International Journal of Scientific Research and Management (IJSRM)
||Volume||08||Issue||08||Pages||EC-2020-423-438||2020||
Website: www.ijsrm.in ISSN (e): 2321-3418
DOI: 10.18535/ijsrm/v8i08.ec03

Forecasting Natural Gas Demand in a Region with Seven Models and Evaluating Their Accuracy Using As Criteria Five Types of Errors/Residuals

[1]E. Stathakis, [2]E. Stambologlou[*]
[1]Diploma of Mechanical Engineer P&A, M.Sc. in Advance Information Systems, a PhD Candidate in Economy of Energy
[2]Diploma of Surveyor Engineer (NTUA 1984), M.Sc. in System Engineering and management, a PhD candidate in Rural Infrastructures

Abstract
The modeling of Natural Gas (NG) demand differs significantly from the demand for electricity in terms of the determinants that affect it, as all fields of economic activities in a modern economy are directly related to electricity but not to NG. But NG is the second energy type after electricity used in all countries in percentages greater than 10% in average terms. NG is going to be installed in the Region of East Macedonia-Thrace (REMTH) the next years. So, we consider it is worth to predict the NG demand in REMTH using eight deterministic forecasting models. In order to do it we used a dataset of 20 years concerning two Greek regions to which the NG is used that period and through them we built the eight forecasting models aiming to find the NG demand in the REMTH. In order to evaluate the reliability and accuracy of them we used four types of statistical errors, Mean Error (ME), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Scale Error (MASE). These are the most widely used measures of evaluating the accuracy of deterministic predictive models, due to their advantages of scale-independency and interpretability. When each of them is used alone has the significant disad-vantage to produce infinite or undefined values for zero or close-to-zero actual values. In order to address this disadvantage, we propose a way to use the same time all of them measuring the accuracy of a model used to forecast the demand of Natural Gas in the Greek region EMTH. The innovation of this paper is that for NG demand forecasting were used seven different models and they are evaluated regarding their reliability /accuracy using five types of residuals or statistical errors.

Keywords: Natural Demand, forecasting models, evaluation of models, time series.

1. Introduction to timeseries analysis and forecasting

Mathematically a timeseries regarding the NG demand is defined by the function of $\mathbf{Y_{ng} = f_t(T_{ed}, I_c, T_{em}, P_{ng}, P_{el}, E)}$ where, $\mathbf{Y_{ng}}$ is dependent variable, $\mathbf{I_c}$ is income per capital, **P** the price of electricity and NG at times $t_1, t_2, \ldots t_n$ and E is elasticities. Therefore, Y is a function of t, and this is denoted as Y = f (t). The analysis of timeseries deals exclusively with the investigation of the overtime behavior of the values of a variable, the observations of which come from a timeseries. The forecast of future prices of the variable according to the timeseries analysis can come from the following categories of forecasting methods, Smoothing Methods, Timeseries Decomposition, ARIMA Analysis. The criteria for evaluating forecasting methods based on timeseries are used to select the appropriate method. These criteria are based on the values of the deviations of the predicted values from the corresponding real values of the time series. For a variable Y, the deviation of the predicted value of Yt from the corresponding real value of Yt for the period t, where t = 1,2,3, .., n, is called forecast Error, is denoted by **et** and is defined as: $\mathbf{e_t = Y_t - Y'_t}$.

To determine the reliability of a particular prediction of a statistical method, we need to study the overtime behavior of prediction error values. This is done by applying various criteria, according to which we evaluate the predicted method used. Each of these criteria is defined by a specific functional relationship of

prediction errors and can be used not only to evaluate a prediction method but also to select the best one between two or more alternative prediction methods. These criteria are:

Mean Squared Error-MSE: is defined as the sum of the squares of the errors divided by the number of time periods n, in which predictions were made $\boldsymbol{MSE} = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{t=1}}^{\boldsymbol{n}} (\boldsymbol{y_t} - \bar{\boldsymbol{y}}_{\boldsymbol{t}})^{\mathbf{2}} = \sum_{\boldsymbol{t=1}}^{\boldsymbol{n}} \boldsymbol{e_t^2}$

Root Mean Squared Error-RMSE: is defined as the square root of MSE, so $\boldsymbol{RMSE} = \sqrt{\boldsymbol{MSE}}$

Mean Absolute Percentage Error-MAPE: is defined as the sum of the absolute values of the prediction errors to the corresponding real time values of the time series divided by the number of time periods n, in which predictions were made. $\boldsymbol{MAPE} = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{t=1}}^{\boldsymbol{n}} \left|\boldsymbol{e_t^2}\right| / \boldsymbol{y_t}$

Mean Percentage Error-MPE: we use it when we are interested in determine if the prediction method is biased, that is, if the predicted values are systematically greater or less than corresponding real. $\boldsymbol{MPE} = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{t=1}}^{\boldsymbol{n}} \left|\boldsymbol{e_t^1}\right| / \boldsymbol{y_t}$

Mean Absolute Deviation-MAD: expresses the average value of the absolute deviations of the predicted values of the time series from the corresponding real ones and is based on the assumption that its severity is linearly related to the magnitude of the error. $\boldsymbol{MAD} = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\boldsymbol{t=1}}^{\boldsymbol{n}} \left|\boldsymbol{e_t^1}\right|$

The timeseries analysis uses tree types for forecasting, the smoothing, the decomposition and the AutoRegressive Average .

1st : Smoothing methods: are techniques that determine the future values of a variable based on how they are applied. Their creation predictions come from the smoothing of the time evolution of the values of the variable, in order to better recognize its mode of behavior. Some of the smoothing methods can also be applied to a small number of observations of the variable, 30 ≤ like our case. The basic smoothing methods are, Simple Moving Average (MA), Simple Exponential Smoothing (SES), Double Moving Average (DMA), Brown, Holt and Winters. If the timeseries shows a trend pattern then we use the double exposure smoothing method, the Brown method, or the Holt method, while if the timeseries shows seasonality then we use the Winters method. If timeseries is stationary the appropriate method of predicting future prices is the method of move averages - MA.

The smoothing models follow the next formulas in brief:

- **Simple Moving Average (MA):** The predictions of a Yt timeseries, for t = 1,2, ...,n, are created by the MA method as follows, $\widehat{\boldsymbol{Y}}_{\boldsymbol{t+1}} = \widehat{\boldsymbol{Y}}_{\boldsymbol{t}} + \boldsymbol{Y_t}/\boldsymbol{m} + \boldsymbol{Y_{t-m}}/\boldsymbol{m}$, where $\mathbf{Y}_{\mathbf{t+1}}$ is the forecast for the period **(t + 1)** and **m** is the number of periods used to calculate the average value.
- **Simple Exponential Smoothing (SES):** The $\acute{\boldsymbol{Y}}_{\boldsymbol{t+1}} = \boldsymbol{aY_t} + (\mathbf{1} - \boldsymbol{a})\acute{\boldsymbol{Y}}_{\boldsymbol{t}}$ the mathematical expression of the SES method and is defined for t = 2, 3, …, n with initial condition $\acute{\boldsymbol{Y}}_{\mathbf{2}} = \boldsymbol{Y_1}$. Applying the SES method to the time series observations for values of α from 0 to 1 we select that value of α that minimizes the value of the MSE criterion.
- **Double Moving Average (DMA):** To configure predictions with this method, a second moving average is calculated from the simple moving average, while in then the linear trend of the timeseries observations having the form $\acute{\boldsymbol{Y}}_{\boldsymbol{t}} + \boldsymbol{h} = \boldsymbol{a_t} + \boldsymbol{hb_t}$, is also taken into account. This method can be used for $\boldsymbol{h} > \mathbf{1}$ to make predictions for more than one future periods, while for $\boldsymbol{h} = \mathbf{1}$ it gives the forecast for the next period. Of course, its use presupposes the existence of a larger number of observations, especially when the value of m is relatively large.
- **Brown or Double Exponential Smoothing (B-DES):** The application of the method is based on the following procedure: **a.** The initial observations of the timeseries are normalized **b.** The smoothed values $\boldsymbol{A_t}$ of the timeseries are complete smoothed **c.** The difference $\boldsymbol{a_t}$ is calculated as: $\boldsymbol{a_t} = \mathbf{2}\boldsymbol{A_t} - \boldsymbol{A'_t}$ **d.** The adjustment factor for the $\boldsymbol{b_t}$. **e.** trend is calculated. The prediction $\boldsymbol{Y_{t+1}}$ for the future **h** period is calculated by calculated $\boldsymbol{Y_{t+1}} = \boldsymbol{a_t} + \boldsymbol{hb_t}$

- **Holt or Exponential Smoothing for Trend Adaptation (H-ESTA):** It has two parameters of smoothing, the a for smoothing the values of the timeseries and the b for smoothing the trend. The Holt is based on the process of smoothing the values of the **At** timeseries, smoothing the trend **Tt** and predicting $\boldsymbol{Y_{t+h}}$ for the future **h** period and is defined as: $\boldsymbol{Y_{t+h} = A_t + hT_t}$ **,** where h = 1,2, 3, ... The Holt method is more applicable in practice, as it has better results than the Brown method.

2nd : Decomposition methods: identify the components of the available timeseries data in order to understand how they behave. The purpose of decomposition methods is to isolate these components as accurately as possible. The components are, trend **(T)**, cyclicity **(C)**, seasonality **(S)** and randomness **(R)**. The trend **(Tt)** is defined as a long-term change of the average price level of the timeseries. In order to determine the trend, there must be a sufficient number of observations and at the same time the appropriate length of the period must be estimated.

The cyclic **(Ct)** of a timeseries represents a wave change due to special exogenous conditions and occurs periodically. The periods are not constant and are usually greater than one year. The timeseries of more economical sizes such as GDP and NG prices show cyclic. The seasonality **(St)** is a periodic variation that has a constant length and can be easily recognized by observing its graphic representation. Because the changes it causes in timeseries data are constant over time, it is easy to deal with its effect by finding seasonality indicators for the corresponding time periods. A typical example of a seasonal timeseries is NG demand in winter months. The randomness **(Rt)** is the timeseries' property according to which the observations vary around a constant mean, have a constant variance, and are probabilistically independent. Observations do not trend upwards or downwards, the variance does not increase overtime, the observations do not tend to be bigger in some periods than in other periods.

In the additive decomposition model the real values of the timeseries for each period are considered as the sum of the four components and are created in the following way: **Yt = Tt + St + Ct + Rt**.

In the Multiplication Decomposition Model (MDM) the real values of the timeseries are determined by the product of the four components, ie as follows: **Yt = Tt* St* Ct* Rt**.

Of the above two models, the additive model is used less frequently in practice because it is difficult to analyze it for computational reasons, and we will use multiple decomposition.

3rd : Autoregression Move Average models (ARIMA): They are stochastic mathematical models used to describe the overtime evolution of a physical quantity. They include the random error or prediction error, the size values that appear at previous times, and other stochastic factors. **ARIMA** models have been extensively studied by **G**. Box and G Jenkins to the extent that their names are almost synonymous with **ARIMA** procedures and their applications in the timeseries analysis and predictions. The Box and Jenkins proposed a family of algebraic prediction models from which we can select the most suitable for our timeseries. The **ARIMA** models based on multiple linear regression are described by the equation: $\boldsymbol{Y_t = b_o + b_1X_1 + b_2X_2 \ldots\ldots b_pX_p + e}$, where **Y** is the dependent variable and $\boldsymbol{X_1, X_2, \ldots, X_p}$ are the independent variables and are values of the timeseries in previous periods. Also, the **ARIMA** models express $\boldsymbol{Y_t}$ as a linear function of the p's of real past values of $\boldsymbol{Y_t}$ and the independent variables $\boldsymbol{X_1, X_2, \ldots, X_p}$**,** are the values of the errors of previous periods as the difference of predicted values from the actual prices. These models can only be used for stationary timeseries, not for nonstationary ones. The "**white oise"** model is expressed as **ARIMA (0,0,0)** while the **"random-walk"** model is expressed as **ARIMA (0,1,0)**. The main things related to ARIMA are:

Το μοντέλο white noise εκφράζεται ως ARIMA (0,0,0) ενώ το μοντέλο random walk ως ARIMA(0,1,0). The main remarks related to ARIMA are:

- The **rk** Auto-correlation coefficient indicates the correlation of the timeseries with itself for observations that are distant from each other and is defined as: $\boldsymbol{r_k = \sum_1^{n-k} (Y_t - \acute{Y}_t)(Y_{t-1} - \acute{Y}_{t-1}) / \sum_1^n \left(Y_t - \acute{Y}_t\right)^2}$
- In a random timeseries, the 95% of the autocorrelation coefficients are in the range defined by the values $\pm \sqrt{(\boldsymbol{1.96/n})}$ where n is the number of observations.

- The coefficient of partial autocorrelation of class **k** is denoted by $\boldsymbol{a_k}$ and can be calculated by applying the method of multiple linear regression with dependent variable $\boldsymbol{Y_t}$ and independent variables $\boldsymbol{Y_{t-1}, \ldots, Y_{t-k}}$: $\boldsymbol{Y_t = b_o + b_1 Y_{t-1} + b_2 Y_{t-2} \ldots\ldots b_k Y_{t-k}}$
- The **"Sliding Operator" B** has no other mathematical meaning than to facilitate the writing of the different types of models that would otherwise be very difficult to be expressed. The "Sliding Operator" is defined as $\boldsymbol{B_\kappa Y_t = Y_{t-k}}$ that is, when an observation is multiplied by the operator, then this results in the observation before **k-time** moments, where k is the exponent of the operator.

We considered worthy to present in tabular form the key-determinants of electricity and NG demand in various countries in weighted average term and how different they are.

Table 1: analysis of t determinants of electricity and NG demand and consumption

Determinants influencing electricity and NG	**Electricity %**	**Natural gas (NG) %**
GDP/capita	64,2	67,8
Structure of GDP	48,7	63,2
Price of electricity and NG	72,4	61,7
Weather	33,4	50,2
Unemployment %	12,4	8,9
Existed infrastructure for NG use	34,2	46,3
Electricity and NG retailers promotion policies	24,8	35,9
Sources: Hodroyiannis 2004, Sanchez-Ubeda and Berzosa 2007, Behrouznia A et al 2010, Toksari 2010, Dombayci 2010, Okajima and Okajima 2013, Dergiades et al 2013 - proper adaptations by authors		

2. Literature review on the issue of NG demand and consumption forecasting

Natural Gas (NG) demand and consumption forecasting was investigated in several different areas, spatial level, time level, gas distribution system levels, sectoral levels- industrial, commercial, and domestic consumers, etc. In addition, there have been some studies on predictions at the giant individual customer level. The above researches proceeded to forecasts for various time horizons. Time horizons ranging from a few hours to a few or many years. Most of these studies, however, dealt with annual, three-year, and multi-year forecasts . The scientific community around the world has been predicting NG demand based on different models and forecasting methods. From very simple statistical models, such as timeseries models, and neural networks to econometric models and various more specialized methods.

Verhulst MJ (1950) studied the NG demand forecasting by the French industry in a sample of 46 companies divided into three groups. He built a model that was defined by the demand equation, the production equation, and the equilibrium equation between price and income. His model achieved rather a good accuracy. The accuracy is estimated via gap analysis, forecasting demand/actual demand X 100. The smaller result, the greater accuracy of model.

Hubbert M.K (1957) explored the life cycle of fossil fuels and establish the so-called "Hubbert curve model" of mathematical relationships included in the complete production cycle of any exhaustive resource and the production rate dQ/dt as a function of time. His model achieved good to moderate accuracy.

Balestra P and Nerlove M (1966) used econometric methods in forecasting the NG demand by domestic and commercial sectors, based on the model of minimal squares. Their model achieved rather a good accuracy. Berndt E.R and Watkins GC (1977) used econometric methods in forecasting the NG demand by domestic and commercial sectors of Columbia and Ontario, based also on the model of minimal squares. Their model achieved good accuracy.

Piggott J.L (1983) predicted the NG demand in sectoral daily and weekly basis in five countries using Box-

Jenkins modeling. His model achieved moderate to good accuracy.

Herbert J.H (1987) used Multiple Regression Modeling Methods to assess the overall monthly industrial demand for NG in the United States. Their model achieved rather a good accuracy.

Werbos P.J (1988) studied a Generalized Back-Propagation application in predicting NG demand in emerging energy markets. His model achieved good accuracy.

Brown R.H et al(1994/1995) developed models based on Feed-Forward ANN to predict NG consumption by households on a daily basis in Wisconsin, USA. Their model achieved a good accuracy.

Smith P et al (1996) used expert systems to predict NG demand and compared their results with them come by traditional forecasting methods.

Bartels R et al (1996) used the statistical method Conditions Demand Analysis (CDA) to calculate the consumption of NG in Australia. Their model achieved good accuracy.

Sailor D, Munoz R (1997) They developed a methodology for assessing the sensitivity of electricity and NG consumption in an area to existed environmental conditions. Their model achieved good accuracy.

Khotanzad and Elragal H (1997) used a combination of ANN forecasters for prediction of NG demand and consumption in different areas. Their model achieved good accuracy.

Al- Fattah SM and Startzman RA (2000) predicted the NG global demand for the next fifty years using an approach to the Multicyclic Hubbert Model with moderate results.

Gumrah F et al (2001) used the idea of the "Degree-Day Model" to model NG demand in sectoral basis to Ankara / Turkey with moderate results.

Siemek et al (2003) used an adaptation of the “Hubbert Model” to predict NG consumption in Poland over the next forty years with moderate TO good results.

Gorucu FB et al (2004) developed a Multi-Linear Regression model to identify the factors influencing NG demand in Ankara and to predict consumption using optimistic and pessimistic scenarios with moderate to good results.

Gil S and Deferrari J (2004) presented a Generalized Model for forecasting the domestic and commercial consumption of NG in urban areas, in the short and medium term forecast horizon, day-ahead up to five years with moderate to good results with moderate to good results .

Elgaral H (2004) proposed a new technique to improve ANN forecasting ttechniques using the "Fuzzy-Genetic” model with good results.

Gutierrez R et al (2005) tested the capabilities of the “Gompertz-type Innovation Diffusion Process GIDP” as a reliable stochastic model for predicting increased NG consumption in Spain with good results.

Potocnik et al (2007/2008), proposed an energy forecasting approach where energy consumption cycles are analyzed, and the information obtained are incorporated into the statistical forecasting model. They studied practices for the construction of models with an explanatory example of the Slovenian economic model that motivates NG distributors to predict their future consumption with the least error delivering the results to a responsible entity. The results were significant satisfactory.

Aras N (2008) introduced an application of Smart Genetic Algorithms to predict the short-term demand of NG by domestic consumers with rather good success.

Jiang et al (2008) investigated three Chinese areas Peking, Guang-Dog and Shanghai, identifying the most important factors leading to NG consumption. Using the MARKAL optimization model, they showed that the NG consumption level is the most sensitive to restrictive scenarios. with rather good success.

Vondracek J et al (2008) presented an OLS statistical approach to forecast NG consumption by individual domestic and small commercial customers with very success.

Aydinalp - Koksal M and Ugursal VI (2008) studied the Conditional Demand Analysis CDA method to model end-use energy consumption by domestic consumers in Canada with very success.

Ma H και Wu Y (2009) used a dynamic "Gray Model" to predict NG consumption in China from 2008 to 2015 with satisfactory results.

Xie Y and Li M (2009) introduced the "Gray Model" optimized with Genetic Algorithms to predict NG consumption in new energy markets with satisfactory results.

Azadeh A et Al (2010) introduced the "Adaptive Network-Based Fuzzy Inference System" (ANFIS) to assess NG demand in Iran using daily consumption data with moderate to good results.

Xie Y and Li M (2009) analyzed NG reserves, distribution system and the sectoral using in China, and predicted the future consumption using the generalized "Weng" and "Gray Model" models with satisfactory results.

Xu G και Wang V (2010) used the "Polynomial Curve and Moving Average Combination Projection PCMACP" model to calculate future NG consumption in China from 2009 to 2015 with moderate to good results.

Erdogdu E (2010/2012/2014) focused on demand characteristics, assessed the short-term and long-term price and revenue elasticities of each NG sector in Turkey, and predicted in 3 different years its future growth using an ARIMA model with good accuracy.

Aramesh A at al (2014) used a general neural and fuzzy-neural algorithm aiming to forecast with great accuracy the NG gas demand in city- gate stations with satisfactory results.

Nick S, Thoenes S (2014) suggested a method to investigate the determinants of what drives natural gas prices in every region or country using a structural VAR approach with satisfactory results.

Azadeh et al (2010/2012/2014/2015) proposed alternative approaches of the «Adaptive Network Based Fuzzy Inference System - Fuzzy Data Envelopment» models for predicting and analysis of NG demand in combination with Neuro-Fuzzy Inference Approach with moderate to good results.

Jiang L.Y (2014) suggested a forecasting model to predict NG gas based on "improved back propagation neural network" with satisfactory success.

Platts I (2016) suggested a methodology and speciation's guide for European natural gas assessments and indices and this guide is used by many energy decision-makers

Lilian M de Menezes at al (2018) investigated the link between the UK natural gas market, and other energy markets, using multivariate GARCH models and data from the spot markets at daily frequency, from January 2000 to May 2015. The proposed bivariate BEKK model allow for spillovers and interactions among energy markets, asymmetries, and dynamics in the fundamental values, as proxied by the interest-adjusted spread between spot and futures prices.

3. Implementation of our models and dataset in timeseries form used

The basic steps toward the solve of our problem are:

- The gathering of overtime 1999-2018 data in timeseries form for the two regions of Attica and Thessaly concerning their total and per capita GDP, electricity, and NG demand/consumption. The same data-exempt NG demand- was collected for the region of EMTH where its citizens predicted to connect with the NG network at the end of 2022.
- Initial approach to define the type of timeseries makes up our datasets through descriptive statistics
- Selection and implementation of seven deterministic forecasting models and their trialed running in order to be checked their functionality.
- Final running the seven deterministic forecasting models with S/W packages SPSS, FORECAST, EVIEWS, WDI
- Saving the prices of the forecasts for their evaluation regarding their accuracy using the four statistical residuals or errors, Mean Squared Error-MSE, Root Mean Squared Error-RMSE, Mean Absolute Percentage Error-MARE, Mean Percentage Error-MPE,

- Calculation and estimation of errors of each model separately.
- Evaluation and comparison of methods based on errors and ranking them according to their accuracy and reliability.

4. Reasons for our correct choice the rather best model

We studied in deep the data before we run the model and interpret results. If we did not do that the results could be strongly driven by outliers and this is especially true for models that minimize denuded squared sums.

We also spent a lot of time understanding the objective function of our models and how the data and models relate to the objective function.

We spent a lot of time understanding dataset and models characteristics and we formed a hypothesis which models are likely to best capture those characteristics. Although the model fanciness, we could forget the dumb way of forecasting data and error metrics are decreasing. Without using many forecasting models and benchmark them , we couldn't have a good absolute comparison for how good our models are. It is explained why we used seven forecasting models, starting from simpler -MA and going ahead with more complicate-ARIMA and then we compare their results using five criteria.

We started working with a small representative sample of total data of timeseries and we saw if we can get something useful out of it.

When we run the seven models in prediction, they get fed with data that is available when we run the model. That data might be different than what we assumed to be available in testing. We made sure that we run our model in realistic out-sample conditions and we understood when it will perform well and when it does not. We made sure you have a true test set free of any leakage from dataset. Especially beware of any time-dependent relationships that could occur in NG demand and use.

We generated test such those accurately reflect data on which we would make predictions. Especially with our timeseries data we likely will have to generate custom cross-validation data or do roll-forward testing.

After we have finished building the model, we tried to find another version of the datasets that could be a surrogate for a true out-sample dataset.

5. Descriptive statistical analysis

The basic data for the region of EMTH are the follow:

Table 2: basic data describing the energy and economic structure of REMTH

Descriptive data of region	**Quantity**
Total population	608.511
Number of families	324.515
Number of buildings	264.167
Houses	187.310
Hotels every type	1.024
Manufacturing buildings	1.739
Schools every level	1.577
Premises and offices	9.995
Hospitals and clinics	125
Various	61.397
Average total energy consumption per family in KWh	13.994
Average electricity consumption per family in KWh	3.750

Average heat-oil consumption per family in KWh	10.994
Estimation of the number of total buildings with strong probability to be connected to NG network	211.334
Source: HELLASTAT adhoc editions 2001-2020	

The tables below gives the necessary and efficient data to make the predictions.

Table 3: overtime data in two Greek regions (Attika /Thessaly) related to NG demand/consumption

Year	**Consumption of NG in M3**	**Users NG**	**GDP total in millions**	**GDP/capita weighted for 2 regions**	**Energy KWh consumption /household**	**Electricity KWh consumption /household**
1999	14.000	15.800	70.250	13.985	14.856	4.322
2000	14.800	16.400	71.684	14.084	14.982	4.120
2001	18.200	19.300	77.031	14.966	14.654	4.352
2002	33.450	34.200	84.098	16.728	14.987	4.532
2003	46.500	47.900	92.377	18.482	14.876	4.423
2004	82.850	83.200	101.212	20.242	15.346	4.236
2005	122.200	123.000	104.269	20.756	15.564	4.122
2006	151.800	152.800	115.548	22.882	15.345	4.098
2007	198.600	200.800	123.760	24.724	15.122	4.087
2008	215.000	218.100	128.903	25.521	15.008	4.124
2009	308.700	309.000	130.815	25.534	14.987	4.002
2010	238.700	239.900	121.297	23.767	15.001	4.108
2011	461.500	463.000	110.925	22.421	15.230	4.020
2012	492.800	495.700	102.188	19.362	14.874	4.234
2013	522.100	525.800	96.527	19.371	14.465	4.356
2014	540.300	544.500	95.161	19.287	14.201	4.567
2015	556.300	559.800	93.669	19.102	14.123	4.543
2016	574.600	577.000	98.886	19.101	14.765	4.632
2017	589.200	602.400	95.129	19.574	14.806	4.652
2018*	598.000	609.100	97.770	20.482	14.808	4.436
2019*	608.000	612.400	100.339	22.153	14.862	4.286

Table 4: overtime same data in Greek region EMTH related to NG demand/consumption exempt NG demand

Year	**GDP in mil €**	**GDP/capita**	**Total energy consumption KWH/household**	**Electricity consumption KWh/household**	**Energy potentially covered by NG in KWh**	**Required heating at least hours/day***

Year	GDP in mil €	GDP/capita	Total energy consumption KWH/household	Electricity consumption KWh/household	Energy potentially covered by NG in KWh	Required heating at least hours/day*
1999	5.673	12.345	3.391	8.954	182	
2000	5.789	10.075	12.383	3.396	8.987	178
2001	6.280	10.701	12.456	3.402	9.054	183
2002	6.686	11.316	12.864	3.765	9.099	165
2003	7.206	12.146	13.005	3.912	9.093	181
2004	7.611	12.772	13.234	3.954	9.280	169
2005	7.868	13.142	13.065	3.863	9.202	193
2006	8.141	13.535	12.986	3.802	9.184	185
2007	8.906	14.741	12.764	3.754	9.010	180
2008	9.450	15.568	12.312	3.456	8.856	177
2009	9.306	15.272	12.763	3.546	9.217	176
2010	9.198	15.057	12.543	3.402	9.141	183
2011	8.150	13.320	12.238	3.322	8.916	185
2012	7.579	12.403	12.432	3.356	9.076	167
2013	7.003	11.498	12.685	3.413	9.272	174
2014	6.878	11.324	12.652	3.410	9.242	175
2015	6.831	11.281	12.386	3.387	8.999	178
2016	6.901	11.432	12.368	3.298	9.070	166
2017	6.946	11.539	12.543	3.345	9.198	164
2018*	7.043	11.701	12.764	3.459	9.305	162
Source: HELLASTAT Special reports, DEH, RAE, DEDHE adhoc reports, EnergyPress adhoc reports, www.oikonomiki.gr						
*The days needed the use of heating systems are based on temperatures lower than 8 Celsius degrees						

For the above timeseries we can calculate the values of variable parameters used in our measures extraction and are listed in next table 4 below:

Table 5: Transformation parameters used in feature measures

Feature	Raw time series data Yt-RAW	Trend and Seasonally Adjusted data Ýt-TSA
Serial Correlation	2, 7.53,0.125	2,7.53,0.125
Non-linearity	1,0.069,2.311	1,0.069,2.311
Skewness	1,1.510,5.991	1,1.510,5.991
Kurtosis	1,2.273,11562	1,2.273,11562

Periodicity	1, 12, 52	----

Continuing the descriptive statistics processing we result to the measures and indices of the following tables 6 and 7:

Table 6: Cross-correlation matrix

Statistical indices	NG M3 consumption	NG Users	Total GDP in m	Weighted GDP/capita for 2 regions	Total Energy consumption /household	Electricity consumption /household	NG Consumption per User
NG Consumption in M3	1.000000	0.999936	0.102020	0.191117	-0.522102	0.409887	0.475341
NG Users	0.999936	1.000000	0.098975	0.188759	-0.522092	0.413696	0.471087
Total GDP in m	0.102020	0.098975	1.000000	0.978134	0.367429	-0.569617	0.661149
Weighted GDP/capita for 2 regions	0.191117	0.188759	0.978134	1.000000	0.336732	-0.525691	0.685335
Energy KWh consumption /household	-0.522102	-0.522092	0.367429	0.336732	1.000000	-0.654082	0.052830
Electricity KWh consumption /household	0.409887	0.413696	-0.569617	-0.525691	-0.654082	1.000000	-0.006266
NG Consumption/ User	0.475341	0.471087	0.661149	0.685335	0.052830	-0.006266	1.000000

Table 7: Descriptive statistics, basic statistical indices, and measures

Statistical indices-measures	NG in M3 consumption	NG users	Total GPD in m €	Weighted GDP/ capita	Total energy consumption/ household	Electricity consumption/ household	NG consumption/ household
Count	20.000000	20.000000	20.000000	20.000000	20.00000	20.000000	20.000000
Mean	318.680000	321.715000	102.079400	20.426950	14.90030	4.296500	0.983216
Std	228.251847	230.408763	15.973852	3.163026	0.35689	0.211692	0.022933
Min	14.800000	16.400000	71.684000	14.084000	14.12300	4.002000	0.902439
25%	112.362500	113.050000	94.764000	19.101750	14.79575	4.117000	0.980854
50%	273.700000	274.450000	99.612500	19.908000	14.92900	4.261000	0.992889
75%	544.300000	548.325000	112.080750	22.536250	15.03650	4.460000	0.994362
Max	608.000000	612.400000	130.815000	25.534000	15.56400	4.652000	0.999029
Pearson r correlation	0.0100	0.0150	0.0240	0.0310	0.0340	0.0420	0.0480
Kendall rank correlation	0.0080	0.0120	0.0220	0.0280	0.0360	0.0430	0.0490
Spearman rank correlation	0.0080	0.0130	0.0230	0.0290	0.0340	0.0410	0.0470

Corr_matrix = NG_cons.corr() corr_matrix

Table 8: Results of white noise from normal distribution of our timeseries concerning NG demand

Year/measures	**Kendall**	**Spearman**	**Pearson ***	**Kendall**	**Spearman**	**Pearson**	**Order Correlation**	**Entropy Correlation**
Year	Parametric	Parametric	Parametric	Non-Param	Non-Param	Non-Param		
1999	**0.0080**	**0.0080**	**0.0100**	**0.0070**	**0.0060**	**0.0100**	**0.0100**	**0.0100**
2000	**0.0120**	**0.0130**	**0.0150**	**0.0150**	**0.0140**	**0.0150**	**0.0160**	**0.0210**
2001	**0.0220**	**0.0230**	**0.0240**	**0.0250**	**0.0260**	**0.0220**	**0.0240**	**0.0290**
2002	**0.0280**	**0.0290**	**0.0310**	**0.0300**	**0.0290**	**0.0320**	**0.0310**	**0.0400**
2003	**0.0360**	**0.0340**	**0.0340**	**0.0370**	**0.0400**	**0.0390**	**0.0410**	**0.0550**
2004	**0.0430**	**0.0490**	**0.0420**	**0.0480**	**0.0480**	**0.0440**	**0.0440**	**0.0640**
2005	**0.0570**	**0.0600**	**0.0500**	**0.0600**	**0.0620**	**0.0510**	**0.0540**	**0.0760**
2006	**0.0650**	**0.0690**	**0.0580**	**0.0680**	**0.0730**	**0.0590**	**0.0600**	**0.0810**
2007	**0.0780**	**0.0760**	**0.0650**	**0.0800**	**0.0800**	**0.0690**	**0.0720**	**0.0930**
2008	**0.0860**	**0.0880**	**0.0740**	**0.0920**	**0.0890**	**0.0810**	**0.0830**	**0.1010**
2009	**0.0940**	**0.0980**	**0.0860**	**0.1040**	**0.1050**	**0.0890**	**0.0980**	**0.1160**
2010	**0.1040**	**0.1020**	**0.0950**	**0.1150**	**0.1140**	**0.1020**	**0.1040**	**0.1270**
2011	**0.1120**	**0.1100**	**0.1000**	**0.1260**	**0.1220**	**0.1090**	**0.1110**	**0.1390**
2012	**0.1210**	**0.1240**	**0.1130**	**0.1340**	**0.1310**	**0.1170**	**0.1230**	**0.1490**
2013	**0.1330**	**0.1380**	**0.1210**	**0.1390**	**0.1410**	**0.1280**	**0.1280**	**0.1600**
2014	**0.1440**	**0.1480**	**0.1280**	**0.1430**	**0.1480**	**0.1380**	**0.1350**	**0.1700**
2015	**0.1570**	**0.1530**	**0.1400**	**0.1490**	**0.1530**	**0.1460**	**0.1440**	**0.1780**
2016	**0.1650**	**0.1660**	**0.1520**	**0.1540**	**0.1610**	**0.1560**	**0.1570**	**0.1880**
2017	**0.1540**	**0.1580**	**0.1490**	**0.1370**	**0.1450**	**0.1320**	**0.1460**	**0.1790**
2018	**0.1490**	**0.1400**	**0.1510**	**0.1370**	**0.1490**	**0.1410**	**0.1460**	**0.1810**

figCorr, axCorr= plt.subplots(figsize=(10,10))

sns.heatmap(corr_matrix, vmin=-1.0, vmax=1.0, ax=axCorr, annot=True, cmap="YlGnBu")

*Pearson r correlation is the most widely used correlation statistic to measure the degree of the relationship between linearly related variables

Further for more reliable and accurate predictions we have to take in account the elasticities of parameters we use in our timeseries. The elasticities for NG demand in Greece for the period under research have as the table below. It is noted that the same elasticities will be in force in the region of EMTH for plausible reasons.

Table 9: elasticities for NG demand

Elasticity*	**Period <12 months**	**Period <12 months**	**Period >12 months**	**Period >12 months**
Type of model	Logarithmic	Linear	Logarithmic	Logarithmic
NG price/m3	-0,163	-0,187	-0,684	-0,751

Income per capita	0,232	0,264	0,652	0,576
Cold days	0,012	0,075	0,073	0,076
all elasticities have been calculated in the means of the timeseries				

From the above datasets in timeseries form we concluded to the follow results using statistical tools of descriptive statistics.

1. The tables show the correlation rank between the variables and are called Cross-Correlation matrices.
2. The correlation between variables is between -1 to 1, so the Ha hypothesis is in force and the Ho one is rejected.
3. The diagonals of the table take values equal to the unit as it represents the correlation between the same variables
4. The outputs of the table are symmetrical with respect to the diagonals.
5. For the possible existence of a strong linear dependence between the interpretive variables, which would give us an erroneous estimate of the regression coefficients, as their dispersion increases, we perform a multicollinearity variability test. The results of E-Views are presented in the table and according to them the coefficients of linear dependence between GDP, GDP / capita and NG demand are quite close to the unit and therefore the interpretive variables of electricity prices are rejected.
6. According to the above analysis, we can conclude that the best adaptation is the non-parametric coefficients of Pearson and Spearman, ie they reject the Ho hypothesis as well as the level of significance.

6. The generalized forecasting model adapted to each prediction method

After all the above, we are led to the generalized form of the prediction model will take the formula $\mathbf{Y_t = b_o + b_1X_1 + b_2X_2 \ldots\ldots b_pX_p + e}$ which of course will give different values for each of the seven forecasting methods since both the residuals and the parameters of the independent variables Xt are differed. Also, they give different demand in $\mathbf{M^3}$ for different periods, since future NG users will install to NG network gradually, as it happens to other two regions. We adapt the timeseries real data to the generalized forecasting model that takes the final formula: $\mathbf{X_t = 0.0046 + 0.35X_{t+1} + 0.38X_{t+2} \ldots\ldots 0.92X_{t+20} + e_t}$ with an estimated standard deviation $\boldsymbol{\sigma_t = 0.0098}$. From the above generalized model and putting each time the different value of $\boldsymbol{e_t}$, the outputs presented to the next table 8 were resulted.

Table 10: the NG demand separately for each forecasting method

Method/period from introduction NG - demand in M^3	**1st 20%**	**5th 65%**	**10th 85%**	**15th 98%**	**20th 100%**
Smoothing methods					
Simple Moving Average (MA)	45.663	137.367	179.634	207.107	228.315
Simple Exponential Smoothing (SES)	43.467	134.256	174.345	204.784	217.334
Double Moving Average (DMA)	42.983	132.875	171.652	201.651	214.902
Double Exponential Smoothing (B-DES)	42.065	132.143	171.003	201.087	210.328
Holt or Exponential Smoothing for Trend Adaptation (H-ESTA)	41.991	130.896	168.783	166.763	209.955
Decomposition methods					
Multiplication Decomposition Model(MDM)	41.345	131.086	169.207	166.206	206.728
Autoregression Average models					
Autoregression Move Average models	41.667	131.387	170.207	166.765	208.339

Method/period from introduction NG - demand in M^3	**1st 20%**	**5th 65%**	**10th 85%**	**15th 98%**	**20th 100%**
(ARIMA)					

To evaluate the prediction accuracy of seven models we use as evaluating criteria the next five errors, Mean Squared Error-MSE, Root Mean Squared Error-RMSE, Mean Absolute Percentage Error-MAPE, Mean Percentage Error-MPE and Mean Absolute Deviation-MAD. Based on the five specific time horizons of the timeseries, the following results are obtained for the errors of the seven forecasting methods that we applied and more specifically for forecasts that concern the next periods.

Table 11: criteria used to evaluate separately each forecasting method using the errors method

Method/criteria/period	**MSE**	**RMSE**	**MAPE**	**MPE**	**MAD**
MA 1st period	-0.6472	9649.6	3.2	1,6	4.5
MA 2nd period	-1.3501	6258.1	1.8	1.1	4.3
SES 1st period	-0.3054	8456.6	3.3	1.5	4.6
SES 2nd period	-1.0221	5986.1	2.2	1.0	4.4
DMA 1st period	-0.3716	7345,4	3.1	1.8	4.2
DMA 2nd period	-1.2181	5343.3	2.0	1.1	3.9
B-DES 1st period	-0.3126	9008.6	2.9	1.5	3,8
B-DES 2nd period	-1.0332	7145.8	1.7	0.9	3.5
H-ESTA 1st period	-0,0962	8240.2	2.7	1.4	3.7
H-ESTA 2nd period	-0.1160	6981.3	1.9	1.0	3.2
MDM 1st period	-0.3125	7876.8	3.2	1.7	4,0
MDM 2nd period	-1.0346	6465.5	2.8	1.4	3.9
ARIMA 1st period	-0.0072	5987.4	2.1	1.1	2.9
ARIMA 2nd period	0.4780	4763.6	1.5	0.8	2.7

All the above results and outputs are presented to the next figure 1.

7. Critics and comments

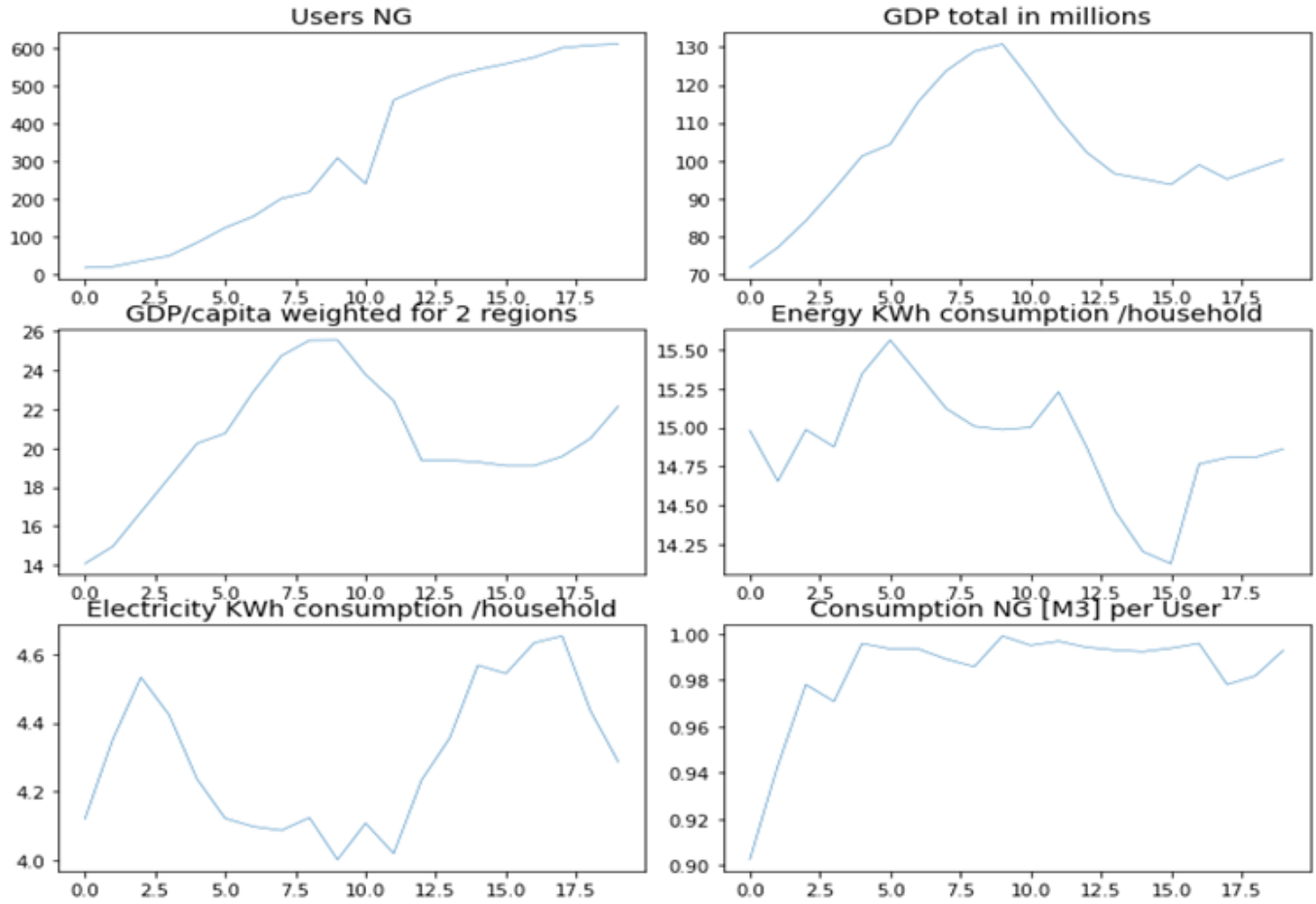

Figure 1: the results of models implemented

After above results and outcomes of the seven forecasting and evaluating models we conclude to:

- Based on the information of table 11, we can tell that the ARIMA method has 16.1% better accuracy performance than all the other six models. ARIMA model is better for producing predictions in future times for which we do not have historical data, as long as timeseries to be cleared from outliers.
- The second better method is the Multiplication Decomposition Model (MDM) which has 6.8% better accuracy performance than all the other five models.
- It should be emphasized the importance of filtering the data of timeseries before producing forecasts. Usually in the initial data there will always be values which do not match the sizes of the neighboring values. A sharp increase or decrease in the values of the testable variable is something common for initial timeseries data, but not something acceptable in the science of predictions. This is because the accumulation of such prices can disorient the adaptation of timeseries models and give unreliable predictions. The same applies if there are empty prices in timeseries. In case the empty values cover a large size of the observations, specialized methods of intermittent demand, such as Croston methods, ADIDA are used, but in our case this was not necessary.
- Continuing the commentary of the results, the importance of filtering the data before the forecasting process has been particularly emphasized, especially when in the initial data there are values-observations which do not keep pace with the sizes of the neighboring values. It is obvious that in the case of processed and filtered data the improvement of the results are better by 12,6%.
- In order to be calculated the correlation measures among independent variables the next formulas were

used:

Measures	Formula
Pearson's r	$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}}$
Spearman rank ρ	$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$
Kendall rank t	$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}$

References

[1.] Al-Fattah S.M, Startzman R.A (2000) "Forecasting world natural gas supply", Journal of Petroleum Technology, 52 (5), pp. 7-19.

[2.] Aramesh A, Ahmadi A (2014) "A general neural and fuzzy-neural algorithm for natural gas flow prediction in city gate stations", Energy and Buildings, 72, pp. 73–79.

[3.] Aras N (2008) "Forecasting residential consumption of natural gas using genetic algorithms", Energy Exploration and Exploitation, 26 (4), pp. 241-266.

[4.] Aydinalp-Koksal M, Ugursal V.I (2008) "Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector", Applied Energy, Volume 85, Issue 4, April 2008, pp 271-296.

[5.] 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, Volume 38, Issue 3, pp 1529-1536.

[6.] Balestra P, Nerlove M (1966) "Pooling cross section and time series data in the estimation of a dynamic model: the demand for natural gas", Econometrica, 34(3), 585-612.

[7.] Berndt ER, Watkins GC (1977) "Demand for natural gas: residential and commercial markets in Ontario and British Columbia", Canadian Journal of Economics, Canadian Economics Association, 10(1) pp 97–111.

[8.] Bartels R, Fiebig D.G, Nahm D (1996) "Regional end use gas demand in Australia", The Economic Record, 72 (219), pp. 319-331.

[9.] Behrouznia A, Saberi M, Azadeh A, Asadzadeh S.M, Pazhoheshfar P (2010), "An Adaptive Network Based Fuzzy Inference System-Fuzzy Data Envelopment Analysis for Gas Consumption Forecasting and Analysis: The Case of South America", In: International Conference on Intelligent and Advanced Systems, ICIAS. Article Number 5716160.

[10.] Brown R. H, Kharouf P, Feng X, Piessens L. P and Nestor D, (1994) "Development of feed-forward network models to predict gas consumption" IEEE International Conference on Neural Networks - Conference Proceedings Volume 2, 1994, Pages 802-805.

[11.] De Menezes L.M, Russo M, Urga G (2018) " The volatility of natural gas prices in the United Kingdom market : drivers and spillover effects," The Energy Journal, 40(1), pp. 143-169.

[12.] Dergiades T, Martinopoulos G, Tsoulfidis L (2013) "Energy consumption and economic growth: Parametric and non-parametric causality testing for the case of Greece," Energy Economics, Elsevier, vol. 36(C), pp. 686-697.

[13.] Dombayci A (2010) "The prediction of heating energy consumption in a model house by using artificial neural networks in Denizli-Turkey", Advances in Engineering Software, 41(2), pp.141-147.

[14.] Elragal H (2004) "Improving neural networks prediction using fuzzy-genetic model", National Radio Science Conference, NRSC, Proceedings Volume 21, pp 393-400.

[15.] Erdogdu E (2010) " Natural gas demand in Turkey", Applied Energy, Volume 87, Issue 1, pp 211-219

[16.] Gil S, Deferrari J (2004) "Generalized model of prediction of natural gas consumption", Journal of Energy Resources Technology, Transactions of the ASME Volume 126, Issue 2, pp. 90-98
[17.] Gorucu F.B, Geris P.U, Gumrah F (2004) "Artificial neural networks modeling for forecasting gas consumption", Energy Sources, 26, pp. 299-307.
[18.] Gutierrez R, Nafidi A, Sanchez R.G, (2005) "Forecasting total natural-gas consumption in Spain by using the stochastic Gompertz innovation diffusion model", Applied Energy, 80 (2), pp115–124.
[19.] Gumrah F, Katircioglu D, Aykan Y, Okumus S, Kilincer N (2001) "Modelling of gas demand using degree day concept: Case study of Ankara", Energy Sources, 23(2), pp. 101-114
[20.] Herbert J.H (1987) "An analysis of monthly sales of natural gas to residential customers in the United States", Energy System and Policy, volume 10, pp. 127–147.
[21.] Hodrogiannis H (2004) "Estimating residential demand for electricity in Greece", Energy Economics, volume 26 (3), pp. 319–334.
[22.] Hubbert M.K (1957) "Nuclear Energy and the Fossil Fuels" Presented before The Spring Meeting of the Southern District Division of Production, American Petroleum Institute, San Antonio, TX, March 8, 1956. Publication No. 95.
[23.] Jiang B, Wenying C. Yuefeng Y, Lemin Z, Victor D (2008) "The future of natural gas consumption in Beijing, Guangdong and Shanghai: An assessment utilizing MARKAL" Energy Policy, Volume 36, Issue 9, pp. 3286-3299.
[24.] Imam A, Startzman R.A, Barrufet M.A, Hubbert M.K (2004) "Multicyclic Hubbert model shows global conventional gas output peaking in 2019", Oil and Gas Journal, 102 (31), pp. 20-28.
[25.] Khotanzad A, Elragal H (1999) "Natural gas load forecasting with combination of adaptive neural networks", In: Proceedings of the International Joint Conference on Neural Networks Volume 6, pp. 4069-4072.
[26.] Ma H, Wu Y (2009) "Grey Predictive on natural gas consumption and production in China" Proceedings of the 2009 2nd Pacific-Asia Conference on Web Mining and Web-Based Application, WMWA 2009 2009, Article number 5232475, pp. 91-94.
[27.] Ma Y, Li Y (2010) "Analysis of the supply-demand status of China's natural gas to 2020", Petroleum Science Volume 7, Issue 1, pp. 132-135.
[28.] Nick S, Thoenes S (2014) «What drives natural gas prices? - A structural VAR approach», Energy Economics, volume 45, pp 517-527.
[29.] Okajima S & Okajima H (2013) "Estimation of Japanese price elasticities of residential electricity demand, 1990–2007," Energy Economics, Elsevier, vol. 40(C), pp. 433-440.
[30.] Platts I (2016) "Methodology and speciation's guide - European natural gas assessments and indices", S&P Global Platts. Available at https://www.platts. com/im.platts.content/methodologyreferences/methodologyspecs.
[31.] Piggott J.L (1983) "Use of Box-Jenkins modelling for the forecasting of daily and weekly gas demand", IEE Colloquium (Digest), (1983 /91), pp. 4-10.
[32.] Potocnik P, Thaler M, Govekar E, Grabec I, Poredos A (2007) "Forecasting risks of natural gas consumption in Slovenia", Energy Policy, 35, pp 4271-4282.
[33.] Potocnik P, Govekar E, Grabec I (2007) "Short-term natural gas consumption forecasting", Proceedings of the 16th IASTED International Conference on Applied Simulation and Modelling, ASM 2007, pp 353-357.
[34.] Potocnik P, Govekar E, Grabec I (2008), "Building forecasting applications for natural gas market", Natural gas research progress, New York: Nova Science Publishers, cop. 2008, pp. 505-530.
[35.] Sailor D, Munoz R (1997) "Sensitivity of electricity and natural gas consumption to climate in the USA—methodology and results for eight states", Energy, 22 (10), pp 987–998.
[36.] Sanchez-Ubeda E.F, Berzosa A (2007) "Modeling and forecasting industrial end-use natural gas consumption", Energy Economics, 29 (4), pp. 710–742.
[37.] Siemek J, Nagy S, Rychlicki S (2003) " Estimation of natural-gas consumption in Poland based on the logistic-curve interpretation", Applied Energy, 75 (1–2), pp. 1–7.
[38.] Smith P, Husein S and Leonard D.T (1996) "Forecasting short term regional gas demand using an

expert system", Expert Systems with Applications, 10 (2): 265–273.
[39.] Toksari D (2010) "Predicting the Natural Gas Demand Based on Economic Indicators: Case of Turkey", Energy Sources Part A Recovery Utilization and Environmental Effects, 32 (6), pp. 559-566.
[40.] Werbos P.J (1988) "Generalization of backpropagation with application to a recurrent gas market model", Neural Networks, 1 (4), pp. 339-356.
[41.] Verhulst M.J (1950) "The theory of demand applied to the French gas industry" Econometrica 1950, 18(1,) pp 45–55.
[42.] Vondracek J, Pelikan E, Konar O, Cermakova J, Eben K, Maly M, Brabec M.A (2008) "Statistical model for the estimation of natural gas consumption", Applied Energy, Volume 85, Issue 5, May 2008, pp 362-370.
[43.] Xie Y, Li M (2009) "Research on prediction model of natural gas consumption based on Grey modeling optimized by genetic algorithm" Proceedings IITA International Conference on Control, Automation and Systems Engineering, CASE 2009 , Article number 5194459, pp 335-337.
[44.] Xu G, Wang W (2010) "Forecasting China's natural gas consumption based on a combination model", Journal of Natural Gas Chemistry, 19, pp. 493–496.