

Article

Exploring an Ensemble of Methods that Combines Fuzzy Cognitive Maps and Neural Networks in Solving the Time Series Prediction Problem of Gas Consumption in Greece

Konstantinos I. Papageorgiou ^{1,*}, Katarzyna Poczeta ² , Elpiniki Papageorgiou ³ ,
Vassilis C. Gerogiannis ^{3,*} and George Stamoulis ¹

¹ Department of Computer Science & Telecommunications, University of Thessaly, 35100 Lamia, Greece; georges@cs.uth.gr

² Department of Information Systems, Kielce University of Technology, 25-541 Kielce, Poland; k.piotrowska@tu.kielce.pl

³ Faculty of Technology, University of Thessaly-Gaiopolis, 41500 Gaiopolis, Larissa, Greece; elpinikipapageorgiou@uth.gr

* Correspondence: konpapageorgiou@uth.gr (K.I.P.); vgerogian@uth.gr (V.C.G.)

Received: 30 September 2019; Accepted: 31 October 2019; Published: 6 November 2019



Abstract: This paper introduced a new ensemble learning approach, based on evolutionary fuzzy cognitive maps (FCMs), artificial neural networks (ANNs), and their hybrid structure (FCM-ANN), for time series prediction. The main aim of time series forecasting is to obtain reasonably accurate forecasts of future data from analyzing records of data. In the paper, we proposed an ensemble-based forecast combination methodology as an alternative approach to forecasting methods for time series prediction. The ensemble learning technique combines various learning algorithms, including SOGA (structure optimization genetic algorithm)-based FCMs, RCGA (real coded genetic algorithm)-based FCMs, efficient and adaptive ANNs architectures, and a hybrid structure of FCM-ANN, recently proposed for time series forecasting. All ensemble algorithms execute according to the one-step prediction regime. The particular forecast combination approach was specifically selected due to the advanced features of each ensemble component, where the findings of this work evinced the effectiveness of this approach, in terms of prediction accuracy, when compared against other well-known, independent forecasting approaches, such as ANNs or FCMs, and the long short-term memory (LSTM) algorithm as well. The suggested ensemble learning approach was applied to three distribution points that compose the natural gas grid of a Greek region. For the evaluation of the proposed approach, a real-time series dataset for natural gas prediction was used. We also provided a detailed discussion on the performance of the individual predictors, the ensemble predictors, and their combination through two well-known ensemble methods (the average and the error-based) that are characterized in the literature as particularly accurate and effective. The prediction results showed the efficacy of the proposed ensemble learning approach, and the comparative analysis demonstrated enough evidence that the approach could be used effectively to conduct forecasting based on multivariate time series.

Keywords: fuzzy cognitive maps; neural networks; time series forecasting; ensemble learning; prediction; machine learning; natural gas

1. Introduction

Time series forecasting is a highly important and dynamic research domain, which has wide applicability to many diverse scientific fields, ranging from ecological modeling to energy [1],

finance [2,3], tourism [4,5], and electricity load [6,7]. A summary of applications regarding forecasting in various areas can be found in a plethora of review papers published in the relevant literature [8–15]. One of the main challenges in the field of time series forecasting is to obtain reasonable and accurate forecasts of future data from analyzing previous historical data [16]. Following the literature, numerous research studies show that forecasting accuracy is improved when different models are combined, while the resulting model seems to outperform all component models [16]. Thus, the combination of forecasts from different models or algorithms becomes a promising field in the prediction of future data.

Researchers can choose between linear and nonlinear time series forecasting methods, depending on the nature of the model they are working on. Linear methods, like autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) [17] are among the methodologies which have found wide applicability in real-world applications. Traffic [18], energy [19–21], economy [22], tourism [23], and health [24] are some example fields in which the ARIMA forecasting technique was used. Considering though that most real-world problems are characterized by a non-linear behavior, many researchers have investigated the use of non-linear techniques for times series-based forecasting and prediction. In particular, machine learning and other intelligent methods have been chosen to address possible nonlinearities in time series modeling, such as nonlinear ARMA time series models, needing though to tackle the type of nonlinearity that is not usually known. Along with the computing power growth and the evolution of data management techniques, there has been a growing interest in the use of advanced artificial intelligence technologies, like artificial neural networks (ANNs) [25] and fuzzy logic systems for forecasting purposes. ANNs and fuzzy logic systems use more sophisticated generic model structures having the ability to incorporate the characteristics of complex data and produce accurate time series models [26], while they also incorporate the advantageous features of nonlinear modeling and data-based learning capabilities. Moreover, unlike traditional time series methods that are not able to adequately capture nonlinear patterns of data, neural networks are usually involved in predicting consumption demand during periods of low or extremely high demand [27]. Among all types of ANN models, the feed-forward network model with backpropagation training procedure (FFN-BP) is one of the most commonly used approaches [28].

The technique of combining the predictions of multiple classifiers to produce a single classifier is referred to as an ensemble [29–32]. Ensemble building has recently attracted much attention among researchers as an effective method that improves the performance of the resulting model for performing classification and regression tasks. It has been demonstrated that an ensemble of individual predictors performs better than a single predictor in the average [33,34], thus achieving better prediction accuracy than that of the individual predictors. Two popular methods for creating accurate ensembles are bagging [29] and boosting [35,36]. There are not many research works concerning potency forecasting of a model in advance, as this is often a difficult task. Trying to avoid the risk of combining models that have poor performance regarding prediction may result in an overall model with deteriorated forecasting accuracy. An ensemble model, however, should rather be formed solely from component models that are rated as adequate, if not good enough, for their forecasting capabilities [37]. Usually, the single neural network (NN) models are combined to create NNs ensembles, to tackle sampling and modeling uncertainties that could probably weaken the forecasting accuracy and robustness of component NN models. Since individual component model could be sensitive under different circumstances, ensembling them results in more powerful outcomes in the context of a decision-making process.

1.1. Related Literature

In recent years, the natural gas market has progressed greatly in terms of fast development and, thus, it has become a very competitive area. Undoubtedly, natural gas is among primary energy sources and has a significant environmental role considering its valuable environmental benefits, such as low-level emissions of greenhouse gases in comparison with other non-renewable energy sources [38,39]. Natural gas demand seems to increase considerably due to several socio-economic and political reasons, while the price and environmental concerns are significant regulatory factors affecting

natural gas demand. Therefore, the prediction of natural gas consumption, as a time series forecasting problem, is becoming important in contemporary energy systems, allowing energy policymakers to apply effective strategies to guarantee sufficient natural gas supplies.

So far, many research papers have tried to give a clear insight regarding natural gas forecasting by suggesting models for predicting the consumption of this non-renewable energy source. A summary of natural gas consumption forecasting regarding prediction methods, input variables used for modeling, as well as prediction area, can be found in many review papers in the relevant literature. Particularly, there is a thorough literature survey of published papers [38–40] that classifies various models and techniques that have been recently applied in the field of natural gas forecasting, with respect to the paradigm that each model/technique is based on, while there has been also an attempt by researchers to classify all models applied in this area according to their performance characteristics, as well as to offer some future research directions. The models presented were developed by researchers to predict natural gas consumption on an hourly, daily, weekly, monthly, or yearly basis, in an attempt to predict natural gas consumption with an acceptable degree of accuracy. Accurate forecasting of natural gas consumption can be particularly important for project planning, engineering design, pipeline operation, gas imports, tariff design, and optimal scheduling of a natural gas supply system [41].

Due to the need for distribution planning, especially in residential areas, the increasing demand for natural gas, and the restricted natural gas network in many countries, the forecasting of consumption on a daily and weekly basis seems to be of high importance. In the relevant literature, many suggestions apply various ANN topologies and methods to support day-ahead natural gas demand prediction [42–50]. For example, [42] used ANN, while [45] used a combination of ANN forecasters for predicting gas consumption at a citywide distribution level. For the same distribution level (citywide), a strategy was proposed in [51] to estimate the forecasting risk by using hourly consumption data. Having as aim to find the best solution for natural gas consumption, the researchers in [52] used linear regression with the sliding window technique, while in [53], a univariate artificial bee colony-based artificial neural network (ANN-ABC) was applied to minimize error in the case of forecasting day-ahead demand. The researchers in [54] also considered various methods for the prediction of daily gas consumption, such as the seasonal autoregressive integrated moving average model with exogenous inputs (SARIMAX), multi-layer perceptron ANN (ANN-MLP), ANN with radial basis functions (ANN-RBF), and multivariate ordinary least squares (OLS).

In the relevant literature, there are also some research works on neuro-fuzzy methods and genetic algorithms applied in natural gas demand, as in [41,55–58]. Specifically, a novel hybrid model that combines the wavelet transform (WT), genetic algorithm (GA), adaptive neuro-fuzzy inference system (ANFIS), and feed-forward neural network (FFNN) was recently examined in [41] and applied to the Greek natural gas grid. Moreover, evolutionary fuzzy cognitive maps (FCMs) were recently used for time series problems modeling and forecasting. FCMs can be understood as recurrent neural networks inheriting many features from them, such as learning capabilities, which elevate the performance of FCMs in modeling and prediction and further helped FCMs to gain momentum over recent years [59,60]. The researchers in [61,62] were the first to examine the application of FCMs to time series modeling, proposing nodes selection criteria in an FCM, which was used to model univariate time series. Further techniques for simplifying FCMs by removing nodes and weights were investigated, while a dynamic optimization of the FCM structure was studied in [63] for univariate time series forecasting. Concerning multivariate interval-valued time series, an evolutionary algorithm for learning fuzzy grey cognitive maps was developed as a nonlinear predictive model [64]. Taking one step further, the researchers in [65] and [66] enhanced the evolutionary FCMs with the structure optimization genetic algorithm (SOGA). These approaches can be used to automatically construct an FCM model after selecting the crucial concepts and defining the relationships between them by taking into consideration any available historical data. An example regarding rented bikes' count prediction was examined, where SOGA-FCM was compared with the multi-step gradient method (MGM) [67] and the real-coded genetic algorithm (RCGA) [68]. A two-stage prediction model for multivariate time series prediction, based

on the efficient capabilities of evolutionary fuzzy cognitive maps (FCMs) and enhanced by structure optimization algorithms and artificial neural networks (ANNs), was introduced in [69]. Furthermore, the researchers in [21,60] recently conducted a preliminary study on implementing FCMs with NNs for natural gas prediction.

1.2. Research Aim and Approach

The purpose of this paper was to propose a new forecast combination approach resulting from FCMs, ANNs, and hybrid models. This ensemble forecasting method, including the two most popular ensemble methods, the Average and the Error-based, is based on ANNs, FCMs with learning capabilities, as well as on a hybrid FCM-ANN model with different configurations, to produce an accurate non-linear time series model for the prediction of natural gas consumption. A real case study problem of natural gas consumption in Greece was performed to show the applicability of the proposed approach. Furthermore, in order to validate the proposed forecasting combination approach, a comparison analysis between the ensemble methods and an innovative machine learning technique, the long short-term memory (LSTM) algorithm (which is devoted to time series forecasting), was conducted, and the results demonstrated enough evidence that the proposed approach could be used effectively to conduct forecasting based on multivariate time series. The LSTM algorithm, as an advanced recurrent NN method, was previously used for short-term natural gas demand forecasting in Greece [70]. In that research paper, LSTM was applied in one day-ahead natural gas consumption, forecasting for the same three Greek cities, which were also examined in the case study presented in the current paper. Many similar works can be found in the literature that examine various forecast combinations in terms of accuracy and error variability but, in the present work, an innovative approach that combines FCMs, ANNS, and hybrid FCM-ANN models, producing a non-linear time series model for the prediction of natural gas consumption, was studied exclusively, contributing to the novelty of the current study. The results demonstrated in a clear way that the proposed approach had attained better accuracies than other individual models. This study justified the superiority of the selective ensemble method over combining the important features and capabilities of the models that consist of the overall approach, making it a useful tool for future work.

The outline of the paper is as follows. Section 2 describes the material and methods of our research study; Section 2.1 describes the case study problem and refers to the datasets of natural gas demand that are used, whereas Section 2.2 presents the analyzed approaches for time series forecasting based on ANNs, FCMs with evolutionary learning algorithms, and their hybrid combinations. The most widely used ensemble methods for forecasting problems (i.e., the error-based and the simple average method) are also presented in Section 2.2. In Section 3, the proposed forecasting combination approach is described. The same Section presents the evaluation criteria, which we have used to analyze the performance of the analyzed approaches for natural gas prediction. Section 4 presents the results of simulation analysis for three different Greek cities, as well as the conducted comparative analysis of the proposed approach with other intelligent techniques. A discussion of the results highlights the main findings of the proposed ensemble forecasts approach. Section 5 summarizes the main conclusions of the paper with further discussion and suggestions about future research expansion.

2. Materials and Methods

2.1. Material-Dataset

In the considered case study, three different prediction datasets of natural gas demand, derived from different districts in Greece, were analyzed from the records of the Hellenic Gas Transmission System Operator S.A. (www.desfa.gr, DESFA). DESFA company is responsible for the operation, management, exploitation, and development of the Greek Natural Gas System and its interconnections in a technically sound and economically viable way. From 2008, DESFA provides historical data of transmission system operation and natural gas deliveries/off-takes. In this research work, historical

data with the values of gas consumption for a period of five years, from 2013 to 2017, were used as initial data to accomplish forecasting. These data were split into training and testing data, where usually the training data came from the first four years and were used for learning models, whereas the data of the last year were used for testing the applied artificial intelligence models.

It is crucial for an efficient forecast to properly select the number and types of inputs. Thus, we emphasized on defining proper input candidates. Six different inputs for time series prediction were considered. The first three inputs were devoted to month indicator, day indicator, and mean temperature. Specifically, concerning the calendar indicators, we used one input for months and one input for days coding. Let $m = 1, 2, \dots, 12$ be the number of months. We considered the following matching: January/1, February/2, ..., December/12. Let $l = 1, 2, \dots, 7$ be the number of days. The day type matching was as follows: Monday/1, Tuesday/2, ..., Sunday/7. The temperature data were obtained by the nearest to the distribution gas point station. The rest three inputs were the previously measured values of natural gas demand, for one-day before, two-day before, and the current day. These six variables were used to form the input pattern of the FCM. The output referred to the total daily demand for the specific distribution point.

The features that were gathered and used in our study to form the FCM model were enough and properly selected according to the relevant literature. From a recent literature review regarding the prediction of natural gas consumption [40], it can be seen that past gas consumption combined with meteorological data (especially temperature) are the most commonly used input variables for the prediction of natural gas consumption. A recent study [41] used past consumption, temperature, months, and days of the week, while in [55], day of week and demand of the same day in the previous year were used as input variables for natural gas forecasting. Considering the above practices described in the literature, it can be concluded that the features used in the current work were enough to predict the consumption of natural gas for the selected areas.

The Greek cities of Thessaloniki, Athens, and Larissa were selected for the conducted simulation analysis and comparison of the best performing algorithms. These different natural gas consumption datasets may offer insight into whether the analyzed algorithms perform equally in different locations, where the energy demand could be completely different for the same days.

2.2. Methods

2.2.1. Fuzzy Cognitive Maps Overview

A fuzzy cognitive map (FCM) is a directed graph in which nodes denote concepts important for the analyzed problem, and links represent the causal relationships between concepts [71]. It is an effective tool for modeling decision support systems. FCMs have been applied in many research domains, e.g., in business performance analysis [72], strategy planning [73], modeling virtual worlds [74], time series prediction [69], and adoption of educational software [75].

The FCM model can be used to perform simulations by utilizing its dynamic model. The values of the concepts change in time as simulation goes on [68]. The new values of the concepts can be calculated based on the popular dynamic model described as follows [59]:

$$X_i(t+1) = F \left(X_i(t) + \sum_{\substack{j=1 \\ j \neq i}}^n X_j(t) \cdot w_{j,i} \right) \quad (1)$$

where $X_i(t)$ is the value of the i th concept at the t th iteration, $w_{j,i}$ is the weight of the connection (relationship) between the j th concept and the i th concept, t is discrete-time, $i, j = 1, 2, \dots, n$, n is the number of concepts, $F(x)$ is the sigmoidal transformation function [58]:

$$F(x) = \frac{1}{1 + e^{-cx}} \quad (2)$$

where c is a parameter, $c > 0$. The weights of the relationships show how causal concepts affect one another. If $w_{j,i} > 0$, then an increase/decrease in the value of the j th concept will increase/decrease the value of the i th concept. If $w_{j,i} < 0$, then an increase/decrease in the value of the j th concept will decrease/increase the value of the i th concept. If $w_{j,i} = 0$, there is no causal relationship between the j th and the i th concepts [74].

The FCM structure is often constructed based on expert knowledge or surveys [74]. We could also use machine learning algorithms and available historical data to construct the FCM model and determine the weights of the relationships between the FCM's concepts.

2.2.2. Fuzzy Cognitive Maps Evolutionary Learning

Evolutionary algorithms are popular techniques for FCMs learning. In this paper, we explored two effective techniques: the real-coded genetic algorithm (RCGA) [68] and the structure optimization genetic algorithm (SOGA) [69].

Real-Coded Genetic Algorithm (RCGA)

The RCGA algorithm defines individual in the population as follows [24]:

$$W' = [w_{1,2}, w_{1,3}, w_{1,4}, \dots, w_{j,i}, \dots, w_{n,n-1}]^T \quad (3)$$

where $w_{j,i}$ is the weight of the relationship between the j th concept and the i th concept.

Individual in the population is evaluated with the use of a fitness function based on data error [66]:

$$fitness_p(MSE_{tr}(l)) = \frac{1}{aMSE_{tr}(l) + 1} \quad (4)$$

where a is a parameter, l is the number of generation, $l = 1, \dots, L$, L is the maximum number of generations, p is the number of individual, $p = 1, \dots, P$, P is the population size, and $MSE_{tr}(l)$ is the data error, described as follows:

$$MSE_{tr}(l) = \frac{1}{N_{tr}} \sum_{t=1}^{N_{tr}} e_t^2 \quad (5)$$

where $t = 1, \dots, N_{tr}$, N_{tr} is the number of training records, and e_t is the one-step-ahead prediction error at the t th iteration, described as follows:

$$e_t = Z(t) - X(t) \quad (6)$$

where $X(t)$ is the predicted value of the output concept, and $Z(t)$ is the desired value of the output concept.

When the maximum number of generations L is reached, or the condition (7) is met, which means that the learning process is successful, then the RCGA stops.

$$fitness_p(MSE_{tr}(l)) > fitness_{max} \quad (7)$$

where $fitness_p(MSE_{tr}(l))$ is the fitness function value for the best individual, and $fitness_{max}$ is a parameter.

Structure Optimization Genetic Algorithm (SOGA)

The SOGA algorithm is an extension of the RCGA algorithm [65,66] that allows the decision-maker to determine the most significant concepts and the relationships between them.

Individual is evaluated based on the fitness function based on new data error, described as follows [66]:

$$MSE'_{tr}(l) = MSE_{tr}(l) + b_1 \frac{n_r}{n^2} MSE_{tr}(l) + b_2 \frac{n_c}{n} MSE_{tr}(l) \tag{8}$$

where b_1, b_2 are the parameters of the fitness function, n_r is the number of the non-zero relationships, n_c is the number of the concepts in the analyzed model, n is the number of all possible concepts, l is the number of generation, $l = 1, \dots, L$, L is the maximum number of generations.

The fitness function that follows (9) calculates the quality of each population.

$$fitness_p(MSE'_{tr}(l)) = \frac{1}{\alpha MSE'_{tr}(l) + 1} \tag{9}$$

where α is an experimentally defined parameter, p is the number of the individual, $p = 1, \dots, P$, P is the population size, and $MSE'_{tr}(l)$ is the new error measure.

We could construct a less complex time series prediction model by removing the redundant concepts and connections between them with the use of a binary vector C and the proposed error function.

The algorithmic steps of the learning and analysis of the FCM in modeling prediction systems with the use of population-based algorithms (SOGA and RCGA) were analytically presented in [69].

For our experiments, the evolutionary operators, a) ranking selection, b) uniform crossover, and c) random mutation were used [76,77]. In addition, we applied elite strategy selection, while a probability of crossover P_c and mutation P_m was assigned to each population.

2.2.3. Artificial Neural Networks

An artificial neural network (ANN) is a collection of artificial neurons organized in the form of layers [25]. Neurons are connected by weighted connections to form a NN. The most widely used ANNs in time series prediction are the multilayer perceptrons with an input layer, an output layer, and a single hidden layer that lies between the input and output layer. The most common structure is an ANN that uses one or two hidden layers, as a feed-forward neural network with one hidden layer is able to approximate any continuous function. Supervised learning algorithms and historical data can be used for the learning process of ANNs. The output of each neuron can be calculated based on the following formula:

$$X(t) = F \left(\sum_{j=1}^m X_j(t) \cdot w_j + b \right) \tag{10}$$

where $X_j(t)$ is the value of the j th input signal, $t = 1, \dots, N_{tr}$, N_{tr} is the number of training records, w_j is the synaptic weight, m is the number of input signals, b is the bias, and F is the sigmoid activation function. Training a neural network needs the values of the connection weights and the biases of the neurons to be determined. There are many neural network learning algorithms. The most popular algorithm for ANN learning is the back-propagation method. In this learning method, the weights change their values according to the learning records until one epoch (an entire learning dataset) is reached. This method aims to minimize the error function, described as follows [14,78,79]:

$$MSE_{tr}(l) = \frac{1}{2N_{tr}} \sum_{t=1}^{N_{tr}} e_t^2 \tag{11}$$

where $t = 1, \dots, N_{tr}$, N_{tr} is the number of training records, l is the number of epoch, $l = 1, \dots, L$, L is the maximum number of epochs, and e_t is the one-step-ahead prediction error at the t th iteration, which is equal to:

$$e_t = Z(t) - X(t) \tag{12}$$

where $X(t)$ is the output value of the ANN, and $Z(t)$ is the desired value.

The modification of the weights in the back-propagation algorithm can be calculated by the formula:

$$\Delta w_{kj}(l) = -\gamma \frac{\partial J(l)}{\partial w_{kj}(l)} \quad (13)$$

where $\Delta w_{kj}(l)$ is a change of the weight w_{kj} at the l th epoch, γ is a learning coefficient.

Backpropagation algorithm with momentum modifies the weights according to the formula:

$$\Delta w_{kj}(l) = -\gamma \frac{\partial J(l)}{\partial w_{kj}(l)} + \alpha \Delta w_{kj}(l-1) \quad (14)$$

where α is a momentum parameter.

2.2.4. Hybrid Approach Based on FCMs, SOGA, and ANNs

The hybrid approach for time series prediction is based on FCMs, the SOGA algorithm, and ANNs [68]. This approach consists of two stages:

1. Construction of the FCM model based on the SOGA algorithm to reduce the concepts that have no significant influence on data error.
2. Considering the selected concepts (data attributes) as the inputs for the ANN and ANN learning with the use of backpropagation method with momentum.

This hybrid structure allows the decision-maker to select the most significant concepts for an FCM model using the SOGA algorithm. These concepts are used as inputs for the ANN model. Such a hybrid approach aims to find the most accurate model for time series prediction problems.

2.2.5. The Ensemble Forecasting Method

The most intuitive and popular way of forecast aggregation is to linearly combine the constituent forecasts [80]. There are various methods proposed in the literature for selecting the combining weights [81]. The most popular and widely used ensemble methods are the error-based and the simple average [82]. The easiest among them is the simple average in which all forecasts are weighted equally, often remarkably improving overall forecasting accuracy [82,83].

Considering that $Y = [y_1, y_2, y_3, \dots, y_N]^T$ is the actual out-of-sample testing dataset of a time series and $\hat{Y}^i = [\hat{y}_1^i, \hat{y}_2^i, \dots, \hat{y}_n^i]^T$ is the forecast for the i th model, the linear combination of n forecasts is produced by [15]:

$$\hat{y}_k = w_1 \hat{y}_k^{(1)} + w_2 \hat{y}_k^{(2)} + \dots + w_n \hat{y}_k^{(n)} = \sum_{i=1}^n w_i \hat{y}_k^{(i)}, \quad \forall k = 1, 2, \dots, N \quad (15)$$

Here, our analysis is based on these most popular ensemble methods. A brief discussion follows for each one.

- The *simple average (AVG)* method [82] is an unambiguous technique, which assigns the same weight to every single forecast. Based on empirical studies in the literature, it has been observed that the AVG method is robust and able to generate reliable predictions, while it can be characterized as remarkably accurate and impartial. Being applied in several models, with respect to effectiveness, the AVG improved the average accuracy when increasing the number of combined single methods [82]. Comparing the referent method with the weighted combination techniques, in terms of forecasting performance, the researchers in [84] concluded that a simple average combination might be more robust than weighted average combinations. In the simple average combination, the weights can be specified as follows:

$$w_i = \frac{1}{n}, \quad \forall i = 1, 2, \dots, n \quad (16)$$

- The *error-based (EB)* method [16] consists of component forecasts, which are given weights that are inversely proportional to their in-sample forecasting errors. For instance, researchers may give a higher weight to a model with lower error, while they may assign a less weight value to a model that presents more error, respectively. In most of the cases, the forecasting error is calculated using total absolute error statistic, such as the sum of squared error (SSE) [80,83]. The combining weight for individual prediction is mathematically given by:

$$w_i = e_i^{-1} / \sum_{i=1}^n e_i^{-1}, \quad \forall i = 1, 2, \dots, n \quad (17)$$

3. The Proposed Forecast Combination Methodology

In the rest of the paper, we explored a new advanced forecasting approach by introducing a different split of dataset in the case of daily, weekly, or monthly forecasting, as well as a combination of forecasts from multiple structurally different models, like ANN and FCM with various efficient learning algorithms and hybrid configurations of them. Also, the two most popular and usually used ensemble methods, the AVG and the EB methods, were applied to the ensemble forecasts to improve the prediction accuracy.

In the described ensemble scheme, the selection of the appropriate validation set, i.e., the selection of the parameter N_{vd} and the group size N_{tr} , is very important. The validation set should reflect the characteristics of the testing dataset that is practically unknown in advance. As such, in this study, we set the following process of data split. The data split takes place by removing 15% of the total dataset N and saving for later use as testing data. The remaining 85% of the dataset is then split again into an 82/18 ratio, resulting in the following portions: 70% for training and 15% for validation. Also, the group size N_{tr} (i.e., the training data) should be appropriately selected so that it is neither too small nor too large.

Due to the problem nature, as we work with time-series data, the most efficient method for resampling is the boosting/bootstrapping method [85]. In boosting, resampling is strategically geared to provide the most informative training data for each consecutive predictor. Therefore, in this study, an appropriate bootstrapping method was applied, so that the training dataset should have the same size at each resampling set, and the validation and testing sets should keep the same size (after excluding the k -values from the in-sample dataset).

The proposed effective forecast combination methodology for time series forecasting, presented in the paper, includes three main processing steps: data pre-processing to handle missing values, normalize the collected time-series data, and split the dataset; the various forecasting methods of ANNs, RCGA-FCMs, SOGA-FCMs, and hybrid SOGA FCM-ANN with their ensembles; and evaluation of the prediction results, implementing the two most popular and used ensemble methods of simple average (AVG) and error-based (EB). Figure 1 visually illustrates the suggested methodology.

In the followed approach, data preprocessing included outlier detection and removal, handling missing data, and data normalization, all of which were in accordance with the principles of Data Science practices described in corresponding literature. For outlier detection, the Z-score was first calculated for each sample on the data set (using the standard deviation value that is presented in the descriptive statistics Tables A1 and A2 in Appendix A). Then, a threshold was specified, and the data points that lied beyond this threshold were classified as outliers and were removed. Mean imputation was performed to handle missing values. Specifically, for numerical features, missing values were replaced by the mean feature value.

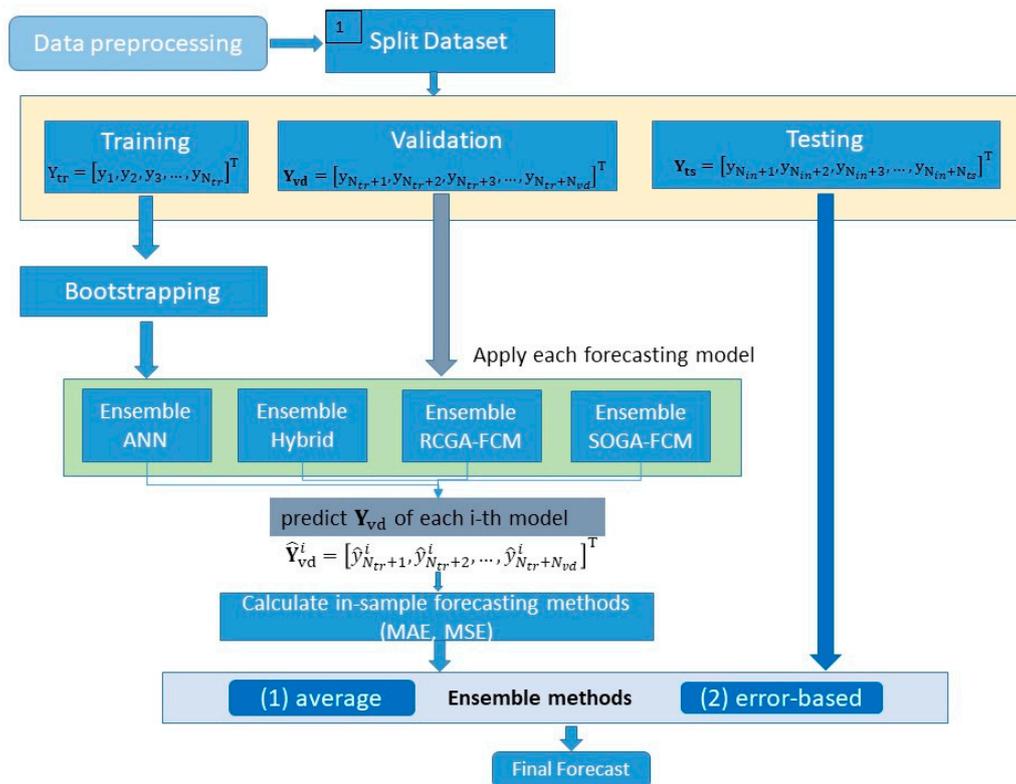


Figure 1. Flowchart of the proposed forecasting combination approach.

Each dataset was normalized to [0,1] before the forecasting models were applied. The normalized datasets were taking again their original values, while the testing phase was implemented. The data normalizations were carried out mathematically, as follows:

$$y_i^{(new)} = \frac{y_i - y^{(min)}}{y^{(max)} - y^{(min)}}, \forall i = 1, 2, \dots, N \tag{18}$$

where $Y = [y_1, y_2, y_3, \dots, y_{N_{train}}]^T$ is the training dataset, and $Y^{(new)} = [y_1^{(new)}, y_1^{(new)}, \dots, y_N^{(new)}]^T$ is the normalized dataset, $y^{(min)}$ and $y^{(max)}$ are, respectively, the minimum and maximum values of the training dataset Y .

We selected the Min-Max normalization method [86] as it is one of the most popular and comprehensible methods, in terms of performance of the examined systems, while several researchers showed that it produces better (if not equally good) results with high accuracy, compared to the other normalization methods [87,88]. In [88], the Min-Max was valued as the second-best normalization method in the backpropagation NN model, justifying our choice to deploy this method for data normalization. Moreover, since the FCM concepts use values within the range [0,1] for the conducted simulations and do not deal with real values, the selected method seemed to be proper for our study. Also, this normalization approach was previously used in [66,69].

Due to our intention to suggest a generic forecasting combination approach (with ANNs, FCMs, and their hybrid structures) able to be applied in any time series dataset, the following steps are thoroughly presented and executed.

Step 1. (Split Dataset) We divided the original time series $Y = [y_1, y_2, y_3, \dots, y_N]^T$ into the in-sample training dataset $Y_{tr} = [y_1, y_2, y_3, \dots, y_{N_{tr}}]^T$, the in-sample validation dataset $Y_{vd} = [y_{N_{tr}+1}, y_{N_{tr}+2}, y_{N_{tr}+3}, \dots, y_{N_{tr}+N_{vd}}]^T$, and the out-of-sample testing dataset

$Y_{ts} = [y_{N_{in}+1}, y_{N_{in}+2}, y_{N_{in}+3}, \dots, y_{N_{in}+N_{ts}}]^T$, so that $N_{in} = N_{tr} + N_{vd}$ is the size of the total in-sample dataset and $N_{in} + N_{ts} = N$, where N is the number of days, or weeks, or months, according to the short- or long-term prediction based on the time series horizon.

Step 2. (Resampling method/Bootstrapping). Let's consider k sets as training sets from the whole dataset every time. For example, in the monthly forecasting, we excluded one month every time from the initial in-sample dataset, starting from the first month of the time series values, and proceeding with next month till $k = 12$, (i.e., this means that 1 to 12 months were excluded from the initial in-sample dataset). Therefore, k subsets of training data were created and used for training. The remaining values of the in-sample dataset were used for validation, whereas the testing set remained the same. Figure 2 shows an example of this bootstrapping method for the ensemble SOGA-FCM approach. In particular, Figure 2a represents the individual forecasters' prediction values and their average error calculation, whereas, in Figure 2b, the proposed forecasting combination approach for SOGA-FCM is depicted for both ensemble methods.

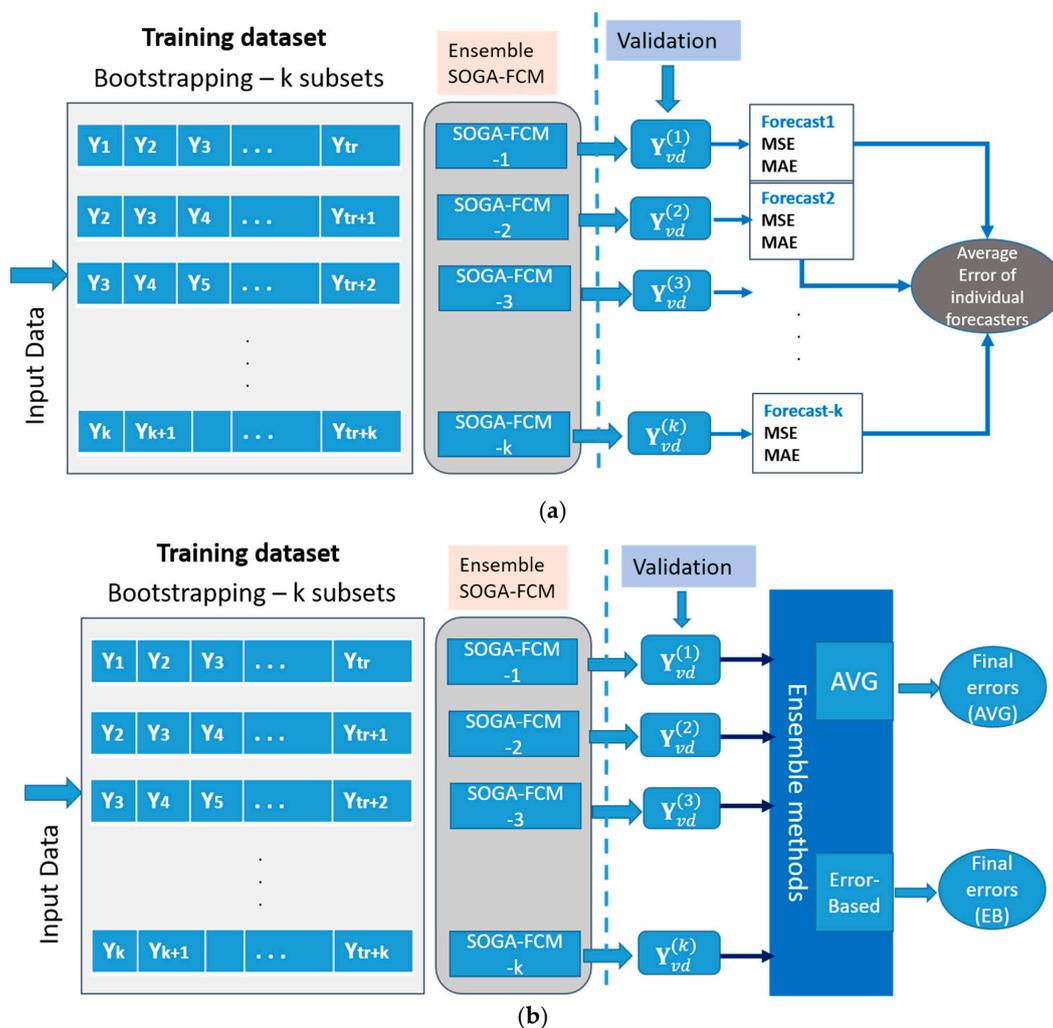


Figure 2. (a) Forecasting approach using individual forecasters of SOGA-FCM and mean average, (b) Example of the proposed forecasting combination approach for SOGA-FCM using ensemble methods. SOGA, structure optimization genetic algorithm; FCM, fuzzy cognitive map.

If we needed to accomplish daily forecasting, then we preselected the number of days excluded at each subset k . For the case of simplicity (as in the case of monthly forecasting), we could consider that one day was excluded at each sub-set from the initial in-sample dataset. The overall approach, including ANN, FCMs, and hybrid configurations of them, is illustrated in Figure 3. In Figure 3, the

four ensemble forecasters were produced after the validation process and used for testing through the proposed approach.

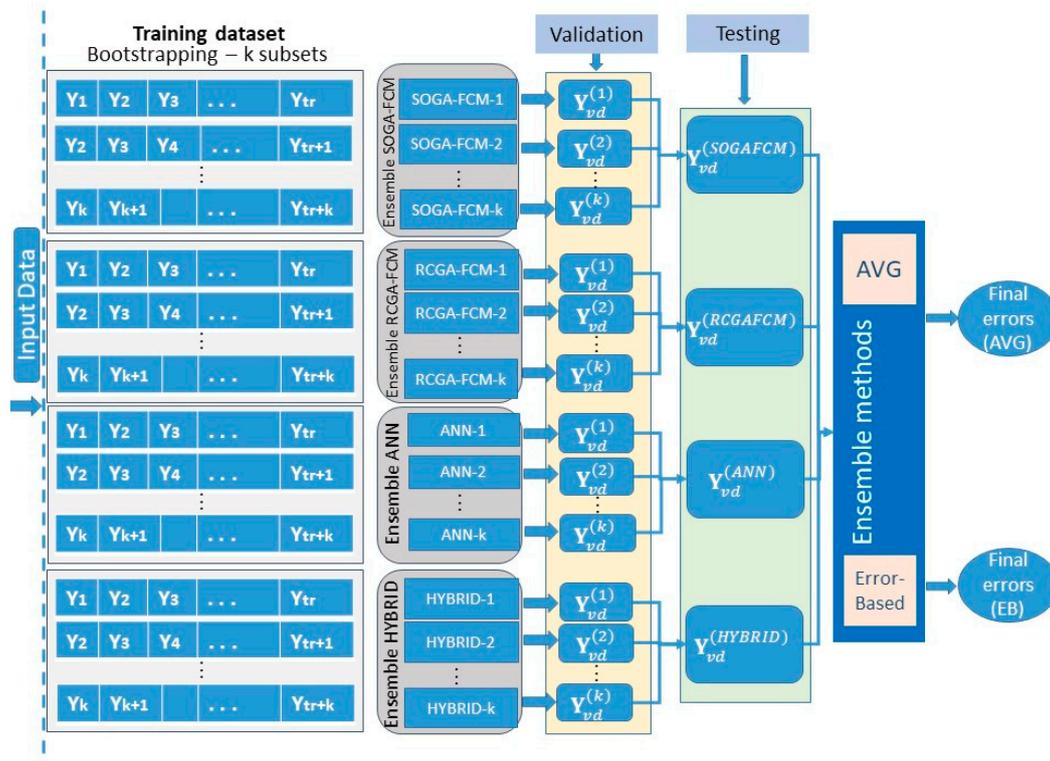


Figure 3. The proposed forecasting combination approach using ensemble methods and ensemble forecasters.

Step 3. We had n component forecasting models and obtained $\hat{Y}_{ts}^i = [\hat{y}_{N_{in}+1}^i, \hat{y}_{N_{in}+2}^i, \dots, \hat{y}_{N_{in}+N_{ts}}^i]^T$ as the forecast of Y_{ts} through the i^{th} model.

Step 4. We implemented each model on Y_{tr} and used it to predict Y_{vd} . Let $\hat{Y}_{vd}^i = [\hat{y}_{N_{tr}+1}^i, \hat{y}_{N_{tr}+2}^i, \dots, \hat{y}_{N_{tr}+N_{vd}}^i]^T$ be the prediction of Y_{vd} through the i^{th} model.

Step 5. We found the in-sample forecasting error of each model through some suitable error measures. We used the mean absolute error (MAE) and the mean squared error (MSE). These are widely popular error statistics [68], and their mathematical formulation is presented below in this paper. In the present study, we adopted the MSE and MAE to find the in-sample forecasting errors of the component models.

Step 6. Based on the obtained in-sample forecasting errors, we assigned a score to each component model as $\gamma_i = \frac{1}{MSE Y_{vd}, \hat{Y}_{vd}^i}, \forall i = 1, 2, \dots, n$. The scores are assigned to be inversely proportional to the respective errors so that a model with a comparatively smaller in-sample error receives more score and vice versa.

Step 7. We assigned a rank $r_i \in 1, 2, \dots, n$ to the i^{th} model, based on its score, so that $r_i \geq r_j$, if $\gamma_i \leq \gamma_j, \forall i, j = 1, 2, \dots, n$. The minimum, i.e., the best rank is equal to 1 and the maximum, i.e., the worst rank is at most equal to n .

Step 8. We chose a number n_r so that $1 \leq n_r \leq n$ and let $I = i_1, i_2, \dots, i_{n_r}$ be the index set of the n_r component models, whose ranks are in the range $[1, n_r]$. So, we selected a subgroup of n_r smallest ranked component models.

Step 9. Finally, we obtained the weighted linear combination of these selected n_r component forecasts, as follows:

$$\hat{y}_k = w_{i_1} \hat{y}_k^{i_1} + w_{i_2} \hat{y}_k^{i_2} + \dots + w_{i_{n_r}} \hat{y}_k^{i_{n_r}} = \sum_{i \in I} w_i \hat{y}_k^i, \forall i = 1, 2, \dots, n \quad (19)$$

Here, $w_{i_k} = \frac{\gamma_{i_k}}{\sum_{k=1}^{n_r} \gamma_{i_k}}$ is the normalized weight to the selected component model, so that $\sum_{k=1}^{n_r} w_{i_k} = 1$.

Step 10. The simple average method could be also adopted, as an alternative to Step 6–9, to calculate the forecasted value.

The validation set was used during the training process for updating the algorithm weights appropriately and, thus, improving its performance and avoiding overfitting. After training the model, we could run it on the testing data, to verify if it has predicted them correctly and, if it has been so, to keep hidden the validation set.

The most popular and widely used performance metrics or evaluation criteria for time series prediction are the following: coefficient of determination (R²), mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The mathematical equations of all these statistical indicators were described in the study [69]. The goodness of fit and the performance of the studied models, when they applied to a natural gas prediction process, were evaluated and compared using two of these five commonly used statistical indicators, namely, the MSE and the MAE [9]. In particular, the performance of the analyzed approaches for natural gas prediction was evaluated based on the following criteria:

1. Mean squared error:

$$\text{MSE} = \frac{1}{T} \sum_{t=1}^T (Z(t) - X(t))^2 \quad (20)$$

2. Mean absolute error:

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T |Z(t) - X(t)| \quad (21)$$

where $X(t)$ is the predicted value of the neural gas at the t th iteration, $Z(t)$ is the desired value of the neural gas at the t th iteration, $t = 1, \dots, N_{ts}$, and N_{ts} is the number of testing records. The lower values of the MSE and MAE indicate that the model performance is better with respect to the prediction accuracy, and the regression line fits the data well.

All the modeling approaches, tests, and evaluations were performed with the use of the ISEMK (intelligent expert system based on cognitive maps) software tool [66], in which all the algorithms based on ANNs, FCMs, and their hybrid combinations were developed. C# programming language has been used for implementing ensemble models and also for developing ISEMK, which incorporates FCM construction from data and learning, both for RCGA and SOGA implementations [69].

4. Results and Discussion

4.1. Case Study and Datasets

The natural gas consumption datasets that were used in this research work to examine the applicability and effectiveness of the proposed forecast methodology corresponded to five years (2013–2017), as described in Section 3. Following the first step of the methodology, we split our dataset into training, validation, and testing ones. For the convenience of handling properly the dataset, we defined the data of the first three years as the training dataset (1095 days), the data of the fourth year as the validation dataset (365 days), and the remaining data (5th year) as the testing dataset (365 days), which approximately corresponded to 60%, 20%, and 20%, respectively, as presented in Section 3. Thus, it was easier for our analysis to handle the above values as annual datasets and have a clearer perception of the whole process.

Out of the three years of the defined training dataset, we used the first two as the initial training dataset, while the third (3rd) year was used as a dataset reservoir for the bootstrapping procedure. This year was properly selected to be part of the initial dataset, as for each value of k (the bootstrapping step), a corresponding number of days/weeks/months was additionally needed to be included in the training dataset during the bootstrapping process, thus, avoiding any possible data shortage and/or deterioration that would lead to inaccurate results.

The proposed forecast combination approach, presented in Section 3, offered generalization capabilities and, thus, it could be applied in various time-series datasets, for a different number of k , according to daily, weekly, or monthly prediction. Taking as an example the case of a month-ahead prediction, for each bootstrapping step k , the training dataset shifted one month ahead, getting one additional month each time from the reserved third year of the initial training dataset. In this case, k more months in total were needed for implementing efficiently this approach. If we considered $k = 12$, then 12 additional months of the initial dataset needed to be added and reserved. This approach justified our case where one year (i.e., the third year) was added to the initial training dataset and was further reserved for serving the purposes of the proposed methodology. Different values of k were also examined without noticing significant divergences in forecasting, compared to the selected k value.

In the next step, the validation procedure (comprising one year of data) was implemented to calculate the in-sample forecasting errors (MSE and MAE) for each ensemble forecasting algorithm (ensemble ANN, ensemble hybrid, ensemble RCGA-FCM, and ensemble SOGA-FCM). The same process was followed for the testing procedure by considering the data of the last year. The two examined ensemble forecasting methods, i.e., the simple average (AVG) and the error-based (EB), were then applied in the calculated validation and testing vectors (Y_{vd}) for each one of the forecast combined methodology (Y_{vd} -ANN, Y_{vd} -Hybrid, Y_{vd} -RCGA, Y_{vd} -SOGA).

4.2. Case Study Results

In this study, we applied both the AVG and the EB method in two different cases: case (A) where scores were calculated for individual forecaster of each one of the methods ANN, hybrid, RCGA-FCM, and SOGA-FCM, and case (B), where scores were calculated for each ensemble forecaster (ANN ensemble, hybrid ensemble, RCGA-FCM ensemble, and SOGA-FCM ensemble).

Considering case (A), Table 1 shows the calculated errors and scores based on the EB method for individual forecaster of the two forecasting methods: ANN and hybrid for the city of Athens. The rest calculated errors and scores, based on the EB method, for individual forecaster for the other two remaining forecasting methods RCGA-FCM and SOGA-FCM for Athens can be found in Appendix A of the paper (Table A3). In Appendix A, parts of the corresponding results for the other two examined cities (Larissa and Thessaloniki) are also presented (Tables A4 and A5).

Table 1. Case (A)-Calculated errors and weights for each ensemble forecaster based on scores for EB (error-based) method (Athens).

| | Validation | | Testing | | | Testing | | | |
|-------|------------|--------|---------|--------|---------|----------|--------|---------|--------|
| | MAE | MSE | MAE | MSE | Weights | MAE | MSE | Weights | |
| ANN1 | 0.0334 | 0.0035 | 0.0350 | 0.0036 | 0.2552 | Hybrid1 | 0.0336 | 0.0034 | 0.2520 |
| ANN2 | 0.0354 | 0.0041 | 0.0387 | 0.0043 | 0 | Hybrid2 | 0.0387 | 0.0043 | 0 |
| ANN3 | 0.0350 | 0.0037 | 0.0375 | 0.0039 | 0.2442 | Hybrid3 | 0.0363 | 0.0037 | 0 |
| ANN4 | 0.0341 | 0.0038 | 0.0365 | 0.0039 | 0 | Hybrid4 | 0.0352 | 0.0035 | 0 |
| ANN5 | 0.0335 | 0.0036 | 0.0358 | 0.0037 | 0.2505 | Hybrid5 | 0.0339 | 0.0034 | 0 |
| ANN6 | 0.0337 | 0.0039 | 0.0355 | 0.0038 | 0 | Hybrid6 | 0.0348 | 0.0036 | 0.2468 |
| ANN7 | 0.0336 | 0.0037 | 0.0362 | 0.0038 | 0 | Hybrid7 | 0.0345 | 0.0035 | 0.2506 |
| ANN8 | 0.0340 | 0.0039 | 0.0360 | 0.0039 | 0 | Hybrid8 | 0.0354 | 0.0036 | 0 |
| ANN9 | 0.0341 | 0.0039 | 0.0367 | 0.0040 | 0 | Hybrid9 | 0.0349 | 0.0036 | 0 |
| ANN10 | 0.0332 | 0.0036 | 0.0355 | 0.0037 | 0.2501 | Hybrid10 | 0.0359 | 0.0038 | 0 |
| ANN11 | 0.0338 | 0.0038 | 0.0365 | 0.0039 | 0 | Hybrid11 | 0.0353 | 0.0038 | 0 |
| ANN12 | 0.0345 | 0.0038 | 0.0349 | 0.0037 | 0 | Hybrid12 | 0.0347 | 0.0033 | 0.2506 |
| AVG | 0.0336 | 0.0037 | 0.0359 | 0.0038 | | AVG | 0.0350 | 0.0036 | |
| EB | 0.0335 | 0.0036 | 0.0358 | 0.0037 | | EB | 0.0340 | 0.0034 | |

MSE: Mean Square Error, MAE: Mean Absolute Error.

Considering case (B), Table 2 presents the calculated weights based on scores for each ensemble forecaster (ANN (ensemble, hybrid ensemble, RCGA ensemble, and SOGA ensemble) for all three cities.

The calculated weights, based on scores for the EB method, were computed using Equation (17). According to this equation, the weights of the component forecasts are inversely proportional to their in-sample forecasting errors, concluding that the model with more error is assigned less weight to it and vice versa [80]. In this work, as the values of errors were high for certain ensemble forecasters, the corresponding weights were approximately zero, so they were considered to have a zero value for further predictions.

Table 2. Case (B)-Calculated weights for each ensemble forecaster based on scores for the EB method.

| | Athens | Thessaloniki | Larissa |
|----------|--------------------------------|--------------|---------|
| | <i>Weights based on scores</i> | | |
| ANN | 0.3320 | 0.34106 | 0.3369 |
| Hybrid | 0.3357 | 0.35162 | 0.3546 |
| RCGA-FCM | 0.3323 | 0 | 0 |
| SOGA-FCM | 0 | 0.30731 | 0.3083 |

ANN: Artificial Neural Network, RCGA-FCM: Real Codded Genetic Algorithm-Fuzzy Cognitive Map, SOGA-FCM: Structure Optimization Genetic Algorithm-Fuzzy Cognitive Map.

The obtained forecasting results of the individual and combination methods are depicted in Tables 3–8, respectively, for the three cities. In each of these tables, the best results (i.e., those associated with the least values of error measures) are presented in bold letters. In Figures 4 and 5, the forecasting results concerning Thessaloniki and Larissa are visually illustrated for both ensemble methods (AVG, EB). Moreover, Figure 6 gathers the forecasting results for all three cities considering the best ensemble method.

Table 3. Calculated errors for individual forecaster based on scores for Athens.

| Validation | ANN | Hybrid | RCGA | SOGA | Ensemble AVG | Ensemble EB |
|------------|--------|--------|--------|--------|---------------|---------------|
| MAE | 0.0328 | 0.0333 | 0.0384 | 0.0391 | 0.0336 | 0.0326 |
| MSE | 0.0036 | 0.0035 | 0.0036 | 0.0037 | 0.0032 | 0.0032 |
| Testing | | | | | | |
| MAE | 0.0321 | 0.0328 | 0.0418 | 0.0424 | 0.0345 | 0.0328 |
| MSE | 0.0033 | 0.0032 | 0.0038 | 0.0040 | 0.0032 | 0.0031 |

Table 4. Calculated errors for each ensemble forecaster based on scores for Athens.

| Validation | ANN Ensemble | Hybrid Ensemble | RCGA Ensemble | SOGA Ensemble | Ensemble AVG | Ensemble EB |
|------------|--------------|-----------------|---------------|---------------|---------------|---------------|
| MAE | 0.0335 | 0.0330 | 0.0388 | 0.0380 | 0.0337 | 0.0337 |
| MSE | 0.0036 | 0.0035 | 0.0036 | 0.0035 | 0.0032 | 0.0032 |
| Testing | | | | | | |
| MAE | 0.0358 | 0.0340 | 0.0422 | 0.0422 | 0.0352 | 0.0352 |
| MSE | 0.0037 | 0.0034 | 0.0038 | 0.0037 | 0.0032 | 0.0032 |

Table 5. Calculated errors for individual forecaster based on scores for Thessaloniki.

| Validation | ANN | Hybrid | RCGA | SOGA | Ensemble AVG | Ensemble EB |
|------------|--------|--------|--------|--------|--------------|---------------|
| MAE | 0.0343 | 0.0341 | 0.0381 | 0.0380 | 0.0347 | 0.0340 |
| MSE | 0.0029 | 0.0028 | 0.0032 | 0.0032 | 0.0028 | 0.0027 |
| Testing | | | | | | |
| MAE | 0.0366 | 0.0381 | 0.0395 | 0.0399 | 0.0371 | 0.0369 |
| MSE | 0.0032 | 0.0033 | 0.0035 | 0.0036 | 0.0032 | 0.0031 |

Table 6. Calculated errors for each ensemble forecaster based on scores for Thessaloniki.

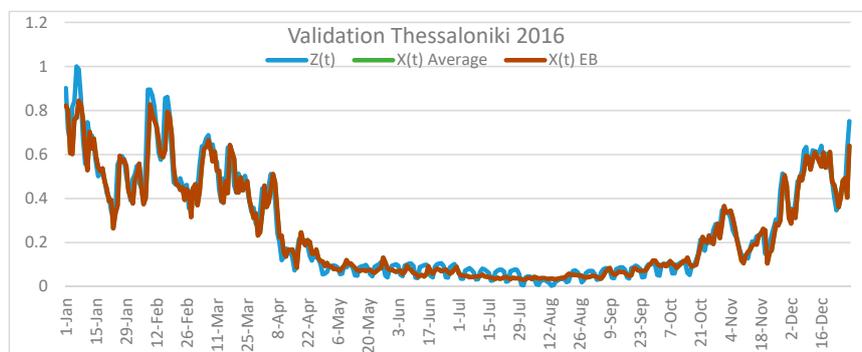
| Validation | ANN Ensemble | Hybrid Ensemble | RCGA Ensemble | SOGA Ensemble | Ensemble AVG | Ensemble EB |
|----------------|--------------|-----------------|---------------|---------------|---------------|---------------|
| MAE | 0.0363 | 0.0361 | 0.0378 | 0.0374 | 0.0355 | 0.0355 |
| MSE | 0.0031 | 0.0031 | 0.0031 | 0.0030 | 0.0028 | 0.0028 |
| Testing | | | | | | |
| MAE | 0.0393 | 0.0394 | 0.0399 | 0.0391 | 0.0381 | 0.0381 |
| MSE | 0.0037 | 0.0037 | 0.0036 | 0.0034 | 0.0034 | 0.0034 |

Table 7. Calculated errors for individual forecaster based on scores for Larissa.

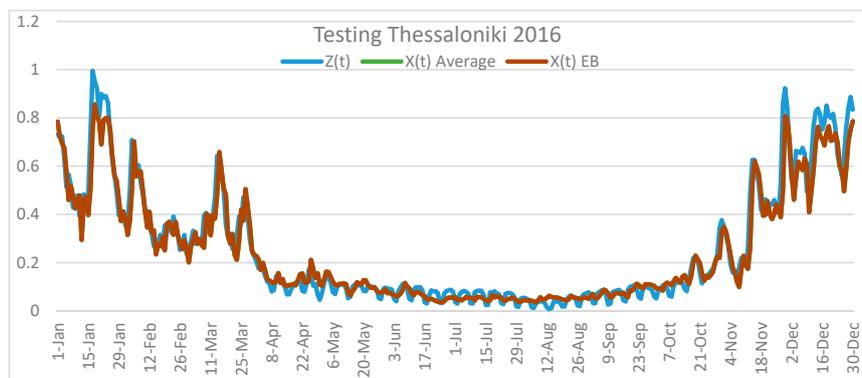
| Validation | ANN | Hybrid | RCGA | SOGA | Ensemble AVG | Ensemble EB |
|----------------|--------|--------|--------|--------|---------------|---------------|
| MAE | 0.0322 | 0.0324 | 0.0372 | 0.0365 | 0.0326 | 0.0319 |
| MSE | 0.0030 | 0.0028 | 0.0033 | 0.0032 | 0.0027 | 0.0027 |
| Testing | | | | | | |
| MAE | 0.0412 | 0.0417 | 0.0466 | 0.0468 | 0.0427 | 0.0417 |
| MSE | 0.0043 | 0.0041 | 0.0047 | 0.0047 | 0.0040 | 0.0040 |

Table 8. Case (B) –Calculated errors for each ensemble forecaster based on scores for Larissa.

| Validation | ANN Ensemble | Hybrid Ensemble | RCGA Ensemble | SOGA Ensemble | Ensemble AVG | Ensemble EB |
|----------------|--------------|-----------------|---------------|---------------|--------------|---------------|
| MAE | 0.0337 | 0.0332 | 0.0371 | 0.0362 | 0.0329 | 0.0326 |
| MSE | 0.0032 | 0.0030 | 0.0032 | 0.0031 | 0.0027 | 0.0026 |
| Testing | | | | | | |
| MAE | 0.0428 | 0.0417 | 0.0458 | 0.0460 | 0.0426 | 0.0423 |
| MSE | 0.0048 | 0.0044 | 0.0045 | 0.0045 | 0.0041 | 0.0040 |

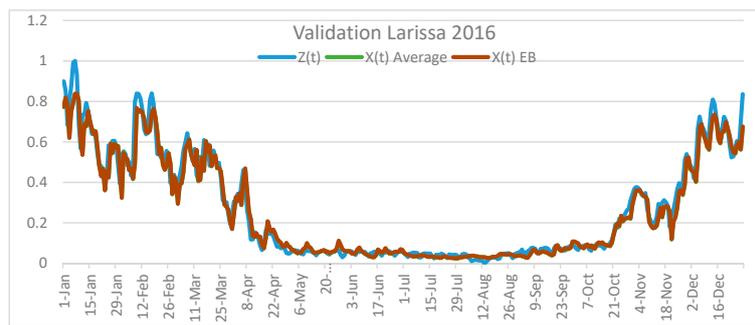


(a)

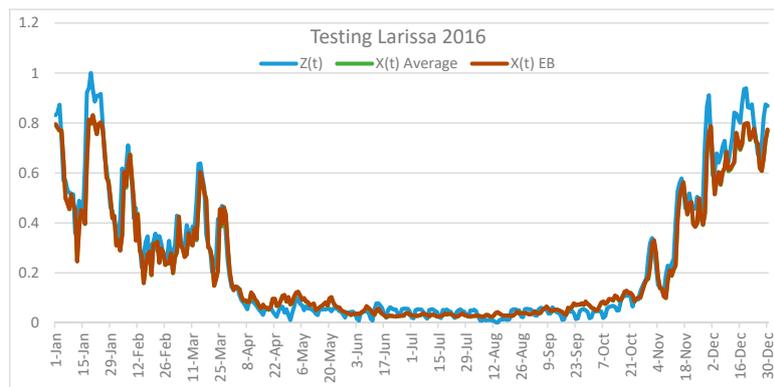


(b)

Figure 4. Forecasting results for Thessaloniki considering the two ensemble methods (AVG, EB) based on scores. (a) Validation, (b) Testing. AVG, simple average; EB, error-based.

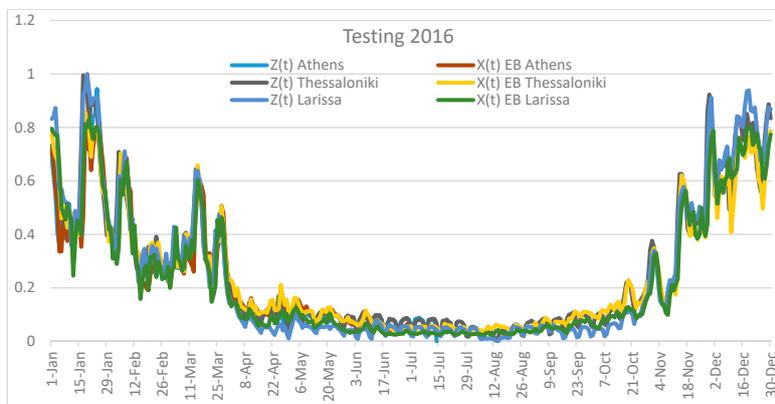


(a)

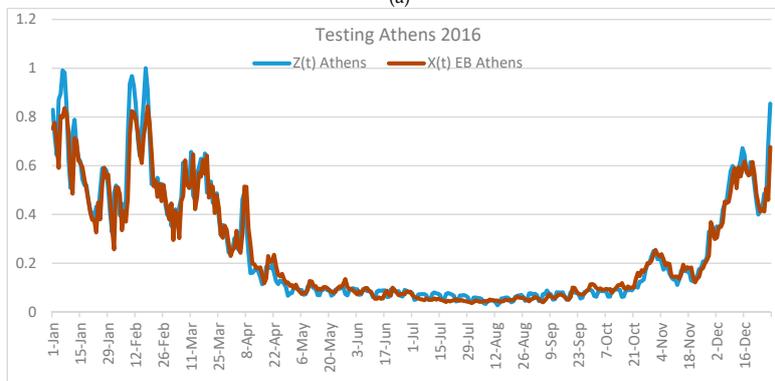


(b)

Figure 5. Forecasting results for Larissa considering the two ensemble methods (AVG, EB) based on scores. (a) Validation, (b) Testing.



(a)



(b)

Figure 6. Cont.

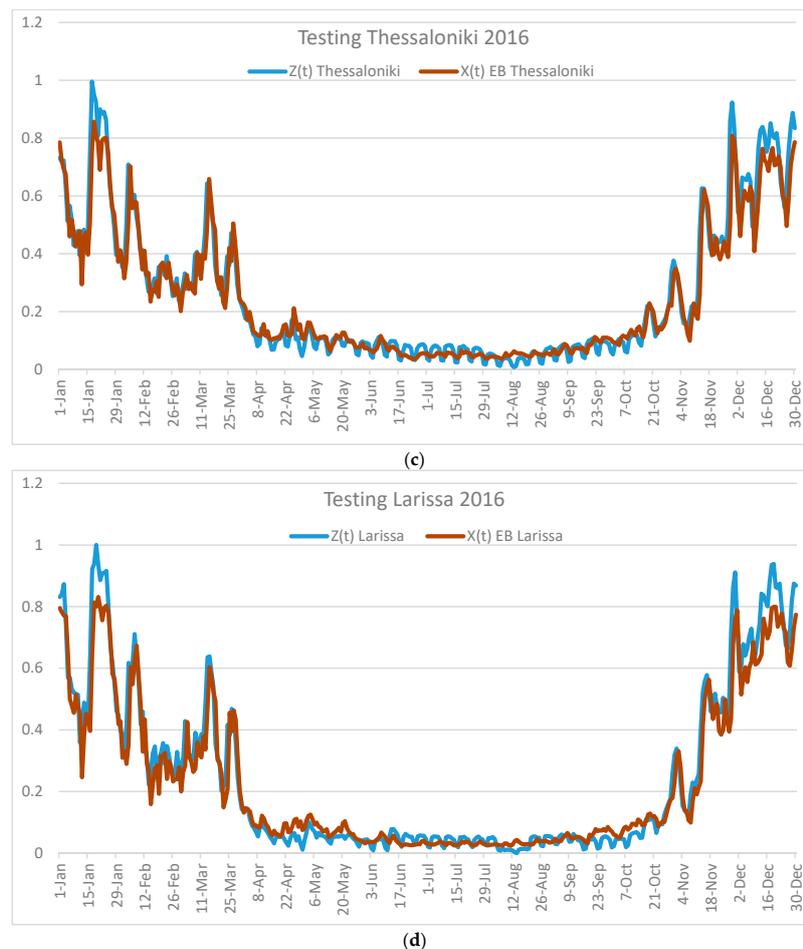


Figure 6. Forecasting results for the three cities considering the best ensemble method. (a) Testing all cities, (b) Testing Athens, (c) Testing Thessaloniki, (d) Testing Larissa.

4.3. Discussion of Results

The following important observations were noticed after careful analysis of Tables and Figures above.

1. After a thorough analysis of the Tables 3–8, on the basis of examining the MAE and MSE errors, it could be clearly stated that the EB method presented lower errors concerning the individual forecasters (ANN, hybrid, RCGA-FCM, and SOGA-FCM) for all three cities (Athens, Thessaloniki, and Larisa). EB seemed to outperform the AVG method in terms of achieving overall better forecasting results when applied to individual forecasters (see Figure 6).
2. Considering the ensemble forecasters, it could be seen from the obtained results that none of the two forecast combination methods had attained consistently better accuracies compared to each other, as far as the cities of Athens and Thessaloniki were concerned. Specifically, from Tables 3–6, it was observed that the MAE and MSE values across the two combination methods were similar for the two cities; however, their errors were lower than those produced by each separate ensemble forecaster.

3. Although the AVG and the EB methods performed similarly for Athens and Thessaloniki datasets, the EB forecast combination technique presented lower MAE and MSE errors than the AVG for the examined dataset of Larissa (see Figure 5).

The fact that when a forecasting method presented lower MAE and MSE errors than another means that the accuracy of the results produced with the first method, in terms of predicting consumption, was higher than the latter forecasting methods examined and compared to, so as the overall performance of the ensemble method was. Regarding the amount of improvement that was presented when a forecasting method was applied, slightly better performance of both ensemble forecasting methods could be noticed, and that constituted strong evidence for the efficiency of the examined method in the domain of natural gas demand forecasting.

In order to examine the efficiency of the proposed algorithm, a statistical test was conducted to reveal no statistical significance. Concerning the individual methods, a *t*-test paired two samples of mean was previously conducted in [60] for the cities of Thessaly (Larissa, Volos, Trikala, and Karditsa), for the year 2016, showing that there was no statistical significance among these techniques. In current work, a *t*-test paired two samples of mean was also performed, regarding the ensemble methods (average and error-based) for the examined cities (Athens, Thessaloniki, and Larissa), regarding the dataset of the same year. The results of the hypothesis tests (Tables A6–A8 in Appendix A) revealed no statistical significance between these techniques. In all cases, the calculated *p*-value exceeded 0.05, so no statistical significance was noticed from the obtained statistical analysis. Therefore, there was no particular need to conduct a post hoc statistical test, since a post hoc test should only be run when you have an overall statistically significant difference in group means, according to the relevant literature [89,90].

Furthermore, for comparison purposes, to show the effectiveness of the proposed forecasting combination approach of multivariate time series, the experimental analysis was conducted with a new and well-known effective machine learning technique for time series forecasting, the LSTM (long short-term memory). LSTM algorithm encloses the characteristics of the advanced recurrent neural network methods and is mainly applied for time series prediction problems in diverse domains [91].

LSTM was applied in one day-ahead natural gas consumption prediction concerning the same dataset of the three Greek cities (Athens, Thessaloniki, and Larissa) in [70]. For the LSTM implementation, one feature of the dataset as a time series was selected. As explained in [70], LSTM was fed previous values, and, in that case, the time-step was set to be 364 values to predict the next 364. For validation, 20% of random data from the training dataset was used, and for testing, the same dataset that was used for the ANN, RCGA-FCM, SOGA-FCM, and hybrid FCM-ANN, as well as with their ensemble structures implementation. In [70], various experiments with different numbers of units, number of layers, and dropout rates were accomplished. Through the provided experimental analysis, the best results of LSTM emerged for one layer, 200 units, and dropout rate = 0.2. These results are gathered in Table 9 for the three cities.

In Table 9, it is clear that both ensemble forecasting methods can achieve high accuracy in the predictions of the energy consumption patterns in a day-ahead timescale. Additional exploratory analysis and investigation of other types of ensemble methods, as well as other types of neural networks, such as convolutional neural networks (CNNs), could lead to a better insight of the modeling the particular problem and achieve higher prediction accuracy.

Table 9. Comparison results with LSTM (long short-term memory) (with best configuration parameters).

| | Best Ensemble | | LSTM (Dropout = 0.2) |
|-------------------|-----------------------|---------------------|----------------------|
| | Case (A) (Individual) | Case (B) (Ensemble) | 1 layer |
| Validation | ATHENS | | |
| MAE | 0.0326 | 0.0337 | 0.0406 |
| MSE | 0.0032 | 0.0032 | 0.0039 |
| Testing | | | |
| MAE | 0.0328 | 0.0352 | 0.0426 |
| MSE | 0.0031 | 0.0032 | 0.0041 |
| Validation | THESSALONIKI | | |
| MAE | 0.0340 | 0.0355 | 0.0462 |
| MSE | 0.0027 | 0.0028 | 0.0043 |
| Testing | | | |
| MAE | 0.0369 | 0.0381 | 0.0489 |
| MSE | 0.0031 | 0.0034 | 0.0045 |
| Validation | LARISSA | | |
| MAE | 0.0319 | 0.0326 | 0.0373 |
| MSE | 0.0027 | 0.0026 | 0.0029 |
| Testing | | | |
| MAE | 0.0417 | 0.0423 | 0.0462 |
| MSE | 0.0040 | 0.0040 | 0.0042 |

5. Conclusions

To sum up, we applied a time series forecasting method for natural gas demand in three Greek cities, implementing an efficient ensemble forecasting approach through combining ANN, RCGA-FCM, SOGA-FCM, and hybrid FCM-ANN. The proposed forecasting combination approach incorporates the two most popular ensemble methods for error calculation in forecasting problems and is deployed in certain steps offering generalization capabilities. The whole framework seems to be a promising approach for ensemble time series forecasting that can easily be applied in many scientific domains. An initial comparison analysis was conducted with benchmark methods of ANN, FCM, and their different configurations. Next, further comparison analysis was conducted with new promising LSTM networks previously used for time series prediction.

Through the experimental analysis, two error statistics (MAE, MSE) needed to be calculated in order to examine the effectiveness of the ensemble learning approach in time series prediction. The results of this study showed that the examined ensemble approach through designing an ensemble structure of various ANN, SOGA-FCM models by different learning parameters and their hybrid structures could significantly improve forecasting. Moreover, obtained results clearly demonstrated that a relatively higher forecasting accuracy was noticed when the applied ensemble approach was compared against independent forecasting approaches, such as ANN or FCM, as well as with LSTM.

Future work is devoted to applying the advantageous forecast combination approach to a larger number of distribution points that compose the natural gas grid of Greek regions (larger and smaller cities) as well as to investigate a new forecast combination structure of efficient convolutional neural networks (CNN) and LSTM networks for time series prediction in various application domains. Furthermore, an extensive comparative analysis with various LSTM structures, as well as with other advanced machine learning and time series prediction methods, will be conducted in future work. The presented research work could also contribute to explainability, transparency, and re-traceability of artificial intelligence (AI) and machine learning systems. These systems are being applied in various fields, and the decisions being made by them are not always clear due to the use of complicated algorithms in order to achieve power, performance, and accuracy. The authors with the use of complicated, but powerful algorithms, such as neural networks and ensemble methods, tried to describe all the steps and models involved in decision-making process to attain explainability and, in future, they would further explore ways to make the best-performing methods more transparent, re-traceable, and understandable, explaining why certain decisions have been made [92].

Author Contributions: For conceptualization, K.I.P.; methodology, K.I.P.; software, K.P.; designed and performed the experiments, K.I.P. and K.P.; analyzed the results, drafted the initial manuscript, and revised the final manuscript, V.C.G., K.I.P., and E.P.; supervised the work, E.P., V.C.G., and G.S.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Descriptive statistics values for real dataset Z(t), forecasting values of AVG and EB methods for the three cities (validation).

| Descriptive Statistics | Athens | | | Thessaloniki | | | Larissa | | |
|------------------------|--------|---------|---------|--------------|---------|---------|---------|---------|---------|
| | Z(t) | X(t)AVG | X(t)EB | Z(t) | X(t)AVG | X(t)EB | Z(t) | X(t)AVG | X(t)EB |
| Mean | 0.2540 | 0.2464 | 0.2464 | 0.2611 | 0.2510 | 0.2510 | 0.2689 | 0.2565 | 0.2575 |
| Median | 0.1154 | 0.1366 | 0.1366 | 0.1335 | 0.1393 | 0.1394 | 0.1037 | 0.1194 | 0.1211 |
| St. Deviation | 0.2391 | 0.2203 | 0.2203 | 0.2373 | 0.2228 | 0.2228 | 0.2604 | 0.2429 | 0.2429 |
| Kurtosis | 0.3610 | -0.2748 | -0.2741 | -0.1839 | -0.5807 | -0.5774 | -0.6564 | -0.8881 | -0.8847 |
| Skewness | 1.1605 | 0.9801 | 0.9803 | 0.9328 | 0.8288 | 0.8298 | 0.8112 | 0.7520 | 0.7516 |
| Minimum | 0.0277 | 0.0367 | 0.0367 | 0.0043 | 0.0305 | 0.0304 | 0.0000 | 0.0235 | 0.0239 |
| Maximum | 1.0000 | 0.8429 | 0.8431 | 1.0000 | 0.8442 | 0.8448 | 1.0000 | 0.8361 | 0.8383 |

Table A2. Descriptive statistics values for real dataset Z(t), forecasting values of AVG and EB methods for the three cities (testing).

| Descriptive Statistics | Athens | | | Thessaloniki | | | Larissa | | |
|------------------------|--------|---------|--------|--------------|---------|---------|---------|---------|---------|
| | Z(t) | X(t)AVG | X(t)EB | Z(t) | X(t)AVG | X(t)EB | Z(t) | X(t)AVG | X(t)EB |
| Mean | 0.2479 | 0.2433 | 0.2433 | 0.2588 | 0.2478 | 0.2478 | 0.2456 | 0.2279 | 0.2291 |
| Median | 0.1225 | 0.1488 | 0.1488 | 0.1179 | 0.1304 | 0.1304 | 0.0695 | 0.0961 | 0.0972 |
| St. Deviation | 0.2159 | 0.2020 | 0.2021 | 0.2483 | 0.2236 | 0.2237 | 0.2742 | 0.2399 | 0.2404 |
| Kurtosis | 0.6658 | 0.2785 | 0.2792 | 0.1755 | -0.1254 | -0.1219 | -0.0113 | -0.1588 | -0.1502 |
| Skewness | 1.2242 | 1.1138 | 1.1140 | 1.1348 | 1.0469 | 1.0479 | 1.1205 | 1.0900 | 1.0921 |
| Minimum | 0.0000 | 0.0359 | 0.0359 | 0.0079 | 0.0358 | 0.0357 | 0.0000 | 0.0233 | 0.0237 |
| Maximum | 0.9438 | 0.8144 | 0.8144 | 0.9950 | 0.8556 | 0.8562 | 1.0000 | 0.8291 | 0.8310 |

Table A3. Case (A)-Calculated errors and weights for each ensemble forecaster (RCGA and SOGA-FCM) based on scores for the EB method (Athens).

| | Validation | | Testing | | | Testing | | | |
|--------|------------|--------|---------|--------|---------|---------|--------|---------|--------|
| | MAE | MSE | MAE | MSE | Weights | MAE | MSE | Weights | |
| RCGA1 | 0.0386 | 0.0036 | 0.0425 | 0.0038 | 0.2531 | SOGA1 | 0.0435 | 0.0037 | 0.2520 |
| RCGA2 | 0.0391 | 0.0038 | 0.0430 | 0.0039 | 0 | SOGA2 | 0.0423 | 0.0038 | 0.2509 |
| RCGA3 | 0.0399 | 0.0039 | 0.0428 | 0.0039 | 0 | SOGA3 | 0.0425 | 0.0038 | 0 |
| RCGA4 | 0.0384 | 0.0036 | 0.0419 | 0.0038 | 0.2522 | SOGA4 | 0.0449 | 0.0042 | 0 |
| RCGA5 | 0.0389 | 0.0037 | 0.0423 | 0.0039 | 0 | SOGA5 | 0.0429 | 0.0040 | 0 |
| RCGA6 | 0.0392 | 0.0036 | 0.0424 | 0.0039 | 0.2472 | SOGA6 | 0.0432 | 0.0038 | 0.2494 |
| RCGA7 | 0.0398 | 0.0038 | 0.0434 | 0.0041 | 0 | SOGA7 | 0.0421 | 0.0039 | 0 |
| RCGA8 | 0.0386 | 0.0037 | 0.0416 | 0.0039 | 0 | SOGA8 | 0.0422 | 0.0039 | 0 |
| RCGA9 | 0.0398 | 0.0036 | 0.0436 | 0.0041 | 0.2472 | SOGA9 | 0.0434 | 0.0042 | 0 |
| RCGA10 | 0.0388 | 0.0037 | 0.0417 | 0.0039 | 0 | SOGA10 | 0.0422 | 0.0040 | 0 |
| RCGA11 | 0.0393 | 0.0038 | 0.0419 | 0.0039 | 0 | SOGA11 | 0.0420 | 0.0038 | 0.2475 |
| RCGA12 | 0.0396 | 0.0037 | 0.0434 | 0.0041 | 0 | SOGA12 | 0.0425 | 0.0040 | 0 |
| AVG | 0.0385 | 0.0036 | 0.0418 | 0.0038 | | AVG | 0.0422 | 0.0039 | |
| EB | 0.0388 | 0.0036 | 0.0422 | 0.0038 | | EB | 0.0422 | 0.0037 | |

Table A4. Case (A)-Calculated errors and weights for each ensemble forecaster based on scores for the EB method (Thessaloniki).

| | Validation | | Testing | | | Testing | | | |
|----------|------------|--------|---------|--------|---------|---------|--------|---------|--------|
| | MAE | MSE | MAE | MSE | Weights | MAE | MSE | Weights | |
| Hybrid1 | 0.0356 | 0.0030 | 0.0390 | 0.0036 | 0.2565 | SOGA1 | 0.0414 | 0.0040 | 0 |
| Hybrid2 | 0.0381 | 0.0036 | 0.0409 | 0.0042 | 0 | SOGA2 | 0.0417 | 0.0040 | 0 |
| Hybrid3 | 0.0371 | 0.0032 | 0.0398 | 0.0039 | 0.2422 | SOGA 3 | 0.0394 | 0.0034 | 0 |
| Hybrid4 | 0.0376 | 0.0032 | 0.0403 | 0.0039 | 0 | SOGA 4 | 0.0406 | 0.0038 | 0 |
| Hybrid5 | 0.0373 | 0.0032 | 0.0401 | 0.0040 | 0 | SOGA 5 | 0.0388 | 0.0033 | 0.2541 |
| Hybrid6 | 0.0375 | 0.0033 | 0.0403 | 0.0040 | 0 | SOGA 6 | 0.0413 | 0.0038 | 0 |
| Hybrid7 | 0.0378 | 0.0033 | 0.0405 | 0.0040 | 0 | SOGA 7 | 0.0415 | 0.0039 | 0 |
| Hybrid8 | 0.0373 | 0.0032 | 0.0402 | 0.0040 | 0 | SOGA 8 | 0.0399 | 0.0036 | 0 |
| Hybrid9 | 0.0378 | 0.0034 | 0.0407 | 0.0041 | 0 | SOGA 9 | 0.0392 | 0.0035 | 0.2448 |
| Hybrid10 | 0.0371 | 0.0032 | 0.0397 | 0.0039 | 0.2410 | SOGA10 | 0.0400 | 0.0037 | 0 |
| Hybrid11 | 0.0370 | 0.0033 | 0.0402 | 0.0040 | 0 | SOGA11 | 0.0403 | 0.0036 | 0.2439 |
| Hybrid12 | 0.0364 | 0.0030 | 0.0406 | 0.0036 | 0.2601 | SOGA12 | 0.0397 | 0.0034 | 0.2569 |
| AVG | 0.0369 | 0.0032 | 0.0398 | 0.0039 | | AVG | 0.0398 | 0.0036 | |
| EB | 0.0361 | 0.0031 | 0.0394 | 0.0037 | | EB | 0.0391 | 0.0034 | |

Table A5. Case (A)-Calculated errors and weights for each ensemble forecaster based on scores for the EB method (Larissa).

| | Validation | | Testing | | | Testing | | | |
|-------|------------|--------|---------|--------|---------|----------|--------|---------|--------|
| | MAE | MSE | MAE | MSE | Weights | MAE | MSE | Weights | |
| ANN1 | 0.0339 | 0.0032 | 0.0425 | 0.0047 | 0.2511 | Hybrid1 | 0.0411 | 0.0043 | 0.2531 |
| ANN2 | 0.0353 | 0.0036 | 0.0438 | 0.0052 | 0 | Hybrid2 | 0.0435 | 0.0051 | 0 |
| ANN3 | 0.0343 | 0.0033 | 0.0433 | 0.0050 | 0 | Hybrid3 | 0.0418 | 0.0045 | 0.2472 |
| ANN4 | 0.0347 | 0.0033 | 0.0429 | 0.0049 | 0 | Hybrid4 | 0.0424 | 0.0048 | 0 |
| ANN5 | 0.0353 | 0.0035 | 0.0436 | 0.0051 | 0 | Hybrid5 | 0.0436 | 0.0051 | 0 |
| ANN6 | 0.0352 | 0.0035 | 0.0432 | 0.0049 | 0 | Hybrid6 | 0.0436 | 0.0051 | 0 |
| ANN7 | 0.0354 | 0.0035 | 0.0441 | 0.0053 | 0 | Hybrid7 | 0.0434 | 0.0050 | 0 |
| ANN8 | 0.0348 | 0.0033 | 0.0427 | 0.0049 | 0 | Hybrid8 | 0.0425 | 0.0047 | 0.2398 |
| ANN9 | 0.0351 | 0.0035 | 0.0439 | 0.0052 | 0 | Hybrid9 | 0.0423 | 0.0047 | 0 |
| ANN10 | 0.0343 | 0.0033 | 0.0431 | 0.0049 | 0.2406 | Hybrid10 | 0.0432 | 0.0050 | 0 |
| ANN11 | 0.0342 | 0.0032 | 0.0436 | 0.0049 | 0.2472 | Hybrid11 | 0.0444 | 0.0053 | 0 |
| ANN12 | 0.0331 | 0.0031 | 0.0428 | 0.0047 | 0.2610 | Hybrid12 | 0.0426 | 0.0043 | 0.2597 |
| AVG | 0.0345 | 0.0033 | 0.0431 | 0.0049 | | AVG | 0.0427 | 0.0048 | |
| EB | 0.0337 | 0.0032 | 0.0428 | 0.0048 | | EB | 0.0417 | 0.0044 | |

Table A6. t-Test: Paired Two Sample for Means between the ensemble methods (AVG and EB) (Athens).

| | X(t) AVG Athens | X(t) EB Athens |
|------------------------------|-----------------|----------------|
| Mean | 0.243342155 | 0.243346733 |
| Variance | 0.040822427 | 0.040826581 |
| Observations | 196 | 196 |
| Pearson Correlation | 0.99999997 | |
| Hypothesized Mean Difference | 0 | |
| df | 195 | |
| t Stat | -1.278099814 | |
| P(T<=t) one-tail | 0.101366761 | |
| t Critical one-tail | 1.65270531 | |
| P(T<=t) two-tail | 0.202733521 | |
| t Critical two-tail | 1.972204051 | |

Table A7. t-Test: Paired Two Sample for Means between the ensemble methods (AVG and EB) (Thessaloniki).

| | X(t) AVG | X(t) EB |
|------------------------------|-------------|-------------|
| Mean | 0.247811356 | 0.247811056 |
| Variance | 0.050004776 | 0.050032786 |
| Observations | 365 | 365 |
| Pearson Correlation | 0.999999788 | |
| Hypothesized Mean Difference | 0 | |
| df | 364 | |
| t Stat | 0.036242052 | |
| P(T<=t) one-tail | 0.485554611 | |
| t Critical one-tail | 1.649050545 | |
| P(T<=t) two-tail | 0.971109222 | |
| t Critical two-tail | 1.966502569 | |

Table A8. t-Test: Paired Two Sample for Means between the ensemble methods (AVG and EB) (Larissa).

| | X(t) AVG Larisa | X(t) EB Larisa |
|------------------------------|-----------------|----------------|
| Mean | 0.227903242 | 0.229120614 |
| Variance | 0.057542802 | 0.05781177 |
| Observations | 365 | 365 |
| Pearson Correlation | 0.999972455 | |
| Hypothesized Mean Difference | 0 | |
| df | 364 | |
| t Stat | -12.44788062 | |
| P(T<=t) one-tail | 3.52722E-30 | |
| t Critical one-tail | 1.649050545 | |
| P(T<=t) two-tail | 7.05444E-30 | |
| t Critical two-tail | 1.966502569 | |

References

- Li, H.Z.; Guo, S.; Li, C.J.; Sun, J.Q. A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm. *Knowl.-Based Syst.* **2013**, *37*, 378–387. [CrossRef]
- Bodyanskiy, Y.; Popov, S. Neural network approach to forecasting of quasiperiodic financial time series. *Eur. J. Oper. Res.* **2006**, *175*, 1357–1366. [CrossRef]
- Livieris, I.E.; Kotsilieris, T.; Stavroyiannis, S.; Pintelas, P. Forecasting stock price index movement using a constrained deep neural network training algorithm. *Intell. Decis. Technol.* **2019**, 1–14. Available online: https://www.researchgate.net/publication/334132665_Forecasting_stock_price_index_movement_using_a_constrained_deep_neural_network_training_algorithm (accessed on 10 September 2019).
- Livieris, I.E.; Pintelas, H.; Kotsilieris, T.; Stavroyiannis, S.; Pintelas, P. Weight-constrained neural networks in forecasting tourist volumes: A case study. *Electronics* **2019**, *8*, 1005. [CrossRef]
- Chen, C.F.; Lai, M.C.; Yeh, C.C. Forecasting tourism demand based on empirical mode decomposition and neural network. *Knowl.-Based Syst.* **2012**, *26*, 281–287. [CrossRef]
- Lu, X.; Wang, J.; Cai, Y.; Zhao, J. Distributed HS-ARTMAP and its forecasting model for electricity load. *Appl. Soft Comput.* **2015**, *32*, 13–22. [CrossRef]
- Zeng, Y.R.; Zeng, Y.; Choi, B.; Wang, L. Multifactor-influenced energy consumption forecasting using enhanced back-propagation neural network. *Energy* **2017**, *127*, 381–396. [CrossRef]
- Lin, W.Y.; Hu, Y.H.; Tsai, C.F. Machine learning in financial crisis prediction: A survey. *IEEE Trans. Syst. Man Cybern.* **2012**, *42*, 421–436.
- Raza, M.Q.; Khosravi, A. A review on artificial intelligence-based load demand techniques for smart grid and buildings. *Renew. Sustain. Energy Rev.* **2015**, *50*, 1352–1372. [CrossRef]
- Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecast.* **2014**, *30*, 1030–1081. [CrossRef]

11. Meade, N.; Islam, T. Forecasting in telecommunications and ICT-A review. *Int. J. Forecast.* **2015**, *31*, 1105–1126. [[CrossRef](#)]
12. Donkor, E.; Mazzuchi, T.; Soyer, R.; Alan Roberson, J. Urban water demand forecasting: Review of methods and models. *J. Water Resour. Plann. Managem.* **2014**, *140*, 146–159. [[CrossRef](#)]
13. Fagiani, M.; Squartini, S.; Gabrielli, L.; Spinsante, S.; Piazza, F. A review of datasets and load forecasting techniques for smart natural gas and water grids: Analysis and experiments. *Neurocomputing* **2015**, *170*, 448–465. [[CrossRef](#)]
14. Chandra, D.R.; Kumari, M.S.; Sydulu, M. A detailed literature review on wind forecasting. In Proceedings of the IEEE International Conference on Power, Energy and Control (ICPEC), Sri Rangalatchum Dindigul, India, 6–8 February 2013; pp. 630–634.
15. Zhang, G. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **2003**, *50*, 159–175. [[CrossRef](#)]
16. Adhikari, R.; Verma, G.; Khandelwal, I. A model ranking based selective ensemble approach for time series forecasting. *Procedia Comput. Sci.* **2015**, *48*, 14–21. [[CrossRef](#)]
17. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C. *Time Series Analysis: Forecasting and Control*, 3rd ed.; Prentice-Hall: Englewood Cliffs, NJ, USA, 1994.
18. Kumar, K.; Jain, V.K. Autoregressive integrated moving averages (ARIMA) modeling of a traffic noise time series. *Appl. Acoust.* **1999**, *58*, 283–294. [[CrossRef](#)]
19. Ediger, V.S.; Akar, S. ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy* **2007**, *35*, 1701–1708. [[CrossRef](#)]
20. Poczeta, K.; Kubus, L.; Yastrebov, A.; Papageorgiou, E.I. Temperature forecasting for energy saving in smart buildings based on fuzzy cognitive map. In Proceedings of the 2018 Conference on Automation (Automation 2018), Warsaw, Poland, 21–23 March 2018; pp. 93–103.
21. Poczeta, K.; Yastrebov, A.; Papageorgiou, E.I. Application of fuzzy cognitive maps to multi-step ahead prediction of electricity consumption. In Proceedings of the IEEE 2018 Conference on Electrotechnology: Processes, Models, Control and Computer Science (EPMCCS), Kielce, Poland, 12–14 November 2018; pp. 1–5.
22. Khashei, M.; Rafiei, F.M.; Bijari, M. Hybrid fuzzy auto-regressive integrated moving average (FARIMAH) model for forecasting the foreign exchange markets. *Int. J. Comput. Intell. Syst.* **2013**, *6*, 954–968. [[CrossRef](#)]
23. Chu, F.L. A fractionally integrated autoregressive moving average approach to forecasting tourism demand. *Tour. Manag.* **2008**, *29*, 79–88. [[CrossRef](#)]
24. Yu, H.K.; Kim, N.Y.; Kim, S.S. Forecasting the number of human immunodeficiency virus infections in the Korean population using the autoregressive integrated moving average model. *Osong Public Health Res. Perspect.* **2013**, *4*, 358–362. [[CrossRef](#)]
25. Haykin, S. *Neural Networks: A Comprehensive Foundation*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 1999.
26. Doganis, P.; Alexandridis, A.; Patrinos, P.; Sarimveis, H. Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. *J. Food Eng.* **2006**, *75*, 196–204. [[CrossRef](#)]
27. Kim, M.; Jeong, J.; Bae, S. Demand forecasting based on machine learning for mass customization in smart manufacturing. In Proceedings of the ACM 2019 International Conference on Data Mining and Machine Learning (ICDMML 2019), Hong Kong, China, 28–30 April 2019; pp. 6–11.
28. Wang, D.M.; Wang, L.; Zhang, G.M. Short-term wind speed forecast model for wind farms based on genetic BP neural network. *J. Zhejiang Univ. (Eng. Sci.)* **2012**, *46*, 837–842.
29. Breiman, L. Stacked regressions. *Mach. Learn.* **1996**, *24*, 49–64. [[CrossRef](#)]
30. Clemen, R. Combining forecasts: A review and annotated bibliography. *J. Forecast.* **1989**, *5*, 559–583. [[CrossRef](#)]
31. Perrone, M. Improving Regression Estimation: Averaging Methods for Variance Reduction with Extension to General Convex Measure Optimization. Ph.D. Thesis, Brown University, Providence, RI, USA, 1993.
32. Wolpert, D. Stacked generalization. *Neural Netw.* **1992**, *5*, 241–259. [[CrossRef](#)]
33. Geman, S.; Bienenstock, E.; Doursat, R. Neural networks and the bias/variance dilemma. *Neural Comput.* **1992**, *4*, 1–58. [[CrossRef](#)]
34. Krogh, A.; Sollich, P. Statistical mechanics of ensemble learning. *Phys. Rev.* **1997**, *55*, 811–825. [[CrossRef](#)]

35. Freund, Y.; Schapire, R. Experiments with a new boosting algorithm. In Proceedings of the 13th International Conference on Machine Learning (ICML 96), Bari, Italy, 3–6 July 1996; pp. 148–156.
36. Schapire, R. The strength of weak learnability. *Mach. Learn.* **1990**, *5*, 197–227. [[CrossRef](#)]
37. Che, J. Optimal sub-models selection algorithm for combination forecasting model. *Neurocomputing* **2014**, *151*, 364–375. [[CrossRef](#)]
38. Wei, N.; Li, C.J.; Li, C.; Xie, H.Y.; Du, Z.W.; Zhang, Q.S.; Zeng, F.H. Short-term forecasting of natural gas consumption using factor selection algorithm and optimized support vector regression. *J. Energy Resour. Technol.* **2018**, *141*, 032701. [[CrossRef](#)]
39. Tamba, J.G.; Essiane, S.N.; Sapnken, E.F.; Koffi, F.D.; Nsouandélé, J.L.; Soldo, B.; Njomo, D. Forecasting natural gas: A literature survey. *Int. J. Energy Econ. Policy* **2018**, *8*, 216–249.
40. Sebalj, D.; Mesaric, J.; Dujak, D. Predicting natural gas consumption a literature review. In Proceedings of the 28th Central European Conference on Information and Intelligent Systems (CECIIS '17), Varazdin, Croatia, 27–29 September 2017; pp. 293–300.
41. Azadeh, I.P.; Dagoumas, A.S. Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model. *Energy* **2017**, *118*, 231–245.
42. Gorucu, F.B.; Geris, P.U.; Gumrah, F. Artificial neural network modeling for forecasting gas consumption. *Energy Sources* **2004**, *26*, 299–307. [[CrossRef](#)]
43. Karimi, H.; Dastranj, J. Artificial neural network-based genetic algorithm to predict natural gas consumption. *Energy Syst.* **2014**, *5*, 571–581. [[CrossRef](#)]
44. Khotanzad, A.; Erlagar, H. Natural gas load forecasting with combination of adaptive neural networks. In Proceedings of the International Joint Conference on Neural Networks (IJCNN 99), Washington, DC, USA, 10–16 July 1999; pp. 4069–4072.
45. Khotanzad, A.; Elragal, H.; Lu, T.L. Combination of artificial neural-network forecasters for prediction of natural gas consumption. *IEEE Trans. Neural Netw.* **2000**, *11*, 464–473. [[CrossRef](#)]
46. Kizilaslan, R.; Karlik, B. Combination of neural networks forecasters for monthly natural gas consumption prediction. *Neural Netw. World* **2009**, *19*, 191–199.
47. Kizilaslan, R.; Karlik, B. Comparison neural networks models for short term forecasting of natural gas consumption in Istanbul. In Proceedings of the First International Conference on the Applications of Digital Information and Web Technologies (ICADIWT), Ostrava, Czech Republic, 4–6 August 2008; pp. 448–453.
48. Musilek, P.; Pelikan, E.; Brabec, T.; Simunek, M. Recurrent neural network based gating for natural gas load prediction system. In Proceedings of the 2006 IEEE International Joint Conference on Neural Networks, Vancouver, BC, Canada, 16–21 July 2006; pp. 3736–3741.
49. Soldo, B. Forecasting natural gas consumption. *Appl. Energy* **2012**, *92*, 26–37. [[CrossRef](#)]
50. Szoplik, J. Forecasting of natural gas consumption with artificial neural networks. *Energy* **2015**, *85*, 208–220. [[CrossRef](#)]
51. Potočnik, P.; Soldo, B.; Šimunović, G.; Šarić, T.; Jeromen, A.; Govekar, E. Comparison of static and adaptive models for short-term residential natural gas forecasting in Croatia. *Appl. Energy* **2014**, *129*, 94–103. [[CrossRef](#)]
52. Akpınar, M.; Yumusak, N. Forecasting household natural gas consumption with ARIMA model: A case study of removing cycle. In Proceedings of the 7th IEEE International Conference on Application of Information and Communication Technologies (ICAICT 2013), Baku, Azerbaijan, 23–25 October 2013; pp. 1–6.
53. Akpınar, M.M.; Adak, F.; Yumusak, N. Day-ahead natural gas demand forecasting using optimized ABC-based neural network with sliding window technique: The case study of regional basis in Turkey. *Energies* **2017**, *10*, 781. [[CrossRef](#)]
54. Taşpınar, F.; Çelebi, N.; Tutkun, N. Forecasting of daily natural gas consumption on regional basis in Turkey using various computational methods. *Energy Build.* **2013**, *56*, 23–31. [[CrossRef](#)]
55. Azadeh, A.; Asadzadeh, S.M.; Ghanbari, A. An adaptive network-based fuzzy inference system for short-term natural gas demand estimation: Uncertain and complex environments. *Energy Policy* **2010**, *38*, 1529–1536. [[CrossRef](#)]
56. Behrouznia, A.; Saberi, M.; Azadeh, A.; Asadzadeh, S.M.; Pazhoheshfar, P. An adaptive network based fuzzy inference system-fuzzy data envelopment analysis for gas consumption forecasting and analysis: The case of South America. In Proceedings of the 2010 International Conference on Intelligent and Advanced Systems (ICIAS), Manila, Philippines, 15–17 June 2010; pp. 1–6.

57. Viet, N.H.; Mandziuk, J. Neural and fuzzy neural networks for natural gas prediction consumption. In Proceedings of the 13th IEEE Workshop on Neural Networks for Signal Processing (NNSP 2003), Toulouse, France, 17–19 September 2003; pp. 759–768.
58. Yu, F.; Xu, X. A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network. *Appl. Energy* **2014**, *134*, 102–113. [[CrossRef](#)]
59. Papageorgiou, E.I. *Fuzzy Cognitive Maps for Applied Sciences and Engineering from Fundamentals to Extensions and Learning Algorithms*; Intelligent Systems Reference Library, Springer: Heidelberg, Germany, 2014.
60. Poczeta, K.; Papageorgiou, E.I. Implementing fuzzy cognitive maps with neural networks for natural gas prediction. In Proceedings of the 30th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2018), Volos, Greece, 5–7 November 2018; pp. 1026–1032.
61. Homenda, W.; Jastrzebska, A.; Pedrycz, W. Modeling time series with fuzzy cognitive maps. In Proceedings of the 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Beijing, China, 6–11 July 2014; pp. 2055–2062.
62. Homenda, W.; Jastrzebska, A.; Pedrycz, W. Nodes selection criteria for fuzzy cognitive maps designed to model time series. *Adv. Intell. Syst. Comput.* **2015**, *323*, 859–870.
63. Salmeron, J.L.; Froelich, W. Dynamic optimization of fuzzy cognitive maps for time series forecasting. *Knowl.-Based Syst.* **2016**, *105*, 29–37. [[CrossRef](#)]
64. Froelich, W.; Salmeron, J.L. Evolutionary learning of fuzzy grey cognitive maps for the forecasting of multivariate, interval-valued time series. *Int. J. Approx. Reason* **2014**, *55*, 1319–1335. [[CrossRef](#)]
65. Papageorgiou, E.I.; Poczeta, K.; Lapidou, C. Application of fuzzy cognitive maps to water demand prediction, fuzzy systems. In Proceedings of the 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Istanbul, Turkey, 2–5 August 2015; pp. 1–8.
66. Poczeta, K.; Yastrebov, A.; Papageorgiou, E.I. Learning fuzzy cognitive maps using structure optimization genetic algorithm. In Proceedings of the 2015 Federated Conference on Computer Science and Information Systems (FedCSIS), Lodz, Poland, 13–16 September 2015; pp. 547–554.
67. Jastrzebow, A.; Poczeta, K. Analysis of multi-step algorithms for cognitive maps learning. *Bull. Pol. Acad. Sci. Technol. Sci.* **2014**, *62*, 735–741. [[CrossRef](#)]
68. Stach, W.; Kurgan, L.; Pedrycz, W.; Reformat, M. Genetic learning of fuzzy cognitive maps. *Fuzzy Sets Syst.* **2005**, *153*, 371–401. [[CrossRef](#)]
69. Papageorgiou, E.I.; Poczeta, K. A two-stage model for time series prediction based on fuzzy cognitive maps and neural networks. *Neurocomputing* **2017**, *232*, 113–121. [[CrossRef](#)]
70. Anagnostis, A.; Papageorgiou, E.; Dafopoulos, V.; Bochtis, D. Applying Long Short-Term Memory Networks for natural gas demand prediction. In Proceedings of the 10th International Conference on Information, Intelligence, Systems and Applications (IISA 2019), Patras, Greece, 15–17 July 2019.
71. Kosko, B. Fuzzy cognitive maps. *Int. J. Man Mach. Stud.* **1986**, *24*, 65–75. [[CrossRef](#)]
72. Kardaras, D.; Mentzas, G. Using fuzzy cognitive maps to model and analyze business performance assessment. In *Advances in Industrial Engineering Applications and Practice II*; Chen, J., Mital, A., Eds.; International Journal of Industrial Engineering: San Diego, CA, USA, 12–15 November 1997; pp. 63–68.
73. Lee, K.Y.; Yang, F.F. Optimal reactive power planning using evolutionary algorithms: A comparative study for evolutionary programming, evolutionary strategy, genetic algorithm, and linear programming. *IEEE Trans. Power Syst.* **1998**, *13*, 101–108. [[CrossRef](#)]
74. Dickerson, J.A.; Kosko, B. Virtual Worlds as Fuzzy Cognitive Maps. *Presence* **1994**, *3*, 173–189. [[CrossRef](#)]
75. Hossain, S.; Brooks, L. Fuzzy cognitive map modelling educational software adoption. *Comput. Educ.* **2008**, *51*, 1569–1588. [[CrossRef](#)]
76. Fogel, D.B. *Evolutionary Computation. Toward a New Philosophy of Machine Intelligence*, 3rd ed.; Wiley: Hoboken, NJ, USA; IEEE Press: Piscataway, NJ, USA, 2006.
77. Herrera, F.; Lozano, M.; Verdegay, J.L. Tackling real-coded genetic algorithms: Operators and tools for behavioural analysis. *Artif. Intell. Rev.* **1998**, *12*, 265–319. [[CrossRef](#)]
78. Livieris, I.E. Forecasting Economy-related data utilizing weight-constrained recurrent neural networks. *Algorithms* **2019**, *12*, 85. [[CrossRef](#)]
79. Livieris, I.E.; Pintelas, P. *A Survey on Algorithms for Training Artificial Neural Networks*; Technical Report; Department of Mathematics, University of Patras: Patras, Greece, 2008.

80. Adhikari, R.; Agrawal, R.K. Performance evaluation of weight selection schemes for linear combination of multiple forecasts. *Artif. Intell. Rev.* **2012**, *42*, 1–20. [[CrossRef](#)]
81. Livieris, I.E.; Kanavos, A.; Tampakas, V.; Pintelas, P. A weighted voting ensemble self-labeled algorithm for the detection of lung abnormalities from X-rays. *Algorithms* **2019**, *12*, 64. [[CrossRef](#)]
82. Makridakis, S.; Winkler, R.L. Averages of forecasts: Some empirical results. *Manag. Sci.* **1983**, *29*, 987–996. [[CrossRef](#)]
83. Lemke, C.; Gabrys, B. Meta-learning for time series forecasting and forecast combination. *Neurocomputing* **2010**, *73*, 2006–2016. [[CrossRef](#)]
84. Palm, F.C.; Zellner, A. To combine or not to combine? Issues of combining forecasts. *J. Forecast.* **1992**, *11*, 687–701. [[CrossRef](#)]
85. Kreiss, J.; Lahiri, S. Bootstrap methods for time series. In *Time Series Analysis: Methods and Applications*; Rao, T., Rao, S., Rao, C., Eds.; Elsevier: Amsterdam, The Netherlands, 2012.
86. Shalabi, L.A.; Shaaban, Z. Normalization as a preprocessing engine for data mining and the approach of preference matrix. In Proceedings of the International Conference on Dependability of Computer Systems (DepCos-RELCOMEX), Szklarska Poreba, Poland, 25–27 May 2006; pp. 207–214.
87. Nayak, S.; Misra, B.B.; Behera, H. Evaluation of normalization methods on neuro-genetic models for stock index forecasting. In Proceedings of the 2012 World Congress on Information and Communication Technologies (WICT 2012), Trivandrum, India, 30 October–2 November 2012; pp. 602–607.
88. Jayalakshmi, T.; Santhakumaran, A. Statistical normalization and back propagation for classification. *Int. J. Comput. Theory Eng.* **2011**, *3*, 89–93. [[CrossRef](#)]
89. Brown, A. A new software for carrying out one-way ANOVA post hoc tests. *Comput. Methods Programs Biomed.* **2005**, *79*, 89–95. [[CrossRef](#)] [[PubMed](#)]
90. Chen, T.; Xu, M.; Tu, J.; Wang, H.; Niu, X. Relationship between omnibus and post-hoc tests: An investigation of performance of the F test in ANOVA. *Shanghai Arch. Psychiatry* **2018**, *30*, 60–64. [[PubMed](#)]
91. Understanding the LSTM Networks. Available online: <http://colah.github.io> (accessed on 10 September 2019).
92. Holzinger, A. From machine learning to explainable AI. In Proceedings of the 2018 World Symposium on Digital Intelligence for Systems and Machines (IEEE DISA), Kosice, Slovakia, 23–25 August 2018; pp. 55–66.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).