the HS feature selection. An extension of the problem that considers the prediction of $CO_2$ emissions from macroeconomic variables is also discussed in this work.

The rest of the paper is structured as follows: Section 2 formally describes the problem under consideration, with specific remarks made on the importance of feature selection in prediction problems tackled by means of computational intelligence algorithms. Next, Section 3 describes the fundamentals of the HS algorithm, with details on the used specific encoding and objective function. The main characteristics of the ELM model are also summarized in this section. Section 4 discusses the performance of the proposed algorithms in a real case of energy demand prediction in Spain, for which a comparison with alternative algorithms in the literature is presented. An extension of the problem to a similar case of one-year ahead $CO_2$ emissions estimation is also discussed in this Section. Finally, Section 5 closes the paper by drawing some ending conclusions.

## 2. Problem definition

Let us consider a time series $\mathbf{E} \triangleq \{E(t)\}_{t=1}^{n}$ of past energy demands for a given country, with $n$ discrete values corresponding to different years; and a set of $m$ predictive variables $\mathbf{X} = \{X_1(t), \ldots, X_m(t)\}$, with $t = 1, \ldots, n$. A model $\mathcal{M}$ provides an estimation $\widehat{\mathbf{E}}$ for $\mathbf{E}$. The problem tackled in this paper consists of finding the best set of $m' \leqslant m$ features out of the $m$ possible variables in $\mathbf{X}$, as well as the values for the components/parameters of the model $\mathcal{M}$ such that a given objective function – usually related to the similarity of the model output to the real energy demand values – is optimized. In this case, we consider that such a function is given by the mean squared error between the observed values and the predicted ones, which is to be minimized, i.e.

$$f(\mathcal{M}) = \frac{1}{n^*}\sum_{j=1}^{n^*}\left(E(j) - \widehat{E}(j)\right)^2, \tag{1}$$

where $n^*$ is the size of a reduced training sample ($n^* < n$).

The above formulation corresponds to a class of the so-called Feature Selection (FS) problem. Feature selection is an important task in supervised classification and regression problems because irrelevant features, used as part of the training procedure can unnecessarily increase the cost and running time of a prediction system, as well as degrade its generalization performance [14]. FS problems can be approached under two different schemes: the first attempts at identifying an appropriate set of features independently of the performance of the model $\mathcal{M}$, which preserve most of the information provided by the original data. This approach is known as the *filter* method for feature selection [14], which is outlined in Fig. 1(a) for the sake of completion.

The second FS approach directly selects a subset of $m'$ features out of the total available in such a way that the performance of the model $\mathcal{M}$ is improved or, at least, not degraded. This approach is usually known as *wrapper* method for the FS. These wrapper methods result to be in general more powerful than filter approaches at the usual penalty of an increased computational cost [15,16]. Fig. 1(b) shows an outline of the wrapper method for feature selection. The search of the best subset of input variables can be performed by means of any search algorithm such as hill-climbing and greedy or evolutionary solvers.

## 3. Materials and methods

As has been advanced in the introduction, this paper evaluates a number of Computational Intelligence approaches when applied to the one-year-ahead energy demand prediction based on

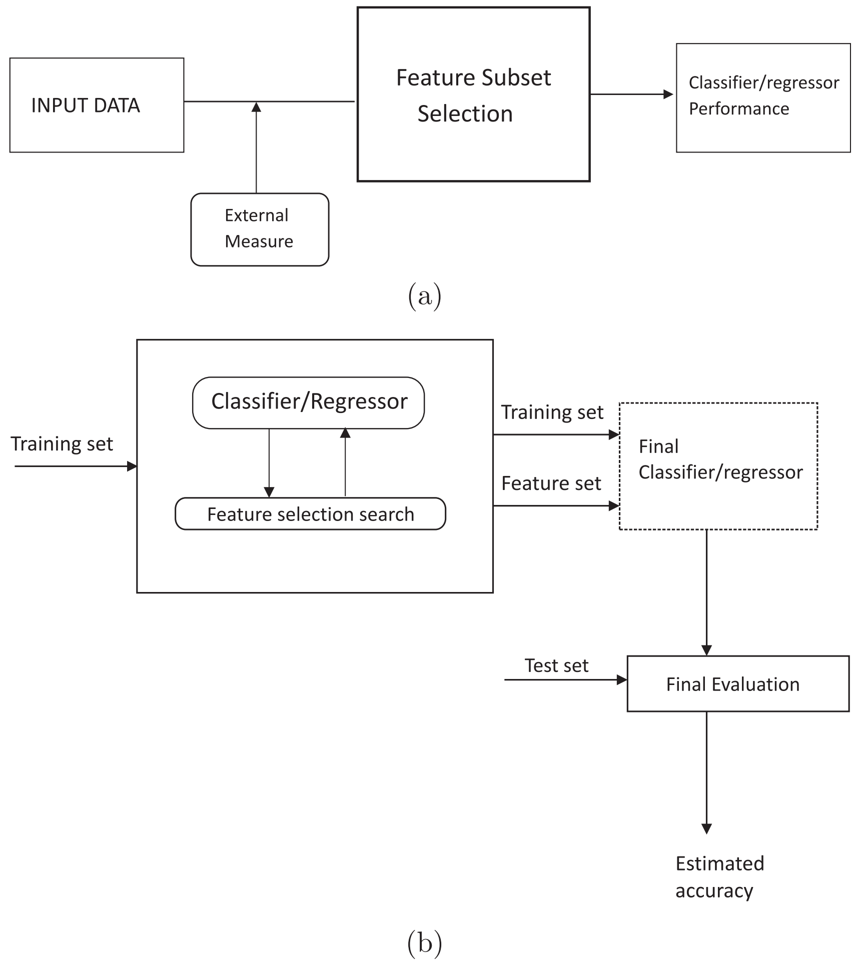


**Fig. 1.** Strategies in FSP: (a) Filter methods. (b) Wrapper methods.

macroeconomic variables. Specifically the scope of this research is twofold: on one hand, the HS evolutionary algorithm is used for the feature selection process and the choice of the best possible predictive model for the problem at hand. On the other hand, an Extreme Learning Machine (ELM) will be used as the predictive engine by leveraging the feature selection made by the HS solver. In this section we describe in detail both approaches and their specific characteristics when used for the prediction problem under analysis.

### 3.1. Hybrid HS-exponential model

First proposed in [12], the Harmony Search algorithm is a derivative-free, meta-heuristic optimization technique that imitates the behavior of a music band when composing music. It hence borrows concepts from this field such as harmony, notes, and pitch and improvisation [17,18]. In essence, this technique iterates on a set of $\varphi$ possible solutions or *harmonies* commonly denoted as Harmony Memory (HM), which are evaluated at each iteration under an *aesthetic* point of view given by the objective function. The HM is updated whenever any of the $\varphi$ improvised harmonies at a given iteration sounds *better* (under a certain fitness criterion) than any of the $\varphi$ harmonies kept from the previous iteration. This procedure is repeated until a maximum number of iterations $\mathcal{I}$ is reached.

At each iteration of the algorithm, the *improvisation* process is applied sequentially to each note of the total set of melodies. The harmony improvisation process of the HS algorithm is driven by two probabilistic parameters: (1) *Harmony Memory Considering Rate*, HMCR and (2) *Pitch Adjusting Rate*, PAR.

(1) The Harmony Memory Considering Rate, HMCR $\in [0,1]$, sets the probability that the new value for a certain note is drawn uniformly from the values of this same note in the other $\varphi - 1$ harmonies. Otherwise, the value of the note is randomly assigned from the valid range of values for that note.
(2) The Pitch Adjusting Rate, PAR $\in [0,1]$, executes subtle adjustments in the chosen harmony. This parameter operates note-wise as follows: the note to which the parameter is being applied will be updated as $x' = x + \epsilon(0,1)$, where $\epsilon$ is a *bandwidth* to adjust the updating done in a given note. Note that this updating is slightly different in discrete-valued notes, where the updating involves alternative valid values for that note.

#### 3.1.1. Problem encoding and HS specific adaptations

The objective function defined in Eq. (1) is considered for guiding the search process of the HS meta-heuristic algorithm. In this case, the model $\mathcal{M}$ used to estimate the energy demand (from variables measured one year in advance) is an exponential model as was previously suggested and used in [3], which is given by:

$$\widehat{E}(t+1) = \sum_{i=1}^{T} w_i X_i(t)^{w_{i+m'}} + w_0 \tag{2}$$

where $m' \leqslant m$ is the number of input variables, and $\mathbf{W} \triangleq \{w_0, w_1, \ldots, w_{2m'}\}$ is a vector of $2m'+1$ weights that describe the model $\mathcal{M}$.

It is important to note that this model imposes conditions in the solution encoding for the HS solver since it must consider two different parts: a first one for the feature selection, and a second one for the values for the model's parameters $\mathbf{W}$. Thus, we consider a binary-valued encoding part for representing the selected features, and a real-valued part for encoding the weights to be used in Eq. (2). Intuitively the number of 1's in the binary part of the encoding – which denote those variables selected as an input to the predictive model – drives the number of weights to be optimized in the real-valued part of the solution vector. In order to avoid a variable-length encoding, we keep the number of 1's in the binary part fixed by applying a restricted search operator [19], which operates by adding or removing 1's when this number is different from the defined one.

Another important point to be highlighted is that the value of the input variables $\mathbf{X}$ is assumed to be normalized in the interval $[-1, 1]$ in order to avoid scale problems with the regression model given in Eq. (2). With this in mind, the weights $\{w_1, \ldots, w_{2m'}\}$ in Eq. (2) are considered to be in the interval $[-1, 1]$, but the bias $w_0$ has been considered to fall in a wider range $\left[-w_0^{min}, w_0^{max}\right] = [-5.0, 5.0]$, for providing the model with some margin for a better fit. A possible example (with randomly chosen values) of the encoding of a given harmony with $m = 15$ input variables and $m' = 4$ features to be selected, is the following:

$$[1\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 1\ 0|\ -3.4\ 0.8\ \ -0.13\ 0.54\ 0.83\ \ -0.64\ 0.41\ 0.02\ \ -0.19]$$

Regarding the parameters controlling the search behavior of the HS solver, their values have been optimized by means of a grid search not shown for the sake of clarity and brevity. The value of the HMCR parameter has been set to linearly increase from $\text{HMCR}_s = 0.7$ to $\text{HMCR}_e = 0.95$ with generations. The PAR parameter is also linearly increased with generations from its starting value $\text{PAR}_s = 0.1$ to its ending value $\text{PAR}_e = 0.5$. In the binary part of each harmony, we consider a different PAR procedure: instead of implementing the conventional mutation based on alphabet vicinity originally proposed in [12], we substitute each bit by the corresponding one in the best solution obtained so far by the algorithm, as has been recently suggested in [20].

### 3.2. Extreme Learning Machines

The second predictive model considered in this work is an Extreme Learning Machine (ELM), which is a fast learning method based on the structure of Multi-Layer Perceptrons (MLP) recently proposed in [21] and successfully applied thereafter to a large number of classification and regression problems [22–24]. The ELM structure is similar to the network given in Fig. 2. The most significant characteristic of the ELM training procedure lies on the fact that it is carried out just by randomly setting the weights of the network, and subsequently computing the inverse of the hidden layer output matrix. This implies an extremely fast, simple training algorithm, as well as an outstanding performance when compared to avant-garde learning methods, usually better than other established approaches such as classical MLPs or support vector machines. Moreover, the universal approximation capability of the ELM network, as well as its classification capability, have been already proven in [25].

Mathematically speaking, the ELM method can be defined by a training set $\aleph \triangleq \{(\mathbf{x}_i, \mathbf{t}_i) | \mathbf{x}_i \in \mathbb{R}^n, \mathbf{t}_i \in \mathbb{R}, i = 1, \ldots, N_T\}$, an activation function $g(x)$ and a number of hidden nodes $\widetilde{N}$. The ELM training algorithm can summarized in the following steps:

1. Randomly assign inputs weights $\mathbf{w}_i$ and bias $b_i$, with $i = 1, \ldots, \widetilde{N}$.
2. Calculate the $N_T \times \widetilde{N}$ hidden-layer output matrix $\mathbf{H}$, defined as

$$\mathbf{H} \triangleq \begin{bmatrix} g(\mathbf{w}_1\mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_{\widetilde{N}}\mathbf{x}_1 + b_{\widetilde{N}}) \\ \vdots & \ddots & \vdots \\ g(\mathbf{w}_1\mathbf{x}_{N_T} + b_1) & \cdots & g(\mathbf{w}_{\widetilde{N}}\mathbf{x}_{N_T} + b_{\widetilde{N}}) \end{bmatrix}. \quad (3)$$

3. Calculate the output weight vector $\beta$ as

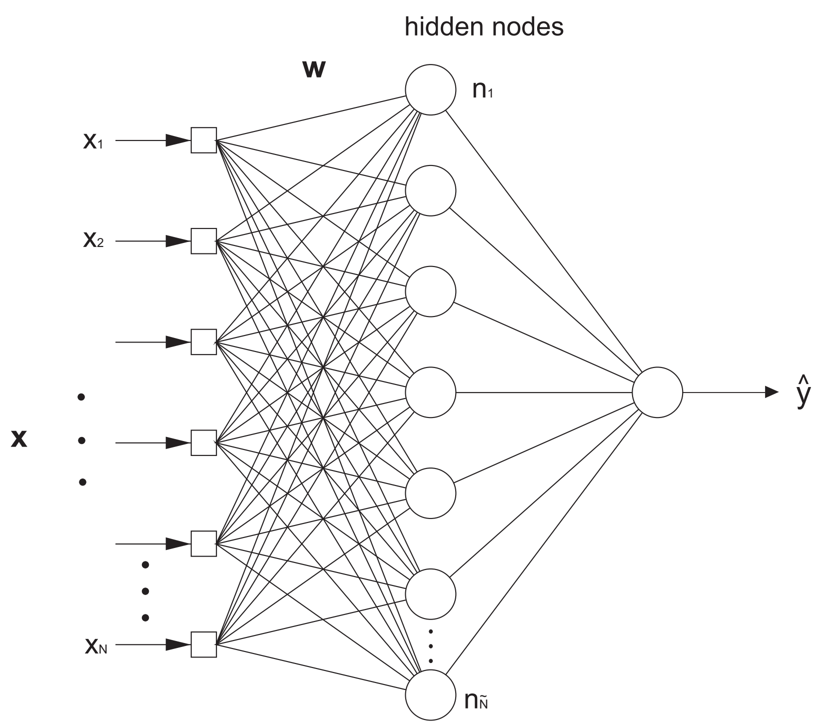


**Fig. 2.** Outline of the Extreme Learning Machine structure.

$$\beta = \mathbf{H}^{\dagger}\mathbf{T}, \tag{4}$$

where $\mathbf{H}^{\dagger}$ stands for the Moore–Pennrose inverse of matrix $\mathbf{H}$ [21], and $\mathbf{T} \triangleq [t_1, \ldots, t_{N_T}]^T$ is the training output vector.

It is important to note that the number of hidden nodes $\widetilde{N}$ is a free parameter of the ELM training that needs to be estimated in order to obtain good results. Usually, scanning an integer range of $\widetilde{N}$ values is the most straightforward and practical method for this estimation. Note that due to their good performance, along with an fast training procedure, ELMs are well suited for hybrid algorithms requiring very fast classifiers and regressors.

## 4. Experimental part

The performance of the proposed approaches has been assessed in a real problem of energy demand estimation in Spain. Data from 1980 to 2011 are available for a total of $m = 14$ predictive variables described in Table 1. A partition of these data into training and test sets is done in such a way that data from $n^* = 15$ years are selected for training and data from $n - n^* = 16$ years for test. Values corresponding to 2010 and 2011 have been kept within the test set in order to check the performance of the predictive model over years of crisis, whereas the rest of data have been randomly chosen to belong to the test or train sets.

First of all, we discuss the results obtained with the HS algorithm. Table 2 shows a comparison of the relative Mean Absolute Error (MAE, in %) obtained with the proposed HS with feature selection, and the result of the GA proposed in [3], in which the first 4 variables in Table 1 are considered. In order to compare both approaches, we have set the number of variables (namely, the number of 1's in the binary part of the HS) to 4. Given their stochastically driven search procedure, 10 runs of each algorithm has been executed, keeping the best and average values of the MAE for the prediction in the test set. As can be seen, the performance of the proposed HS is better than the algorithm in [3], in terms of both its average value over the 10 runs and the best result obtained. Fig. 3 shows the energy demand prediction versus the real values in the considered test set for the algorithms under comparison. Recall that the energy demand is predicted for $t + 1$ from input variables measured in $t$. As can be seen, the prediction fits really well with the real curve, even in the years under economic crisis (from 2008 onwards). It is easy to see how the proposed HS approach outperforms the GA scheme in the years near the crisis (from 2008). Another interesting analysis is related to the best 4 variables selected by the HS (indexes 1, 3, 7 and 10). Variable 1 is the GDP of the country the year before the energy estimation is considered ($t$). Variable 3 is the export figure of the country in $t$, 7 is the electricity production in $t$, and finally variable 10 stands for the fossil fuel consumption in $t$. Interestingly yet reasonably, these 4 variables seems to be better predictive inputs than those proposed in [3] (export, import, GDP and population), since the error made by the proposed algorithm with these variables is lower than the one obtained by the GA in [3]. In this sense, the proposed algorithm has more degrees of freedom than the proposal in [3], i.e., it can choose input variables from a pool of possibilities whereas the GA in [3] was constrained to deal with the four predictive variables mentioned before, what makes the proposed HS algorithm to perform better in this problem.

The next step in the evaluation of the HS in this problem of energy demand estimation is to consider the effect of including a different number of input variables in the model. Table 3 shows the obtained results, where it can be noticed that the best result is obtained with 7 input variables (yielding a best MAE of 2.36%). Furthermore, it can be seen that this set of 7 features includes 3 of the 4 variables discussed before: in this context the reader should note that the HS does not only select variables, but also optimizes the parametric weights of the predictive model, thus direct comparison between different cases would not be fair in any case. Fig. 4 shows the energy demand prediction versus the real values (test set), using the model $\mathcal{M}$ obtained with the HS with 7 input variables (15 weights in $\mathbf{W}$). Fig. 5 shows the histogram (over 10 runs) of the variables included in the prediction for the HS with 7 input variables. As can be seen, variables 11 and 12 are the two most considered by the algorithm in the 10 runs performed, whereas variables 2, 6, 8 and 13 are the less used. However, note that depending on the run, groups of different variables may arise that provide better results, so in general it is difficult to see a clear pattern of feature selection for the HS, and it depends on the run that the algorithm will select a different set of features.

Finally, Table 4 lists the results obtained with the alternative ELM predictive model when applied to the same set of Spanish variables (1) without feature selection (i.e. all predictive variables are considered as inputs to the model); (2) with the 4 variables selected by the HS; and (3) with the 7 variables selected by the HS (see Table 3). The performance of the ELM without feature selection is worse than the HS. On the other hand, the results of the ELM with 4 and 7 features show that this process improves its performance. In the case of 4 features, the ELM result is still worse than the HS, though still better than the solution obtained with a GA in [3]. The result rendered by the ELM with 7 predictive variables is better than the one by the HS and GA and, in fact, results in the best solution obtained through all the experiments carried out in this work. Fig. 6 shows the energy demand prediction versus the real values in the test set, for the case with 4, 7 and all input features to the ELM model. The prediction provided by the ELM is accurate in all cases, even in the crisis years (2010 and 2011), where the energy demand decreased with respect to the one in previous years.

**Table 1**
Variables considered in this problem of energy demand estimation. The first 4 variables correspond to the study in [3].

| # | Variable |
|---|---|
| 1 | GDP |
| 2 | Population |
| 3 | Export |
| 4 | Import |
| 5 | Energy production (kTOE) |
| 6 | Electricity power transport (kW h) |
| 7 | Electricity production (kW h) |
| 8 | GDP per unit of energy use |
| 9 | Energy imports net (% of use) |
| 10 | Fossil fuel consumption (% of total) |
| 11 | Electric power consumption (kW h) |
| 12 | $CO_2$ emissions total (Mtons) |
| 13 | Unemployment rate |
| 14 | Diesel consumption in road (kTOE) |

**Table 2**
Comparison of the relative MAE (%) obtained with the proposed HS and the algorithm in [3] for the problem of energy demand estimation in Spain.

| Algorithm | Relative MAE best in (%) | Relative MAE average of 10 runs in (%) | Selected variables (best run) |
|---|---|---|---|
| HS with FSP | 2.60 | 4.03 | 1, 3, 7, 10 |
| GA from [3] | 2.89 | 4.43 | 1, 2, 3, 4 |

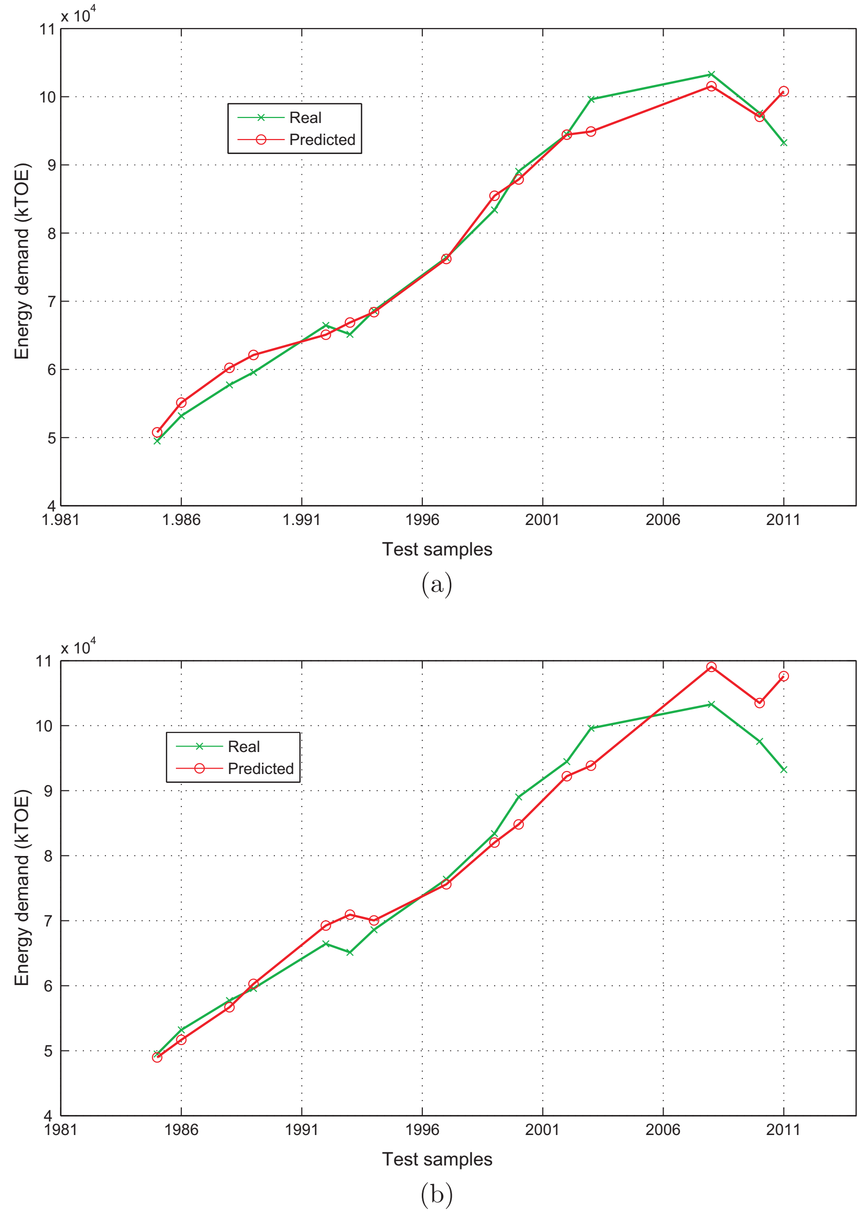


**Fig. 3.** Real energy demand versus prediction in the test set considered: (a) HS with FSP and (b) GA from [3].

**Table 3**
Relative MAE (%) obtained with the proposed HS with different number of input variables for the problem of energy demand estimation in Spain.

| Number variables | Relative MAE best in (%) | Relative MAE average of 10 runs in (%) | Selected variables (best run) |
|---|---|---|---|
| 4 | 2.60 | 4.03 | 1, 3, 7, 10 |
| 5 | 2.54 | 3.21 | 2, 6, 8, 10, 11 |
| 6 | 3.40 | 4.63 | 5, 8, 9, 10, 11, 12 |
| 7 | **2.36** | 4.05 | 1, 2, 3, 7, 8, 9, 12 |
| 8 | 2.62 | 4.55 | 2, 4, 6, 8, 9, 11, 12, 14 |
| 9 | 2.89 | 4.11 | 1, 2, 6, 7, 10, 11, 12, 13, 14 |

Boldface value stands for the best MAE obtained.

### *4.1. Further analysis: application to $CO_2$ emission estimation*

The increasing in energy demand at worldwide scale, has brought a continuous increasing of greenhouse gasses emissions, specially $CO_2$, with important environmental consequences for many different habitats and natural systems [26,27]. The methodology proposed in this paper can be directly applied to the $CO_2$ emission estimation in a given country: The model considered for this problem is similar to the one shown in Section 2, but in this case Energy demand substitutes $CO_2$ emissions as predictive variable (variable 12, Table 1), and the total $CO_2$ emissions in the country is now the objective variable. We have applied the proposed HS with FSP and the ELM to this problem for the case of Spain, and we use again the GA in [3] for comparison purposes. First, the HS with four predictive variables is considered in order to directly compare with the GA in [3]. Table 5 shows the results obtained in this problem. As can be seen, the accuracy of both approaches is lower than in the case of energy demand estimation, what means that the $CO_2$ emissions is a more complex problem to model with macro-economic variables. The best solution, obtained with the proposed HS algorithm, obtains a 5.7% of error in terms of MAE, improving the 6.15 obtained by the GA for comparison. It is quite interesting that two features selected by the HS algorithm in this problem were also selected by the algorithm in the problem of energy

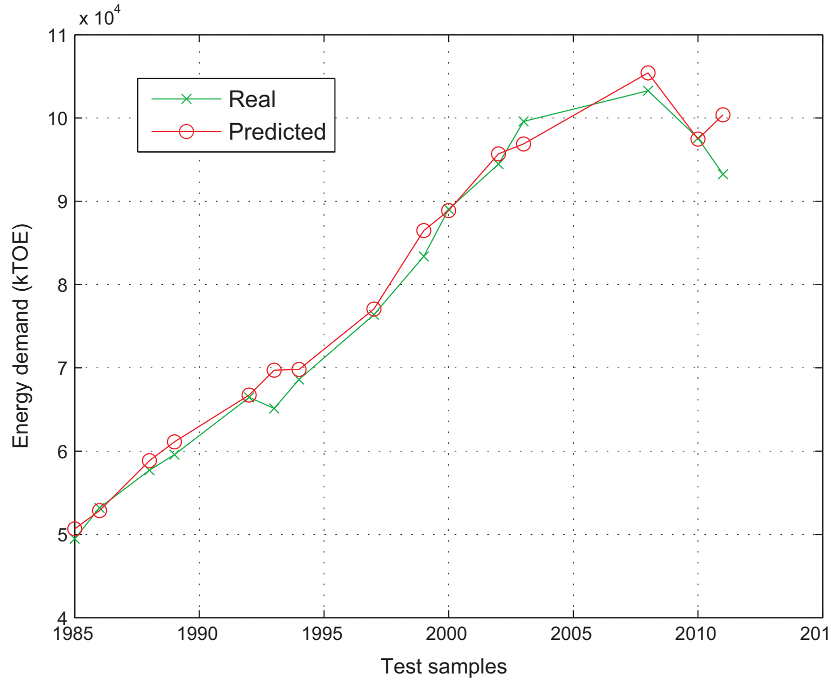


**Fig. 4.** Real energy demand versus prediction with the HS (7 input variables and 15 values in $\mathcal{W}$).

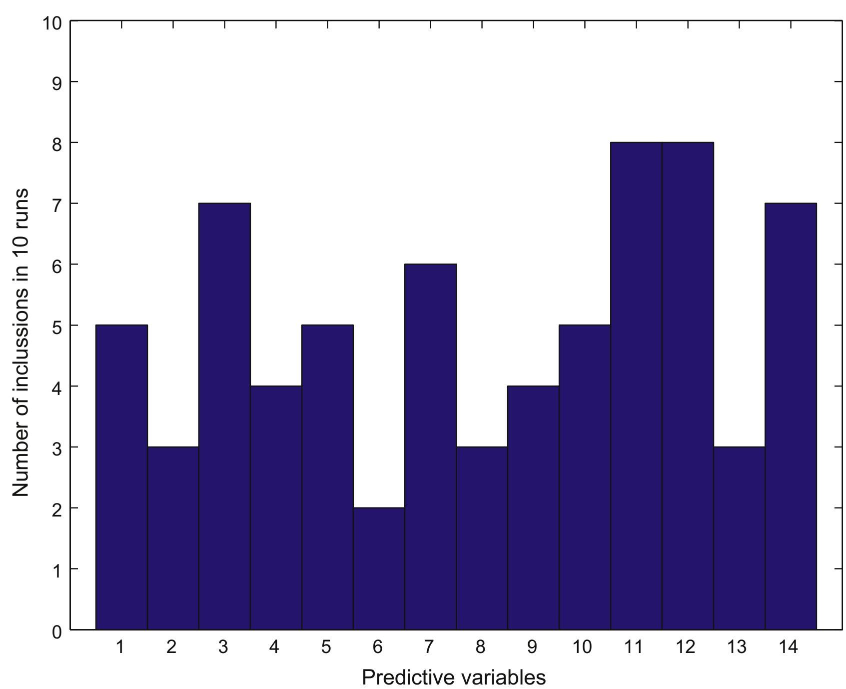


**Fig. 5.** Histogram obtained from 10 runs of the HS approach considering 7 input variables.

**Table 4**
Relative MAE (%) obtained with the proposed ELM using different number of input variables, for the problem of energy demand estimation in Spain.

| Number input variables | Relative MAE best in (%) | Relative MAE average of 10 runs in (%) | Selected variables |
|---|---|---|---|
| 4 | 2.78 | 3.27 | 1, 3, 7, 10 |
| 7 | **2.16** | 3.72 | 1, 2, 3, 7, 8, 9, 12 |
| 14 | 2.93 | 4.43 | All |

Boldface value stands for the best MAE obtained.

demand estimation (variables 3 and 7, Exports and Electricity production), whereas the other two are different, and in this case of $CO_2$ emissions prediction are related to energy imports and electric power consumption. The results obtained by the ELM in the $CO_2$ emission estimation problem are shown in Table 6. As can be seen in this case the ELM is able to obtain better results than the HS an GA counterparts. This table also shows the ELM performance when the best 4 input variables selected by the HS are used: an improvement of the network performance is obtained, with a best result of 4.75% of error in terms of MAE.

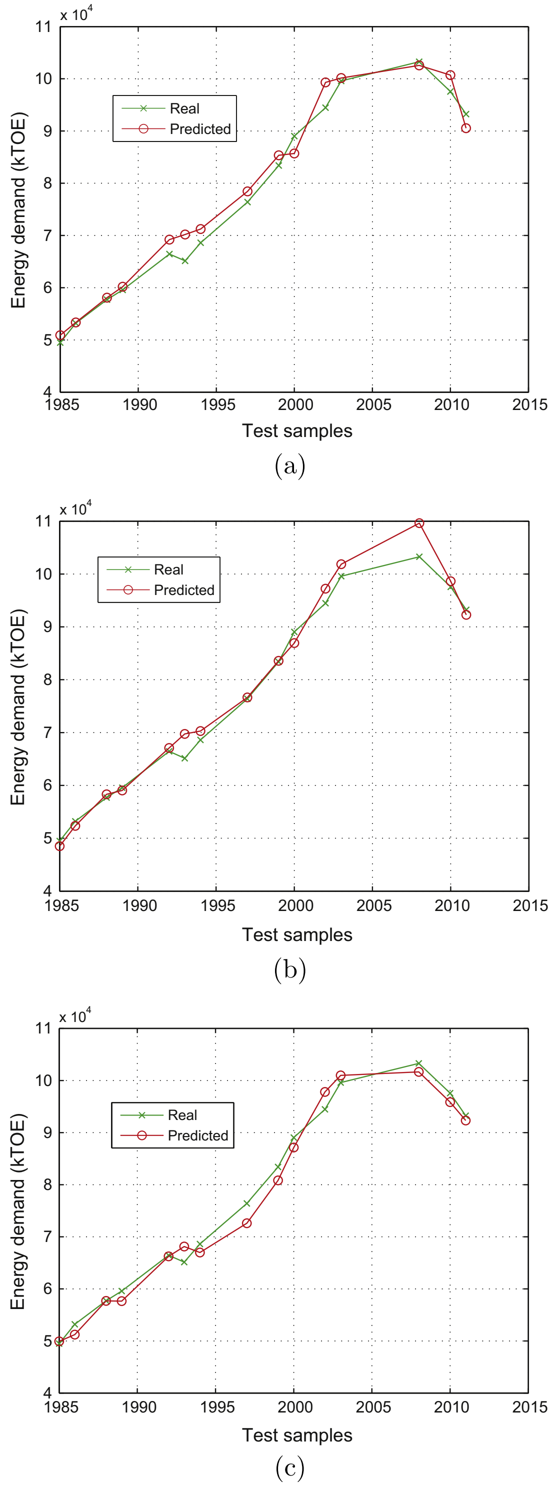


**Fig. 6.** Real energy demand versus prediction in the test set considered, with the ELM and different number of predictive variables: (a) 4 variables; (b) 7 variables; (c) 14 variables.

Fig. 7 depicts the prediction obtained with the proposed HS, ElM with all the variables and ELM with feature selection using the HS algorithm. As can be seen, all the models discussed in this paper are able to catch the trend of the $CO_2$ emissions, though the ELM is able to obtain the best result. However, the performance of the algorithms in this problem is worse than that of these approaches in the case of the Energy demand estimation.

**Table 5**
Comparison of the relative MAE (%) obtained with the proposed HS and the algorithm in [3] for the problem of $CO_2$ emissions estimation in Spain.

| Algorithm | Relative MAE best in (%) | Relative MAE average of 10 runs in (%) | Selected variables (best run) |
|---|---|---|---|
| HS with FSP | 5.70 | 6.18 | 3, 7, 9, 11 |
| GA from [3] | 6.15 | 7.21 | 1, 2, 3, 4 |

**Table 6**
Relative MAE (%) obtained with the proposed ELM with the best 4 variables selected by the HS and with all variables considered (without feature selection), for the problem of $CO_2$ emissions estimation in Spain.

| Algorithm | Relative MAE in (%) (4 input variables) | Relative MAE in (%) all variables |
|---|---|---|
| ELM | 4.75 | 5.23 |

### 4.2. Final discussion: extension of the proposed algorithms to alternative problems

The approaches presented in this paper can be extended to different problems with similar characteristics as the prediction of energy demand or $CO_2$ tackled in this work. First, there are many prediction problems in which the predictive variables are related to macroeconomic structure and situation of countries. In this sense, the algorithms proposed in this paper can be of general application to these problems. For example, the prediction of the companies' revenue in different sectors (energy, telecommunications, insurance, etc.) can be modelled with this approach, since they strongly depend on the macroeconomic situation of a country. The proposed approach can be also useful to evaluate the effect of constructing new renewable energy facilities (wind or solar farms) at a national level, in terms of their effect in the countries' economy. Note that the fact that the algorithm includes a part of feature selection and the model's parameters in the same encoding opens the possibility of extending its application to some more specific problems (not so dependent on macroeconomic variables) such as wind or solar resource prediction. The outcome of the HS may then be included in a different regression such as an ELM in the same way we proceed here for obtaining more accurate results. One specific issue to be tackled is the inclusion of different number of features in the HS encoding. As has been mentioned before in this paper, it is a difficult point, since the number of parameters in the prediction model depends on the number of features, so it will imply to manage variable-length individual within the HS. The extension of the proposed approach to alternative problem may need to solve this issue, by considering modified implementations of the HS to cope with variable length encodings. It will provide a more general approach, able to explore a larger number of possibilities in terms of the best features for a given problem. Finally, let us give a brief note about the performance of the proposed algorithms in problems where the objective variable dramatically changes among different years. It is not the case of the problems at hand (energy demand nor $CO_2$ prediction) where the variation from one year to the following is moderate. Thus, in these problems the proposed algorithms are able to catch the trend in the prediction very well. However, in alternative problems, such as the prediction of telecommunication revenue for example, the differences of revenue from one year to the following may be dramatically different. In this case, the proposed algorithm may have problems to catch the trend of the objective function in the prediction, and some data treatment would be needed (such as objective data transformation with a non-linear filter in order to make it smoother).

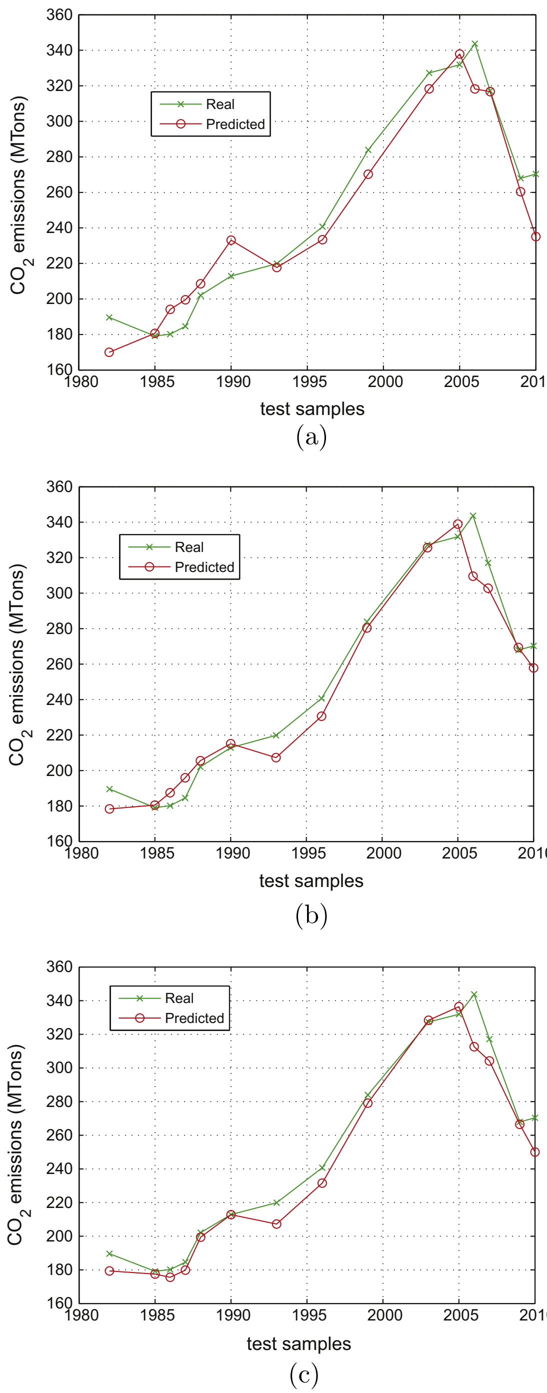


**Fig. 7.** $CO_2$ emissions estimation with the proposed HS and ELM: (a) HS (4 input variables and 9 values in $\mathcal{W}$); (b) ELM without feature selection; (c) ELM with the 4 input variables form the HS.

## 5. Conclusions

In this paper we have presented two different approaches to estimate one-year-ahead energy demand from macroeconomic variables. Specifically, we have introduced a new algorithm based on a modified Harmony Search (HS) algorithm, in which several improvements have been included to enhance its performance. Unlike previous approaches to the same application, we consider a large number of predictive variables along with a feature selection mechanism, which is realized by including a binary part in the solution encoding of the HS solver. We have also shown the performance of an Extreme Learning Machine neural network in the same application scenario, when incorporating the aforementioned HS-based feature selection mechanism. In order to evaluate the performance of the described approaches we have utilized real macroeconomic data of Spain from the last 30 years for predicting the energy demand of this country. The comparison is carried out against an existing approach (genetic algorithm) with four fixed predictive variables (i.e. without feature selection). The obtained results show that the proposed approaches are able to obtain very accurate energy demand predictions – outperforming those from previous works – even in years of crisis, where there is a change of trend in the energy consumption of the country. Finally, we have extended the experimental part of the paper by tackling a similar problem, in this case one-year ahead $CO_2$ emissions estimation from macro-economic variables, where it has also been shown the good performance of the proposed approaches in these kind of problems.

## Acknowledgements

This work has been partially supported by the Spanish Ministry of Science and Innovation under Project Number ECO2010-22065-C03-02.

## References

[1] Suganthi L, Samuel AA. Energy models for demand forecasting – a review. Renew Sustain Energy Rev 2012;16:1223–40.
[2] CSIRO and the Natural Edge Project. Energy transformed: sustainable energy solutions for climate change mitigation; 2007. p. 6.
[3] Ceylan H, Ozturk HK. Estimating energy demand of Turkey based on economic indicators using genetic algorithm approach. Energy Convers Manage 2004;45:2525–37.
[4] Ünler A. Improvement of energy demand forecasts using swarm intelligence: the case of Turkey with projections to 2025. Energy Policy 2008;36:1937–44.
[5] Kiran MS, Özceylan E, Gündüz M, Paksoy T. Swarm intelligence approaches to estimate electricity energy demand in Turkey. Knowl-Based Syst 2012;36:93–103.
[6] Yu S, Zhu KJ. A hybrid procedure for energy demand forecasting in China. Energy 2012;37:396–404.
[7] Kiran MS, Özceylan E, Gündüz M, Paksoy T. A novel hybrid approach based on particle swarm optimization and ant colony optimization to forecast energy demand of Turkey. Energy Convers Manage 2012;53:75–83.
[8] Yu S, Wei YM, Wang K. A PSO-GA optimal model to estimate primary energy demand of China. Energy Policy 2012;42:329–40.
[9] Rager M, Gahm C, Denz F. Energy-oriented scheduling based on evolutionary algorithms. Comput Oper Res 2015;54:218–31.
[10] Yu S, Zhu K, Zhang X. Energy demand projection of China using a path-coefficient analysis and PSO-GA approach. Energy Convers Manage 2012;53:142–53.
[11] Piltan M, Shiri H, Ghaderi SF. Energy demand forecasting in Iranian metal industry using linear and nonlinear models based on evolutionary algorithms. Energy Convers Manage 2012;58:1–9.
[12] Geem ZW, Hoon Kim J, Loganathan GV. A new heuristic optimization algorithm: harmony search. Simulation 2001;76(2):60–8.
[13] Manjarres D, Landa-Torres I, Gil-Lopez S, Del Ser J, Bilbao MN, Salcedo-Sanz S, et al. A survey on applications of the harmony search algorithm. Eng Appl Artif Intell 2013;26:1818–31.
[14] Blum A, Langley P. Selection of relevant features and examples in machine learning. Artif Intell 1997;97:245–71.
[15] Weston H, Mukherjee S, Chapelle O, Pontil M, Poggio T, Vapnik V. Feature selection for SVMs. In: Advances in NIPS 12. MIT Press; 2000. p. 526–32.
[16] Kohavi R, John GH. Wrappers for features subset selection. Int J Digit Libr 1997;1:108–21.
[17] Geem ZW. Novel derivative of harmony search algorithm for discrete design variables. Appl Math Comput 2008;199(1):223–30.
[18] Geem ZW, Sim KB. Parameter-setting-free harmony search algorithm. Appl Math Comput 2010;217(8):3881–9.
[19] Salcedo-Sanz S, Camps-Valls G, Pérez-Cruz F, Sepúlveda-Sanchis J, Bousoño-Calzón C. Enhancing genetic feature selection through restricted search and Walsh analysis. IEEE Trans Syst Man Cybernet – Part C 2004;34(4).
[20] Wang L, Yang R, Xu Yin, Niu Q, Pardalos PM, Fei M. An improved adaptive binary harmony search algorithm. Inform Sci 2013;232:58–87.
[21] Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. Neurocomputing 2006;70(1):489–501.
[22] Huang GB, Wang DH, Lan Y. Extreme learning machines: a survey. Int J Mach Learn Cybernet 2011;2(2):107–22.

[23] Huang GB, Chen L. Convex incremental extreme learning machine. Neurocomputing 2007;70:3056–62.
[24] Huang GB, Chen L. Enhanced random search based incremental extreme learning machine. Neurocomputing 2008;71:3460–8.
[25] Huang GB, Chen L, Siew CK. Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Trans Neural Netw 2006;17(4):879–92.
[26] Meng M, Niu D, Shang W. A small-sample hybrid model for forecasting energy-related $CO_2$ emissions. Energy 2014;64:673–7.
[27] Pao HT, Fu HC, Tseng CL. Forecasting of $CO_2$ emissions, energy consumption and economic growth in China using and improved grey model. Energy 2012;40:400–9.