Chemical and Process Engineering: New Frontiers, 2023, 44(3), e19
DOI: 10.24425/cpe.2023.146721

Using an artificial neural network model for natural gas heat combustion forecasting

Jolanta Szoplik*, Paulina Muchel

West Pomeranian University of Technology in Szczecin, Faculty of Chemical Technology and Engineering, Department of Chemical and Process Engineering, Piastów 42, 71-065 Szczecin, Poland

*** Corresponding author:**
e-mail:
joanta.szoplik@zut.edu.pl

Presented at 24th Polish Conference of Chemical and Process Engineering, 13–16 June 2023, Szczecin, Poland.

Article info:
Received: 10 May 2023
Revised: 26 June 2023
Accepted: 4 July 2023

Abstract
One of the parameters characterizing the quality of the gaseous fuel transported in gas pipeline network to consumers and being the basis for the classification of gaseous fuels is the heat of combustion. The main research hypothesis of this paper is the analysis of the possibility of using MLP 18-y_i-1 neural network model to forecast the natural gas heat of combustion with a forecast error smaller than in case it calculates the heat of combustion based on the composition of natural gas predicted using the MLP 18-65-5 (Szoplik and Muchel, 2023). The training of the models was carried out on the basis of 8760 real data, presenting the hourly heat of natural gas combustion at one of the measurement points of this parameter in the pipeline network. The model takes into account the influence of calendar factors (month, day of the month, day of the week and hour of the day) and weather factors (ambient temperature) on the amount of heat of natural gas combustion in a given location of the gas network. Many MLP 18-y_i-1 models were trained, differing in the number of neurons in the hidden layer and activation functions of neurons in the hidden and output layers.

Keywords
heat of combustion variability, heat of combustion forecasting, artificial neural network model, risk analysis

1. INTRODUCTION

One of the classification parameters of gaseous fuels and determining the usefulness of a given fuel in industry or household is the heat of combustion. Natural gas transported to consumers via pipeline network is a mixture of natural gas extracted in the country, delivered from other countries by land, or regasified LNG gas transported to the country by sea from various suppliers. The natural gas composition depends on the location of extraction and the degree of purification before transport, as well as the dynamics of the gas flow in the network due to the variable distribution of gas from the network. The detailed composition and heat of combustion of natural gas are monitored on an ongoing basis at many points of the gas network throughout the country with a frequency of several minutes. The analysis of such data showed that at a given point of measurement of natural gas parameters in the gas network, the variability of the heat of natural gas combustion in time is observed, depending mainly on the gas composition (direct factor) and network load (indirect factor) (Szoplik and Muchel, 2023). Based on the composition of natural gas, the exact value of the heat of combustion of the gas H_v [MJ/m^3] can be determined from Eqs. (1) and (2) from the PN-EN ISO 6976:2016-11 standard:

$$H_v(t_1, t_2, p_2) = \frac{\sum_{j=1}^{N}\left(x_j\left[(H_c)_G^o\right]_j(t_1)\right)}{V} = \frac{\sum_{j=1}^{N}\left(x_j\left[(H_c)_G^o\right]_j(t_1)\right)}{\frac{Z_{(t_2,p_2)}RT_2}{p_2}} \quad (1)$$

$$Z_{(t_2,p_2)} = 1 - \left(\frac{p_2}{p_0}\right)\left[\sum_{j=1}^{N} x_j s_j(t_2, p_0)\right]^2 \quad (2)$$

where: $[(H_c)_G^o]_j(t_1)$ – ideal gross molar-basis calorific value of component j, kJ/mol, V – real-gas molar volume of the mixture, m^3/mol, x_j – mole fraction of component j of natural gas, N – number of components in a mixture, R – molar gas constant, J/molK, T – thermodynamic temperature, K, t – Celsius temperature, °C, p – pressure, kPa, p_0 – atmospheric pressure, kPa, $Z_{(t_2,p_2)}$ – compression factor at the metering reference conditions, s_j – summation factor.

The knowledge of the heat of combustion of the fuel in the gas network is necessary for the correct transport of gaseous fuel through the network and for the settlement and balancing of the system. In addition to time-consuming and costly laboratory testing of natural gas composition using gas chromatography, simulations of the flow of gas mixtures in network pipelines are frequently used methods for tracking the composition of natural gas. This method, however, requires a detailed description of the operations taking place in the system (compression, jet pressure reduction) and the network topology. In (Alves Jr. and Fontes, 2022), a model was proposed to simulate the dynamics of changes in the composition and heat of combustion of gas being a mixture of gases supplied by several suppliers to the Brazilian gas network. Studies have shown that knowledge of gas composition and analysis of its changes can lead to a reduction in gas distribution costs. Models for simulating the flow of gas mixtures in the network can be used to analyze changes in gas composition depending on the network load (Bermúdez and Shabani,

2022; Chaczykowski and Zarodkiewicz, 2017; Chaczykowski et al., 2018; Fan et al., 2021; Osiadacz and Chaczykowski, 2020) or in the case of adding alternative fuels to the network, such as hydrogen (Saedi et al., 2021; Zhang et al., 2022). The authors of the study (Chaczykowski et al., 2018) indicate the need to monitor changes in the composition of natural gas in the gas network, especially in the conditions of introducing energy from unconventional sources into the pipelines, performed under short-term contracts.

Another method of obtaining information on the composition and heat of combustion of natural gas at a given point in the network of pipelines distributing gas to consumers can be forecasting using an artificial neural network (ANN) model. ANN models make it possible to take into account the influence of many factors on the forecast value without the need to describe these relationships in detail. Numerous studies confirm the usefulness of artificial neural network models for forecasting demand for natural gas in a given production process (Androjić and Dolaček-Alduk, 2018), for a given city (Szoplik, 2015), for a given country (Merkel et al., 2018), continent or world (Hafezi et al., 2019). The models took into account various factors (calendar, economic, demographic or meteorological) having a significant impact on gas demand. ANN models have also been used to monitor the safe transport of natural gas through the pipeline, taking into account corrosion (Wen et al., 2019) or cracking (Wang et al., 2022) of the pipeline walls. The transport of natural gas through the pipeline is accompanied by changes in the temperature and pressure of the gas stream, which affect the change of important physical parameters of the gas. ANN models were also used to predict the compressibility factor (Farzaneh-Gord et al., 2021; Okoro et al., 2022; Sanjari and Nemati Lay, 2012), which is a measure of the deviation of a real gas from an ideal gas. Chu et al. (2021) predicted the compressibility factor (Z-factor), speed of sound and viscosity of natural gas, while Farzaneh-Gord et al. (2018) predicted the compressibility factor and speed of sound of the natural gas as parameters necessary for the correct calibration of flowmeters using sonic nozzles. ANN models were also used to predict the viscosity (AlQuraishi and Shokir, 2011; Nemati Lay et al., 2012) and natural gas density (AlQuraishi and Shokir, 2011; Wood and Choubineh, 2020) depending on the composition, pressure and temperature of the gas stream. The number of natural gas components included in the ANN models as input data varied from 10 to 21, and the gas composition was expressed as molar fractions. On the other hand, in (Szoplik and Muchel, 2023), an MLP model of a neural network was proposed for forecasting five key components of natural gas depending on selected calendar factors and ambient temperature. The number of neurons in the hidden layer of the network and the activation functions of neurons in the hidden and output layers were selected experimentally.

Depending on the complexity of the analyzed problem, various ANN models were used. Forecasting one or more output parameters was the most often performed using the MLP or RBF model of a neural network (Bagheri et al., 2019; Mohandes et al., 2000; Okoro et al., 2022; Wang et al., 2018; Szoplik and Muchel, 2023; Yang et al., 2020; Zhang et al., 2016). In the case of forecasting many parameters, two different approaches were used simultaneously: many identical ANN network models were designed to forecast each parameter individually (AlQuraishi and Shokir, 2011; Chu et al., 2021; Farzaneh-Gord et al., 2018; Kharitonova et al., 2019; Piotrowski et al., 2003; Postawa et al., 2022; Sharifi et al., 2020; Shenbagaraj et al., 2021) or one network model with multiple outputs was designed (George et al., 2018; Park et al., 2020; Postawa et al., 2022; Singh et al., 2007; Szoplik and Muchel, 2023). Models with multiple neurons in the output layer of the network have been used to predict the composition of mixtures when interdependence between the components of the mixture is observed. However, in the case where there is no relationship between the forecasted parameters, several forecasting models were used and a smaller forecast error of individual parameters was obtained compared to the forecast error made using one model with multiple outputs.

However, regardless of the ANN model adopted in the research, in each case many network models were trained and models with the highest correlation coefficients and the smallest forecast errors (MAPE, MSE, RMSE) were selected experimentally. Various network learning algorithms (LM, SDRP, BFGS and others) were used in the research, various activation functions of neurons (logistic, linear, exponential and others), and training of models was carried out on different sizes of real or theoretical data sets.

The use of the MLP model of a neural network for forecasting selected parameters is particularly useful when the forecasted value depends on many factors, and their influence is difficult to describe with known dependencies. The value of the gas heat combustion directly depends on the composition of the gas, while in the case of transporting gas through a network of pipelines fed by different suppliers and delivered to customers who consume gas in different quantities, it is difficult to predict the heat of combustion using simple forecasting models. The choice of the MLP model for forecasting the heat of combustion of gas transported through a network of pipelines supplying gas to customers allows taking into account the influence of factors directly (gas composition) and indirectly (ambient temperature, month, day of the week or hour of the day) affecting its value.

The main research hypothesis of this paper is the analysis of the possibility of using the feed-forward model of the MLP 18-y_i-1 neural network to forecast the heat of natural gas combustion with a forecast error smaller than in case it calculates the heat of combustion based on the composition of natural gas predicted using the MLP 18-65-5. The training of the neural network models was carried out on the basis of 8760 real data, presenting the hourly heat of combustion of natural gas at one of the measurement points of this parameter in the pipeline network distributing gas in Poland.

The model takes into account the influence of calendar factors (month, day of the month, day of the week and hour of the day) and weather factors (ambient temperature) on the amount of heat of natural gas combustion in a given location of the gas pipeline network. Many MLP 18-y_i-1 models were trained, differing in the number of neurons in the hidden layer of the network and the activation functions of neurons in the hidden and output layers. The evaluation of the quality of the trained models and the selection of the model for forecasting the heat of combustion of natural gas were made on the basis of the correlation coefficient of the test set and the calculated MAPE error of the forecast. It was assumed that the best quality model is characterized by the smallest MAPE forecast error. The trained MLP model can be used to forecast the heat of combustion value of natural gas at a given measurement point in the gas network for any day of the month and week, as well as ambient temperature and hour of day. On the other hand, the analysis of the risk of incorrect forecast was used to estimate the probability of obtaining a forecast with error greater than the average MAPE forecast error.

2. RESEARCH METHODOLOGY

The research methodology and the scope of analyses of the heat of combustion forecasting results are schematically shown in Fig. 1. The first stage of the research included the selection of factors influencing the variability of the natural gas heat of combustion H_v. On the basis of 8760 data presenting the variability of the heat of combustion value over time, resulting from the variability of the natural gas composition and the load on the pipeline network, five parameters influencing the heat of combustion at a given time were selected. A set of 8760 (data for 2018) results of the heat of combustion of natural gas calculated on the basis of the real composition of natural gas (example data provided by the GAZ-SYSTEM S.A. – Gas Transmission Operator of the Polish natural gas network) or heat of combustion value calculated on the basis of the gas composition predicted using the MLP 18-65-5 model (Szoplik and Muchel, 2023) was prepared. These values were used to assess the quality of heat of combustion predictions obtained with the new trained MLP models.

The second stage consisted in training MLP models with a different number of neurons in the hidden layer or different activation functions of neurons for forecasting natural gas heat of combustion. The MLP 18-65-1 model (activation functions: logistic-logistic) had the same number of neurons in the hidden network layer and activation functions as in the model used to predict the composition of natural gas (Szoplik and Muchel, 2023). Subsequent models differed from the previous ones in the number of neurons in the hidden layer or in the assumed activation functions of MLP 18-y_i-1 neurons. The selection of the best quality MLP model from among the trained ones was made on the basis of the correlation coefficient R and the MAPE error of the forecast, defined by Eq. (3):

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{H_{V(\mathrm{real})} - H_{V(\mathrm{MLP})}}{H_{V(\mathrm{real})}}\right| \tag{3}$$

where: $H_{V(\mathrm{real})}$ – real natural gas heat of combustion [MJ/m^3], $H_{V(\mathrm{MLP})}$ – forecasting natural gas heat of combustion [MJ/m^3], N – number of all cases in new data set for forecasting ($N = 8760$).

The best quality MLP 18-y_i-1 models were used to prepare forecasts of natural gas heat of combustion value for the new input data set ($N = 8760$ data for 2019). In the third stage

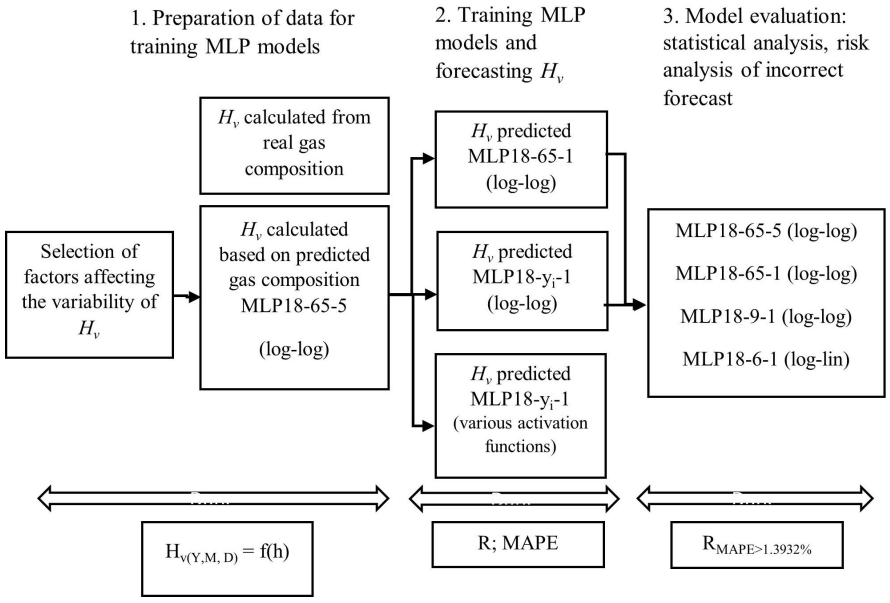

Figure 1. Research methodology.

of the research, the obtained results of the forecasts were subjected to a statistical analysis of the significance of differences between the average values of the heat of combustion of gas, forecasted by the best quality MLP models. The obtained results of the heat of combustion forecasts were also compared with the values of the heat of combustion calculated on the basis of the composition of natural gas predicted by the MLP 18-65-5 model (Szoplik and Muchel, 2023). The quality of the results of the heat of combustion forecasted and calculated on the basis of the forecasted composition was assessed according to the probability (risk analysis) of obtaining a forecast with a MAPE error $> 1.3932\%$.

2.1. Analysis of heat of combustion variability over time

The value of the natural gas heat of combustion depends on the composition of the gas, i.e. the content of the most abundant component (methane) and the content of the component with the highest heat of combustion value (ethane). The contents of other gas components have a smaller impact on the heat of combustion of natural gas transported through the pipeline network. Fig. 2 shows the variability of the methane and ethane content and the heat of combustion H_v of natural gas in subsequent hours in 2018 at one of the measurement points for the properties of natural gas. There is a clear impact of the content of these components in the gas on the heat of combustion value, which decreases with the increase in the methane content and increases with the increase in the ethane content. A detailed statistical analysis of the variability of natural gas composition over time is included in the paper (Szoplik and Muchel, 2023), where it was shown that this variability results from supplying the network with gas of different composition and additionally depends on the dynamics of gas in the network (gas consumption from the network by consumers).

Based on the analysis of real data on the demand for gas by consumers, an analysis of the variability of the load of the sample network pipeline was carried out and it was shown that the ambient temperature, time of day and day of the week and month have the greatest impact on gas consumption.

Figure 3 shows the results presenting the differences between the average values of the monthly, daily or hourly gas flow and the average values of the annual, monthly or daily gas pipeline load, respectively. Figure 3a shows the variability of the gas pipeline load in relation to the average annual value in subsequent months of the year and the average temperature of the month. There are clear differences in the monthly variation of the gas pipeline load, which depend on the ambient temperature (Fig. 3b). In addition, the load on the pipeline transporting gas depends on the hour of the day and day of the week (Fig. 3c). On days with low ambient temperatures (winter months) and during the morning rush hours (7 am on weekdays or 10 am on holidays), the load

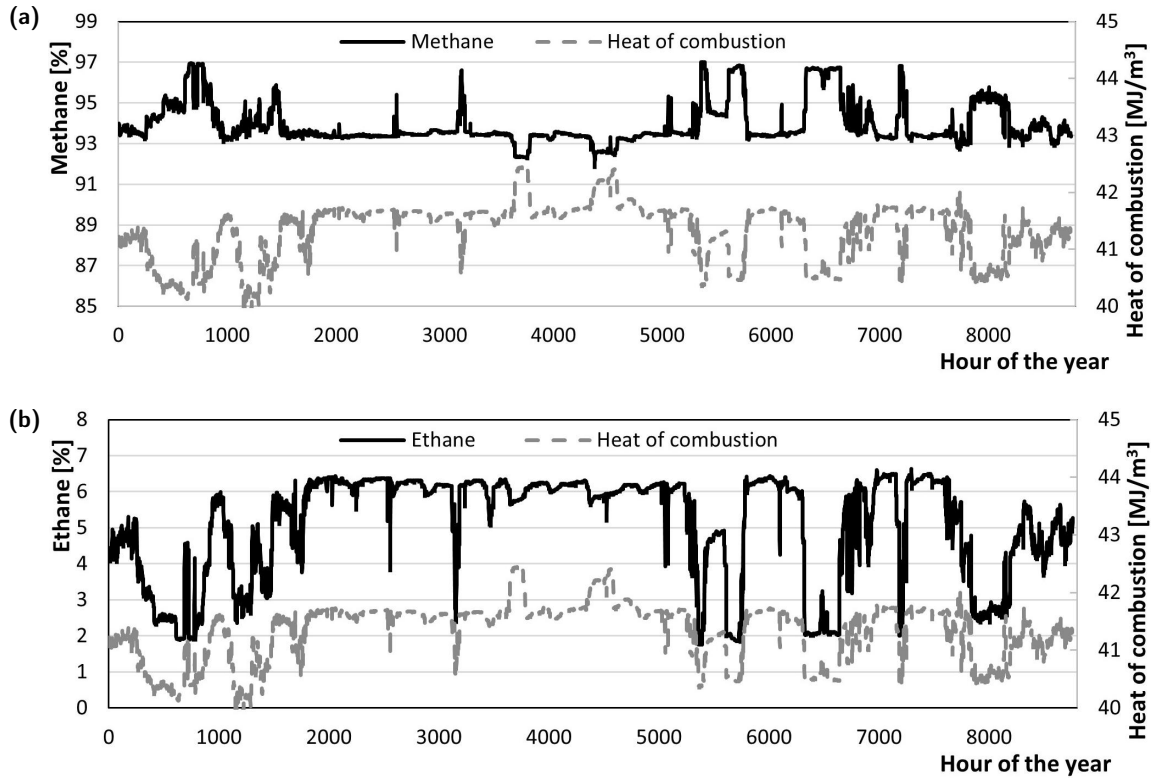

Figure 2. Variability of the content of two selected components and the heat of combustion value of natural gas over time; (a) results for methane; (b) results for ethane.

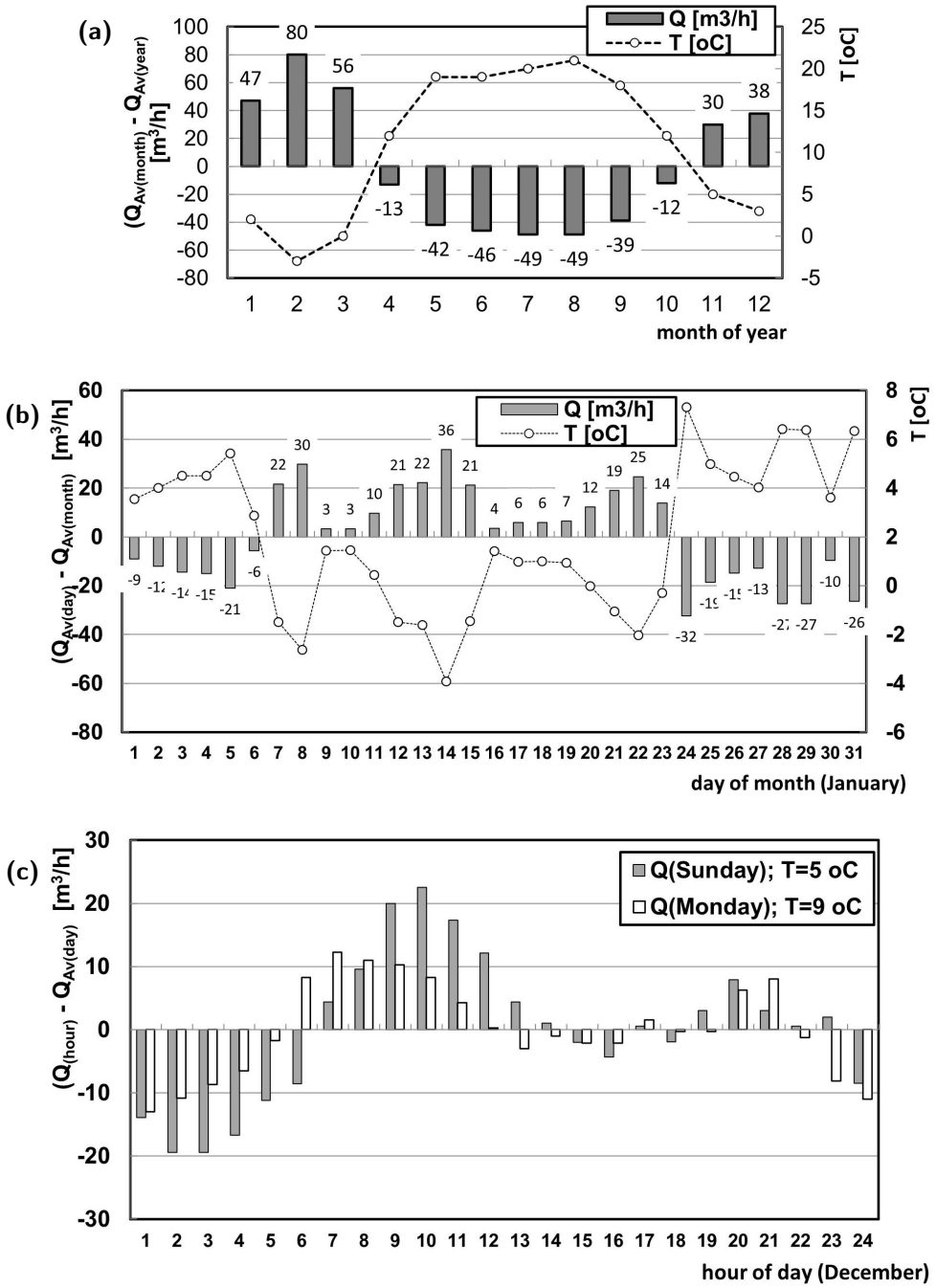

Figure 3. Natural gas pipeline load variation; (a) monthly, (b) daily, (c) hourly.

on the network pipeline is clearly higher than the average. On the other hand, in the months of the summer season (high ambient temperatures), the load on the network's gas pipelines is much lower than the average. A clear influence of the ambient temperature on the load on the gas pipelines of the network is particularly visible on the days of the winter months (low ambient temperatures). Based on the analysis of the results (Fig. 3) of the variation of the gas network load, resulting from the changing demand for gas by consumers taking gas from the network, calendar factors (day of the week, day of the month, month and hour of the day) and meteorological factors (ambient temperature) were selected. Therefore, these factors affect the change in the load on the gas network in subsequent hours of the year. The variability of gas demand by consumers and the load on network pipelines were also analyzed elsewhere (Szoplik and Muchel, 2023; Szoplik and Stelmasińska, 2019; Szoplik, 2015).

2.2. Training the MLP network model

The MLP (MultiLayer Perceptron) model is one of the most frequently used artificial neural network models in forecasting. The model is built of many neurons arranged in at least

three layers (Fig. 4), and the flow of signals takes place only in one direction from the input layer (X) to the output layer (Z). The number of neurons in the hidden layer (Y) of the network depends on the number of variables included in the model that affect the predicted value. In the case of training the MLP model of the artificial neural network with one hidden layer (Fig. 4) for forecasting the heat of combustion of natural gas 15 qualitative variables (3 types of the day of the week D1, D2, D3 and 12 months M1, ..., M12) were taken into account in the input layer and 3 quantitative variables (air temperature T, hour of the day h and day of the month D_M), while the output layer contained only 1 quantitative variable (heat of combustion of gas H_V). The number of neurons in the hidden layer of the MLP model, which depends on the number of neurons in the input layer, was determined experimentally. Fig. 4 shows a diagram of a neural network model built of an input layer X containing 18 neurons, one hidden layer Y containing y_i neurons and 1 neuron in the output layer Z. Signals received in neurons of the input layer in the form of information about the month, day of the month, the type of the day of the week, the hour of the day and ambient temperature are subject to weighting (w_i), and then they are sent to all neurons of the hidden layer, where their weighted summation takes place and the output signal is activated. Then, after assigning successive weights (w_j), the signals from the neurons of the hidden layer are sent to the neuron of the output layer Z. In the neurons of the output layer, the weighted signals are added up, and after activation, they constitute the predicted value of the heat of combustion of natural gas H_V.

The training of the MLP model was carried out on a set of 8760 cases, which were randomly divided into 3 subsets: training – 70%, test – 20% and validation – 10% of the entire data set. The learning process was carried out using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) variable metric method. The study tested various functions of neurons activation in the hidden and output layers: linear, logistic, hyperbolic tangent and exponential. A total of several hundred different structures of the MLP 18-y_i-1 network were trained, changing the number of neurons in the hidden layer y_j in the range from 5 to 65, of which several dozen were selected to prepare forecasts of natural gas heat of combustion for the new set of input data. Data for training the network in the form of real example heat of combustion of gas recorded every hour in the period from 1 January 2018 to 31 December 2018 were made available by the operator of the Polish natural gas transmission network GAZ-SYSTEM S.A., while weather data were read from the meteorological database.

2.3. Results of natural gas heat of combustion forecasts

The best quality MLP models of the network (trained on data specific for 2018) differing in the number of neurons in the hidden layer and the logistic activation function of the neurons of the hidden and output layers, and with different functions of neurons activation were used to prepare forecasts of the natural gas heat of combustion for the new set of input

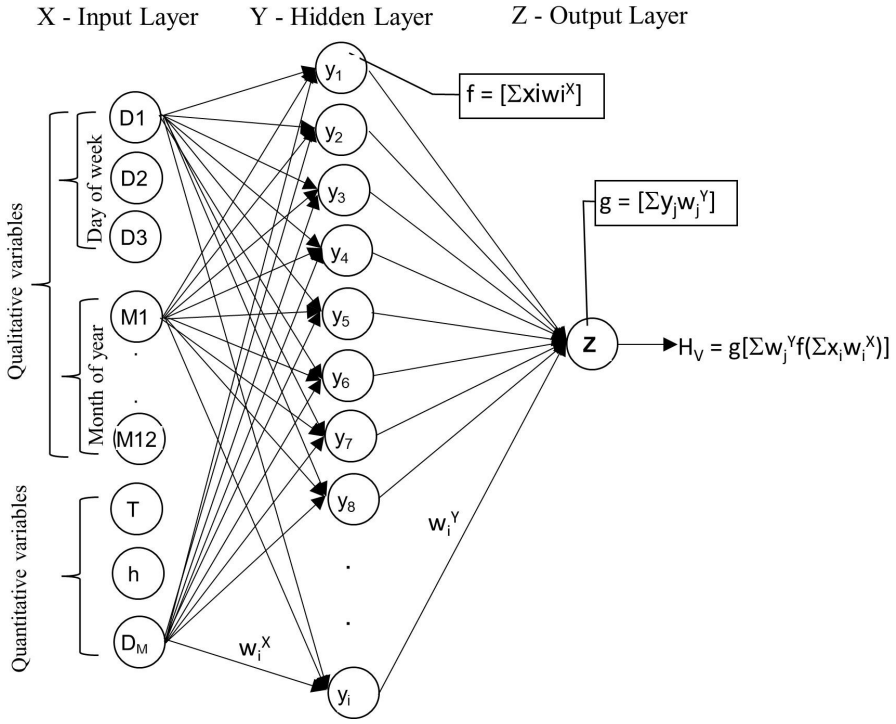

Figure 4. MultiLayer Perceptrons (MLP) with eighteen neutrons in input layer X, y_i neutrons in hidden layer Y and one neutron in output layer Z; D1 – workday, D2 – weekday, D3 – holiday, M1 – January, ..., M12 – December, T – temperature, h – hour of day, D_M – day of Month.

data (data for 2019). On the basis of the determined MAPE errors of H_V forecasts determined individually for each MLP model of the network, 10 models with the smallest MAPE average values were selected.

Figure 5 shows the determined MAPE errors for the predictions obtained using MLP models with different numbers of neurons in the hidden layer and with the logistic activation function in the hidden and output layers, and the values of the correlation coefficient for the R_T test set for each network model are additionally marked. Increasing the number of neurons in the hidden layer of the network increases the R_T coefficient, but does not improve the quality of combustion heat forecasts. This is evidenced by the MAPE error values of the forecasts made for the new sets of input data. Among the MLP models tested in the research, the MLP 18-9-1 model was considered the best quality, for which the MAPE error = 1.3811%.

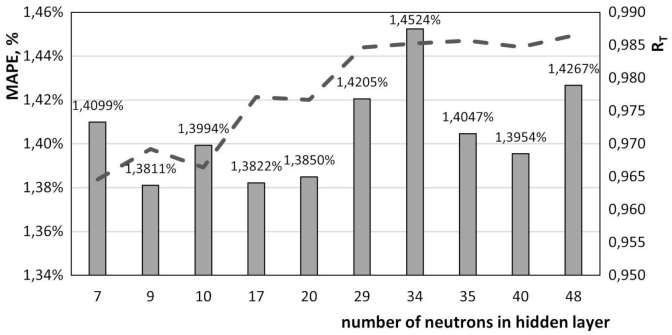

Figure 5. Summary of coefficients R_T (dashed line) and error MAPE (columns) for different structure of MLP networks models and logistic activation function of neurons in the hidden and output layers (results for 2019).

Similar research results, however, obtained using MLP models with a different number of neurons in the hidden layer and different activation functions of the neurons of the hidden and output layers are summarized in Table 1. The change of the activation function of neurons slightly affects the value of the MAPE error calculated on the basis of forecasts made for the new set of input data. Among the network models listed in Table 1, the MLP 18-6-1 model with the functions of neuron activation: logistic and linear was considered the best quality. The forecast error for this model MAPE = 1.3609% is smaller compared to the MLP 18-9-1 model with logistic activation functions.

Comparison of the results of H_V forecasts obtained using the best quality MLP 18-9-1 model with the activation functions (logistic-logistic) with the real values of the heat of combustion of natural gas is shown in Fig. 6.

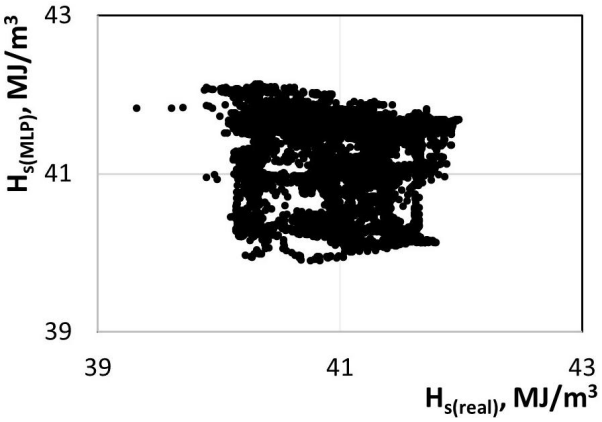

Figure 6. Comparison of the values predicted by the MLP 18-9-1 model and the real value of heat of combustion of natural gas.

Table 1. Summary of coefficients R and MAPE error for different structure of MLP networks models and different activation function of neurons in the hidden and output layers.

Structure of MLP	R_L	R_T	R_V	The function of activating neurons in the hidden layer	The function of activating neurons in the output layer	MAPE, %
MLP 18-5-1	0.9537	0.9512	0.9486	logistic	linear	1.3850
MLP 18-6-1	**0.9635**	**0.9603**	**0.9602**	**logistic**	**linear**	**1.3609**
MLP 18-7-1	0.9650	0.9612	0.9604	logistic	linear	1.3874
MLP 18-9-1	0.9626	0.9614	0.9618	logistic	linear	1.3744
MLP 18-8-1	0.9754	0.9727	0.9708	logistic	tanh	1.3787
MLP 18-10-1	0.9716	0.9695	0.9664	logistic	tanh	1.3830
MLP 18-17-1	0.9829	0.9806	0.9814	tanh	logistic	1.3845
MLP 18-20-1	0.9813	0.9793	0.9792	tanh	logistic	1.3820
MLP 18-9-1	0.9703	0.9670	0.9632	tanh	linear	1.3829
MLP 18-19-1	0.9775	0.9750	0.9750	tanh	linear	1.3869

3. EVALUATION OF THE QUALITY OF NATURAL GAS HEAT OF COMBUSTION FORECASTS

The assessment of the quality of heat of combustion of natural gas forecasts was made on the basis of a statistical analysis of the average values of heat of combustion value of gas forecast using various MLP models and the results of the probability of obtaining a forecast with an error greater than the MAPE error = 1.3932% of the heat of combustion value calculated on the basis of the forecast composition of natural gas.

3.1. Statistical analysis of the average values of gas combustion heat forecasts

The results of calculations of the mean values and the standard deviation of the predicted heat of combustion of natural gas obtained using the best MLP models differing in the number of neurons in the hidden layer or activation functions of neurons in the hidden and output layers are summarized in Table 2. The statistical significance of differences between successive pairs of H_V means was assessed using t-Student test or Cochran–Cox test at the significance level $\alpha = 5\%$. The choice of the test depended on the fulfillment of the condition of insignificance (t-Student test) or significance (Cochran–Cox test) of differences between the standard deviations of the predicted heat of combustion of gas in the compared sets of results.

Table 2. Calculation results of the standard deviation and mean heat of combustion value predicted by selected MLP models.

Model MLP	MAPE$_{HV}$ [%]	H_V (average) [MJ/m^3]	(s_{HV}) [MJ/m^3]
MLP 18-65-5	1.3932 [log–log]	41.223	0.5419
MLP 18-65-1	1.4507 [log–log]	41.230	0.5658
MLP 18-9-1	1.3811 [log–log]	41.226	0.5399
MLP 18-6-1	1.3609 [log–lin]	41.227	0.5285

The Bartlett test was used to assess the significance of differences between the standard deviation values of the heat of combustion value forecasts, in which the statistic of the test X^2 is described in Eq. (4):

$$X^2 = \frac{2.303}{1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k}\frac{1}{n_i - 1} - \frac{1}{N-k}\right)} \cdot \left[(N-k)\lg\left(\frac{1}{N-k}\sum_{i=1}^{k} s_i^2(n_i-1)\right) - \sum_{i=1}^{k}(n_i-1)\lg\left(s_i^2\right)\right] \quad (4)$$

where: k – number of sets compared with each other, n_i – number of cases in i set of results, N – number of all cases in all sets of results, s_i – standard deviation for the i set of results.

The value of X^2_{kr} was read from a dedicated table for the significance level $\alpha = 0.05$ and the number of degrees of freedom $f = k - 1$. The calculation results in the form of the test statistics and critical values of the test, presented in Table 3, indicate that the standard deviations of the predicted heat of combustion values differ statistically significantly for the pair of models MLP 18-65-5 and MLP 18-65-1, and then MLP 18-65-1 and MLP 18-9-1. Therefore, the average heat of combustion values will be compared using the Cochran–Cox test. In the case of using the MLP 18-65-5 model, the heat of combustion of the natural gas was calculated from the Eq. (1) based on the predicted content of five components of natural gas, while the use of the MLP 18-65-1 model enabled direct forecasting of the heat of combustion value. The lack of a statistically significant difference in the values of the standard deviation of the heat of combustion value for the remaining pairs of models (Table 3) made it possible to use the t-Student test to compare the average values of the predicted heat of combustion value. In this case, the significance of differences between standard deviations of H_V predictions obtained using models with a similar number of neurons in the hidden layer and different activation functions was compared.

The average values of the predicted heat of combustion were compared using the Cochran–Cox test, in which the statistics of the test C and the critical value of the test C_{kr} were calculated from Eqs. (5) and (6), respectively:

$$C = \frac{|x_{av1} - x_{av2}|}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}} \quad (5)$$

$$C_{kr} = \frac{\frac{s_1^2}{n_1 - 1}t_1 + \frac{s_2^2}{n_2 - 1}t_2}{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}} \quad (6)$$

where: x_{av1}, x_{av2} – average values in sets 1 and 2, s_1^2, s_2^2 – variance values in sets 1 and 2, n_1, n_2 – number of cases in sets 1 and 2, t_1, t_2 – critical values read from the t-Student distribution table for the number of degrees of freedom resulting from the equations: $f_1 = n_1 - 1$, $f_2 = n_2 - 1$ and for the adopted significance level α for the set 1 and 2.

In the second case, the significance of the differences between the average values of the forecast heats of combustion was assessed on the basis of the t-Student test using Eq. (7):

$$t = \frac{|x_{av1} - x_{av2}|}{\sqrt{s_1^2 + s_2^2}}\sqrt{n} \quad (7)$$

where: $n = n_1 = n_2$ – the size of the set. The critical value t_{kr} was read from the t-Student distribution table for the

Table 3. Summary of Bartlett's test results.

Compared MLP models		The value of the X^2 Bartlett test statistic	Critical value of the X^2_{kr} Bartlett test	Significance of differences between standard deviations (variances) $H_0 : X^2 < X^2_{kr}$ $H_1 : X^2 > X^2_{kr}$
MLP 18-65-5 [log–log]	MLP 18-65-1 [log–log]	6.9669	3.84	significant
MLP 18-65-1 [log–log]	MLP 18-9-1 [log–log]	8.2365	3.84	significant
MLP 18-9-1 [log–log]	MLP 18-6-1 [log–lin]	1.6957	3.84	insignificant
MLP 18-9-1 [log–log]	MLP 18-9-1 [log–lin]	0.2381	3.84	insignificant
MLP 18-9-1 [log–log]	MLP 18-8-1 [log–tan]	0.3687	3.84	insignificant
MLP 18-9-1 [log–log]	MLP 18-20-1 [tan–log]	0.8504	3.84	insignificant

H_0: the averages for the compared sets do not differ statistically significantly;
H_1: the averages for the compared sets differ statistically significantly

significance level $\alpha = 0.05$ and the number of degrees of freedom $f_1 = n_1 + n_2 - 2$.

The results of the Cochran–Cox and t-Student test statistics calculations and the relevant critical values of the tests are presented in Table 4. It has been shown that in each analyzed case there are no statistically significant differences between the compared pairs of average values of the predicted natural gas heat of combustion. The results of this stage of research show that simplifying the MLP model (e.g. reducing the number of neurons in the hidden or output layer of the network or changing the activation function of neurons) reduces the MAPE error of the heat of combustion value forecast from 1.3932% (for MLP 18-65-5) to 1.3609% (for MLP 18-6-1). The average heat of combustion forecast by successive MLP models did not differ statistically significantly, but increasingly simpler MLP models were obtained without deteriorating the quality of H_V forecasts.

3.2. Risk analysis for incorrect of natural gas heat of combustion forecast

The MLP forecasting models were also compared according to the probability of receiving the heat of combustion value forecast with a MAPE error $> 1.3932\%$. It was assumed that the factors adopted in the forecasting model as input data (day of the week, month, temperature or time of day) influence the probability of incorrect forecast to a different extent and partial risk coefficients $R_{p,\alpha}$ were determined from Eq. (8):

$$R_{p,\alpha} = P\left(\text{MAPE} \geq 1.3932\%,\ \alpha\right) = \frac{n_{\text{MAPE}\geq 1.3932\%,\alpha}}{n_\alpha} \quad (8)$$

Table 4. Summary of Cochran–Cox and t-Student test results.

Compared MLP models		The value of the C Cochran–Cox test statistic	Critical value of the C_{kr} Cochran–Cox test	Significance of differences between averages $H_0 : C < C_{kr}$ $H_1 : C > C_{kr}$
MLP 18-65-5 [log–log]	MLP 18-65-1 [log–log]	0.8538	4.0964	insignificant
MLP 18-65-1 [log–log]	MLP 18-9-1 [log–log]	0.4572	3.7447	insignificant
		The value of the t t-Student test statistic	Critical value of the t_{kr} t-Student test	$H_0 : t < t_{kr}$ $H_1 : t > t_{kr}$
MLP 18-9-1 [log–log]	MLP 18-6-1 [log–lin]	0.1352	1.9600	insignificant
MLP 18-9-1 [log–log]	MLP 18-9-1 [log–lin]	0.3815	1.9600	insignificant
MLP 18-9-1 [log–log]	MLP 18-8-1 [log–tan]	0.6373	1.9600	insignificant
MLP 18-9-1 [log–log]	MLP 18-20-1 [tan-log]	0.0017	1.9600	insignificant

H_0: the averages for the compared sets do not differ statistically significantly;
H_1: the averages for the compared sets differ statistically significantly

where: α – factor type, $n_{\mathrm{MAPE}\geq 1.3932\%,\alpha}$ – number of cases for factor α for which MAPE error $\geq$ 1.3932%, n_α – total number of cases for factor α.

Calculations were made for the results obtained for three models of the MLP network separately for each factor influencing the heat of gas combustion. The partial risk ratios listed in Table 5 enable the estimation of the cumulative risk of incorrect forecasting (MAPE error > 1.3932%) of the heat of combustion of natural gas for any day of the week and month, as well as ambient temperature and time of day.

The value of the cumulative risk $R_t$was estimated on the basis of Eq. (9):

$$R_t = P\left(\mathrm{MAPE} \geq 1.3932\%,\ \alpha\right) = \frac{1}{N_\alpha}\sum R_{p,\alpha} \qquad (9)$$

where: N_α – number of factors taken into account in the risk model.

The results of R_t risk calculations for three models (MLP 18-65-5, MLP 18-65-1, MLP 18-6-1) for twelve assumed sets of input data (day of the week, month, ambient temperature and time of day) are shown in Fig. 7. The MAPE error

Table 5. Values of partial risk $R_{p,\alpha}$ for each factor subset for different MLP models.

No.	Factor type	Factor subset	Partial risk [MLP 18-65-5 log–log], %	Partial risk [MLP 18-65-1 log–log], %	Partial risk [MLP 18-6-1 log–lin], %
1.	Day of the week	Working day	42	46	41
2.		Weekend day	46	46	43
3.		Holiday day	22	17	19
4.		January	71	66	67
5.		February	46	51	48
6.		March	53	50	46
7.		April	27	35	33
8.		May	20	23	21
9.		June	43	43	43
10.		July	42	39	39
11.		August	41	44	36
12.		September	65	71	68
13.		October	44	47	41
14.		November	51	51	44
15.		December	6	17	3
16.	Hour of the day	5 am–10 am	42	45	40
17.		11 am–3 pm	42	44	41
18.		4 pm–10 pm	43	45	40
19.		11 pm–4 am	42	45	40
20.		–10–5 °C	43	45	40
21.		6–18 °C	43	46	42
22.		19–37 °C	41	41	38

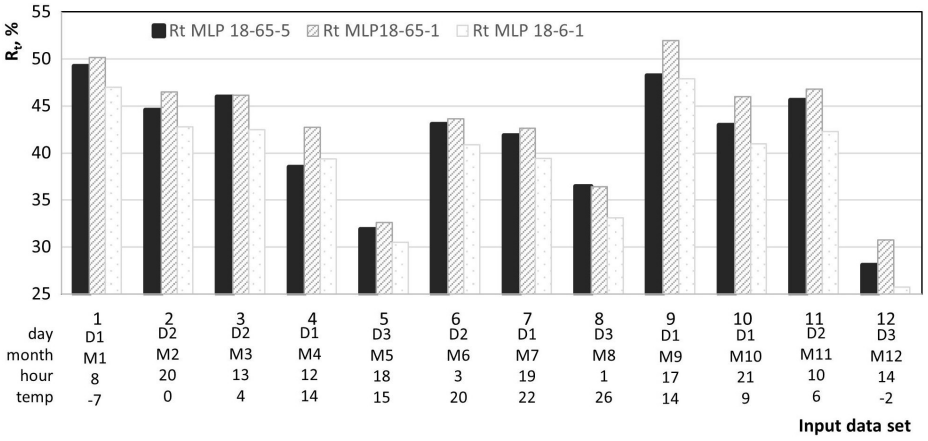

Figure 7. R_t cumulative risk values determined for 12 selected sets of input data (*D1* – workday, *D2* – weekday, *D3* – holiday, *M1* – January, ..., *M12* – December) and 3 models of the MLP network used in forecasting the natural gas heat of combustion.

$= 1.3932\%$ was assumed in the calculations, which is characteristic for the heat of combustion calculated on the basis of the gas composition predicted using the MLP 18-65-5 model. Lower cumulative risk R_t means lower probability of obtaining a forecast with MAPE error $> 1.3932\%$. The analysis of the results presented in Fig. 7 showed that among all the analyzed cases, the lowest risk of obtaining the heat of combustion value of natural gas forecast with the MAPE error $> 1.3932\%$ is for the MLP 18-6-1 model. However, the use of the MLP 18-65-1 model is the least favorable (out of the 3 compared models).

4. SUMMARY

Based on a set of 8760 real data presenting the natural gas heat of combustion values and calendar factors (month, day of the week and month, hour of the day) and weather (ambient temperature), many models MLP $18 - y_i - 1$ for forecasting natural gas heat of combustion were trained, differing in the number of neurons in the hidden layer or activation functions of neurons. The best quality MLP models were then used to prepare 8760 heat of combustion value forecasts (for a new set of 8760 input data) and the MLP 18-6-1 model was selected. MLP 18-6-1 model is characterized by a smaller forecast error of MAPE $= 1.3609\%$ compared to the heat of combustion forecast error of MAPE $= 1.3932\%$ for MLP 18-65-5 model, which was calculated on the basis of the predicted natural gas composition. Therefore, the main goal of the article has been achieved – the MLP forecasting model has been designed, which directly forecasts the heat of combustion with a smaller error than in the case of the MLP model, which forecasts the composition of natural gas, and the heat of combustion is calculated based on the forecasted content of individual components in the gas. Based on the cumulative risk analysis, it was also shown that the probability of obtaining a forecast (using the MLP 18-6-1 model) of the heat of combustion of natural gas H_V with the MAPE error $> 1.3932\%$ is much lower compared to the H_V value calculated on the basis of the composition of natural gas forecasted by the MLP 18-65-5 model.

Taking into account the dynamics of changes in the energy market (changes of suppliers, replacement of fossil fuels with other types of fuels or new possibilities of using natural gas), periodic verification of the model and possible training of the model with a new set of input data should be taken into account.

ACKNOWLEDGMENTS

The calculations presented in the paper were made on the basis of real sample data provided by the GAZ-SYSTEM S.A. – Gas Transmission Operator of the Polish natural gas network. The authors would like to thank the company for cooperation.

SYMBOLS

C	value of the statistic of the Cochran–Cox test
C_{kr}	critical value of the statistic of the Cochran–Cox test
$D1$	workday,
$D2$	weekday,
$D3$	holiday,
D_M	day of Month
f	number of degrees of freedom
h	hour of day,
H_0	null hypothesis
H_1	alternative hypothesis
$\left[(H_c)_G^\circ\right]_j(t_1)$	ideal gross molar-basis calorific value of component j, kJ/mol
H_v	heat of combustion of the gas, MJ/m^3
$H_{V(\mathrm{MLP})}$	forecasting natural gas heat of combustion
$H_{V(\mathrm{real})}$	real natural gas heat of combustion
k	number of sets compared with each other
M1, ..., M12	January, ..., December
N	number of all cases in all sets of results
N	number of components in a mixture
n_i	number of cases in i set of results
p	pressure, kPa
p_0	atmospheric pressure, kPa
R	molar gas constant, J/molK
R_L	coefficient of learning set
R_T	coefficient of test set
R_V	coefficient of validation set
s_i	standard deviation for the i set of results
s_i^2	variance values in i set
s_j	summation factor
t	Celsius temperature, ^{o}C
T	temperature, K
t	value of the statistic of the t-Student test
t_i	critical values read from the t-Student distribution table for the number of degrees of freedom and for the adopted significance level α for the i set
t_{kr}	critical value of the statistic of the t-Student test
V	real-gas molar volume of the mixture, m^3/mol
x_{avi}	average values in i set
x_j	mole fraction of component j of natural gas,
$Z_{(t2,p2)}$	compression factor at the metering reference conditions,
α	significance level (in statistical analysis)
α	factor type (in risk analysis)
n_α	total number of cases for factor α (in risk analysis)
N_α	number of factors taken into account in the risk model
$n_{MAPE\geq 1.3932\%,\alpha}$	number of cases for factor α for which MAPE error $\geq 1.3932\%$ (in risk analysis)
R_t	cumulative risk
R_p	partial risk

X^2_{kr} critical value of the statistic of the Bartlett test
X^2 value of the statistic of the Bartlett test

Abbreviations

ANN Artificial Neural Network
BFGS Broyden–Fletcher–Goldfarb–Shanno training algorithm
LM Levenberg–Marquardt training algorithm
LNG Liquefied Natural Gas
MAPE Mean Absolute Percentage Error
MLP MultiLayer Perceptrons
MSE Mean Squared Error
RMSE Root-Mean-Square Error
SDBP Steepest Descent Back Propagation training algorithm

REFERENCES

AlQuraishi A.A., Shokir E.M., 2011. Artificial neural networks modeling for hydrocarbon gas viscosity and density estimation. *J. King Saud Univ. Eng. Sci.*, 23, 123–129. DOI: 10.1016/j.jksues.2011.03.004.

Alves Jr. O., Fontes C., 2022. Modeling and optimization of natural gas distribution networks for new supplier projects. *Energy Convers. Manage.: X*, 15, 100240. DOI: 10.1016/j.ecmx.2022.100240.

Androjić I., Dolaček-Alduk Z., 2018. Artificial neural network model for forecasting energy consumption in hot mix asphalt (HMA) production. *Constr. Build. Mater.*, 170, 424–432. DOI: 10.1016/j.conbuildmat.2018.03.086.

Bagheri M., Akbari A., Mirbagheri S.A., 2019. Advanced control of membrane fouling in filtration systems using artificial intelligence and machine learning techniques: A critical review. *Process Saf. Environ. Prot.*, 123, 229–252. DOI: 10.1016/j.psep.2019.01.013.

Bermúdez A., Shabani M., 2022. Numerical simulation of gas composition tracking in a gas transportation network. *Energy*, 247, 123459. DOI: 10.1016/j.energy.2022.123459.

Chaczykowski M., Sund F., Zarodkiewicz P., Hope S.M., 2018. Gas composition tracking in transient pipeline flow. *J. Nat. Gas Sci. Eng.*, 55, 321–330. DOI: 10.1016/j.jngse.2018.03.014.

Chaczykowski M., Zarodkiewicz P., 2017. Simulation of natural gas quality distribution for pipeline systems. *Energy*, 134, 681–698. DOI: 10.1016/j.energy.2017.06.020.

Chu J., Liu X., Zhang Z., Zhang Y., He M., 2021. A novel method overcomeing overfitting of artificial neural network for accurate prediction: Application on thermophysical property of natural gas. *Case Stud. Therm. Eng.*, 28, 101406. DOI: 10.1016/j.csite.2021.101406.

Fan D., Gong J., Zhang S., Shi G., Kang Q., Xiao Y., Wu C., 2021. A transient composition tracking method for natural gas pipe networks. *Energy*, 215, 119131. DOI: 10.1016/j.energy.2020.119131.

Farzaneh-Gord M., Mohseni-Gharyehsafa B., Ebrahimi-Moghadam A., Jabari-Moghadam A., Toikka A., Zvereva I., 2018. Precise calculation of natural gas sound speed using neural networks: An application in flow meter calibration. *Flow Meas. Instrum.*, 64, 90–103. DOI: 10.1016/j.flowmeasinst.2018.10.013.

Farzaneh-Gord M., Rahbari H.R., Mohseni-Gharesafa B., Toikka A., Zvereva I., 2021. Accurate determination of natural gas compressibility factor by measuring temperature, pressure and Joule-Thomson coefficient: Artificial neural network approach. *J. Pet. Sci. Technol.*, 202, 108427. DOI: 10.1016/j.petrol.2021.108427.

George J., Arun P., Muraleedharan C., 2018. Assessment of producer gas composition in air gasification of biomass using artificial neural network model. *Int. J. Hydrogen Energy*, 43, 9558–9568. DOI: 10.1016/j.ijhydene.2018.04.007.

Hafezi R., Akhavan A.N., Zamani M., Pakseresht S., Shamshirband S., 2019. Developing a data mining based model to extract predictor factors in energy systems: Application of global natural gas demand. *Energies*, 12, 4124. DOI: 10.3390/en12214124.

Kharitonova O.S., Bronskaya V.V., Ignashina T.V., Al-Muntaser A.A., Khairullina L.E., 2019. Modeling of absorption process using neural networks. *IOP Conf. Ser.: Earth Environ. Sci.*, 315, 032025. DOI: 10.1088/1755-1315/315/3/032025.

Merkel G.D., Povinelli R.J., Brown R.H., 2018. Short-term load forecasting of natural gas with deep neural network regression. *Energies*, 11, 2008. DOI: 10.3390/en11082008.

Mohandes M., Balghonaim A., Kassas M., Rehman S., Halawani T.O., 2000. Use of radial basis functions for estimating monthly mean daily solar radiation. *Sol. Energy*, 68, 161–168. DOI: 10.1016/S0038-092X(99)00071-7.

Nemati Lay E., Peymani M., Sanjari E., 2012. Prediction of natural gas viscosity using artificial neural network approach. *Int. J. Chem. Mol. Eng.*, 6, 577–583. DOI: 10.5281/zenodo.1333845.

Okoro E.E., Ikeora E., Sanni S.E., Aimihke V.J., Ogali O.I., 2022. Adoption of machine learning in estimating compressibility factor for natural gas mixtures under high temperature and pressure applications. *Flow Meas. Instrum.*, 88, 102257. DOI: 10.1016/j.flowmeasinst.2022.102257.

Osiadacz A.J., Chaczykowski M., 2020. Modeling and simulation of gas distribution networks in a multienergy system environment. *Proc. IEEE*, 108, 1580–1595. DOI: 10.1109/JPROC.2020.2989114.

Park J., Cho J., Choi H., Park J., 2020. Prediction of reformed gas composition for diesel engines with a reformed EGR system using an artificial neural network. *Energies*, 13, 5886. DOI: 10.3390/en13225886.

Piotrowski K., Piotrowski J., Schlesinger J., 2003. Modelling of complex liquid–vapour equilibria in the urea synthesis process with the use of artificial neural network. *Chem. Eng. Process. Process Intensif.*, 42, 285–289. DOI: 10.1016/S0255-2701(02)00060-0.

PN-EN ISO 6976:2016-11. *Natural gas – Calculation of calorific values, density, relative density and Wobbe indices from composition.*

Postawa K., Fałtynowicz H., Pstrowska K., SzczygiełJ., Kułażyński M., 2022. Artificial neural networks to differentiate the composition and pyrolysis kinetics of fresh and long-stored maize. *Bioresour. Technol.*, 364, 128137. DOI: 10.1016/j.biortech.2022.128137.

Saedi I., Mhanna S., Mancarella P., 2021. Integrated electricity and gas system modelling with hydrogen injections and gas composition tracking. *Appl. Energy*, 303, 117598. DOI: 10.1016/ j.apenergy.2021.117598.

Sanjari E., Nemati Lay E., 2012. Estimation of natural gas compressibility factors using artificial neural network approach. *J. Nat. Gas Sci. Eng.*, 9, 220–226. DOI: 10.1016/ j.jngse.2012.07.002.

Sharifi K., Sabeti M., Rafiei M., Mohammadi A.H., Ghaffari A., Asl M.H., Yousefi H., 2020. A good contribution of computational fluid dynamics (CFD) and GA-ANN methods to find the best type of helical wire inserted tube in heat exchangers. *Int. J. Therm. Sci.*, 154, 106398. DOI: 10.1016/ j.ijthermalsci.2020.106398.

Shenbagaraj S., Sharma P.K., Sharma A.K., Raghav G., Kota K.B., Ashokkumar V., 2021. Gasification of food waste in supercritical water: An innovative synthesis gas composition prediction model based on Artificial Neural Networks. *Int. J. Hydrogen Energy*, 46, 12739–12757. DOI: 10.1016/j.ijhydene. 2021.01.122.

Singh V., Gupta I., Gupta H.O., 2007. ANN-based estimator for distillation using Levenberg–Marquardt approach. *Eng. Appl. Artif. Intell.*, 20, 249–259. DOI: 0.1016/j.engappai.2006. 06.017.

Szoplik J., 2015. Forecasting of natural gas consumption with artificial neural networks. *Energy*, 85, 208–220. DOI: 10.1016/ j.energy.2015.03.084.

Szoplik J., Muchel P., 2023. Using an artificial neural network model for natural gas compositions forecasting. *Energy*, 263, 126001. DOI: 10.1016/j.energy.2022.126001.

Szoplik J., Stelmasińska P., 2019. Analysis of gas network storage capacity for alternative fuels in Poland. *Energy*, 172, 343–353. DOI: 10.1016/j.energy.2019.01.117.

Wang B., Guo Y., Wang D., Zhang Y., He R., Chen J., 2022. Prediction model of natural gas pipeline crack evolution based on optimized DCNN-LSTM. *Mech. Syst. Signal Process.*, 181, 109557. DOI: 10.1016/j.ymssp.2022.109557.

Wang M., Zhao L., Du R., Wang C., Chen L., Tian L., Stanley H.E., 2018. A novel hybrid method of forecasting crude oil prices using complex network science and artificial intelligence algorithms. *Appl. Energy*, 220, 480–495. DOI: 10.1016/j.apenergy.2018.03.148.

Wen K., He L., Liu J., Gong, J., 2019. An optimization of artificial neural network modeling methodology for the reliability assessment of corroding natural gas pipelines. *J. Loss Prev. Process Ind.*, 60, 1–8. DOI: 10.1016/j.jlp.2019.03.010.

Wood D.A., Choubineh A., 2020. Transparent machine learning provides insightful estimates of natural gas density based on pressure, temperature and compositional variables. *J. Nat. Gas Geosci.*, 5, 33–43. DOI: 10.1016/j.jnggs.2019.12.003.

Yang Z., Mourshed M., Liu K., Xu X., Feng S., 2020. A novel competitive swarm optimized RBF neural network model for short-term solar power generation forecasting. *Neurocomputing*, 397, 415–421. DOI: 10.1016/j.neucom.2019.09.110.

Zhang C., Wei H., Xie L., Shen Y., Zhang K., 2016. Direct interval forecasting of wind speed using radial basis function neural networks in a multi-objective optimization framework. *Neurocomputing*, 205, 53–63. DOI: 10.1016/j.neucom.2016. 03.061.

Zhang Z., Saedi I., Mhanna S., Wu K., Mancarella P., 2022. Modelling of gas network transient flows with multiple hydrogen injections and gas composition tracking. *Int. J. Hydrogen Energy*, 47, 2220–2233. DOI: 10.1016/j.ijhydene.2021. 10.165.