Expert Systems With Applications 92 (2018) 380–389

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Energy consumption forecasting based on Elman neural networks with evolutive optimization

L.G.B. Ruiz*, R. Rueda, M.P. Cuéllar, M.C. Pegalajar

Department of Computer Science and Artificial Intelligence, University of Granada, c/ Periodista Daniel Saucedo Aranda, s.n., 18071 Granada, Spain

ARTICLE INFO

Article history:
Received 12 July 2017
Revised 6 September 2017
Accepted 23 September 2017
Available online 29 September 2017

Keywords:
Energy efficiency
Neural networks
Time series forecasting
Evolutionary algorithm

ABSTRACT

Buildings are an essential part of our social life. People spend a substantial fraction of their time and spend a high amount of energy in them. There is a grand variety of systems and services related to buildings, in order to better control and monitoring. The prompt taking of decisions may prevent costs and contamination. This paper proposes a method for energy consumption forecasting in public buildings, and thus, achieve energy savings, in order to improve the energy efficiency, without affecting the comfort and wellness. The prediction of the energy consumption is indispensable for the intelligent systems operations and planning. We propose an Elman neural network for forecasting such consumption and we use a genetic algorithm to optimize the weight of the models. This paper concludes that the proposed method optimizes the energy consumption forecasting and improves results attained in previous studies.

1. Introduction

Energy efficiency is an area of increasing importance because of the rising cost of energy and growing environmental concerns. The building sector consumes one-third of the global consumption and records close to 40% of global CO_2 emissions (International Energy Agency, 2013). The operations of building systems produce approximately 90% of these pollutant gases, in order to maintain a certain comfort level, closely associated with the heating, cooling and lighting systems. The worries over energy consumption and related noise, light pollution and safety problems have motivated the sustainable building facilities by means of construction practices and context-sensitive design (Olubunmi et al., 2016).

Due to the increased awareness of environmental issues and energy security, building regulations and polices related to new and refurbished building have been established in many countries: USA with the LEED —Leadership in Energy and Environmental Design— and BREEAM —Building Research Establishment Environmental Assessment Methodology— and Europe with the Energy Performance of Buildings Directive (Lord et al., 2016). Architects, planners and engineers are increasingly requiring that consider energy codes for minimizing environmental impact and resource consumption.

The most significant decisions linked to sustainable design are usually made in the early design stages, determining its environmental impact and its energy costs (Basbagill et al., 2013). Commonly, energy analysis is habitually performed after the architectural design and related documents have been produced. This practice into the design process leads to an inefficient way of retroactively modifying the design afterwards to achieve a set of performance criteria (Jalaei & Jrade, 2014). Energy efficiency is a decisive quality in order to reach environmentally friendly buildings, and what's more, is an effective strategy for reducing energy consumption and related gas emissions, with the consequent economic savings this can represent.

In more recent years, the new sensing technologies are continually being developed and integrated in the most diverse environments (Ekwevugbe et al., 2013). These provide useful and descriptive information of the building if we know how to take advantage of the powerful data. Nevertheless, there is a marked diversity in the data flows, owing to its irregular and varied source, coming from heating, ventilation, air-conditioning and lighting systems obtaining information such as internal and external temperature, sound level, carbon-dioxide, energy consumed, intensity, maximum demand, lighting state, wind speed, wind direction, pressure, precipitation (Khosravani et al., 2016; Ruiz et al., 2016) or even occupancy (Balaji et al., 2013); making its treatment difficult. The monitoring systems offers a possibility of collecting and storing a vast quantity of data. Processing all this information is not a trivial undertaking, this task frequently requires the combination of

Abbreviations: ANN, Artificial Neural Network; ARIMA, Auto-Regressive Integrated Moving Average; ENN, Elman Neural Network; GA, Genetic Algorithm; LM, Levenberg–Marquadt; MA, Memetic Algorithm; MLP, Multilayer Perceptron Model; MSE, Mean Squared Error; NAR, Nonlinear Autoregressive Model; NARX, Nonlinear Autoregressive Model with Exogenous Inputs; NN, Neural Network; SVR, Support Vector Regression; UGR, University of Granada.

* Corresponding author.
E-mail addresses: bacaruiz@ugr.es (L.G.B. Ruiz), ramonrd@ugr.es (R. Rueda), manupc@decsai.ugr.es (M.P. Cuéllar), mcarmen@decsai.ugr.es (M.C. Pegalajar).

https://doi.org/10.1016/j.eswa.2017.09.059
0957-4174/

different datasets that might be not related a priori. The crucial need for analysing big amount of data has revolutionized the Machine Learning, Data Mining and Statistics using prediction, classification, regression, clustering and dimensionality reduction techniques (Balón-Canedo et al., 2017); many tools have been developed in genomics —enabling inexpensive and high-throughput measurement of the genome—, neuroscience —important diseases have been shown to be related to bran connectivity networks—, economics and finance —implementing specialized data analytics programs to identify key business insights that can be exploited to support better decision making—, social network —data analysis of data produced by Twitter, Facebook, LinkedIn and YouTube using these data to predict influenza epidemic or stock market trend— (Fan et al., 2014). By extension, to make an efficient use of energy, in view of achieving remarkable reductions in consumption and significant economic saving, becomes an important and challenging issue. Lately, specific applications for building efficient energy have been investigated, as is the case of the energy forecasting (Andrade & Bessa, 2017) and consumption patterns (Chou & Ngo, 2016).

Forecasting models for energy consumption furnish intelligence with in a building for improving energy use, cost saving and reducing environmental impact without the need to compromise on performance and comfort. Predictive management of a building system can reduce peak power demands which translates into energy savings (Dhillon et al., 2016).

An energy prediction model represents an essential role in smart buildings. It has been proven that a small increase in forecasting accuracy would save millions of dollars in operation costs (Bunn & Farmer, 1985). The time series prediction is habitually handled as a hard paradigm because there may be diverse influencing factors, like weather conditions, social and economic conditions. There is an abundant research literature focused on time series prediction, the most popular methods are collected by Palli's book (Pallit & Popovic, 2005), some examples are Regressive Models, Artificial Neural Network —ANN—, Trees, Fuzzy methods and Support Vector Regression.

Kaur and Sachin (Kaur & Ahuja, 2017) predict the electricity consumption using autoregressive moving average model —ARIMA—. Ma and Liu use the grey system —system with partial information known, it has two part: system with completely known information and system with completely unknown information (Julong, 1989)— theory to forecast the natural gas consumption of China (Ma & Liu, 2017). Simple and multiple linear regression is applied by Fumo and Rafe Bismas to predict energy consumption in family houses (Fumo & Rafe Biswas, 2015). Davlea and Teodorescu present a neuro fuzzy model to develop a middle-term load forecaster (Davlea & Teodorescu, 2016). Dhillon et al. employ Support Vector Regression —SVR— for short term load forecasting (Dhillon et al., 2016). The ANN is the mostly used machine learning method and present great results, such as the Adaptive Network Based Inference System model to forecast building energy consumption in a cold region of Ekici (Bektas Ekici & Aksoy, 2011), Rodger's study which uses the fuzzy logic coupled with regression, nearest neighbour and artificial neural network to create a predictive model to make predicting demand for natural gas and energy cost savings in public buildings (Rodger, 2014) and many other works (Benedetti et al., 2016; Egrioglu et al., 2016; Kanarachos et al., 2017; Pino-Mejías et al., 2017; Rodrigues et al., 2017).

However, the main disadvantage of ANN is its slow convergence and easy local minimum stagnation. This leads to the idea of using a technique to avoid these problems, and those are the Genetic Algorithm —GA— which is a global search and an optimization method. GA is widely used for optimizing models in time series forecasting for building energy consumption (Bhandari & Gill, 2016; Zhang, Deb, Lee, Yang, & Shah, 2016).

This paper is a straight continuation of a previous work (Ruiz et al., 2016), and it proposes a method for predicting energy consumption by using ANN and the GA to improve the accuracy of these models. The main objective of this paper is to provide a methodology to analyse historical energy consumption, and perform the daily prediction with such models. Furthermore, a comparison is made between ANNs and identifies if the energy consumption forecasting can improve with external information or it depends entirely on historical consumption. This research has used data of the faculties, centres and schoolrooms of the University of Granada —UGR—. The energy management systems are relatively new in UGR, they have been implemented and introduced in the most recent years. However, sufficient data are available to carry out this work and analyse building's behaviours.

This paper deals with energy efficiency in public and distributed buildings analysing a new proposed model of ANN with previous forecasting methods for energy consumption applied.

The present paper is divided as follows: Section 2 presents the suggested methodology, the employed artificial neural networks and the description of the genetic algorithms. Section 3 describes the proposed system containing the data processing and noise treatment. Section 4 introduces experiments performed, the description of the real data used, parameters, results and discussion achieved. The paper ends with some practical implications and concluding comments.

2. Proposed system

The bulk of energy time-series modelling is represented in Fig. 1 where input is provided by building automation systems which stores all raw information in a database, and the output is forecasted consumption. The details of each component are outlined below.

A. Building automation system

Software characterized by a number of digital controllers, provides an asynchronous communication architecture for interacting with distributed building automation devices. It collects and presents building data, so that it can be interpreted.

B. Historical energy consumption database

The resources database is in charge of the information of the registered energy use. Besides, the database includes other extra knowledge from varying distributed sensors, such as power demand and temperature monitored. This constitutes a means for studying relations between energy and temperature, if exists.

C. Data treatment

Initially, database saves raw data which normally contains noise, incomplete, unreliable and missing data. It is important at this stage to transform the data to convert them to a suitable form. In this paper the time granularity energy consumption used is daily according previous works (Ruiz et al., 2016).

Tangible and solid sensors are the link between the real world and digital world. And these devices sometimes present failures due to broken device, transmission errors or any other issues caused by the impairment. To solve this question two method have been applied: a) Energy time series consumption is filtered with a moving average filter and a sliding windows technique to eliminate breaks or other irregular patterns in the data (Smith, 1997), according to the next Eq. (1), b) Linear interpolation based on the immediate neighbours values at grid points to fill missing values.

$$y(n) = \frac{1}{windowsSize} \cdot \Big(x(n) + x(n-1) + \ldots + x(n - (windosSize - 1))\Big) \tag{1}$$

Finally, the data are normalized between [0, 1] to standardize variables into same range using Eq. (2). And guaranteeing that

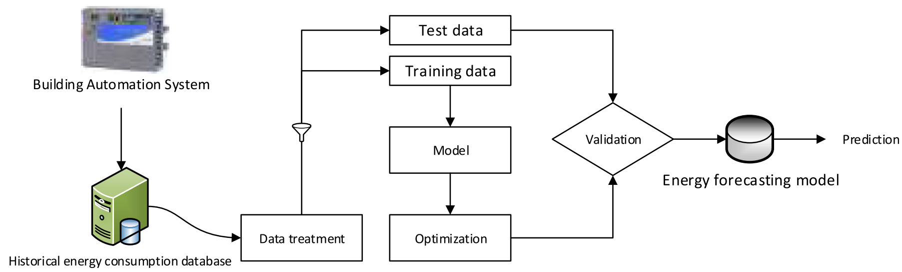

Fig. 1. Proposed system flowchart.

there are no attributes which are more important than others, and also eases a stable convergence of network weights and biases.

$$y_{normalized} = \frac{y - y_{min}}{y_{max} - y_{min}} \tag{2}$$

D. Train and test

The dataset is split into two sets, the training set is 70% and the testing set is the remaining 30%. Both are selected randomly from the total energy consumption available.

E. Model and optimization

To predict the energy consumption at a determined time, we use three kind of ANN: NAR, NARX and ENN. These models are trained using LM algorithm and hyperbolic tangent sigmoid transfer function for the hidden layers (Vogl et al., 1988). The learning stage is an iterative process. At this point, GA is performed to optimize ANN, these were all depicted in Section 3.3 (Genetic Algorithm). All parameters are explained in the experiments section.

F. Validation and model

The results are validated through comparison test data forecasting with the 30% of the data isolated before training. If the model's response is similar to test data, then it is assumed that the learning has been successful.

3. Methodology

The proposed methodology which follows has been developed in four stages. The first stage is data capture and preparation. Once the data have been compiled, the second stage is the ANN forecasting model. The next stage is genetic optimizing. And the final phase involves the analysis review and the use of the optimized ENN, achieving this through experimentation which appears in the next section.

The present study derives from previous researches done (Ruiz et al., 2016) utilizing two well-known models for energy consumption forecasting: the non-linear autoregressive neural networks —NAR— for modelling the data process of one dimensional time series using past values (Ferlito et al., 2015); and the non-linear autoregressive neural network with exogenous inputs —NARX including another external series which might provide relevant information (Cadenas et al., 2016). This paper proposed Elman Neural Network —ENN— and using genetic optimization to enhance preceding results. There are numerous studies in literature for solving time series prediction with neural networks (Benedetti et al., 2016; Bhandari & Gill, 2016; Egrioglu et al., 2016; Pino-Mejías et al., 2017).

These models can be listed in order of complexity as: 1) NAR. 2) NARX. 3) ENN. The latter two networks offer the possibility of incorporating extra information to enhance the forecasting accuracy. In the real world, not all buildings are able to stock large quantities of data or register, sometimes only the energy meter is saving. On other occasions, management systems handle more information such as temperature. The suggested models are appropriate for both strategies, and the aim of this work is to determine which model is the better choice for this problem.

Subsequent sections describe employed models and analyses its pros and cons to solve the problem of energy consumption forecasting.

3.1. NAR and NARX neural network

Time series are a sequence of data, observations or numerical values usually recorded at uniform time intervals (Brockwell & Davis, 2013). Typically measured every second, minute, hour, day, week or even each year; although another time interval is valid: every 30 seconds, 12 hours, etc. A time series associated to the variable Z over the time set T is denoted by:

$$Z = \{Z_t : t \in T\} \tag{3}$$

Where Z_t is the value of Z at time t. For example, the energy consumption of a building defines a time series, indicating in each instant t the consumption spent by the building.

In many instances, the data depend not only on the total amount spent but also on other possible influence factors. In order to tackle this issue, the NAR and NARX neural networks have proved to be a very helpful tool in time series environments (Cadenas et al., 2016; Wang et al., 2016). These models are a kind recurrent system which are able to learn by itself, improving the approximation of the ANN by reducing the output error. NAR uses past values for the actual time series to predict next values as determined by the following equation (Ibrahim et al., 2016):

$$y(t) = h(y(t-1),\ y(t-2),\ \ldots,\ y(t-p)) + \epsilon(t) \tag{4}$$

Where h is a nonlinear function which depends on p past values of the output y as shown in Fig. 2 excluding the exogenous input x —in grey colour— and ε represents random error sequence and independent distributed (Wang et al., 2016). In much the same way as NAR, NARX equation is defined as:

$$y(t) = h\Big(x(t-1),\ x(t-2),\ \ldots,\ x(t-p_x),\ y(t-1),\ y(t-2),\ \ldots,\ y(t-p_y)\Big) + \epsilon(t) \tag{5}$$

Which includes external input x of the neural network. It is zero in case of NAR model.

The basic structure of recurrent neural networks is presented in Fig. 2. There is an input layer with two time series. A hidden layer with p delays. And an output layer with an activation function. The Levenberg-Marquardt —LM— (Ampazis & Perantonis, 2000) method is used for optimizing the learning rate based on this gradient, which combines the local convergence properties of Gauss-Newton method near a minimum with the consistent error decrease provided by gradient descendent far away from a solution; this error is a mean squared error based on a learning sample. The LM procedure computes the Jacobian matrix of the error function which takes great use of memory. Nevertheless this computational cost is well worth because it increase the rate of convergence of the algorithm (Hagan & Menhaj, 1994).

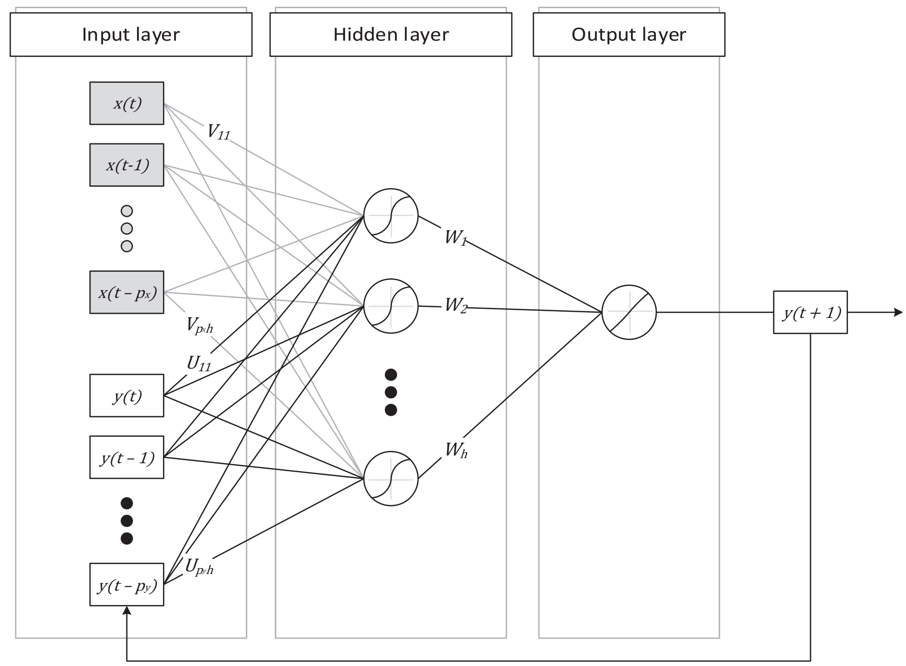

Fig. 2. Representation of the structure of non-linear autoregressive neural network —NAR (without grey part)— with exogenous input —NARX (with the grey part)—.

The NAR is used with one input —energy consumption at the previous time $\{y(t-1), \ldots, y(t-p)\}$— and one output —the predicted value $y(t)$—. Similarly, NARX network models the same energy consumption $\{y(t-1), \ldots, y(t-p)\}$ with the external input —temperature— given the p past values. The parameters of the neurons and delays are set in accordance with the best results achieved in antecedent studies (Ruiz et al., 2016). In the next sections, all these parameters are specified.

V_{ij} represents the weight between exogenous input i and the hidden neuron j, W_i is the weight between hidden neuron i and the output neuron and U_{ij} is the weight for the connection between input i and the hidden neuron j. p_x and p_y are the past values introduced of the exogenous and input series respectively. h is the number of neurons in the hidden layer.

3.2. Elman neural network

Due to the present problem involves working in historical data, *memory* is an essential feature to process temporal information. The Elman Neural Network —ENN— (Elman, 1990) introduces this concept of *memory*. Often, the past contains rich information and this is why it is stored in memory neurons. In ENN positive feedback is used to build this memory structure, and it is illustrated in Fig. 3. These specialized units are called *context* units or *state* units store preceding outputs of hidden layer by using a positive feedback mechanism. For each unit in the hidden layer an extra context unit is fully connected with all the hidden neurons in a forward mode. State neurons are only connected to internal nodes of the network, and not with the outside world.

The following equations defines the ENN with n inputs, h hidden neurons, and o outputs:

$$S_j(t) = f\left(\sum_{r=1}^{h}\sum_{j=1}^{h} U_{jr}S_r(t-1) + \sum_{i=1}^{n}\sum_{j=1}^{h} V_{ji}X_i(t)\right) \quad (6)$$

$$O_k(t) = g\left(\sum_{j=1}^{h}\sum_{k=1}^{o} W_{kj}S_j(t)\right) \quad (7)$$

Here, $S_j(t)$ is the output of hidden neuron $j \in [1, h]$, $X_i(t)$ is the input data to neuron $i \in [1, n]$, $O_k(t)$ is the output $k \in [1, o]$ at time t. U, V, W are matrix with the network's weights. Thus, V_{ji} is the weight for the connection between input neuron i and the hidden neuron j. U_{jr} is the weight between the recurrent connection r and the hidden neuron j. And W_{kj} is the weight between hidden neuron j and output neuron k. f and g are activation functions.

x_i is the input i, and the previous state in time $t-1$ of the hidden neuron j is represented by $q_j(t-\lambda)$ where $\lambda \geq 1$ is the number of previous hidden states stored.

The main dissimilarity between ENN and NAR(X) models is the recurrent link that appears from hidden layer to *Context or State Units*. In ENN, the *Context Units* store hidden neuron values at previous time step. However, there is another architecture whose *state* stores output neuron values at different past time: this is the Jordan networks (Jordan, 1997). The use of ENN in this study is supported by multiple works (Bao et al., 2016; Cuéllar et al., 2005; Cuéllar et al., 2007; Delgado et al., 2006a; Delgado et al., 2006b; Qin et al., 2016) adhering excellent results.

3.3. Genetic algorithm

Genetic Algorithms —GA— have shown outstanding degrees of success in task related to neural network training (Cuéllar et al., 2005; Cuéllar et al., 2007; Delgado et al., 2006aa, 2006b). And for this reason, GA are used in this study in order to improve the accuracy of ANN prediction, because an improvement of a few percentages in the forecasting accuracy would bring benefits worth large amounts of money (Sadat Hosseini & Gandomi, 2012).

In essence, GAs simulate the mechanics of biological evolution. Following the philosophy of the famous naturalist Charles Darwin, the GA is based on natural selection or best adapted survival. In nature, individuals must adapt to their environment through a process name evolution. This evolving keeps positive aspects of an individual over time, and features that undermine the chromosome was ruled out. *The genetic algorithm is a highly parallel mathematical algorithm that transforms a set —population—of individual mathematical objects, each with an associated fitness value, into a new population* —i.e., the next generation— *using operations pat-*

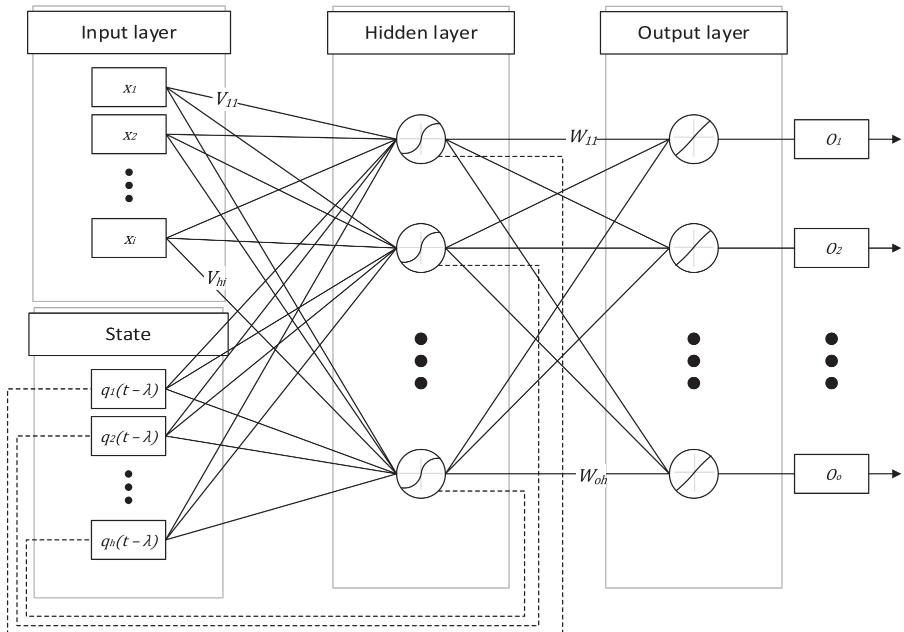

Fig. 3. Architecture of the Elman Neural Network.

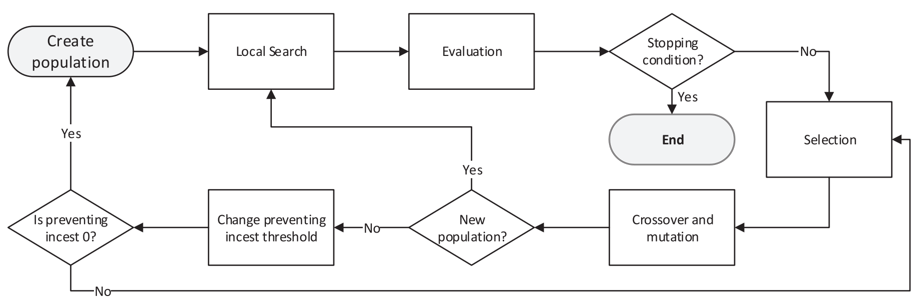

Fig. 4. General Flowchart of the memetic algorithm —MA— based on binary-coded CHC schema.

terned after the Darwinian principle of reproduction and survival of the fittest and after naturally occurring genetic operations —notably sexual recombination—. (Koza, 1992)

This study suggests the adaptation of the binary CHC —Cross generational elitist selection, Heterogeneous recombination, and Cataclysmic mutation— algorithm for a real-coded problem (Blanco et al., 2001; Cordón et al., 2006). The flowchar of the GA is shown in Fig. 4.

This algorithm strikes a balance between diversity and convergence thanks to an elitist selection of individuals, invest prevention and initialization procedure of the population (Eshelman, 1991). For a better understanding of the algorithm components, before, there is a need to describe in detail who is an individual in this population, and then define fitness function, selection operation and local search. The hybridization of genetic algorithm and local search leads to a new kind of evolutionary algorithm commonly known as memetic algorithm.

An individual represents the structure of a NN. It would be made up of all weight which compose the network. Following the structure presented in Fig. 4, an ENN with one input, one output and two hidden neurons storing its previous state $t-1$, Fig. 5 illustrates an example of NN encoding.

The algorithm starts creating a first set of individuals in a totally random way. Afterward, it performs a local search to improve individual characteristics based on LM method. Thereupon, all individual in the population are evaluated by the fitness function, assigning everyone a Mean Square Error —MSE— according the network's goodness of fit. If $\hat{y}$ is a vector of m predictions, and y is the vector of observed values corresponding to the inputs, then the MSE can be estimated by Eq. (8):

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(\widehat{y_i} - y_i)^2 \tag{8}$$

If the stopping criteria is false, individuals who are going to cross are selected, usually known as parents. The incest prevention enables CHC to delay premature convergence: distance between the original parents must be exceed a certain limit, this limit is called as incest threshold. The Hamming distance is used in the original CHC. In this case, it makes no sense for real-coded problems, because weights of the NN are real values. Euclidean distance has been adopted instead, due to its widespread use and remarkable resolution (Han et al., 2015).

When the crossover operator was able to build a new population, then a new local search is performed for these fresh individuals. The recombination of the ANN weights is illustrated in Fig. 6. And the process repeats itself.

Otherwise, incest threshold is decreased, in favour of further crossing. At that time, parents are selected once again. A re-initialization is performed when the incest limit reaches zero, because the algorithm has fallen into local minimum. To gain an understanding of CHC algorithm, see the Eshelman's work (Eshelman, 1991).

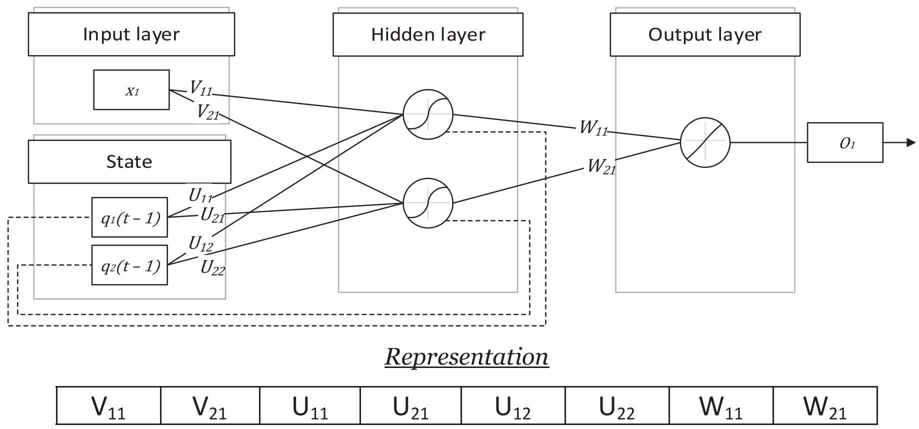

Fig. 5. Encoding Elman Neural Network architecture for the Genetic Algorithm with one input, two hidden neurons, one output and one delay in the memory state.

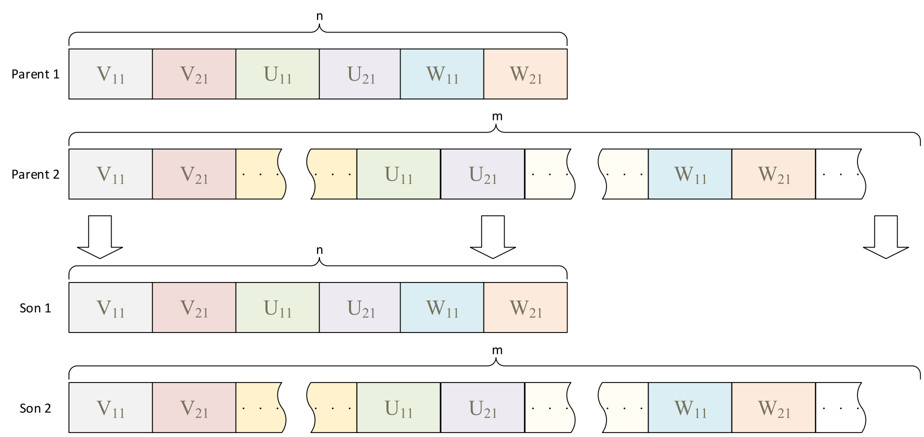

Fig. 6. Example of the crossover procedure for two networks with different sizes.

The crossover procedure combines two parents and creates two children. If both parents contain that gene, the genetic recombination of this gene is carried out based on BLX-α operator, which combines two parents p^1 and p^2 to generate offspring s by sampling a new value in the range $[min_i - I \bullet \alpha,\ max_i + I \bullet \alpha]$ at each gene i. Where min_i and max_i are smaller and larger parent values at position i. I is $max_i - min_i$. And α value has been set to 0,2 (Picek et al., 2013). The remaining connections of the largest son are directly inherited from the biggest parent.

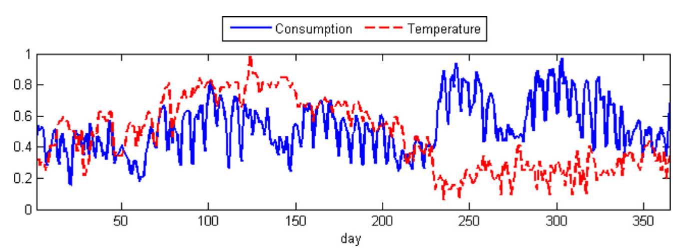

Fig. 7. Example of normalized consumption and temperature recorded during one year.

4. Experiments

4.1. Dataset

This work uses collected data from energy consumption of several University of Granada's building (UGR, Granada, Andalucía, Spain), such as classrooms, lecture rooms, laboratories and research centres. UGR owns a smart management system in its distributed facilities. It collects and monitored information of the building in real time from diverse sensors, for the purpose of being analysed and better understanding buildings behaviours.

Because of management systems implementation is relatively new, not all building dispose the same sensors. Therefore, not all buildings record the same information. Most of them collect energy consumption and climatic data. The UGR is made up of five campuses: *Centro, Cartuja, Fuentenueva, Aynadamar and Ciencias de la Salud*. These campuses are spread around different places of the city. Autonomous cities of Melilla and Ceuta also contain UGR's centres, placed in separate campus. As a whole, the UGR is assembled of 22 faculties, 5 schools, 8 training centres and 5 culture, sport and service centres. Due to Data Protection Act this study cannot reveal details over the facilities, and hence, buildings consumptions are labelled with a number. Eight edifices have been chosen, two representative building from the campuses.

Fig. 7 presents the temperature and consumption pattern during 1 year. Both series have been normalized because they have distinct units. Energy consumption is recorded in kW, and temperature is in degrees Celsius. This picture gives evidence of a peculiar behaviours, for example, highest consumption is performed at lowest temperatures, securely caused by air conditioners and heating. For this reason, it is interesting to include temperature in our predictive models. But on the other hand, incorporating too many variables, makes the model much more complicated and inserts uncertainty into the system because it would depend on known variables. The temperature is a sustainable dependence thanks to its smooth and regular behaviour.

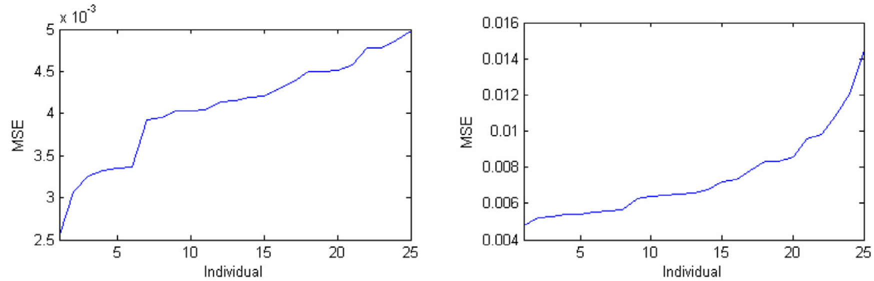

Fig. 8. Two examples of Mean Square Error of each individual of a population once GA have finished.

We study models without temperature too, in order to have a possible alternative and how good it was compared to previous predictors and the new proposed.

4.2. Parameters

This section summarizes the parameters used for each model, and may safely be skipped by readers who are easily bored.

The number of neurons for NAR and NARX models have been gathered from the best results obtained in preceding study (Ruiz et al., 2016). In summary: 14 neurons for the buildings 1 and 4; 12 neurons for the buildings 2, 3 and 7; 9 neurons for buildings 5, 6 and 8. These parameters have been taken to allow the comparison between outcomes with the same structure.

In the ENN case, it is not necessary establish the same number of neurons as NAR(X) because it does not follow the same architecture as NAR and NARX networks. The best number of neurons and delay have been set by testing experimentally different sizes and using cross-validation: 10 hidden neurons have been set and it has a memory of 5 past values.

The training function used in all cases is the LM backpropagation optimization. Minimum gradient is 10^{-7}, the training gain, μ, is a parameter which measures the adapting and learning rate of the model, its range is $[10^{-3}, 10^{10}]$ and the μ decrease and increase ratio are 0.1 and 10 respectively. The nets stop training at many different epochs and use validation sets to select the best model. There is an important aspect which should be considered with the ANN: a low number of hidden neurons was used to avoid overfitting to obtain a model as simple as possible, but too few neurons may lead to a negligent and ineffective learning. Furthermore, cross validation is performed in order to deal with this issue.

The population size has been set to 25 individuals. The error of each individual is illustrated in Fig. 8, it shows the corresponding error curves for two executions, and in both cases the difference between the best and the worst individual is more than double. Therefore, to increase the population size makes the GA works with worse models and to decelerate the optimization process. The interval value for a gene is $[-10, 10]$, stop criteria is set in 100 generations, crossover and mutation probability are 90% and 10%, respectively.

4.3. Results and discussion

The proposed method was programmed in Matlab software run on Intel® Core™ i7-6700 CPU @ 3.40 GHz. In order to confirm the robustness of the achieved results, for each experimental the simulation was run 5 times. Table 1 gathers the outcomes achieved by using GA and previous ones. The first column lists the buildings by id, remember that each building has been selected as specified in Section 4.1 Dataset. Second column relates to the MSE obtained with NAR models optimized by GA + LM. The next column illustrates the errors of NARX models —with temperature— and improved by GA + LM. In the last two columns, the table indicate the results of the new suggested networks, without and with temperature respectively.

In all buildings, previous results are worse than new ones. Indeed, in each of these cases, they are well above the worst of the proposed model, ENN. This table illustrates the excellent performance of the GA + LM. This provides an optimization for both models, non-autoregressive and Elman. Thus, after considering the results, it may be concluded that the adapted CHC algorithm is a good method for neural network optimization.

Besides, in order to facilitate comprehension and clarity, Fig. 9, illustrates a comparison between all models. The graphic views provide a quick overview of the improvements. The proposed method achieves a significant enhancement whichever model. In all but one case, the models are ranked as follows: 1) the ENN with temperature, 2) ENN without temperature, 3) NAR network and 4) NARX network with temperature. On the one hand, considering NAR and NARX networks only, the results are satisfactory, optimized networks are much better than previous one. These models correspond to an average of 35% improvement, with a 16% in the worst case and a 52% in the best case, acquired with NAR model and building 6. On the other and, the new proposed network, ENN, is able to obtain an even better fit. With an average of 61% improvement, 51% in the worst case and up to 82% in the best one. The propounded method with ENN provides much better results for all probe sets and considerably higher average score, compared to all other models.

Fig. 10 depicts an example of the forecasted consumption with all model used. The blue series is the original data normalized, so it is the desired value which models should be adjusted. The grey sequence is the response of the best models reached in previous work. Yellow and orange series concern to optimized NAR and NARX respectively thanks to GA + LM. And the green crosses are the Elman forecasting. This figure does not show two different series with and without temperature ENN because the variations are imperceptible and would make it difficult for the graph display. As shown in that picture all forecasted models follows the trend of the data quite well. However, the most faithful of the real data is Elman series which accurately predicts future values. And the worst is the non-optimized model with GA + LM of the preceding study where there are a few values which are more separated from the original data, for example in the days 25, 35, 60 and 77. The NAR and NARX models offer similar behaviour. Indeed, Fig. 9 details how optimized NAR models achieve better MSE than NARX with temperature. This is not the case of ENN where the predictor using temperature enhance a bit closer.

5. Conclusion

In this paper, we have introduced a new methodology to energy consumption forecasting and achieve optimum models. The GA has proven to be a useful and a key factor for optimizing ANN, and it helps to significantly enhance in NAR and NARX models too, used in previous works. Besides, the ENN have been very effective and

Table 1
Results of 8 buildings including results of previous works and the present method. Best results in bold.

Building	Previous results	NAR	NARX temperature	Elman	Elman temperature
1	0.018200	0.012060	0.011442	0.007585	**0.005288**
2	0.014200	0.007681	0.008305	0.005203	**0.004698**
3	0.013000	0.009608	0.010558	0.005706	**0.005439**
4	0.017000	0.009736	0.010079	0.006966	**0.006604**
5	0.006700	0.005173	0.005598	0.003266	**0.002627**
6	0.006200	0.002953	0.003324	0.001376	**0.001059**
7	0.013500	0.007535	0.007998	0.006258	**0.005660**
8	0.009300	0.006830	0.007162	0.004320	**0.003929**

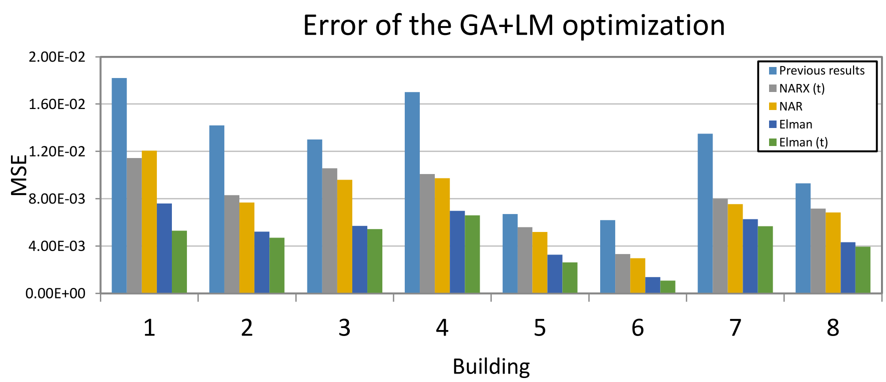

Fig. 9. Comparative evaluation between preceding outcomes and new optimized models by using MA.

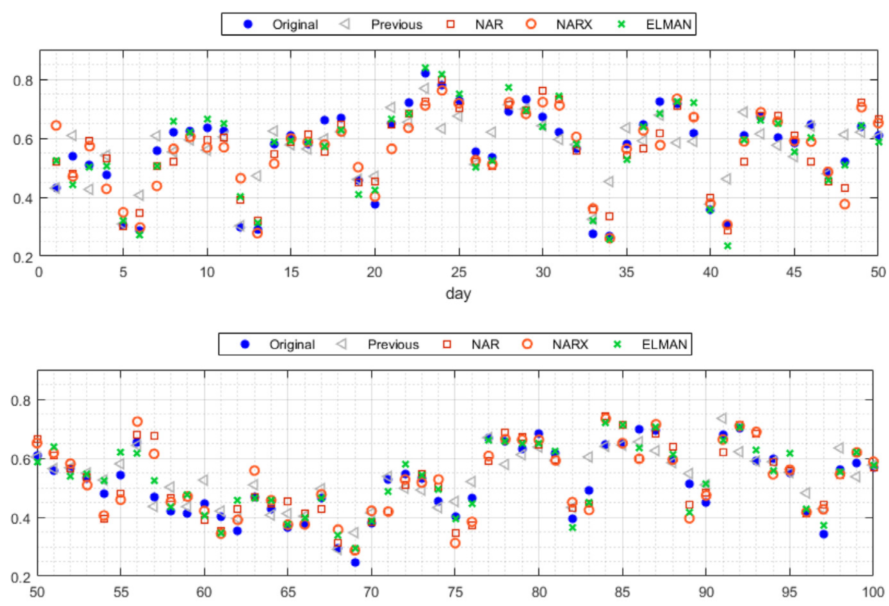

Fig. 10. Forecasted and original values for energy time series consumption for 100 days. The first 50 days in the above graphic and the 50 last days in the chart below.

it has demonstrated to be the best network in all test performed, obtaining an average improvement of 61%.

The main advantage in using NAR and NARX networks is their simplicity. However, this advantage limits its accuracy. Likewise, the major problem of the ENN lies in its complexity, because increasing components of a neural network —that is to say: including memory layer— implies increasing the number of connections, and the results is a much more complex model. This problem is known as the «curse of dimensionality», a kwell-know problem in statistical learning, this expression is used in phenomena that appear with high-dimensional data, and that have most often unfortunate consequences on the behaviour and performances of learning algorithms (Korn et al., 2001). In this respect, it is better to seek a compromise between predictor model complexity and an acceptable level of error in the results. Although, a complex model maybe is not an important variable to consider if it can save a lot of money.

In our approach, we assume that each building has a device integrated to capture energy consumption and store it. Sometimes, external data, such as temperature, is not available, because it depends on the sensors implemented and the building's budget. Thus, this study works this two approaches which lend support to both cases, achieving a good degree of MSE: 0.005085 and 0.004413 for models without temperature and including temperature respectively. Given the importance of relationship between current and past data, further studies will focus upon developing a system to find time relations in the building consumption by using clustering methods.

Acknowledgments

This work has been developed with the support of the Department of Computer Science and Artificial Intelligence of the University of Granada, TIC111, and the project TIN201564776-C3-1-R.

Bibliography

Ampazis, N., & Perantonis, S. J. (2000). Levenberg–Marquardt algorithm with adaptive momentum for the efficient training of feedforward networks. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium: Vol. 1* (pp. 126–131). vol. 121. http://dx.doi.org/10.1109/IJCNN.2000.857825.

Andrade, J. R., & Bessa, R. J. (2017). Improving Renewable Energy Forecasting with a Grid of Numerical Weather Predictions. *IEEE Transactions on Sustainable Energy, 8*(4). 1–1. http://dx.doi.org/10.1109/TSTE.2017.2694340.

Balaji, B., Xu, J., Nwokafor, A., Gupta, R., & Agarwal, Y. (2013). Sentinel: Occupancy based HVAC actuation using existing WiFi infrastructure within commercial buildings. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems* (pp. 1–14). ACM.

Balón-Canedo, V., Remeseiro, B., Sechidis, K., Martinez-Rego, D., & Alonso-Betanzos, A.. *Algorithmic challenges in Big Data analytics.* http://www.cs.man.ac.uk/~sechidik/publications/2017/esann2017.pdf.

Bao, G., Lin, Q., Gong, D., & Shao, H. (2016). Hybrid Short-term Load Forecasting Using Principal Component Analysis and MEA-Elman Network. In D.-S. Huang, K. Han, & A. Hussain (Eds.), *Intelligent computing Methodologies: 12th international Conference, ICIC 2016, Lanzhou, China, august 2-5, 2016, Proceedings, part III* (pp. 671–683). Cham: Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-42297-8_62 .

Basbagill, J., Flager, F., Lepech, M., & Fischer, M. (2013). Application of life-cycle assessment to early stage building design for reduced embodied environmental impacts. *Building and Environment, 60*, 81–92. http://dx.doi.org/10.1016/j.buildenv.2012.11.009.

Bektas Ekici, B., & Aksoy, U. T. (2011). Prediction of building energy needs in early stage of design by using ANFIS. *Expert Systems with Applications, 38*, 5352–5358. http://dx.doi.org/10.1016/j.eswa.2010.10.021.

Benedetti, M., Cesarotti, V., Introna, V., & Serranti, J. (2016). Energy consumption control automation using Artificial Neural Networks and adaptive algorithms: Proposal of a new methodology and case study. *Applied Energy, 165*, 60–71. http://dx.doi.org/10.1016/j.apenergy.2015.12.066.

Bhandari, R., & Gill, J. (2016). An Artificial Intelligence ATM forecasting system for Hybrid Neural Networks. *International Journal of Computer Applications, 133*, 13–16. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.734.7389&rep=rep1&type=pdf.

Blanco, A., Delgado, M., & Pegalajar, M. C. (2001). A real-coded genetic algorithm for training recurrent neural networks. *Neural Networks, 14*, 93–105. http://dx.doi.org/10.1016/S0893-6080(00)00081-2.

Brockwell, P. J., & Davis, R. A. (2013). *Time series: theory and methods*. Springer Science & Business Media. *http://dx.doi.org/10.1007/978-1-4419-0320-4.*

Bunn, D., & Farmer, E. D. (1985). *Comparative models for electrical load forecasting.* John Wiley and Sons Inc., New York, NY.

Cadenas, E., Rivera, W., Campos-Amezcua, R., & Cadenas, R. (2016). Wind speed forecasting using the NARX model, case: La Mata, Oaxaca, México. *Neural Computing and Applications, 27*, 2417–2428. http://dx.doi.org/10.1007/s00521-015-2012-y.

Cordón, O., Damas, S., & Santamaría, J. (2006). Feature-based image registration by means of the CHC evolutionary algorithm. *Image and Vision Computing, 24*, 525–533. http://dx.doi.org/10.1016/j.imavis.2006.02.002.

Cuéllar, M., Delgado, M., & Pegalajar, M. (2005). Multiobjective evolutionary optimization for Elman recurrent neural networks, applied to time series prediction. *Fuzzy Economic Review, 10*, 17–33. http://search.proquest.com/docview/228956321.

Cuéllar, M. P., Delgado, M., & Pegalajar, M. C. (2007). Topology optimization and training of recurrent neural networks with pareto-based multi-objective algorithms: A experimental study. In F. Sandoval, A. Prieto, J. Cabestany, & M. Graña (Eds.), *Computational and ambient Intelligence: 9th international work-conference on artificial neural Networks, IWANN 2007, san Sebastián, Spain, june 20-22, 2007. Proceedings* (pp. 359–366). Berlin, Heidelberg: Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-540-73007-1_44.

Chou, J.-S., & Ngo, N.-T. (2016). Time series analytics using sliding window metaheuristic optimization-based machine learning system for identifying building energy consumption patterns. *Applied Energy, 177*, 751–770. http://dx.doi.org/10.1016/j.apenergy.2016.05.074.

Davlea, L., & Teodorescu, B. (2016). A neuro-fuzzy algorithm for middle-term load forecasting. In *2016 International Conference and Exposition on Electrical and Power Engineering (EPE)* (pp. 005–009). http://dx.doi.org/10.1109/ICEPE.2016.7781292.

Delgado, M., Pegalajar, M. C., & Cuéllar, M. P. (2006a). Evolutionary training for dynamical recurrent neural networks: An application in finantial time series prediction. *Mathware and Soft Computing, 13*, 89–110. http://eudml.org/doc/41874.

Delgado, M., Pegalajar, M. C., & Cuéllar, M. P. (2006b). Memetic evolutionary training for recurrent neural networks: An application to time-series prediction. *Expert Systems, 23*, 99–114. http://dx.doi.org/10.1111/j.1468-0394.2006.00327.x.

Dhillon, J., Rahman, S. A., Ahmad, S. U., & Hossain, M. J. (2016). Peak electricity load forecasting using online support vector regression. In *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)* (pp. 1–4). http://dx.doi.org/10.1109/CCECE.2016.7726784.

Egrioglu, E., Bas, E., Aladag, C. H., & Yolcu, U. (2016). Probabilistic fuzzy Time series method based on artificial neural network. *American Journal of Intelligent Systems, 6*, 42–47. http://dx.doi.org/10.5923/j.ajis.20160602.02.

Ekwevugbe, T., Brown, N., Pakka, V., & Fan, D. (2013). Real-time building occupancy sensing using neural-network based sensor network. In *2013 7th IEEE International Conference on Digital Ecosystems and Technologies (DEST)* (pp. 114–119). http://dx.doi.org/10.1109/DEST.2013.6611339.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211. http://dx.doi.org/10.1207/s15516709cog1402_1.

Eshelman, L. J. (1991). The CHC adaptive search algorithm: how to have safe search when engaging in nontraditional genetic recombination. In R. Gregory J. E (Ed.). In *Foundations of genetic algorithms: Volume 1* (pp. 265–283). Bloomington Campus, Indiana, USA: Elsevier. http://dx.doi.org/10.1016/B978-0-08-050684-5.50020-3.

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National Science Review, 1*, 293–314. http://dx.doi.org/10.1093/nsr/nwt032.

Ferlito, S., Atrigna, M., Graditi, G., Vito, S. D., Salvato, M., & Buonanno, A. (2015). Predictive models for building's energy consumption: An Artificial Neural Network (ANN) approach. In *2015 XVIII AISEM Annual Conference* (pp. 1–4). http://dx.doi.org/10.1109/AISEM.2015.7066836.

Fumo, N., & Rafe Biswas, M. A. (2015). Regression analysis for prediction of residential energy consumption. *Renewable and Sustainable Energy Reviews, 47*, 332–343. http://dx.doi.org/10.1016/j.rser.2015.03.035.

Hagan, M. T., & Menhaj, M. B. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks, 5*, 989–993. http://dx.doi.org/10.1109/72.329697.

Han, X., Liang, Y., Li, Z., Li, G., Wu, X., & Wang, B. (2015). An Efficient genetic algorithm for optimization Problems with time-consuming fitness Evaluation. *International Journal of Computational Methods, 12*, 1350106. http://dx.doi.org/10.1142/s0219876213501065.

Ibrahim, M., Jemei, S., Wimmer, G., & Hissel, D. (2016). Nonlinear autoregressive neural network in an energy management strategy for battery/ultra-capacitor hybrid electrical vehicles. *Electric Power Systems Research, 136*, 262–269. http://dx.doi.org/10.1016/j.epsr.2016.03.005.

International Energy Agency.. *Transition to sustainable building. Strategies and opportunities to 2050* http://www.iea.org/publications/freepublications/publication/transition-to-sustainablebuildings.html.

Jalaei, F., & Jrade, A. (2014). An automated BIM model to conceptually design, analyze, simulate, and assess sustainable building projects. *Journal of Construction Engineering, 2014*, 21. http://dx.doi.org/10.1155/2014/672896.

Jordan, M. I. (1997). Serial Order: A parallel distributed processing approach. *Advances in Psychology, 121*, 471–495. http://dx.doi.org/10.1016/S0166-4115(97)80111-2.

Julong, D. (1989). Introduction to grey system theory. *The Journal of Grey System, 1*, 1–24. https://pdfs.semanticscholar.org/3013/67cf1d3f0aa5248a5b8a9ffa4813ef861620.pdf.

Kanarachos, S., Christopoulos, S.-R. G., Chroneos, A., & Fitzpatrick, M. E. (2017). Detecting anomalies in time series data via a deep learning algorithm combining wavelets, neural networks and Hilbert transform. *Expert Systems with Applications, 85*, 292–304. http://dx.doi.org/10.1016/j.eswa.2017.04.028.

Kaur, H., & Ahuja, S. (2017). Time series analysis and prediction of electricity consumption of health care institution using ARIMA model. In K. Deep, J. C. Bansal, K. N. Das, A. K. Lal, H. Garg, A. K. Nagar, & M. Pant (Eds.). In *Proceedings of sixth international conference on soft computing for problem Solving: SocProS 2016: Volume 2* (pp. 347–358). Singapore: Springer Singapore. http://dx.doi.org/10.1007/978-981-10-3325-4_35.

Khosravani, H., Castilla, M., Berenguel, M., Ruano, A., & Ferreira, P. (2016). A comparison of energy consumption prediction models based on neural networks of a bioclimatic building. *Energies, 9*, 57. http://www.mdpi.com/1996-1073/9/1/57.

Korn, F., Pagel, B. U., & Faloutsos, C. (2001). On the “dimensionality curse” and the “self-similarity blessing&rdquo. *IEEE Transactions on Knowledge and Data Engineering, 13*, 96–111. http://dx.doi.org/10.1109/69.908983.

Koza, J. R. (1992). *Genetic Programming: on the programming of computers by means of natural selection*. Massachusetts Institute of Technology. http://www.cs.bham.ac.uk/~wbl/biblio/cache/bin/cache.php?koza:book, http___mitpress.mit.edu_books_genetic-programming, http://mitpress.mit.edu/books/genetic-programming. ISBN 0-262-11170-5. Printed and bound in the United States of America.

Lord, S.-F., Noye, S., Ure, J., Tennant, M. G., & Fisk, D. J. (2016). Comparative review of building commissioning regulation: A quality perspective. *Building Research & Information, 44*, 630–643. http://dx.doi.org/10.1080/09613218.2016.1181955.

Ma, X., & Liu, Z. (2017). Application of a novel time-delayed polynomial grey model to predict the natural gas consumption in China. *Journal of Computational and Applied Mathematics, 324*, 17–24. http://dx.doi.org/10.1016/j.cam.2017.04.020.

Olubunmi, O. A., Xia, P. B., & Skitmore, M. (2016). Green building incentives: A review. *Renewable and Sustainable Energy Reviews, 59*, 1611–1621. http://dx.doi.org/http://doi.org/10.1016/j.rser.2016.01.028.

Pallit, A., & Popovic, D. (2005). *Computational intelligence in time series forecasting*. Heidelberg: Springer. *http://dx.doi.org/10.1007/1-84628-184-9*.

Picek, S., Jakobovic, D., & Golub, M. (2013). On the recombination operator in the real-coded genetic algorithms. In *2013 IEEE Congress on Evolutionary Computation* (pp. 3103–3110). http://dx.doi.org/10.1109/CEC.2013.6557948.

Pino-Mejías, R., Pérez-Fargallo, A., Rubio-Bellido, C., & Pulido-Arcas, J. A. (2017). Comparison of linear regression and artificial neural networks models to predict heating and cooling energy demand, energy consumption and CO_2 emissions. *Energy, 118*, 24–36. http://dx.doi.org/10.1016/j.energy.2016.12.022.

Qin, S., Wang, J., Wu, J., & Zhao, G. (2016). A hybrid model based on smooth transition periodic autoregressive and Elman artificial neural network for wind speed forecasting of the Hebei region in China. *International Journal of Green Energy, 13*, 595–607. http://dx.doi.org/10.1080/15435075.2014.961462.

Rodger, J. A. (2014). A fuzzy nearest neighbor neural network statistical model for predicting demand for natural gas and energy cost savings in public buildings. *Expert Systems with Applications, 41*, 1813–1829. http://dx.doi.org/10.1016/j.eswa.2013.08.080.

Rodrigues, F., Cardeira, C., Calado, J. M. F., & Melício, R. (2017). Family houses energy consumption forecast tools for smart grid management. In P. Garrido, F. Soares, & A. P. Moreira (Eds.), *CONTROLO 2016: proceedings of the 12th portuguese conference on automatic control* (pp. 691–699). Cham: Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-43671-5_58.

Ruiz, L. G. B., Cuellar, M. P., Delgado, M., & Pegalajar, M. C. (2016). An Application of non-linear autoregressive neural networks to predict energy consumption in public buildings. *Energies, 9*, 21. http://dx.doi.org/10.3390/en9090684.

Sadat Hosseini, S. S., & Gandomi, A. H. (2012). Short-term load forecasting of power systems by gene expression programming. *Neural Computing and Applications, 21*, 377–389. http://dx.doi.org/10.1007/s00521-010-0444-y.

Smith, S. W. (1997). Moving average filters. *The Scientist and Engineer's Guide to Digital Signal Processing*, 277–284. http://www.dspguide.com/ch15.htm.

Vogl, T. P., Mangis, J. K., Rigler, A. K., Zink, W. T., & Alkon, D. L. (1988). Accelerating the convergence of the back-propagation method. *Biological Cybernetics, 59*, 257–263. http://dx.doi.org/10.1007/bf00332914.

Wang, X., Song, H., Bai, S., Zeng, T., Zhao, S., Wu, W., Li, W., & Liu, Y. (2016). Study of the monitored method of atomic clock data exception based on the model of dynamic neural network time series-NAR. In *2016 IEEE International Frequency Control Symposium (IFCS)* (pp. 1–4). http://dx.doi.org/10.1109/FCS.2016.7546764.

Zhang, F., Deb, C., Lee, S. E., Yang, J., & Shah, K. W. (2016). Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique. *Energy and Buildings, 126*, 94–103. http://dx.doi.org/10.1016/j.enbuild.2016.05.028.