TEKA. COMMISSION OF MOTORIZATION AND ENERGETICS IN AGRICULTURE – 2014, Vol. 14, No. 4, 93–98

Short-Term Forecasting of Natural Gas Demand by Rural Consumers Using Regression Models

Krzysztof Nęcka, Małgorzata Trojanowska

Department of Power Engineering and Agricultural Processes Automation, Agricultural University of Cracow
Balicka Str. 116B, 30-149 Kraków, Poland
e-mail: krzysztof.necka@ur.krakow.pl; malgorzata.trojanowska@ur.krakow.pl

Received December 02.2014; accepted December 17.2014

Summary. The study used various kinds of multiple regression models (standard and non-parametric) to predict the daily demand for natural gas by rural consumers. The construction of forecasting models was performed in the workspace of *STATISTICA Data Miner*, using the data mining technique. The analysis of forecast errors showed that all models applied in the study i.e. standard regression models, neural networks, regression trees, models based on the *MARS*, *SVM* and *k-Nearest Neighbours* methods generate good-quality forecasts, while artificial neural networks turned out to be the most effective method.
Key words: short-term forecasts, multiple regression, data mining.

INTRODUCTION

For natural gas companies, like for other enterprises in the energy sector, the forecasting of demand for energy carriers is one of the basic activities required by the management of these businesses, and the quality of forecasts directly affects the energy security of consumers as well as the profits of these enterprises. In natural gas companies, forecasts of demand for natural gas are prepared for various time horizons.

The short-term forecast of demand for gas includes hourly and daily forecasts for a maximum of seven days in advance. These activities facilitate the operational management of an enterprise, and, in particular, serve to plan daily purchases of gas.

Medium-term forecasts are daily and monthly with a time horizon of between one week and a year. These are used for planning purchases of gas in longer periods, and for planning gas transmission operations.

Long-term forecasts are annual predictions with the perspective of a forecast of between 5 to 10 years. These are used by gas companies in the long-term planning of gas purchases, and for planning the development of gas infrastructure.

Under the progressing liberalization of the natural gas market, short-term forecasts are becoming increasingly important. In recent years, an increasing number of studies have emerged, aimed at improving the quality of this type of forecast. The short-term prediction of demand for natural gas uses both traditional methods and methods based on artificial intelligence, as well as combined methods [14, 15, 17].

Traditional forecasting methods can be divided into those which use linear [6, 9, 11, 12, 13, 16] and non-linear models (logistic model, exponential model, Gompertz model, and the like) [4, 12] or models of auto-regression and means (*ARIMAX*, *SARIMAX*, recursive autoregression models *RARX*) [3, 9, 10, 11, 18]. In recent years, there has been an increase in the number of published studies which use neural networks [2, 5, 7, 8, 10, 16, 18, 20], recursive neural networks [11], fuzzy neural networks [1], and neural networks optimized by using genetic algorithm [19] to forecast the daily consumption of natural gas.

Among Polish publications in this field, most studies deal with long-term forecasts of the demand for gas. Publications dealing with short-term forecasts are few [6, 7].

The objective of this study was to use various types of regression models to predict the daily demand for natural gas by rural consumers, and to carry out comparisons in this area.

MATERIALS AND METHODS

The consumer demand for natural gas depends on a number of factors. For this reason, multiple regression models are very suitable for such forecasts. The regression function can be given not only in the form of a mathematical formula, but also as an algorithm. In the study, the construction of forecasting models was performed using standard linear regression models, but also using algorithms in the form of neural networks, regression trees, as well as applying *MARS*, *SVM* and *k-Nearest Neighbours* methods.

The construction of the models was carried out in the workspace of *STATISTICA Data Miner*, using data mining methodologies. Developing models was preceded by a correlation analysis aimed at finding the factors of greatest effect on the daily usage of natural gas by rural consumers.

The calculations and analyses were carried out on the basis of hourly measurements of gas usage by consumers from rural areas of southwestern Poland, who are supplied by a low-pressure network via a selected gas pressure reduction station. Data from the 2008–2011 period was used. Data from 2008-2010 was used as a training set, whereas data from 2011- was used as a testing set. The quality of the models was evaluated on the basis of values for ex-post forecast errors, using the following formula:

$$\omega_1 = \frac{\left|G_t - G_t^p\right|}{G_t} \cdot 100, \quad (1)$$

$$\omega_2 = \frac{1}{n} \cdot \sum_{t=1}^{n} \frac{\left|G_t - G_t^p\right|}{G_t} \cdot 100, \quad (2)$$

$$\omega_3 = \frac{\sum_{t=1}^{n} \left|G_t - G_t^p\right|}{G_c} \cdot 100, \quad (3)$$

where:
G_t – actual daily consumption of gas,
G_t^p – forecast of daily gas consumption,
G_c – actual consumption of gas during *n* days of observation.

RESULTS

Among other forecasts, natural gas enterprises in Poland are obliged to make forecasts of daily gas consumption one day in advance. A day of consumption is defined as 24 hrs from 6AM to 6AM the next day.

The course of hourly and daily natural gas consumption by rural consumers within the study period is presented in Figs 1 and 2.

Meteorological and time factors were considered among the factors affecting gas consumption. The meteorological factors taken into account included: temperature (daily – mean, maximum, minimum), humidity (daily – mean, maximum, minimum), wind velocity (daily – mean), wind direction (daily – mean), pressure (daily – mean), with these values taken with delay to the day of analyzed gas consumption.

Time factors include the correlation between the values for daily gas consumption with the values of the variable delayed with the time and the day of the week. A total of 144 variables were considered in the correlation analysis.

It was found that the greatest effect on the value of daily gas consumption by rural consumers was exerted by this value delayed by one day, the mean temperature of the previous day, and the day of the week. In further statistical analyses, these values were adopted as the variables explaining the daily demand for natural gas.

In order to develop prediction models which permit finding the daily demand for natural gas, a project was created in the graphic environment of *Statistica 10 Data Miner* (Fig. 3). Among other elements, the nodes were placed there to permit drawing forecasts based on: *Neural Networks (NN*), *Classification and Regression Trees (CART*), *Multivariate Adaptive Regression Splines (MARS)*, *Support Vector Machine (SVM*), *k Nearest Neighbours (k-NN*), as well as standard *Multiple Regression (MR*).

Standard multiple regression requires meeting many assumptions pertaining to sample size, explained and explaining variables, residuals of the model, although in data mining these assumptions are tested less restrictively than in traditional statistics. The other procedures used in this study are non-parametric procedures which do not require meeting any initial assumptions.

Among the latter procedures, the oldest and best described is the method of artificial neural networks. It has been dynamically developing for several decades and it is now regarded as a very refined modelling technique, capable of mapping very complex relationships.

Classification and Regression Trees are other advanced prediction tools of *Data Mining*. The tree is a graphical model created as a result of the recurrent division of a set of observations into *n* of disjoint subsets. In most general terms, the objective of the analysis with the applied *CART* algorithm is to find the set of logical conditions for division, of the *if-so,* leading to the unambiguous classification of objects.

The *MARS* method can be regarded as an extension of regression trees and multiple regression. The *Multivariate Adaptive Regression Splines* algorithm utilises the method of the recurrent division of trait space in the nodes of basic functions to construct the regression model in the form of spline curves. The relationship between variables is modelled with the use of a set of basic coefficients and functions generated on the basis of data. The *SVM* method is the next one implemented in the *Statistica* software, which serves to solve regression and classification problems. It consists of constructing non-linear decision borders, separating areas in the space of predictors, corresponding to different values of the dependent variable.

K Nearest Neighbours is a method where instead of matching the model, similar objects are sought. The basis of the method is an intuitive conviction that similar objects will fall into the same class. The predictions of the *k-NN* method are determined on the basis of *k* objects from the training set which are most similar to the object for which the value of the dependent variable is determined.

Mean errors were calculated for the training set (Table 1) and testing set (Table 2) for forecasts found with the use of particular methods. The most accurate method turned out to be the neural networks, whereas the standard regression was the least accurate.

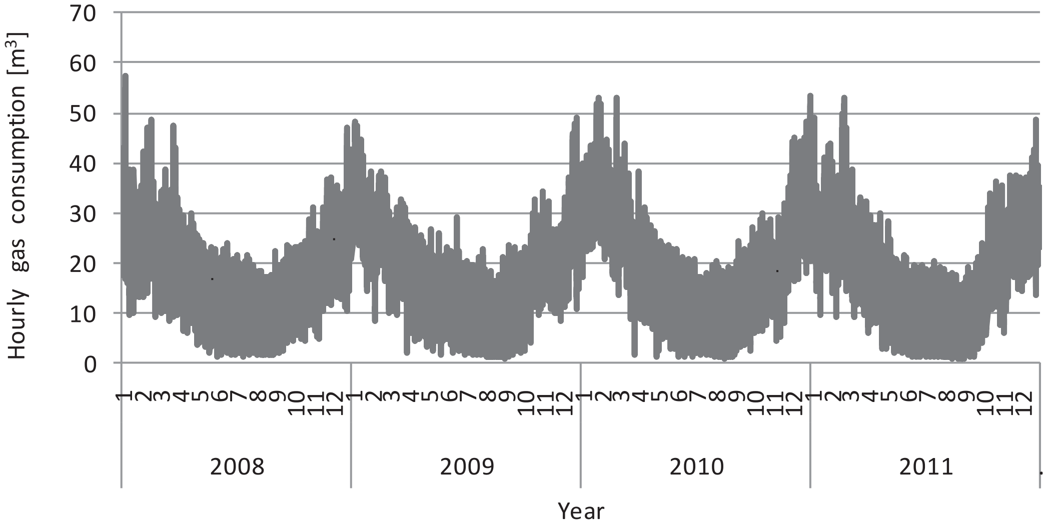

Fig. 1. Hourly natural gas consumption

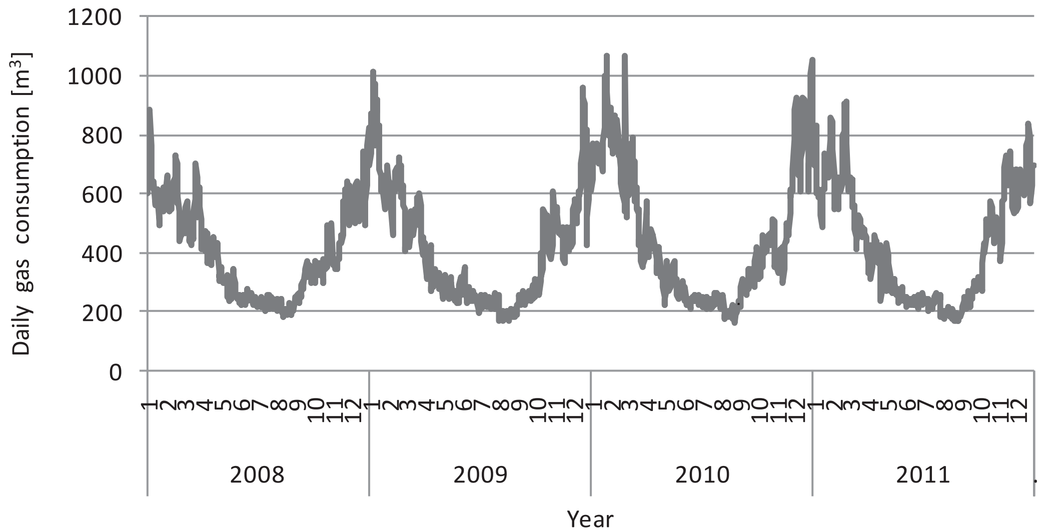

Fig. 2. Daily natural gas consumption

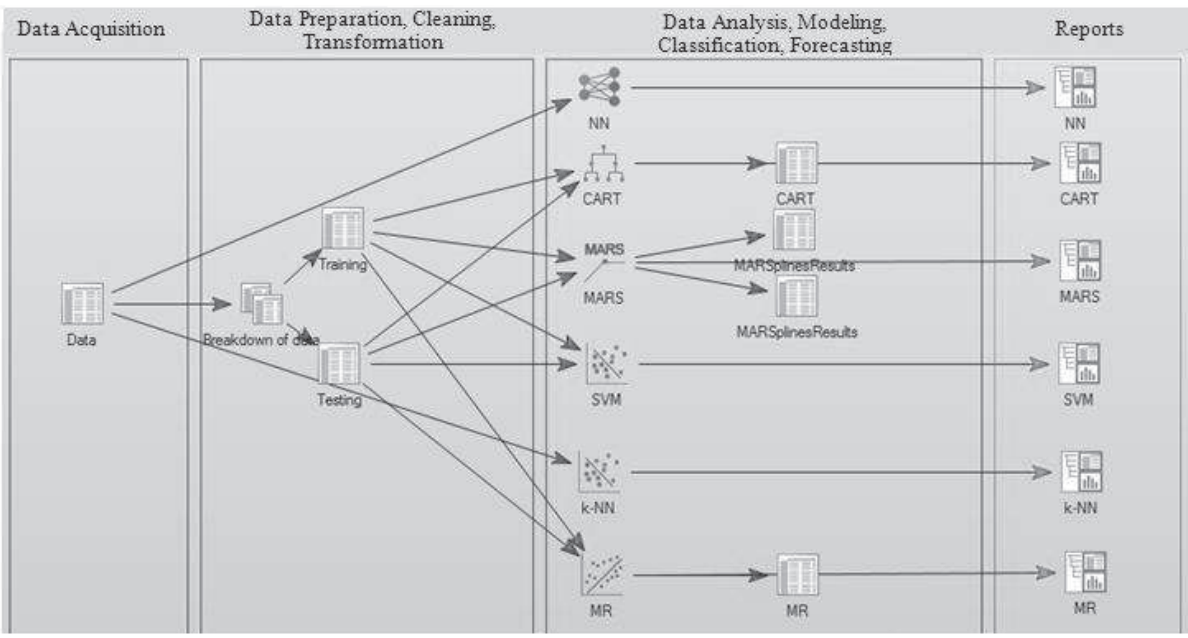

Fig. 3. View of models constructed in the workspace of *Data Miner*

Table 1. Forecast errors found for training set

Model / Error	*NN*	*CART*	*MARS*	*SVM*	*k–NN*	*MR*
ω_2 [%]	5.44	6.88	6.39	6.81	5.54	7.39
ω_3 [%]	5.52	6.62	6.27	6.44	5.26	6.98

Table 2. Forecast errors found for testing set

Model / Error	*NN*	*CART*	*MARS*	*SVM*	*k–NN*	*MR*
ω_2 [%]	5.39	6.05	6.25	6.25	6.38	6.78
ω_3 [%]	5.39	5.64	5.97	5.96	5.97	6.30

For the purpose of analyzing daily gas consumption forecasts, the distribution functions of relative errors w_1 were determined and their courses for the testing set are presented in Fig.4. As shown in the figure, irrespective of the method used, the proportion of smallest errors is similar (11-14%). Greater differences between the w_1 frequencies of occurrence for the studied models were seen when this proportion increased. In the *k–NN* model, the proportion of errors amounting to less than 10% was 70% of observations whereas for the *NN* model it was 83%. The advantage of the *NN* model over the other models is also visible in the proportion of highest observed errors. They did not exceed 20% in as many as 99% of the observations.

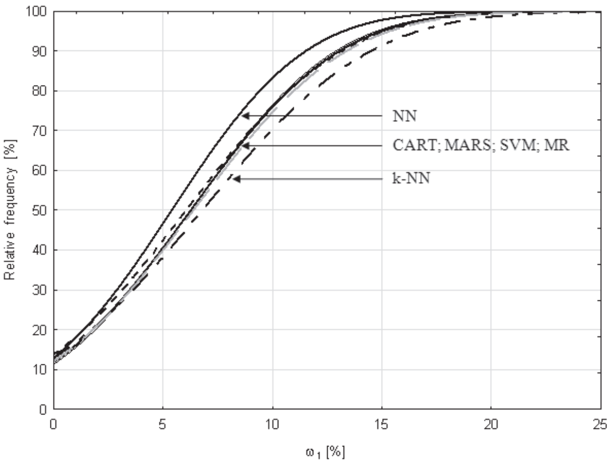

Fig. 4. Empirical distribution functions of w_1 forecast errors

CONCLUSIONS

The greatest effect on the value of daily gas consumption by rural consumers was exerted by this value delayed by one day, mean temperature of the previous day, and the day of the week.

In view of the requirements of the quality of short-term forecasts posed by gas enterprises, the forecasts determined on the basis of multiple regression models, both standard and non-parametric i.e. *NN, CART, MARS, SVM,* and *k–NN* can be considered admissible ($5.44\% \leq w_2 \leq 7.39\%$) and accurate ($5.39\% \leq w_2 \leq 6.78\%$).

The analysis of forecast errors proved that the most effective methods are the neural networks, whereas the greatest errors in forecasts are generated in standard regression.

REFERENCES

1. **Azadeh A., Asadzadeh S.M., Ghanbari A. 2010:** An adaptive network-based fuzzy inference system for short-term natural gas demand estimation: Uncertain and complex environments. Energy Policy 38, 1529-1536.
2. **Azari A.; Shariaty-Niassar M., Alborzi M. 2012:** Short-term and medium-term gas demand load forecasting by neural networks. Iranian Journal of Chemistry and Chemical Engineering 31(4), 77-84.
3. **Brabec M., Konar O., Pelikan E., Maly M. 2008:** A nonlinear mixed effects model for the prediction of natural gas consumption by individual customers. International Journal of Forecasting 24, 659-678.
4. **Brabec M., Konar O., Maly M., Pelikan E., Vondracek J. 2008:** A statistical model for natural gas standardized load profiles. J Roy Statist. Soc. Series C: Applied Statistics 58(1), 123-139.
5. **Chen Q., She Y., Xu X. 2013:** Combination model for short-term load forecasting. The Open Automation and Control Systems Journal 5, 124-132.
6. **Dittman P., Szabela-Pasierbińska E. 2007:** Short-term sales forecasts in management of natural gas distributing works. Management 11(1), 147-154.
7. **Kelner J.M. 2003:** Prognozowanie krótkoterminowe poborów gazu z sieci przesyłowych metodą sztucznych sieci neuronowych. Gaz, Woda i Technika Sanitarna 6, 196-204.
8. **Kizilaslan R., Karlik B.2008:** Comparison neural networks models for short term forecasting of natural gas consumption in Istanbul. In: Applications of digital information and web technologies, ICADIWT 2008, 4-6 August 2008, 448-453.
9. **Potocnik P., Govekar E. 2010:** Practical results of forecasting for the natural gas market. http://www.intechopen.com/books/natural-gas/practical-results-of-forecasting-for-the-natural-gas market.
10. **Potocnik P., Govekar E., Grabec I. 2007:** Short-term natural gas consumption forecasting. In: Proceedings of the 16th IASTED International Conference on Applied Simulation and Modelling – ASM 2007, 353-357.
11. **Potocnik P., Soldo B., Simunovic G., Saric T., Jeromen A., Govekar E. 2014:** Comparison of static and adaptive models for short-term residential natural gas forecasting in Croatia. Applied Energy 129, 94-103.
12. **Sabo K., Scitovski R., Vazler I., Zekic-Susac M. 2011:** Mathematical models of natural gas consumption. Energy Conversion and Management 52, 1721-1727.
13. **Simunek M., Pelikan E. 2008:** Temperatures data preprocessing for short-term gas consumption forecast. In: IEEE International Symposium on Industrial Electronics; 1192-1196.
14. **Smith P., Husejn S. 1996:** Forecasting short term regional gas demand using an expert system. Expert Systems with Applications 10(2), 265-273.

15. **Soldo B. 2012:** Forecasting natural gas consumption. Applied Energy 92, 26-37.
16. **Soldo B., Potocnik P., Simunovic G., Saric T., Govekar E. 2014:** Improving the residential natural gas consumption forecasting models by using solar radiation. Energy and Buildings 69, 498-506.
17. **Suganthi L., Anand A. Samuelb A.A. 2012:** Energy models for demand forecasting-A review. Renewable and Sustainable Energy Reviews 16, 1223-1240.
18. **Taspınar F., Celebi N., Tutkun N. 2013:** Forecasting of daily natural gas consumption on regional basis in Turkey using various computational methods. Energy and Buildings 56, 23- 31.
19. **Yu F., Xu X. 2014:** A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network. Applied Energy 134, 102-113.
20. **Zhou H., Su G., Li G. 2011:** Forecasting daily gas load with OIHF-Elman Neural Network. Procedia Computer Science 5, 754-758.

PROGNOZOWANIE KRÓTKOOKRESOWE ZAPOTRZEBOWANIA ODBIORCÓW WIEJSKICH NA GAZ ZIEMNY Z WYKORZYSTANIEM MODELI REGRESYJNYCH

Streszczenie. W pracy wykorzystano różnego rodzaju modele regresji wielorakiej (standardowe i nieparametryczne) do predykcji dobowego zapotrzebowania odbiorców wiejskich na gaz ziemny. Budowę modeli predykcyjnych przeprowadzono w przestrzeni roboczej *STATISTICA Data Miner*, przy użyciu metod *data mining*. Analiza błędów prognoz wykazała, że wszystkie zastosowane w pracy modele tj. standardowej regresji, sieci neuronowych, drzew regresyjnych, oparte na metodzie *MARS*, wektorów nośnych oraz k–najbliższych sąsiadów generują prognozy dobrej jakości, przy czym metodą najbardziej efektywną okazały się sztuczne sieci neuronowe.

Słowa kluczowe: prognozy krótkookresowe, regresja wieloraka, data mining.