# PREDICTION OF NATURAL GAS CONSUMPTION BY NEURAL NETWORKS

**Dario Sebalj**
*J. J. Strossmayer University of Osijek, Faculty of Economics in Osijek,
31000 Osijek, Croatia
dario.sebalj@efos.hr*

**Josip Mesaric**
*J. J. Strossmayer University of Osijek, Faculty of Economics in Osijek,
31000 Osijek, Croatia
josip.mesaric@efos.hr*

**Ana Pap**
*J. J. Strossmayer University of Osijek, Faculty of Economics in Osijek,
31000 Osijek, Croatia
ana.pap@efos.hr*

***ABSTRACT***

*Due to its environmental benefits, natural gas has become one of the most popular energy sources. Natural gas is the third largest energy source in 2020, after oil and coal, accounting for nearly 25%. The consumption of natural gas has been increasing in recent years, except for last year when consumption decreased by 2.3%. The aim of this paper is to present a neural network model (using Multilayer Perceptron algorithm) that could predict natural gas consumption on an hourly basis. The dataset consists of hourly natural gas consumption data obtained from natural gas supplier and distributor, and meteorological data. There have been many studies in which researchers have attempted to predict gas consumption, and the accuracy of these models is important for decision making, especially for gas nominations (gas orders). The results show that the statistical correlation between the actual and predicted values is very high, but the relative absolute error and root relative squared error are about 25% which cannot be considered satisfactory for this type of prediction. The comparison between the actual and predicted values shows that the model appears to be good at predicting gas consumption in the winter months, but predicts lower values than actual ones for the summer months.*

***Keywords:*** *Algorithms, Energy, Natural gas, Multilayer Perceptron, Machine learning*

## 1. INTRODUCTION

According to BP (2021), during the COVID-19 crisis, global energy demand fell by 4.5% in 2020, which is the largest decline since the end of World War II. The decline was mainly due to falling oil demand as a result of the global lockdowns. However, natural gas showed greater resilience than oil. Its consumption fell by 2.3% but the share of natural gas in primary energy continued to rise, reaching 24.7%. This is not surprising as natural gas is becoming more popular due to its positive impact on the environment. Natural gas consumption is the main topic of this article. Consumers are constantly consuming natural gas as it flows through the pipelines and passes through different supply chain participants - from producer (or importer), transporter and distributor to consumers. The natural gas system operates on balancing principles. The total amount of gas that is injected into the transmission system (from import or production) must also be withdrawn (from the distribution system as consumption). Suppliers, who have signed the agreement with consumers to supply the gas on their behalf, are obliged to make a so-called nomination (or order) of natural gas.

The nomination is the quantity of gas required for the following day (or even broader period of time). If the actual consumption exceeds the nominated quantity, the supplier has to pay a certain penalty. For this reason, accurate prediction models of natural gas consumption are crucial due to financial reasons. The main goal of this paper is to develop a neural networks model that can accurately predict the hourly natural gas consumption. The input variables will consist of meteorological data as well as historical gas consumption data. This article is organised as follows: Section 2 provides an overview of previous research in this field, in Section 3 methodology and data are described, results are presented in Section 4, while the discussion and conclusion are given at the end of the paper.

## 2. LITERATURE REVIEW

Natural gas prediction methods are the subject of research of numerous researchers. Šebalj et al. (2017) conducted the systematic analysis of the methods for predicting natural gas consumption. They analyzed 39 articles from year 2013 to 2017. The results show that the prediction of natural gas (NG) consumption can be done on different areas and in different horizons, using various methods and input variables. Prediction areas refers to country, region, city or individual consumer area. The largest number of authors were predicting the natural gas consumption on the country or city level. Regarding prediction horizon, 2/3 of the papers deal with the predictions on yearly or daily level. In predicting natural gas consumption several methods were used. The most common method is Neural network (or methods based on the same principles, like ANFIS), followed by various Mathematical and statistical models or Time series analysis methods. For modeling, researchers often use past NG consumption and weather data (temperature, wind speed, wind velocity) as input variables. Karabiber and Xydis (2020) also conducted a literature review analysis. Their results show that the most used variables are Heating Degree Days (HDD), day type (weekend or weekday), meteorological conditions (wind speed, wind chill, rain amounts, solar radiation, max and min temperature etc.). Of the more recent works, several can be highlighted. For example, Arik (2019) suggested an Artificial Bee Colony (ABC) algorithm and linear regression for predicting natural gas consumption in Turkey and results show that ABC outperforms the linear regression model. Es (2020) presented a new grey seasonal forecast model to predict monthly natural gas demand in Turkey. The accuracy of the model was measured by the mean absolute error (MAE), the root-mean-square error (RMSE), the mean absolute percentage error (MAPE) and the post-error ratio (C). The proposed model produced a MAPE of 8.67% which considers to be excellent. Anagnostis et al. (2019) used a Long Short-Term Memory (LSTM) algorithm to forecast a day-ahead natural gas demand in Greece. The LSTM algorithm is a time series method with characteristics similar to neural network. The performance measures were mean square error (MSE), MAE, MAPE and coefficient of determination ($R^2$). They tested and evaluated 15 combinations of number of layers and compared the results with different neural networks structures where the proposed model shows its efficiency. Karabiber and Xydis (2020) presented four day-ahead forecasting models (three neural network models and one ARIMA model) to forecast the NG consumption in Denmark. They wanted to develop a more accurate forecasting model than the current forecaster. As the input variables they used historical consumption, heating degree days, wind speed, biogas production, type of day, electricity pricing, gas pricing, solar radiation, minimum and maximum hourly temperature for day, lagged consumption and temperature, and fourier terms (FT) for seasonality. The proposed model has better MAPE, ranging from 34% to 72% reduction in comparison with the current forecasting model. Min et al. (2020) compared two method of natural gas demand forecasting based on the monthly dataset of NG demand and meteorological factors of Beijing. They proposed a novel intelligent prediction model (EMD_BP model) and BP neural network algorithm. The results show that the fitting error of the EMD_BP and BP model were 3.7% and 4.6% respectively.

An integrated approach of classical least square time series approach with neural network models was proposed by Pradhan et al. (2018). That model performed better than the classical least square model and neural network model. Machine learning techniques for predicting natural gas consumption were used in research conducted by Sharma et al. (2021). To capture the intra-day variability in NG demand, they used a block-wise approach where separate model was developed for each block of the day. Four different forecasting models were developed using the block-wise technique (a gradient boosting model - GB, a gradient boosting model using features from Principal Component Analysis – GB-PCA, ANN-CG model using features from sensitivity analysis and ANN-CG model using features from PCA (ANN-CG-PCA)). They also developed three hybrid forecasting models by combining the forecasts from the four individual models. The combined models outperformed the individual models and the MAPE was around 15%.

## 3. METHODOLOGY

This research was conducted in several stages, as shown in Figure 1.

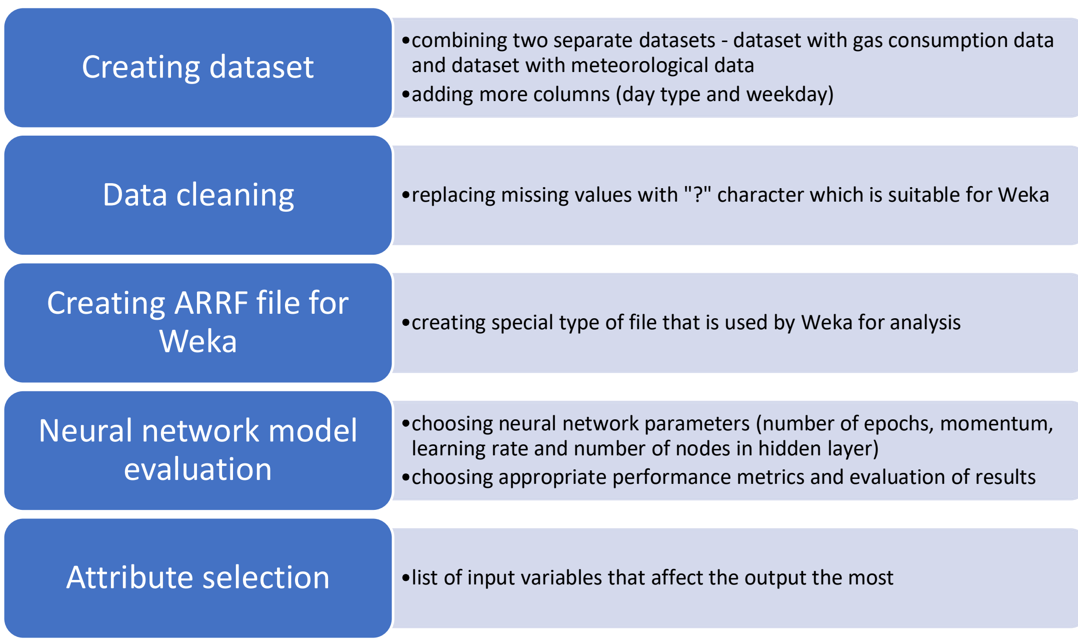


*Figure 1: Research design stages*
*(Source: Authors)*

The neural network model was created using the data mining tool Weka. According to Witten et al. (2011), Weka workbench is a collection of state-of-the-art machine learning algorithms that includes methods for the main data mining problems: regression, classification, clustering, association rules and attribute selection.

### 3.1. Neural networks

According to Müller and Renhardt (1990), "neural network models are algorithms for cognitive tasks, such as learning and optimization, which are in a loose sense based on concepts derived from research into the nature of the brain" and typically are used in problems of classification or forecasting (Gurney, 1997). Simply put, a neural network is a machine learning method that simulates the behavior of the human brain. The neural network contains several types of layers - input layer, hidden layer(s) and output layer, which are connected by weighted connections. Each of these layers contains a different number of nodes (neurons).

The number of nodes in the input layer refers to a number of input variables, while the number of nodes in the output layer represents the class (output) variable(s). In this paper, the performance of the neural network model was tested using the multilayer perceptron (MLP) algorithm. The multilayer perceptron algorithm is a general-purpose feed forward network and one of the most popular and widely used neural network algorithms. It is a modification of the standard linear perceptron, introduced by Rosenblatt in the late 1950s, and it uses three or more layers of nodes (neurons) with nonlinear activation functions. Its advantage over perceptron is that it can distinguish data that is not linearly separable (Zekić-Sušac et al., 2009; Alsmadi et al., 2009). To optimize the error function, it uses the back propagation algorithm which searches for the minimum of the error function in the weight space using the method of gradient descent. The combination of weights which minimizes the error function is considered to be a solution of the learning problem. It also needs to use an activation function. One of the most popular activation functions for backpropagation networks is the sigmoid function (Rojas, 1996). The general architecture of an artificial neural network can be seen in Figure 2.

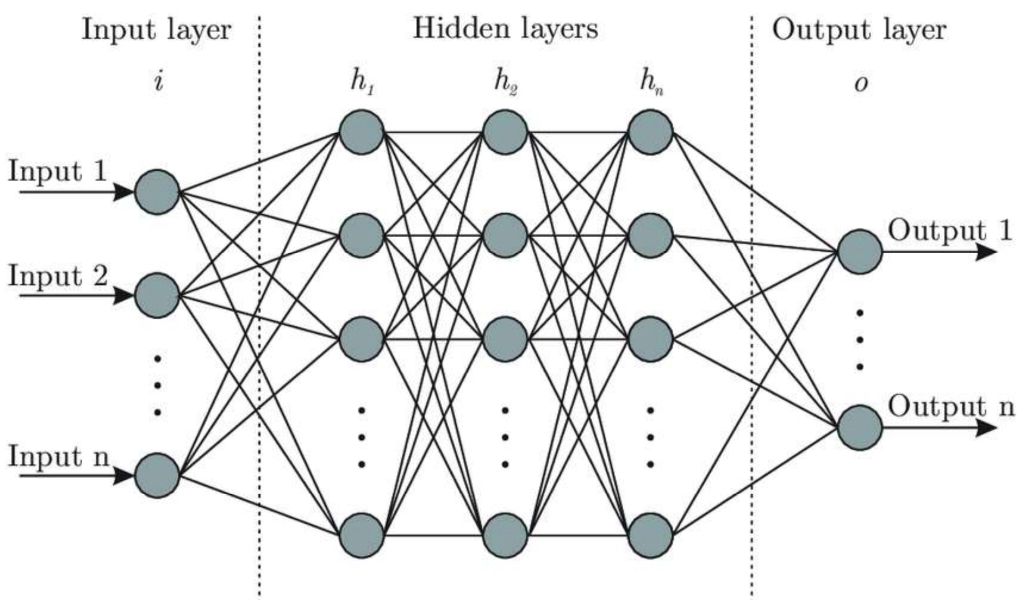


*Figure 2: Artificial neural network architecture*
*(Source: Bre and Gimenez, 2017)*

### 3.2. Data

The dataset for creating a neural network model consists of real historical data of natural gas consumption and meteorological data. The consumption data was received from a natural gas supplier and distributor in Croatia and represents the hourly gas consumption of the household sector, for the period from January 1 to December 31, 2017. The meteorological data is provided by the Croatian Meteorological and Hydrological Service. This dataset contains a total of 8,754 records (one record for each hour in the year 2017). The total number of input variables is 10, as shown in Table 1.

*Table following on the next page*

| Variable no. | Variable description | Descriptive statistics |
|---|---|---|
| 1 | Hour (1-24) | |
| 2 | Day (1-31) | |
| 3 | Month (1-12) | |
| 4 | Day type {HOL="Holiday", WD="Working day", DAH="Day after holiday", WE="Weekend"} | HOL=3.50%, WD=68.55%, DAH=0.55%, WE=27.40% |
| 5 | Weekday {Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday} | Monday-Saturday=14.25%, Sunday=14.46% |
| 6 | Temperature (in °C) | Min: -18.9, Max: 38.3, Mean: 12.14, StdDev: 10.036 |
| 7 | Humidity (in %) | Min: 21, Max: 100, Mean: 73.715, StdDev: 20.856 |
| 8 | Pressure (in hPa) | Min: 979, Max: 1028, Mean: 1006, StdDev: 7.051 |
| 9 | Wind speed (in m/s) | Min: 0.1, Max: 8.5, Mean: 2.098, StdDev: 1.233 |
| 10 | Wind direction (00 to 32) | Min: 2, Max: 32, Mean: 19.32, StdDev: 8.705 |

*Table 1: Input variables*
*(Source: Authors)*

The variables **Hour** (1-24)**, Day** (1-31)**, Month** (1-12) are used to examine the difference in natural gas consumption for each hour, day and month of the year. It was assumed that the **Day type** (holiday – HOL, working day – WD, day after holiday – DAH and weekend – WE) has a certain impact to the output. On holidays or weekends, the consumption should be increased. If a holiday falls, for example on Thursday, there is a possibility that the holiday will be combined with a weekend, so gas consumption will be higher on Friday as well. Therefore, the value "day after holiday" is introduced. The variable **Weekday** (Monday – Sunday) is used to study the impact of weekdays on gas consumption. The last five variables are meteorological variables, such as: **Temperature, Humidity, Pressure, Wind speed, Wind direction**. The output variable refers to the natural gas hourly consumption. Figure 3 shows fluctuations in hourly gas consumption through the year 2017. As expected, the lowest natural gas consumption was recorded in the summer months. There were no sudden fluctuations in natural gas demand during the observed year.

*Figure following on the next page*

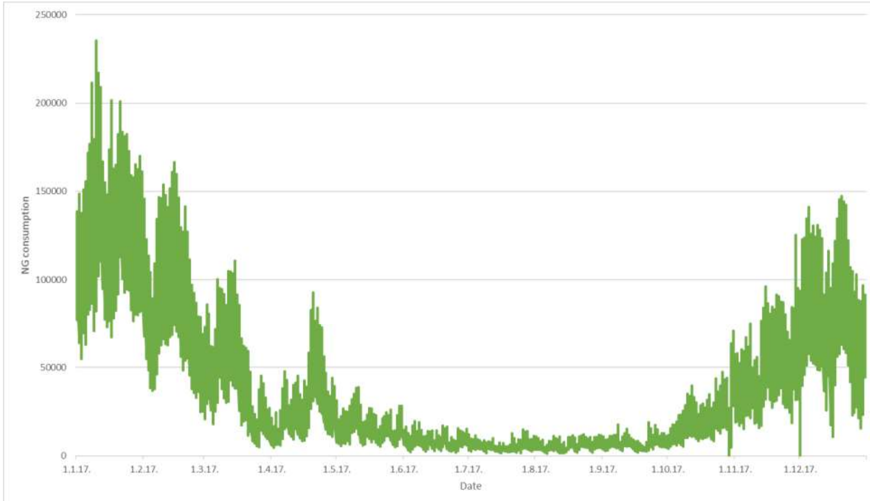


*Figure 3: Hourly natural gas consumption through year 2017*
*(Source: Authors)*

### 3.3. Neural network setup

For the purpose of this paper, the multilayer perceptron (MLP) algorithm was used. The neural network consisted of three layers. The input layer contains 19 nodes – one node for each numeric variable, and one node for each value of nominal variables. The output layer contains one node representing the hourly historical gas consumption, and there are 10 neurons in one hidden layer. According to Murtagh (1991), the number of hidden layers and the number of nodes in each layer can vary for a given problem. Heaton (2017) considers that the most problems require only one hidden layer. As for the number of neurons in the hidden layer, the same author suggests that it should be between the size of the input layer and the size of the output layer. Weka suggests that 10 is the optimal number of nodes in the hidden layer which is consistent with Heaton's research. The neural network architecture is presented in Figure 4.

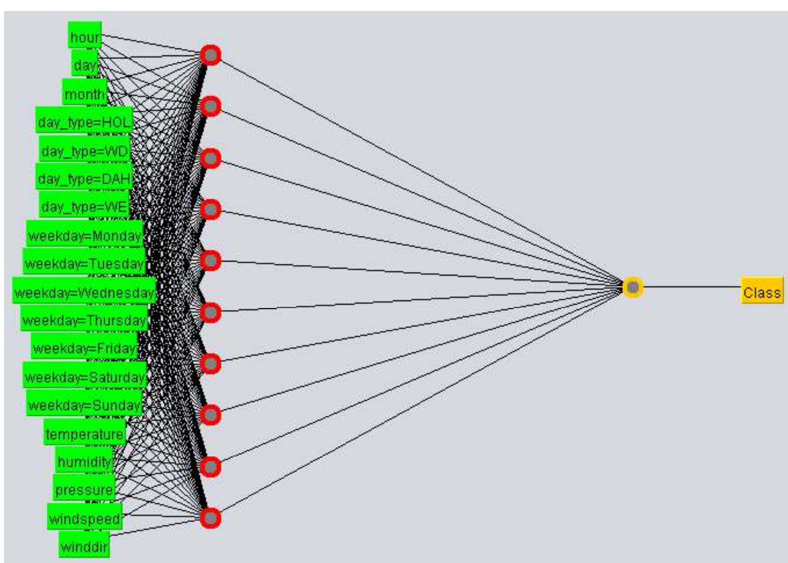


*Figure 4: Neural network architecture*
*(Source: Authors)*

The number of epochs to train through was set to 500, learning rate was 0.3 and momentum 0.2.

## 4. RESULTS

Usually, the neural network modeling dataset is divided into two subsamples – the training set and the test set. Since this dataset contains data for only one year, splitting it into a training set and a test set will make the sample unrepresentative (the test set will contain more data from one season). Therefore, 10-fold cross-validation was used to implement the neural network. This means that the data was randomly divided into 10 parts, one of which was used for testing and the remaining nine parts were used for training. This process was repeated 10 times and then the 10 error estimates were averaged. The accuracy of prediction was measured using five of the most common evaluation metrics – correlation coefficient, mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE) and root relative squared error (RRSE), calculated as follows (Witten et al., 2011):

$$MAE = \frac{|p_1 - a_1| + \cdots + |p_n - a_n|}{n}$$

$$RMSE = \sqrt{\frac{|p_1 - a_1|^2 + \cdots + |p_n - a_n|^2}{n}}$$

$$RAE = \frac{|p_1 - a_1| + \cdots + |p_n - a_n|}{|a_1 - \bar{a}| + \cdots + |a_n - \bar{a}|}$$

$$RRSE = \sqrt{\frac{|p_1 - a_1|^2 + \cdots + |p_n - a_n|^2}{|a_1 - \bar{a}|^2 + \cdots + |a_n - \bar{a}|^2}}$$

where $p_1, p_2, \dots, p_n$ are the predicted values, and $a_1, a_2, \dots, a_n$ are the actual values.

Low values of error metrics signify higher accuracy of the model, but correlation coefficient value close to 1 is preferred, signifying better performance of the model and that the regression curve fits well on the data (Anagnostis et al., 2020). After 10-fold cross-validation, the neural network was built, and the results are shown in Table 2.

| | |
|---|---|
| Correlation coefficient | 0.9648 |
| Mean absolute error (MAE) | 9005.72 |
| Root mean squared error (RMSE) | 12215.48 |
| Relative absolute error (RAE) | 24.53% |
| Root relative squared error (RRSE) | 26.90% |

*Table 2: Neural network results*
*(Source: Authors)*

The results show that the statistical correlation between the actual and predicted values is very high and that these values are perfectly correlated. The RAE and RRSE errors are almost the same (about 25%) and usually this value can be considered as reasonable (see Ma and Liu, 2017). However, for predicting natural gas consumption, the model would be considered successful if the error rate is up to 10%. Figure 4 shows the comparison between the predicted (blue) and actual (orange) values of natural gas consumption on different basis. The first image shows the comparison on an hourly basis. The model appears to be good at predicting gas consumption in the winter months but predicts lower values than actual ones for the summer months. If the data is summarized on a daily basis, it can be seen that the line showing the predicted values follows the line with the actual values quite well. Even better results can be seen by comparing actual and projected values on a monthly basis, as shown in the third image.

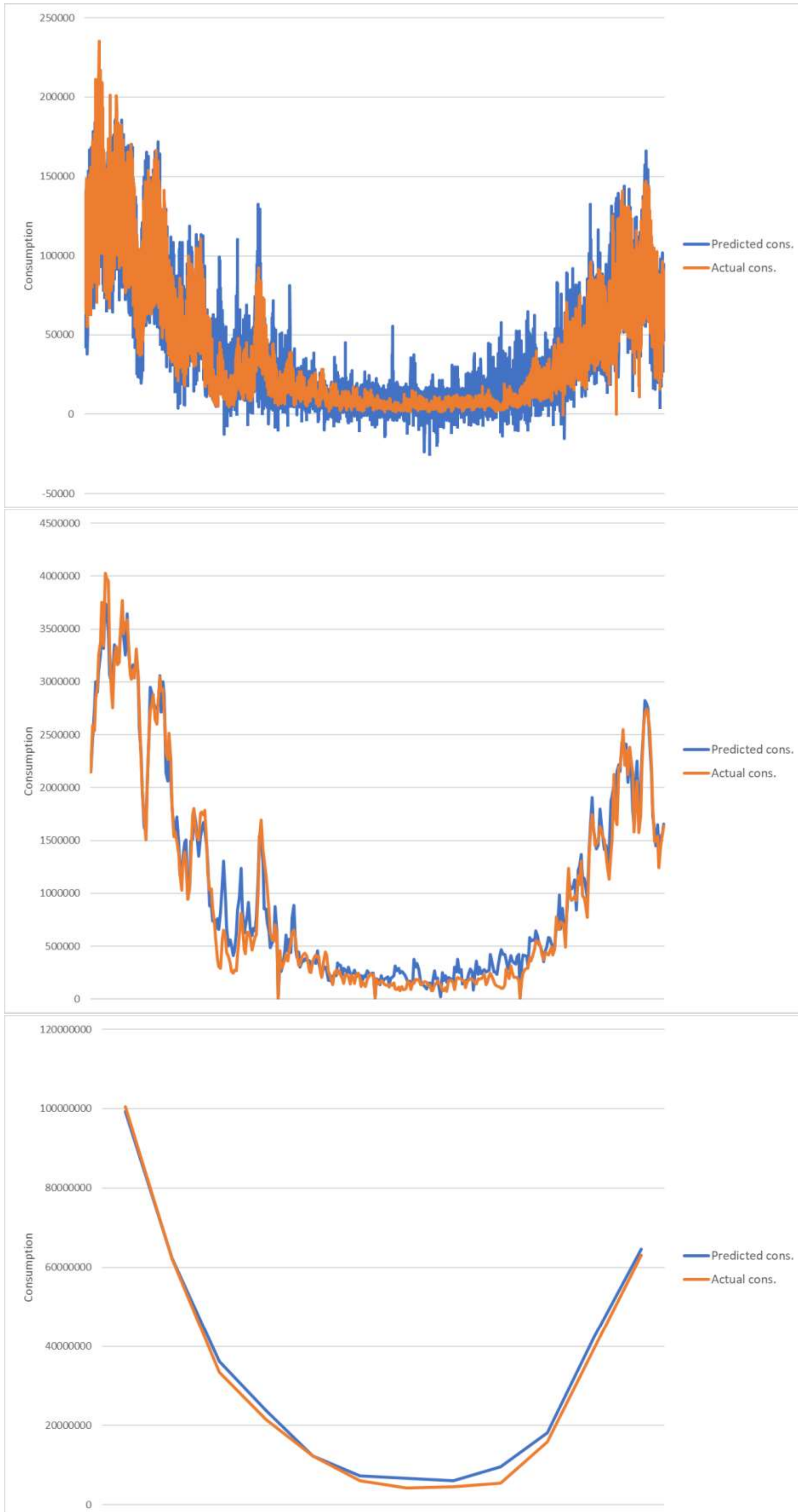


*Figure 5: Comparison of the predicted and actual values of natural gas consumption on the hourly, daily and monthly basis*
*(Source: Authors)*

This confirms the research conducted by Šebalj et al. (2018) who concluded that it is very difficult to predict gas consumption and, thus, submit nominations, since the difference between nominations and actual consumption rise (accuracy drops) when the time period is narrowed. This is also the case here. The predicted consumption for each month is very close to the actual consumption. When the time period is narrowed down (e.g., to an hourly level), the difference is much larger.

### 4.1. Attribute selection

After model evaluation, the attribute selection process was conducted. It includes searching through all possible combinations of input variables in order to find the subset of attributes that works best for prediction. First step is to choose the attribute evaluator in Weka. It determines what method is used to assign a worth to each subset of attributes (Bouckaert et al., 2016). Weka contains several attribute evaluators, and the comparison of the attribute selection results is shown in Table 3. The variables are ranked by the importance and the given values.

| **Attribute selection evaluator** | | | |
|---|---|---|---|
| **CfsSubset** | **Classifier** | **Correlation** | **Relief** |
| month | wind direction | pressure | hour |
| day type | month | humidity | month |
| temperature | day | hour | pressure |
| wind speed | wind speed | windspeed | temperature |
| | day type | day type | day |
| | weekday | weekday | weekday |
| | temperature | day | day type |
| | humidity | wind direction | wind speed |
| | pressure | month | wind direction |
| | hour | temperature | humidity |

*Table 3: Attribute selection results*
*(Source: Authors)*

After the attribute selection process, the authors conducted four additional experiments (one for each attribute selection result) in which only the five highest ranked variables were included. The results were worse than those of the first experiment in which all variables were included in building the neural network.

## 5. DISCUSSION AND CONCLUSION

Forecasting natural gas consumption is of great importance for suppliers who must nominate (order) a certain amount of gas each day, one day in advance (for each hour). The nominated amount of gas should be equal to the amount consumed by the consumers in order for the transmission system to be in balance. Since it is very difficult to accurately predict future consumption based only on past experience and historical data, advanced forecasting methods can play an important role. Previous research has shown that neural networks are one of the machine learning methods that can accurately predict natural gas consumption. Therefore, the authors of this paper chose this method to predict NG consumption. The dataset consisted of data related to the historical natural gas consumption and meteorological data for the year 2017. To build the neural network model, 10 input variables were included - hour, day, month, day type, weekday, temperature, humidity, pressure, wind speed and wind direction. The Multilayer Perceptron (MLP) algorithm was used, and the network consisted of three layers - input layer, hidden layer with 10 nodes and output layer. The prediction accuracy was measured using the five most common metrics - correlation coefficient, MAE, RMSE, RAE and RRSE.

The results showed that the relative prediction error (RAE and RRSE) was around 25% which cannot be considered satisfactory in this case as the error should be as low as possible (below 10%). Also, a graphical comparison of the actual and predicted consumption can show that the neural network model can predict the consumption relatively well in the winter months, while it is slightly worse in the summer months. A limitation of this work can be the fact that the gas consumption data was collected from only one distributor and from one measuring-reduction station (which, however, has the highest gas flow). For future research, the authors plan to include more input variables (such as solar radiation, etc.) in the model. Also, the development of two separate models is planned - one for forecasting consumption in the winter months, and the other for forecasting consumption in the summer months. This will show whether the accuracy of the forecast (at least for the winter months) will increase.

**LITERATURE:**

1. Alsmadi, M. K., Bin Omar, K., Noah, S. A., and Almarashdah, I. (2009). Performance Comparison of Multi-layer Perceptron (Back Propagation, Delta Rule and Perceptron) algorithms in Neural Networks. *IEEE International Advance Computing Conference*, pp. 296-299.
2. Anagnostis, A., Papageorgiou, E., Dafopoulos, V. and Bochtis, D. (2019). Applying Long Short-Term Memory Networks for natural gas demand prediction. *10th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pp. 14-20, Patras (Greece).
3. Arik O. A. (2019). Artificial bee colony algorithm to forecast natural gas consumption of Turkey. *SN Applied Sciences*, 1, 1138.
4. Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A. and Scuse, D. (2016). *WEKA Manual for Version 3-8-0*.
5. BP (2021). *Statistical Review of World Energy* (70th edition).
6. Bre, F. and Gimenez, J. M. (2017). Prediction of wind pressure coefficients on building surfaces using Artificial Neural Networks. *Energy and Buildings*, 158.
7. Es, H. A. (2020). Monthly natural gas demand forecasting by adjusted seasonal grey forecasting model. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*. Volume 43.
8. Gurney, K. (1997). *An introduction to neural networks*. London; NewYork: Routledge.
9. Heaton, J. (2017). *The number of hidden layers*. Retrieved 10.11.2021 from https://www.heatonresearch.com/2017/06/01/hidden-layers.html
10. Karabiber, O. A. and Xydis, G. (2020). A review of the day-ahead natural gas consumption in Denmark: starting point towards forecasting accuracy improvement. *International Journal of Coal Science & Technology*, 8(1), pp. 1-22.
11. Karabiber, O. A. and Xydis, G. (2020). Forecasting day-ahead natural gas demand in Denmark. *Journal of Natural Gas Science and Engineering*, 76, 103193.
12. Ma, X., and Liu, Z. (2017). Application of a novel time-delayed polynomial grey model to predict the natural gas consumption in China. *Journal of Computational and Applied Mathematics*, 324, 17-24.
13. Min, J., Dong, Y., Wu, F., Li, N. and Wang, H. (2020). Comparative Analysis of Two Methods of Natural Gas Demand Forecasting. *2020 Asia Conference on Geological Research and Environmental Technology*, 632, 032033.
14. Müller, B. and Reinhardt, J. (1990). *Neural Networks: An Introduction*. Berlin: Springer-Verlag.
15. Murtagh, F. (1991). Multilayer perceptrons for classification and regression. *Neurocomputing*, 2(5-6), 183–197.

16. Pradhan, P. K., Dhal, S., and Kamila, N. K. (2018). Artificial Neural Network Conventional Fusion Forecasting Model for Natural Gas Consumption. *7th International Conference on Computing, Communications and Informatics (ICACCI)*. Bangalore (India).
17. Rojas, R. (1996). The Backpropagation Algorithm. In: *Neural Networks: A Systematic Introduction*. Berlin: Springer.
18. Šebalj, D., Mesarić, J. and Dujak, D. (2018). Analysis of natural gas trading system in Croatia: A preliminary research. In: Soliman, K. S. (Ed.), *Proceedings of the 32nd International Business Information Management Association Conference (IBIMA),* pp. 1524-1535, Seville (Spain).
19. Šebalj, D., Mesarić, J. and Dujak, D. (2019). Analysis of Methods and Techniques for Prediction of Natural Gas Consumption: A Literature Review. *Journal of Information and Organizational Sciences*, 43(1), pp. 99-117.
20. Sharma, V., Cali, Ü., Sardana, B., Kuzlu, M., Banga, D., Pipattanasomporn, M. (2021). Data-driven short-term natural gas demand forecasting with machine learning techniques, *Journal of Petroleum Science and Engineering*, 206, 108979.
21. Witten, I. H., Frank, E. and Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington: Morgan Kaufmann Publishers.