

Article

Day-Ahead Natural Gas Demand Forecasting Using Optimized ABC-Based Neural Network with Sliding Window Technique: The Case Study of Regional Basis in Turkey

Mustafa Akpinar [†], M. Fatih Adak ^{*,†} and Nejat Yumusak [†]

Computer Engineering Department, Computer and Information Sciences Faculty, Sakarya University, 2nd Ring Street, Esentepe Campus, Serdivan, 54187 Sakarya, Turkey; akpinar@sakarya.edu.tr (M.A.); nyumusak@sakarya.edu.tr (N.Y.)

* Correspondence: fatihadak@sakarya.edu.tr; Tel.: +90-264-295-7049

† These authors contributed equally to this work.

Academic Editor: Moran Wang

Received: 18 April 2017; Accepted: 1 June 2017; Published: 6 June 2017

Abstract: The increase of energy consumption in the world is reflected in the consumption of natural gas. However, this increment requires additional investment. This effect leads imbalances in terms of demand forecasting, such as applying penalties in the case of error rates occurring beyond the acceptable limits. As the forecasting errors increase, penalties increase exponentially. Therefore, the optimal use of natural gas as a scarce resource is important. There are various demand forecast ranges for natural gas and the most difficult range among these demands is the day-ahead forecasting, since it is hard to implement and makes predictions with low error rates. The objective of this study is stabilizing gas tractions on day-ahead demand forecasting using low-consuming subscriber data for minimizing error using univariate artificial bee colony-based artificial neural networks (ANN-ABC). For this purpose, households and low-consuming commercial users' four-year consumption data between the years of 2011–2014 are gathered in daily periods. Previous consumption values are used to forecast day-ahead consumption values with sliding window technique and other independent variables are not taken into account. Dataset is divided into two parts. First, three-year daily consumption values are used with a seven day window for training the networks, while the last year is used for the day-ahead demand forecasting. Results show that ANN-ABC is a strong, stable, and effective method with a low error rate of 14.9 mean absolute percentage error (MAPE) for training utilizing MAPE with a univariate sliding window technique.

Keywords: demand forecasting; day-ahead forecast; natural gas; neural network; artificial bee colony (ABC)

1. Introduction

The global demand for clean energy resources that meet the increasing need of energy demands is rising day by day. Since the early 1990s, natural gas is used more for these energy resources. While household users consume natural gas for heating, cooking and hot water, factory users utilize them for power generation, transportation, processing, heating, cooling and cooking. The cost and selling price of natural gas are affected by natural gas consumption of high-use industrial subscribers with expenditure items of energy. Therefore, forecasting year ahead natural gas demands close to actual consumption is important to industrial subscribers.

Although industrial subscribers' consumption needs to be predictable, household and low-consuming subscribers do not have to know in advance. This makes consumption estimation

for low-consuming subscribers difficult. Demand forecasting methods have been developed and continue to be developed, in order to perform the difficult estimation for low-consuming subscribers' consumption. Decision makers in the energy sector use these methods to make predictions about future demand, and supply-and-demand must overlap as much as possible. The supply-and-demand should be balanced with high accuracy. As a result, the stabilization process becomes a very important sub-discipline of energy sectors including electricity, gas, water and wind.

Privatization of the electricity and natural gas sectors in Turkey brought the formation of a market structure. In the market structure, high errors in demand forecast result in penalties. Operation of the market and penalties will be discussed in the following section. There are various demand forecast ranges in natural gas, like year ahead monthly, month-ahead daily capacity reservation and day-ahead forecasting. The most difficult range among these demands is day-ahead forecasting, since it is hard to implement and has a low error rate in prediction. In this study, day-ahead natural gas demand forecast using low consuming subscribers' data is predicted. The hybrid method applying the artificial bee colony (ABC) algorithm for training the artificial neural network (ANN) structure was used.

1.1. Related Work

The literature for this study can be roughly grouped into two categories, according to the methods applied. The first category is the daily natural gas consumption demand forecast, and the second is the ANN and hybrid methods used for energy demand. There are many studies based on daily natural gas demand forecasting [1–12]. Khotanzad et al. worked on a combination of ANN forecasters for predicting natural gas consumption at a citywide distribution level [1]. Gorucu et al. used ANN to forecast gas consumption [2] at a citywide distribution level. Potocnik et al. proposed a strategy to estimate the forecasting risk for the citywide distribution level [3] using hourly consumption data. Akpinar and Yumusak divided consumption monthly by season, and tried to forecast consumption [4]. Sanchez-Ubeda and Berzosa presented a novel prediction model that provides forecasting for the end-use of industrial consumption in Spain for a medium-term horizon (1–3 years) with a very high resolution (days) based on a decomposition approach on a national level [5]. Yokoyama et al. proposed a global optimization method called the model trimming method to identify the model parameters [6]. They used neural networks with predicted air temperature and relative humidity as input and energy demand forecasted. Akpinar and Yumusak used a linear regression in their study with the sliding window technique [8]. They slid different sized windows data, and researched the best solution for natural gas demand. Natural gas consumption is forecasted using daily gas consumption data through different methods including 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) [9]. They found that SARIMAX gives more accurate results than the others. Soldo et al. used the linear autoregressives model with exogenous inputs (ARX), ANN and support vector machines (SVM) to forecast daily natural gas consumption with solar radiation [11]. Their results confirm that solar radiation improves forecast accuracy. Similar to Soldo's study, a simulation work was done on energy consumption and valuable results were obtained [13].

ANN and hybrid methods are frequently used in the energy sector [1,2,6,7,9–12,14–33]. The adaptive network-based fuzzy inference system (ANFIS) for estimating natural gas demand is one of the hybrid methods in these studies [7]. Azadah et al. used historical data in their study. Karimi and Dastranj used an ANN-based genetic algorithm (GA) to predict natural gas consumption [10]. They used GA to optimize the parameters of neural network topology. Yalcinoz and Eminoglu forecasted the electricity load of Nigde province in Turkey [15]. They used past data for mid-term monthly forecasting, and weather data along with historical data to forecast daily loads with ANN. Amjadi attempted one-day hourly price forecasting of electricity markets by a new Fuzzy NN [16]. They examined a proposed method for the Spanish electricity market, and this proposed technique was more accurate than the autoregressive integrated moving average (ARIMA), wavelet-ARIMA,

multilayer perceptron (MLP) and radial basis function NN (RBF). Saini researched feedforward ANN based on steepest descent, Bayesian regularization, resilient and adaptive backpropagation (BP) learning methods to forecast seven-day peak load with weather, and past peak load information [25]. Best performance is accomplished with adaptive BP learning for peak load forecasting. Azadeh et al. proposed ANFIS-fuzzy data envelopment analysis (FDEA) [30]. FDEA is used to examine the behavior of gas consumption, and the algorithm is capable of dealing with both complexity and uncertainty. Szoplik analyzed seasonal and diurnal variation [31]. In this research, the design and training of the MLP model to forecast the hourly demand for natural gas in the city has been studied. In another study, Azadeh et al. showed how to model sharp drops/jumps in natural gas consumption [32]. They proposed an emotional learning-neuro-fuzzy inference approach for optimum training and forecasting of gas consumption estimation.

1.2. Motivation

Optimization algorithms, mentioned previously, are used in ANN. The new optimization technique called Artificial Bee Colony (ABC) is one of them, and has a wide range of usage for optimizing, mentioned below. The ABC algorithm could be used as combination with other algorithms [12,33–35]. Akpinar et al. forecasted day-ahead natural gas demand using Hybrid ANN-ABC and ANN-BP [12]. They used various ANN structures and hidden layers. They found 18% mean absolute percentage error (MAPE) and 0.891 coefficient of determination. Uzlu et al. estimated hydroelectric generation using ANN with the ABC algorithm for Turkey [33]. They found that the ANN-ABC model is more accurate than classical ANN. Li et al. studied optimal power flow problems using differential evolution (DE) and ABC algorithms [34]. They mentioned that the DE algorithm solves problems with a large population size, as opposed to ABC and a proposed hybrid DE-ABC algorithm. They found DE-ABC convergence time took less than DE, and the DE-ABC algorithm was effective. An energy efficient optimal deployment strategy is studied in their research. Adak and Yumusak studied classification of the aroma data for four fruits using ABC [35]. They discovered ANN trained by ABC was successful in classifying aroma data.

Suganthi presented a survey about energy demand forecasting [36]. A review of forecasting natural gas demand was performed by Soldo [37], and another review of the ABC algorithm proposed in this study was also presented by Karaboga [38].

1.3. Our Contribution

This paper studies forecasting day-ahead natural gas demand. ANN has a wide range usage area for predicting energy demand as mentioned in previous studies, and the ABC algorithm has been used for optimization in several fields. In our paper, as an alternative of using the BP algorithm, the ABC algorithm is applied for the training stage of the ANN.

The main reasons for selecting ABC are that ABC is easier to apply, and it requires less parameters than other algorithms. Since the exploration feature of ABC is more successful than others, it will reach the global minimum without getting stuck in a shallow local minimum. In addition, the use of the ANN-ABC algorithm with univariate data, and use of the sliding window technique are important points of study while predicting day-ahead demand. ABC optimized feedforward-ANN (ANN-ABC) and ANN-BP with three different hidden layer structures and various neurons are applied for the day-ahead demand prediction of natural gas, which is a sub-branch of the energy sector. These methods have not been used together previously for natural gas and day-ahead predictions. The data does not contain any information with other variables except its own past data.

The rest of the paper is organized as follows: the natural gas market of Turkey and collected natural gas, and the data are presented in Section 2. A theoretical description of methods is provided in Section 3. Section 4 gives detailed information about modeling, definitions, scenarios and results. The key findings and next studies are given at the end of the paper as conclusions in Section 5.

2. Experimental Background

2.1. Natural Gas Consumption in Turkey

Turkey's natural gas market is shown in Figure 1. The companies make an annual, month-ahead daily capacity reservation, and day-ahead forecast based on regulations placed on dotted lines [39]. The producer is generally outside of Turkey [40]. Import/export and wholesale companies could import natural gas to Turkey through pipelines or as liquid natural gas (LNG). Companies report their year-ahead monthly, month-ahead daily capacity reservation and day-ahead forecasts in a hierarchical manner to the companies that have a contract, excluding the bottom level. Import or wholesale companies make a final estimation using data collected from the bottom-up. Each month and each day, these forecasts are inspected, and if a mean absolute percent error is higher than 10% (depends on consumption range), penalties occur [39]. If month-ahead daily capacity reservation passes an acceptable error limit six times in a month, a penalty occurs for all of the remaining days of the month. The penalty rate exhibits 10% MAPE distribution throughout the year and varies between 8% and 12% according to the consumption-estimation amount. The fact that estimations are around 15% MAPE during the year indicates that the amount of penalty to be paid will be low and the forecast is well done. The Energy Market Regulatory Authority (EMRA, also known by the acronym EPDK) inspects the natural market controlled by the Petroleum Pipeline Corporation (PPC is also known as BOTAS). According to the EMRA report, in 2014, imported natural gas by nine long-term and two spot (LNG) import licensed entities was 49.262 billion m³ [40]. It was stated in the same report that 14.78% of total imports was LNG with 7.281 billion m³. At the national level, total consumption was nearly 20% for the household ratio [40]. This consumption amount that affects penalties was noticeably high, and it was forecasted from the residential/and small commercial end users who are subscribers of the city distribution company based on the bottom part of the market. As mentioned in the report, nearly 26% of total consumption at the national level comprises the sum of household and low-consumption subscribers [40].

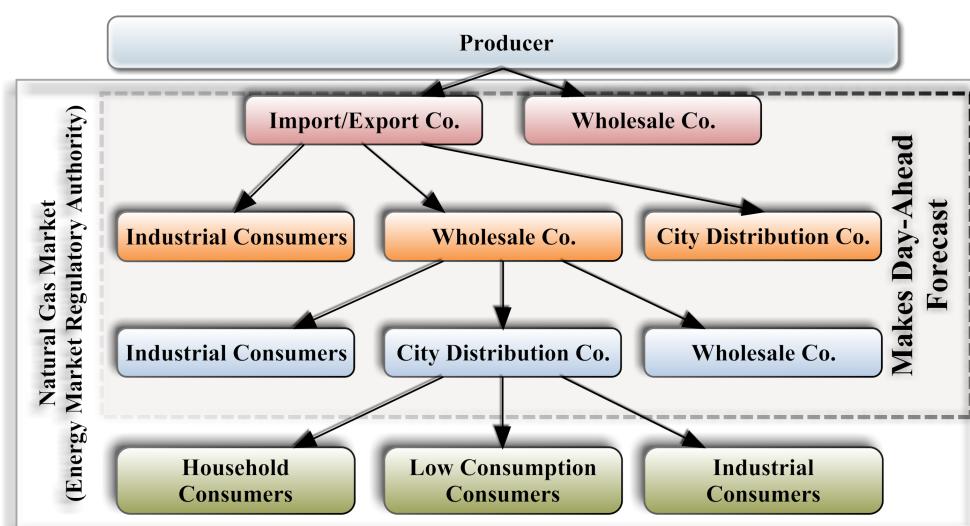


Figure 1. Turkey natural gas market and users.

2.2. The Preparation of the Data

Natural gas is distributed to end-users through pipelines in Turkey [41]. Reduction and measuring stations (RMSs) connect the cities to the pipelines. The pressure of natural gas decreases and the consumption volume of natural gas is calculated at these stations. There are three types of RMS. They are RMS-A, RMS-B and RMS-C. National distribution to regional distribution pipelines connect

through the RMS-A. The consumption range in an hour is 10,000 m³ to 300,000 m³ for RMS-A, and the pressure reduces from 40–75 bar to 12–25 bar. City natural gas distribution companies administer RMS-A type stations. The other two kinds of RMS are B and C, and they reduce pressure from 6–25 bar to 4 bar, and from 1–4 bar to 0.3 bar. Steel lines are used to connect RMS-A and RMS-B due to the high pressure. Distinct from RMS-A and RMS-B, polyethylene lines are used before and after RMS-C. Hourly consumptions are measured and calculated for all RMSs. The natural gas consumption data is collected from the natural gas distribution company of Sakarya province (AGDAS), Sakarya, Turkey. Daily resolution consumption data is converted from hourly resolution consumptions for all RMSs. Telemetry systems were established so high-consuming industrial subscribers have access to RMS for remote measurement. In addition, 90% of industrial subscribers have RMS and a telemetry system in the city. Telemetry consumptions for RMS-B and RMS-C are subtracted from RMS-A consumptions daily. The remaining 10% of industrial subscribers are low-consuming industries. Thus, this daily consumption of household and low-consuming subscribers are found. Finally, daily time series consumption data is prepared and ready for forecasting in this study. Time series forecasting is an important area of forecasting, in which past observations of the same variable area are collected and analyzed to develop a model describing the underlying relationship [42]. In this study, four years of natural gas consumption data was collected (2011–2014) and used. The first three years are used for training the ANN. The consumption data from 2014 is used for testing. Column values in the dataset are daily consumption. Seven inputs and one output are used to create the network structure. One week of data from before the estimation date is used to forecast. As in reality, the estimation day (forecasting day) does not contain any consumption data. The forecasted day is a day ahead of the forecasting day. Thus, the previous seven days of data before the forecasting day is added as an input for the dataset (Figure 2). As an example, forecasting for 22 April 2014 is done on 21 April 2014 by using data between 14 April 2014 and 20 April 2014. The amount of data used for training is 1088 days of data, and the amount of data used for testing is 365 days of data.

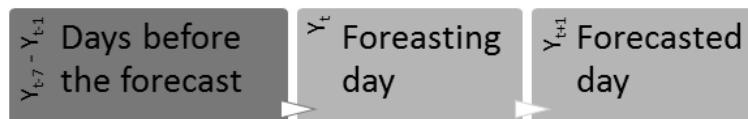


Figure 2. Day-ahead forecasting dataset.

The real consumption values are too big to use because the sigmoid function in ANN training requires values between zero to one. Generally, the min-max normalization function is applied to normalize values between zero and one. However, normalization is done in the [0.1–0.9] range, in order to avoid numerical flowing and zero value tendency for learning [43]. The min-max normalization equation is shown in Equation (1):

$$x_{0.1-0.9} = 0.8 \left[\frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \right] + 0.1. \quad (1)$$

Table 1 shows descriptive statistics of consumption with normalization. The first column C gives real consumption statistics, the second column C_{0-1} gives consumption with a normalized [0–1] range, and the third column $C_{0.1-0.9}$ gives consumption with a normalized [0.1–0.9] range. As seen in Table 1, the skew is positive and the median is smaller than mean, thus the right tail of series is longer. The sum of consumptions has been about 425 billion m³ for four years, and the range of consumption was about 950,000 m³.

Table 1. Descriptive statistics of consumption and normalized consumption.

Descriptive Statistics	C	C_{0-1}	$C_{0.1-0.9}$
Mean	290,914	0.28	0.32
Standard Error	6604	0.01	0.01
Median	202,069	0.18	0.25
Mode	45,254	0.02	0.11
Standard Deviation	252,417	0.27	0.21
Sample Variance	63,714,170,871	0.07	0.05
Kurtosis	-0.90	-0.90	-0.90
Skewness	0.65	0.65	0.65
Range	947,195	1.00	0.80
Minimum	27,765	0.00	0.10
Maximum	974,960	1.00	0.90
Sum	425,025,317	405.89	470.82
Count	1461	1461	1461

3. Method

This study examines forecasting with a back propagation learning based neural network and an artificial bee colony learning based feedforward network. These methods are briefly introduced in this section.

3.1. Artificial Neural Network (ANN)

Artificial neural networks are developed to apply the working principles of the brain. They use neurons, which carry out simple processes and have the ability to process data and make calculations through an interconnected network [44]. ANN is composed of layers, neurons and weights. Neurons build up the layers and are connected to each other through links called weights. A Backpropagation (BP) algorithm is used traditionally in the learning phase. ANN consists of three basic layers. These are the input, hidden, and output layers. The hidden layer may consist of one or more layers. ANN weights are adjusted by using a BP algorithm traditionally.

3.1.1. Feedforward Algorithm

A sample is chosen from the training set, either randomly or in a specific order, to apply the feedforward process to the network. The NET value is calculated using Equation (2) weights, W_{kj} , of the related neurons multiplied by the output values, O_k^i , and each are summed up in the equation:

$$NET_j^a = \sum_{k=1}^n W_{kj} O_k^i \quad (2)$$

NET values are put into an activation function, and hidden output values are calculated. The hidden outputs are multiplied by the weights which connect the hidden layer with the output layer, and net output values are found. Similarly, these net output values are put into the activation function given in Equation (3) to find the output of the network. β_j^a represents the threshold value of the related neuron:

$$O_j^a = \frac{1}{1 + e^{-(NET_j^a + \beta_j^a)}} \quad (3)$$

3.1.2. Backpropagation Algorithm

BP algorithm is a traditional method for training a neural network. Weights and threshold values of a network are determined randomly in the beginning of the training. The period of this random choice can be limited by the user. At the end of the feedforward, the network output, O_m , differs to

an extent from the expected output, T_m , and this difference is referred to as an error, E_m . The error is calculated at the end of each iteration Equation (4):

$$E_m = T_m - O_m \quad (4)$$

Epoch, on the other hand, is the case when all of the training data set is given to the network. The error value of each epoch is the sum of the squares of the iteration errors Equation (5). The error value mean squared error (MSE) of the last epoch is the error at the end of the training, and it is expected to be low:

$$TE = \frac{1}{2} \sum_m E_m^2 \quad (5)$$

If the error value of an iteration is above an acceptable level, weights and threshold values are updated and feedforward is repeated. The weights connecting the hidden and output layers are updated first. The amount of change for each weight is calculated beforehand using Equation (6), in order to make the update. Momentum α and the learning coefficient λ are taken into account during this calculation. The learning coefficient determines the learning level in each iteration. The smaller it is, the slower the learning phase is. Momentum, however, is intended to help the network not get stuck in local minima. δ_m represents the local gradient, and is calculated by using Equation (7):

$$\Delta W_{jm}^a(t) = \lambda \delta_m O_j^a + \alpha \Delta W_{jm}^a(t-1) \quad (6)$$

$$\delta_m = f'(NET) E_m \quad (7)$$

The $f'(NET)$ term in Equation (7) is the derivative of the activation function. If the activation function is chosen to be the sigmoid function, the term is expanded as in Equation (8):

$$\delta_m = O_m(1 - O_m)E_m \quad (8)$$

3.2. Artificial Bee Colony Algorithm (ABC)

The artificial bee colony algorithm is a swarm optimization algorithm, and it simulates the nectar searching behavior of bees. The first ABC algorithm, which was developed by Karaboga, includes three different types of bees [45]. These are the employee, onlooker and scout bees. There are some assumptions in the algorithm. Each source is assumed to be controlled by only one employee bee. Numbers of employed and unemployed bees are equal. Any employee bee turns into a scout bee if its source is depleted. Figure 3 gives the work procedure of the ABC algorithm.

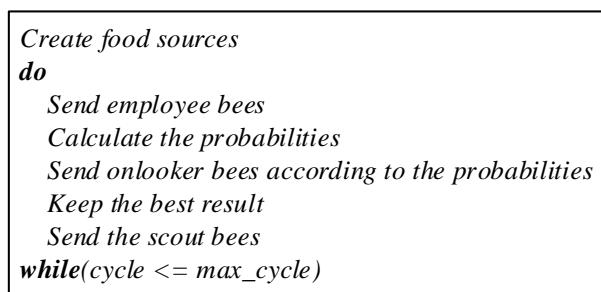


Figure 3. The procedure of the artificial bee colony (ABC) algorithm.

Nectar found in the food sources is collected by employee bees. Onlooker bees go out to find new nectar sources in light of the information shared by employee bees. Scout bees, on the other hand, randomly search for new nectar sources and turn into an employee bee once they find a source. Food sources are produced at the first step of the ABC algorithm within the given upper and lower

limits. The *rand* function is used in this food source generation process to generate random numbers as shown in Equation (9), i represents the food source, and j represents the parameter to be optimized in the equation:

$$x_{ij} = x_{\min j} + \text{rand}(0, 1) (x_{\max j} - x_{\min j}) \quad (9)$$

Employee bees turn into scout bees when the food source is depleted, and they determine new nectar sources by referring to present nectar source information that onlooker bees give. Equation (10) is used in this process. v_{ij} represents the new nectar source. i is a random index that represents the food source, and j represents the parameter to be optimized. ϕ_{ij} is a random number that controls the generation of neighbor food sources:

$$v_{ij} = x_{ij} + \phi_{ij} (x_{ij} - x_{kj}) \quad (10)$$

The fitness value of the new food source is determined by the fitness function in Equation (11). If the nectar amount of the new food source is lower than the present one, the food source is not changed and the searching process continues:

$$\text{fitness}_i = \left\{ \begin{array}{ll} \frac{1}{1+f_i} & f_i \geq 0 \\ \frac{1}{abs(f_i)} & f_i < 0 \end{array} \right\} \quad (11)$$

f_i in Equation (11) is the fitness value of the i^{th} solution, and is related to the nectar amount of the food source in the i^{th} location. The ABC algorithm runs for a predetermined number of iterations, and it intends to find the global minimum.

3.3. ABC Based ANN (ANN-ABC)

Even though artificial neural networks have been extensively applied in various fields, they still have weaknesses, such as over-fitting the training data or getting stuck in a shallow local minimum [46]. There are numerous proposals in existing literature for bypassing these problems. As analyzed, most of the solutions contain hybrid algorithms [47]. In this study, as an alternative to using the BP algorithm, the ABC algorithm is applied for the training stage of the artificial neural network. The main reasons for selecting ABC are that ABC is easier to apply, and it requires less parameters than other algorithms. Since the exploration feature of ABC is more successful than others, it will reach the global minimum without hooking the local minimum. The flow diagram presented in Figure 4 shows the ABC training process of the ANN. First, the ABC inputs randomly generated weights and thresholds into an array and optimizes the values, after which it determines the most fitted weight and threshold values. Fitness in this context is computed by placing weight and threshold values, which are calculated in each iteration into the network, and by feedforward movement.

The ABC will try to optimize the weights and thresholds of the artificial neural network stated in Figure 5. The array to be given for the algorithm of Figure 5 will be [a, b, c, d, b1, b2, e, f, g, h, b3, b4]. The error in each iteration resulting from the feedforwarding is the return value of the fitness function, and computed according to the specified error types.

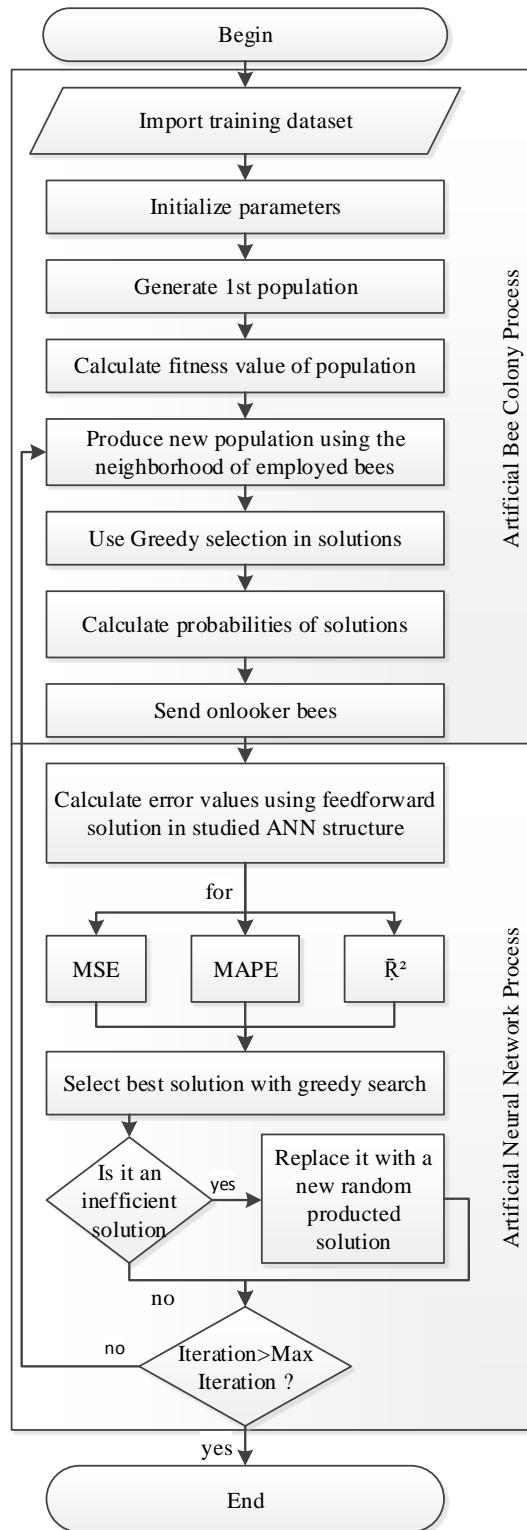


Figure 4. Flow chart of artificial neural network (ANN)-ABC.

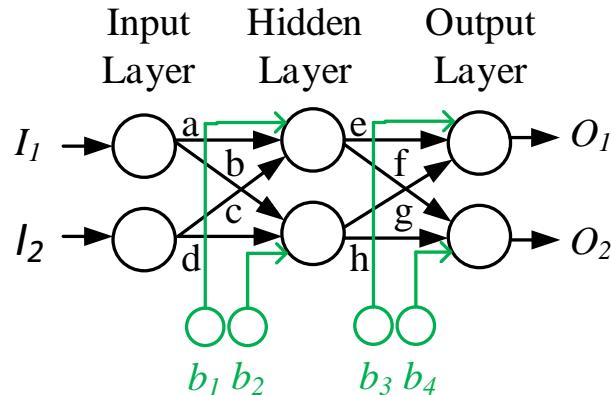


Figure 5. Sample artificial neural network for ANN-ABC.

3.4. Different Training Error Parameters

The training for feedforward ANN is processed related to the MSE error values. It is essentially based on decreasing the MSE errors in each epoch. However, in the BP algorithm training, the back differentiate error is propagated into the network, and the calculated MSE value is only used for informational purposes. For natural gas demand forecasting, the MSE decrement is not appropriate. MAPE is used for natural gas forecasting [39]. Thus, the data error in total is obtained individually for each data on average. Another significant property of MAPE is absolute value. Due to the absolute value, error directions are not considered, and the error is reduced from both the negative and positive side. Since the global minimum is searched for while training, there is no sense in multiplying the error by 100. Therefore, the percentage for the MAPE application is not included in this work. Another expression used in the training is the coefficient of determination (\bar{R}^2). The strength of the relationship between training and realization is determined by this term. This value can vary between -1 and 1 . If either the training or realization value decreases while the other one is increasing, the relationship is considered to be negative. If both of them increase, the relationship is considered positive. If the coefficient of the determination value is close to zero, it refers to the weakness. If this value is close to -1 or 1 , the relationship is described as strong. The training with \bar{R}^2 value prevents the aggregation of the error for the entire series in a particular region. Therefore, the average error remains stable in the summer, winter, spring and fall seasons on average. Since the global minimum is searched for in this study; \bar{R}^2 must be defined in a different way. This expression is defined as $|1 - \bar{R}^2|$ and represented as $\bar{\bar{R}}^2$. Thus, a high \bar{R}^2 value will approach the global minimum, which is zero. Eventually, the MSE, MAPE and \bar{R}^2 values are used during the training stage in this work.

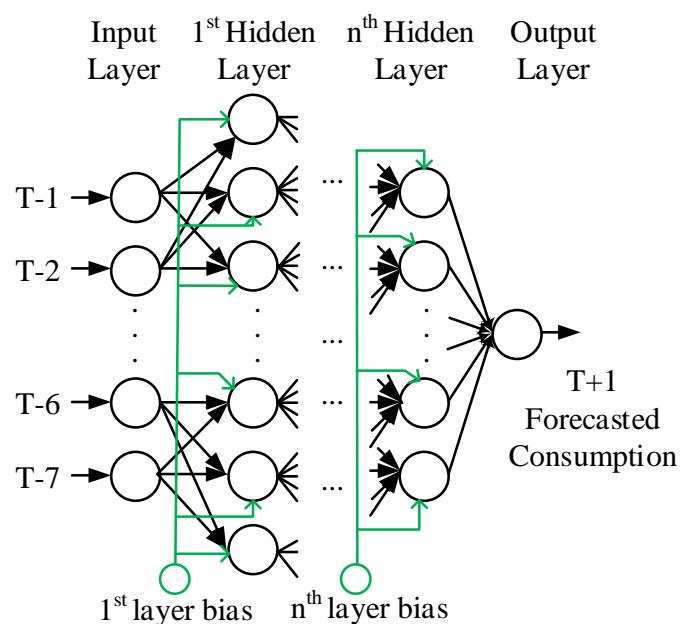
4. Scenarios and Results

In this study, different scenarios are prepared to forecast natural gas consumption. These scenarios have two parts. The first part is training the ANN using the BP algorithm, and the second part is training the ANN using the ABC algorithm. All trainings are realized with the same numbers of neurons and hidden layers. The MSE, MAPE and \bar{R}^2 measurements are selected in the training stage for the ABC. The MSE, MAPE and \bar{R}^2 measurements are also calculated for BP training. ANNs are tested for the year 2014 using calculated training weights. Table 2 shows the training parameters for the ABC and BP. In this table, the food source limit is the values of natural gas consumption, and the value is 365 due to the daily forecast.

Table 2. ABC and back propagation (BP) parameters.

Parameter (ABC)	Value	Parameter (BP)	Value
Lower bound	-10	Learning rate	0.2
Upper bound	10	Momentum	0.8
Colony size	100	Weights' lower bound	-1
Food source limit	365	Weights' upper bound	1

The ANN structure is shown in Figure 6. Seven days before the forecasting day were used for input values for the network, and day-ahead consumption is forecasted. The training stage for ANNs are prepared individually by MSE, MAPE and \bar{R}^2 error calculations. Normalized consumptions are transformed to real consumptions before calculating errors. The error calculations (MSE, MAPE and \bar{R}^2) are done with real consumptions. Thus, training is done with real data, and various ANNs are designed to determine the best-performing ANN. For this purpose, one hidden layer network with 20, 40, 60, 80 and 100 neurons is trained with 500, 1000, 3000, 5000, 7000, and 10,000 epochs. Thirty different results are found for BP training, and 90 different results are found for ABC training for one hidden layer. The test data in the hidden layer network that provides suitable network performance are selected. The suitable network performance criterion is 0.2 or less MAPE. The networks having suitable performance are redesigned with two hidden layers, and the training is repeated with the same number of epochs used for one hidden layer. In the two hidden layers structure, the second hidden layer's neuron number is incremented in tens from 10 to 60. For each MSE, MAPE and \bar{R}^2 training error for ABC training and MSE for BP training, 120 individual network structures are prepared and trained. Unlike the suitable performance of one hidden layer network structure, two hidden layers' network structures performance criteria is 0.16 or less MAPE. The networks are redesigned for the performance criteria of two hidden layers and five, 15, and 30 neurons are used in the third hidden layer with the same numbers of epochs used for one and two hidden layers.

**Figure 6.** The structure of ANN chosen for training.

The numbers of prepared ANN structures within the performance criteria are given in Table 3. The rows indicate the number of hidden layers, and the columns indicate the number of epochs. It is seen that the most efficient training is with 7000 epochs.

Table 3. Number of prepared ANN structures within the performance criteria.

Hidden Layer Epochs	500	1000	3000	5000	7000	10,000
One Hidden Layer	5	5	5	5	5	5
Two Hidden Layers	6	12	6	12	18	6
Three Hidden Layers	-	-	3	3	12	9

The error terms MSE, MAPE and \bar{R}^2 are used for training the ANN-ABC. Due to the nature of BP, only MSE is used in the training stage of ANN-BP. The network structures with the best results from MAPE for the test dataset and the abbreviations for them are given in Table 4. The best network structures are shown for three criteria. These three criteria are the number of hidden layers, training algorithms for the ABC, and the test results for BP and ABC. The ANN models with the lowest MSE, MAPE and \bar{R}^2 values in the test dataset are shown in Table 4. In the test dataset for one hidden layer BP training model, 40 neurons and 1000 epochs are the best solutions for the MSE value. For ABC training using the MSE error values in one hidden layer, the best test dataset result is obtained with 20 neurons and 3000 epochs. Other situations are seen in the table. Abbreviations are written along with the type of training and two parameters in parentheses. The first parameter is the error type for ANN at the end of training with the BP algorithm (the training can be done only with MSE). If the training type is ABC, then the first parameter shows an error type used in training and at the end of training. The first parameter *S* stands for MSE value, *M* stands for MAPE value, and *R* stands for \bar{R}^2 value. The second parameter in parentheses is the number of hidden layers.

The test dataset based on the best results is found with a different ANN structure and different number of epochs (Table 4). All results mentioned in this paragraph are based on the best test dataset outcomes (Figure 7). The best results when MSE is used for error criteria are 500 epochs for ABC training (Figure 7(a2)) and 1000 epochs for BP training. The lowest MSE value is found as 646,201,826.5 in three hidden layers on BP training, while the lowest MSE value for ABC training is 2,185,385,306 in three hidden layers. As can be seen, the BP training MSE value is 3.5 times less than the lowest ABC training MSE value (Figure 7(a3)). For MAPE error criteria, the lowest epoch number is 7000 for training the ABC, and 10,000 epochs for BP training (Figure 7(b1–b3)). The lowest MAPE values are 0.0894 and 0.1412 with three hidden layers for BP and ABC training, respectively (Figure 7(b3)). At the end of the training, the MAPE value for BP is two times lower than ABC. The lowest epoch numbers of \bar{R}^2 in training are 7000 and 3000 for BP and ABC, respectively (Figure 7(c1,c2)). The \bar{R}^2 value is too low for BP training with 1.78×10^{-6} with three hidden layers. 0.0734 is the lowest value of \bar{R}^2 for ABC training with three hidden layers. \bar{R}^2 is close to 1 in BP training, and the \bar{R}^2 value for ABC training is 0.9266 (Figure 7(c3)). During the training stage, differentiated errors are propagated to the network weights as the structure of BP. Thus, BP training presents less errors than ABC training on both error criteria and hidden layer structures (Figure 7). The error criteria results indicate that BP has a significant success in training. However, ABC training results are less successful compared to the BP training. This statement can be also applied to forecasting results for BP and ABC.

Table 4. ANNs with the lowest MAPEs (mean absolute percentage error) and abbreviations.

Types of Structures		MSE		MAPE		\bar{R}^2	
		BP	ABC	BP	ABC	BP	ABC
One hidden layer	Neurons/Epochs	$40/10^3$	$20/3 \times 10^3$	$20/10^4$	$20/7 \times 10^3$	$20/7 \times 10^3$	$40/5 \times 10^3$
	Abbreviation	BP(S,1)	ABC(S,1)	BP(M,1)	ABC(M,1)	BP(R,1)	ABC(R,1)
Two hidden layers	Neurons/Epochs	$20 + 50/10^4$	$20 + 40/500$	$20 + 30/10^4$	$20 + 20/7 \times 10^3$	$20 + 60/10^4$	$20 + 60/10^4$
	Abbreviation	BP(S,2)	ABC(S,2)	BP(M,2)	ABC(M,2)	BP(R,2)	ABC(R,2)
Three hidden layers	Neurons/Epochs	$20 + 60 + 15/10^4$	$20 + 10 + 15/10^4$	$20 + 60 + 30/10^4$	$20 + 10 + 5/7 \times 10^3$	$20 + 60 + 30/10^4$	$20 + 10 + 5/3 \times 10^3$
	Abbreviation	BP(S,3)	ABC(S,3)	BP(M,3)	ABC(M,3)	BP(R,3)	ABC(R,3)

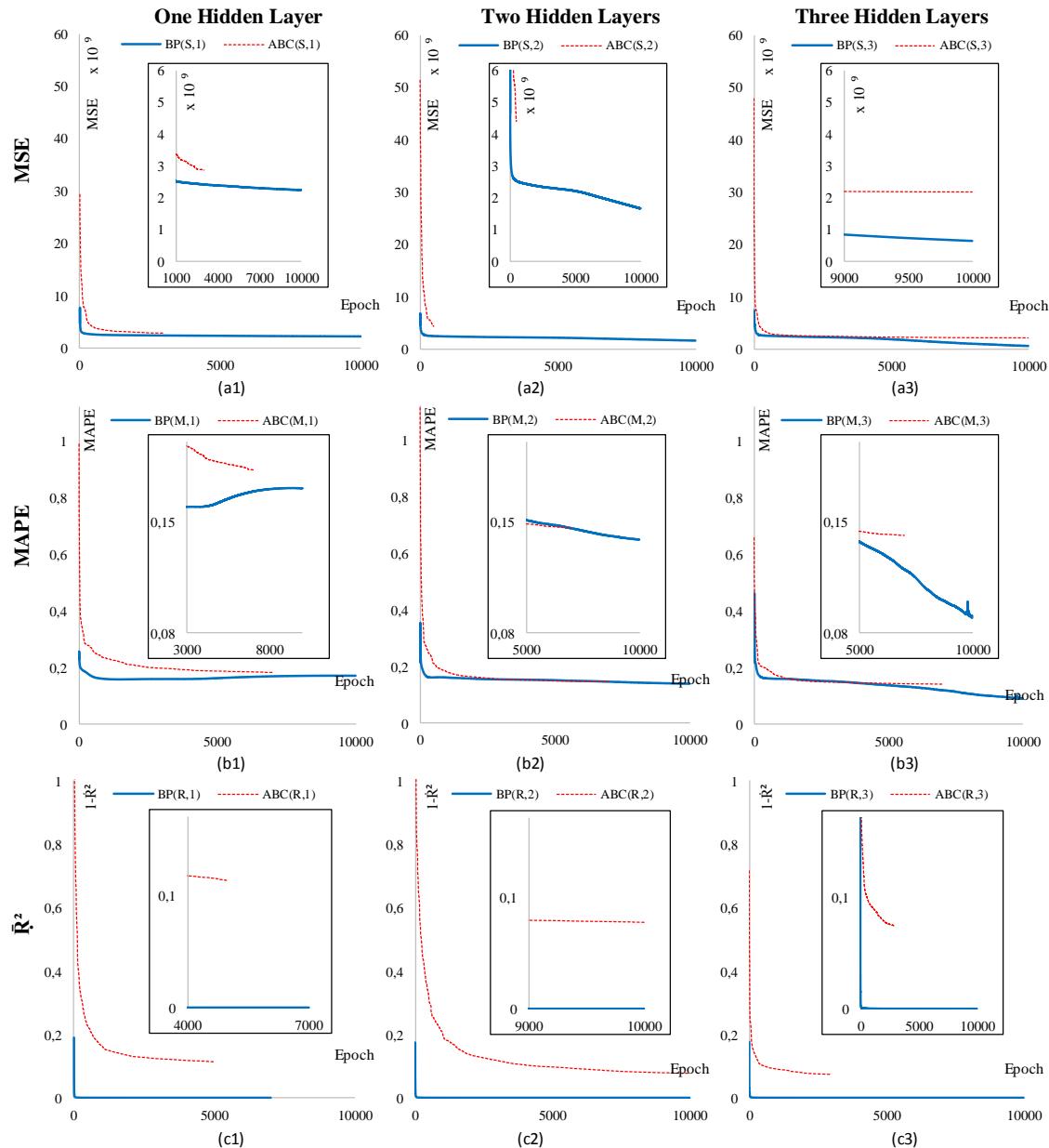


Figure 7. ANN training states with the lowest MAPEs (mean absolute percentage error). (a) Training with MSE (mean squared error); (b) Training with MAPE (mean absolute percentage error); (c) Training with \bar{R}^2 .

As depicted in Figure 7, most of the training takes place in the first 100 epochs of BP training. After 100 epochs, the training slope goes down to nearly zero. Even with the MAPE criterion on one hidden layer, training errors increase (Figure 7(b1)). Unlike BP training, error values in ABC decreases in every epoch step continuously. This situation expresses that, after finding the most effective nectar, the ABC searches for better sources nearby. This is so that it continues to make improvements in learning. Comparatively, BP tends to over-fit the training data. After the training step, day-ahead forecasts are done from 01.01.2014 to 31.12.2014 by trained networks. Seven-day consumptions before the forecasting day are used for day-ahead prediction. Based on the training error criteria, forecasting results for the lowest error ANN are showed in Figure 8 for three different hidden layer structures. The most noticeable point that is independent from the error criteria is that the BP and ABC errors visibly drop as the number of hidden layers increase. For BP, one and two hidden layers have high

errors independent from the error criteria. The tested network, which is trained by BP, previously had 100% MAPE with a single hidden layer, approximately 63% MAPE with two hidden layers, and 33% with three hidden layers, regardless of the error criteria. It can be said that each added hidden layer decreases 33% MAPE in the BP algorithm. The negligibility of the layer amount of the ABC trained network is also notable. Independent from the error criteria, both one and two hidden layers have 16.8% MAPE on average, while three hidden layers have nearly 16.6% MAPE. In other words, in the ABC trained network, hidden layer number does not vary regardless of the error criteria. The fact that different results obtained in the same structure by changing the network weights shows the success level of the training. Since the ABC trained network has better results than the BP trained network, the accuracy of the approach applied in the study is confirmed.

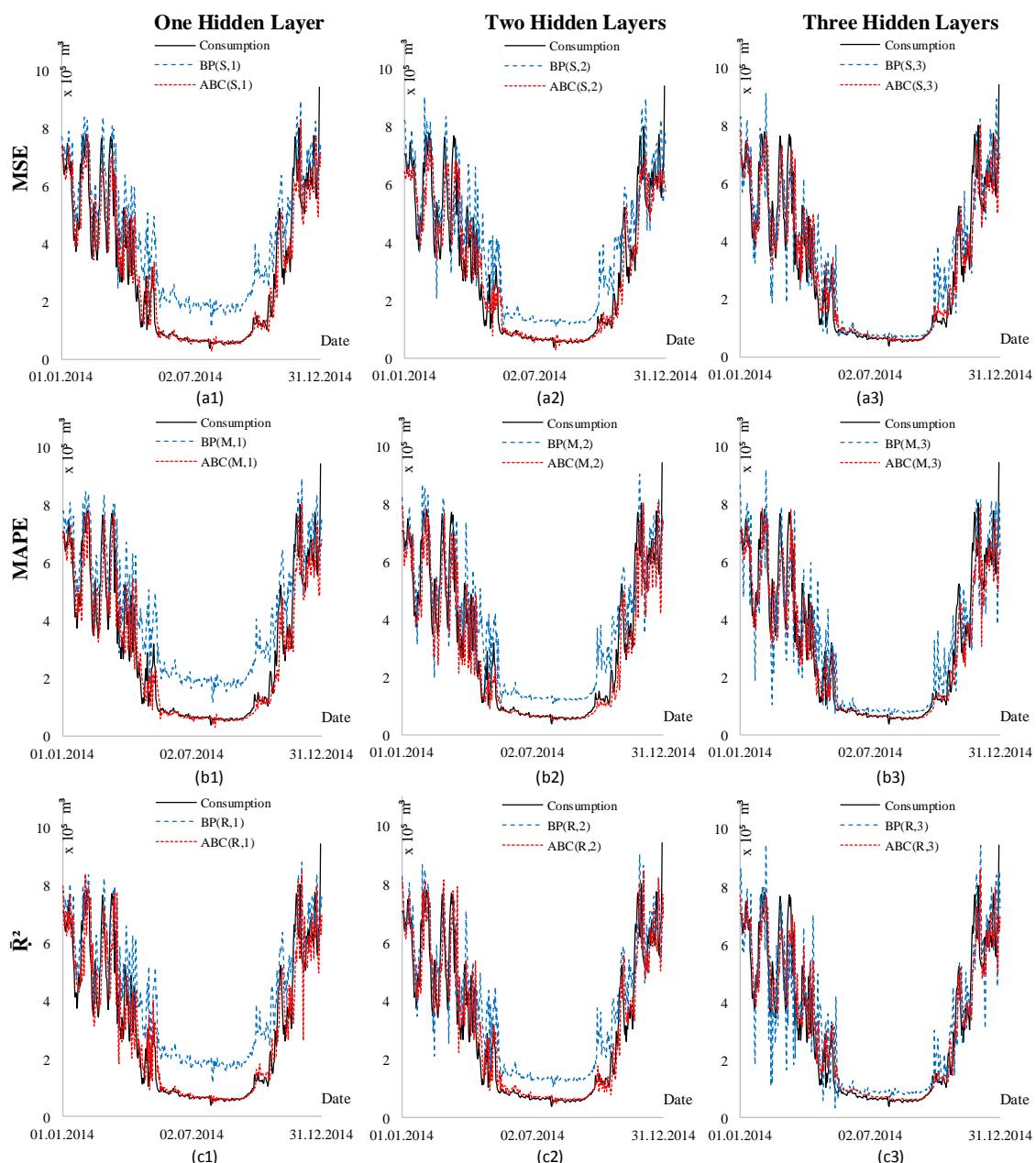


Figure 8. Day-ahead forecasts of the lowest MAPEs. (a) Forecast of MSE (mean squared error); (b) Forecast of MAPE (mean absolute percentage error); (c) Forecast of \bar{R}^2 .

Tests in all scenarios showed that ANN-ABC generates lower error values than ANN-BP. The error values at the end of the test process are presented in Table 5 according to the hidden layers. Among single hidden layer structures for all scenarios, the network that has 7000 epochs with 20 neurons has the best performance, with a MAPE value of 16.29%. Among the two hidden layer networks, the best performing network with a 15.36% MAPE value is the network that ran with 7000 epochs and has 20 + 20 neurons. In addition, the best network of the three hidden layers structures presents a 14.94% MAPE value run with 7000 epochs and 20 + 10 + 5 neurons. All of these values are for ANN-ABC. Since ANN-BP performs worse compared to ANN-ABC in all scenarios, the ANN is trained by a different BP examination. In this case, the ANN-BP has one, two and three hidden layers and a normalized dataset between [0, 1], where these values are not converted to real numbers during training. The lowest error value obtained with a single hidden layer network is 41.7%, two hidden layers is 30.21%, and three hidden layers is 29.97%.

Table 5. Day-ahead forecasting results.

Training Type	One Layer		Two Layers		Three Layers	
	BP	ABC	BP	ABC	BP	ABC
MSE	Model	BP(S,1)	ABC(S,1)	BP(S,2)	ABC(S,2)	BP(S,3)
	MAPE	99.2%	16.4%	63.8%	17.6%	30.2%
MAPE	Model	BP(M,1)	ABC(M,1)	BP(M,2)	ABC(M,2)	BP(M,3)
	MAPE	99.9%	16.3%	61.7%	15.4%	33.9%
\bar{R}^2	Model	BP(R,1)	ABC(R,1)	BP(R,2)	ABC(R,2)	BP(R,3)
	MAPE	97.6%	17.8%	63.5%	17.4%	34.3%

MSE: mean squared error; MAPE: mean absolute percentage error; BP: backpropagation; ABC: artificial bee colony.

The lowest MAPE values based on the predictions for the test data of BP and ABC trainings are given in Figure 9. During the forecasting series, the BP algorithm error is usually higher than the ABCs. The high prediction values for the summer period indicate that the BP algorithm effect is stronger in the winter. The high influence in winter months from BP leads to a sharp increase or decrease in predictions for summer, spring, and fall seasons, when the previous days' consumption values are used as input. The sudden increase or decrease in the BP algorithm can be seen clearly from February–March and November–December. For instance, in October, the BP algorithm forecast is totally different from the actual consumption. In this month, the slow increase in consumption behavior causes a very high consumption response, and this implies that BP uses memorization instead of learning. In contrast, it can be said that all seasons have a similar effect on the ABC algorithm. It can also be clearly seen that the prediction and realization overlap for the summer. The fact that a small amount of increase in October has almost the same level of consumption prediction, clearly showing the success of ABC training. Different from the BP algorithm, towards the end of April, the ABC algorithm predictions are proportionally similar to the real consumption decrease.

Figure 10 shows states for the lowest MAPE values of the BP and ABC models during training. The lowest error for the BP algorithm is found in the MSE value, while, for the ABC algorithm, it is the MAPE value. Therefore, the graph has a two-sided y -axis, where the left axis represents MAPE, and the right axis represents MSE values. In the detailed BP algorithm graph, from the 5.21×10^9 MSE value to the 7.5×10^9 , MSE value is reached at the 10th epoch, and the local minimum point is reached only one time. The 110th epoch value has a 3×10^9 MSE value. While the training stage after this time stays almost unchanged, after 4000 epochs, the training increased, and at the end of 10,000 epochs, it reached a 6.46×10^8 MSE value. The BP algorithm reduced the MSE error by one tenth in training. In the BP algorithm, 7/10 of the total reduced errors take place until 4000 epochs, while 3/10 of reduced errors occur between 4000 and 10,000 epochs. This state indicates that the majority of the training is completed until the first half of training. However, in the ABC model, MAPE decreased for each epoch during the training stage, which had a value of 0.67 at the beginning. It is also analyzed that, in training, the ABC has below 15% MAPE for the 6097th epoch, and 14.68% MAPE at the end of

the 7000th Epoch. In the test, the ANN-ABC gives 14.9% MAPE for one-day forecasting, which proves that the ABC does not memorize the consumption data, and the training with the ABC succeeds.

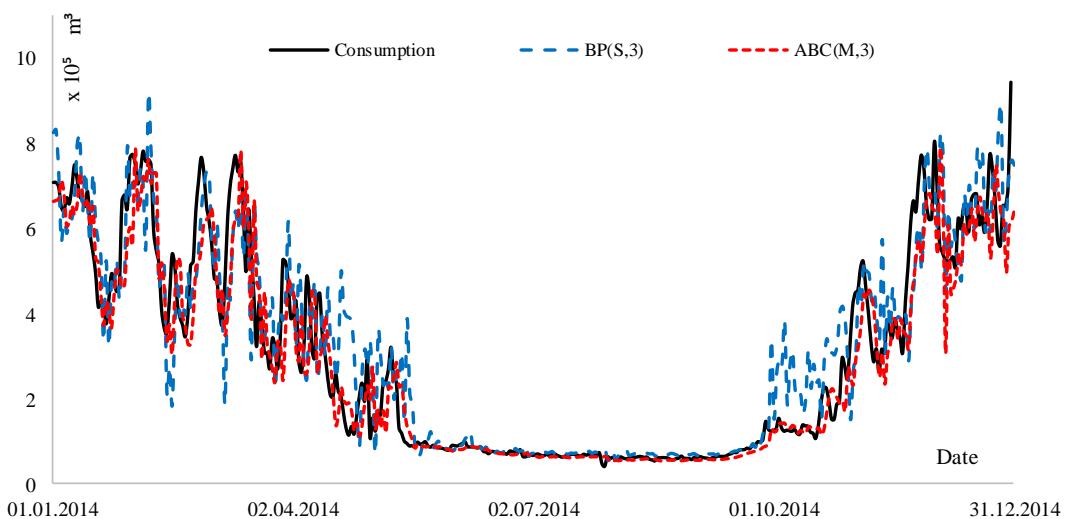


Figure 9. The best ANN structures of BP (backpropagation) and ABC day-ahead forecasts for test data.

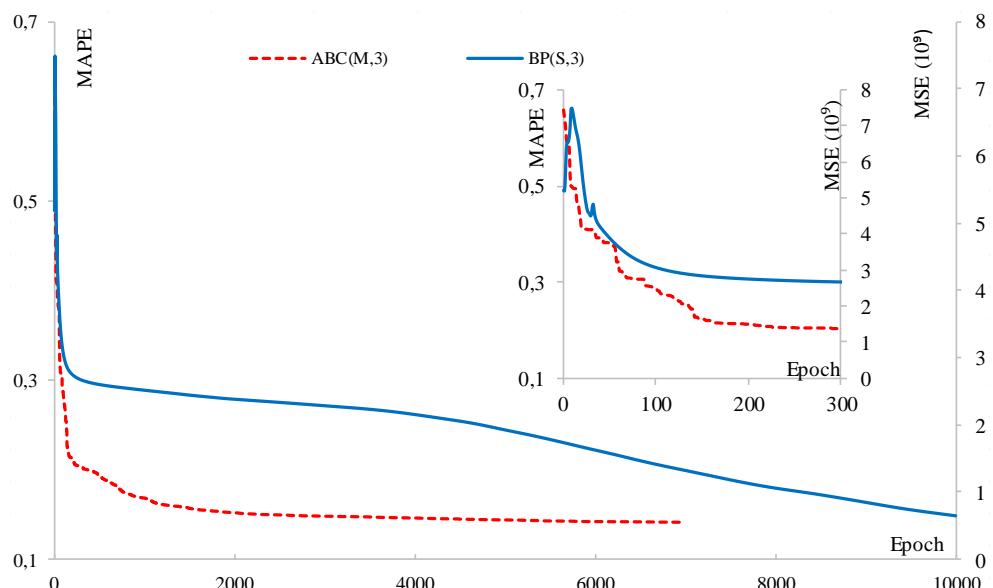


Figure 10. The best ANN structures of BP and ABC day-ahead forecasts for training data.

5. Conclusions

The study researches natural gas demand forecasting by applying BP and ABC learning to neural networks. Three different criteria in ABC learning and one criteria for BP learning are prepared. The MSE criteria is used for training with BP and ABC algorithms. MAPE and \bar{R}^2 are coded for ABC training. Even though coding the program is difficult, adaptation would be easy for companies. Decision makers can use the natural gas demand forecasting results obtained from forecasting models as decision support systems. Therefore, they can comfortably use this support system for determining day-ahead demand and show the consistency of forecasts by comparing their predictions and neural network results. Based on the results, the main conclusions of this paper are as follows: in the testing stage, the ABC training gives better results than BP training, and ABC training with MAPE criteria gives the best results for each number of hidden layers. The ABC gave 14.9% MAPE with three hidden

layers. The MAPE value obtained is very close to the 10% MAPE value, which causes penalties. This shows that the amount of penalty to be paid will be lower than with the method trained with BP. In the BP training, 30.2% MAPE is high; thus, the penalty will be much higher than the ABC performed prediction. The ABC training network confirms the possibility of forecasting natural gas demand.

Acknowledgments: The authors would like to acknowledge the AGDAS (Adapazari Turkey Natural Gas Distribution Company), Sakarya, Turkey for information about sector and valuable supports to prepare the data.

Author Contributions: Mustafa Akpinar obtained the energy consumption data, Mustafa Akpinar and M. Fatih Adak analysed and normalized energy consumption data. Mustafa Akpinar, M. Fatih Adak, and Nejat Yumusak work on learning model and develop a new learning model together. Case studies were done by M. Fatih Adak and Mustafa Akpinar.

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

Abbreviations

The following abbreviations are used in this manuscript:

ABC	Artificial Bee Colony
ANFIS	Adaptive Neuro Fuzzy Inference System
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
BP	Back Propagation
DE	Differential Evolution
EMRA	Energy Market Regulatory Authority
FDEA	Fuzzy Data Envelopment Analysis
GA	Genetic Algorithm
LNG	Liquid Natural Gas
MAPE	Mean Absolute Percentage Error
MLP	Multi Layer Perceptron
MSE	Mean Squared Error
OLS	Ordinary Least Squares
PPC	Petroleum Pipeline Corporation
RBF	Radial Basis Functions
RMS	Reduction and Measuring Stations
SVM	Support Vector Machines

References

1. 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.
2. Gorucu, F.B. Artificial Neural Network Modeling for Forecasting Gas Consumption. *Energy Sources* **2004**, *26*, 299–307.
3. Potočnik, P.; Thaler, M.; Govekar, E.; Grabec, I.; Poredoš, A. Forecasting risks of natural gas consumption in Slovenia. *Energy Policy* **2007**, *35*, 4271–4282.
4. Akpinar, M.; Yumusak, N. Estimating household natural gas consumption with multiple regression: Effect of cycle. In Proceedings of the IEEE 2013 International Conference on Electronics, Computer and Computation (ICECCO), Ankara, Turkey, 7–9 November 2013; pp. 188–191.
5. Sánchez-Úbeda, E.F.; Berzosa, A. Modeling and forecasting industrial end-use natural gas consumption. *Energy Econ.* **2007**, *29*, 710–742.
6. Yokoyama, R.; Wakui, T.; Satake, R. Prediction of energy demands using neural network with model identification by global optimization. *Energy Convers. Manag.* **2009**, *50*, 319–327.
7. Azadeh, A.; Asadzadeh, S.; 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.
8. Akpinar, M.; Yumusak, N. Naïve forecasting household natural gas consumption with sliding window approach. *Turk. J. Elec. Eng. Comp. Sci.* **2017**, *25*, 30–45.
9. Taşpinar, 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.

10. Karimi, H.; Dastranj, J. Artificial neural network-based genetic algorithm to predict natural gas consumption. *Energy Syst.* **2014**, *5*, 571–581.
11. Soldo, B.; Potočnik, P.; Šimunović, G.; Šarić, T.; Govekar, E. Improving the residential natural gas consumption forecasting models by using solar radiation. *Energy Build.* **2014**, *69*, 498–506.
12. Akpinar, M.; Adak, M.F.; Yumusak, N. Forecasting Natural Gas Consumption with Hybrid Neural Networks—Artificial Bee Colony. In Proceedings of the IEEE 2016 2nd International Conference on Intelligent Energy and Power Systems, Kiev, Ukraine, 7–11 June 2016; pp. 101–106.
13. Adak, M.F.; Duru, N.; Duru, H.T. Elevator simulator design and estimating energy consumption of an elevator system. *Energy Build.* **2013**, *65*, 272–280.
14. Yao, S.; Song, Y.; Zhang, L.; Cheng, X. Wavelet transform and neural networks for short-term electrical load forecasting. *Energy Convers. Manag.* **2000**, *41*, 1975–1988.
15. Yalcinoz, T.; Eminoglu, U. Short term and medium term power distribution load forecasting by neural networks. *Energy Convers. Manag.* **2005**, *46*, 1393–1405.
16. Amjadi, N. Day-Ahead Price Forecasting of Electricity Markets by a New Fuzzy Neural Network. *IEEE Trans. Power Syst.* **2006**, *21*, 887–896.
17. PAO, H. Comparing linear and nonlinear forecasts for Taiwan's electricity consumption. *Energy* **2006**, *31*, 2129–2141.
18. Ermis, K.; Midilli, A.; Dincer, I.; Rosen, M. Artificial neural network analysis of world green energy use. *Energy Policy* **2007**, *35*, 1731–1743.
19. Sözen, A.; Arcaklıoglu, E. Prediction of net energy consumption based on economic indicators (GNP and GDP) in Turkey. *Energy Policy* **2007**, *35*, 4981–4992.
20. Hamzaçebi, C. Forecasting of Turkey's net electricity energy consumption on sectoral bases. *Energy Policy* **2007**, *35*, 2009–2016.
21. Tso, G.K.; Yau, K.K. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy* **2007**, *32*, 1761–1768.
22. González-Romera, E.; Jaramillo-Morán, M.; Carmona-Fernández, D. Monthly electric energy demand forecasting with neural networks and Fourier series. *Energy Convers. Manag.* **2008**, *49*, 3135–3142.
23. Azadeh, A.; Ghaderi, S.; Sohrabkhani, S. Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. *Energy Convers. Manag.* **2008**, *49*, 2272–2278.
24. Ying, L.C.; Pan, M.C. Using adaptive network based fuzzy inference system to forecast regional electricity loads. *Energy Convers. Manag.* **2008**, *49*, 205–211.
25. Saini, L.M. Peak load forecasting using Bayesian regularization, Resilient and adaptive backpropagation learning based artificial neural networks. *Electr. Power Syst. Res.* **2008**, *78*, 1302–1310.
26. Pao, H. Forecasting energy consumption in Taiwan using hybrid nonlinear models. *Energy* **2009**, *34*, 1438–1446.
27. Geem, Z.W.; Roper, W.E. Energy demand estimation of South Korea using artificial neural network. *Energy Policy* **2009**, *37*, 4049–4054.
28. Sözen, A. Future projection of the energy dependency of Turkey using artificial neural network. *Energy Policy* **2009**, *37*, 4827–4833.
29. Kankal, M.; Akpinar, A.; Kömürcü, M.h.; Özşahin, T.ü. Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables. *Appl. Energy* **2011**, *88*, 1927–1939.
30. Azadeh, A.; Saberi, M.; Asadzadeh, S.M.; Hussain, O.K.; Saberi, Z. A neuro-fuzzy-multivariate algorithm for accurate gas consumption estimation in South America with noisy inputs. *Int. J. Electr. Power Energy Syst.* **2013**, *46*, 315–325.
31. Szoplik, J. Forecasting of natural gas consumption with artificial neural networks. *Energy* **2015**, *85*, 208–220.
32. Azadeh, A.; Asadzadeh, S.; Mirseraji, G.; Saberi, M. An emotional learning-neuro-fuzzy inference approach for optimum training and forecasting of gas consumption estimation models with cognitive data. *Technol. Forecast. Soc. Chang.* **2015**, *91*, 47–63.
33. Uzlu, E.; Akpinar, A.; Özturk, H.T.; Nacar, S.; Kankal, M. Estimates of hydroelectric generation using neural networks with the artificial bee colony algorithm for Turkey. *Energy* **2014**, *69*, 638–647.
34. Li, Y.; Wang, Y.; Li, B. A hybrid artificial bee colony assisted differential evolution algorithm for optimal reactive power flow. *Int. J. Electr. Power Energy Syst.* **2013**, *52*, 25–33.

35. Adak, M.; Yumusak, N. Classification of E-Nose Aroma Data of Four Fruit Types by ABC-Based Neural Network. *Sensors* **2016**, *16*, 304.
36. Suganthi, L.; Samuel, A.A. Energy models for demand forecasting—A review. *Renew. Sustain. Energy Rev.* **2012**, *16*, 1223–1240.
37. Soldo, B. Forecasting natural gas consumption. *Appl. Energy* **2012**, *92*, 26–37.
38. Karaboga, D.; Gorkemli, B.; Ozturk, C.; Karaboga, N. A comprehensive survey: Artificial bee colony (ABC) algorithm and applications. *Artif. Intell. Rev.* **2014**, *42*, 21–57.
39. BOTAS. *Iletim Sebekesi Isleyis Duzenlemelerine Iliskin Esaslar (The Basis of Regulatory Process on Transmission Network)*; The Official Gazette of the Turkish Republic: Ankara, Turkey, 2013.
40. Energy Market Regulatory Authority. *Turkish Natural Gas Market Report 2014*; Technical report; Energy Market Regulatory Authority: Ankara, Turkey, 2015.
41. Akpinar, M.; Yumusak, N. Year Ahead Demand Forecast of City Natural Gas Using Seasonal Time Series Methods. *Energies* **2016**, *9*, 727.
42. Zhang, G. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **2003**, *50*, 159–175.
43. Chen, T. Estimating job cycle time in a wafer fabrication factory: A novel and effective approach based on post-classification. *Appl. Soft Comput.* **2016**, *40*, 558–568.
44. Schalkoff, R.J. *Artificial Neural Networks*, 1st ed.; McGraw Hill: New York, NY, USA, 1997.
45. Karaboga, D. *An Idea Based on Honey Bee Swarm for Numerical Optimization*; Technical Report TR06; Erciyes University: Kayseri, Turkey, 2005; p. 10.
46. Bishop, C. *Pattern Recognition and Machine Learning*, 1st ed.; Springer-Verlag: New York, NY, USA, 2006; p. 738.
47. Vasant, P.M. *Meta-Heuristics Optimization Algorithms in Engineering, Business, Economics, and Finance*; IGI Global: Hershey, PA, USA, 2013; p. 734.



© 2017 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/>).

Copyright of Energies (19961073) is the property of MDPI Publishing and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.