DOI 10.2478/ama-2023-0004

acta mechanica et automatica, vol.17 no.1 (2023)
Special Issue "Machine Modeling and Simulations 2022"

TIME SERIES ANALYSIS OF FOSSIL FUELS CONSUMPTION IN SLOVAKIA BY ARIMA MODEL

Mária MICHALKOVÁ*, Ivana POBOČÍKOVÁ*

*Faculty of Mechanical Engineering, Department of Applied Mathematics, University of Žilina,
Univerzitná 8215/1, 010 26 Žilina, Slovakia

maria.michalkova@fstroj.uniza.sk, ivana.pobocikova@fstroj.uniza.sk

received 28 June 2022, revised 23 October 2022, accepted 24 October 2022

Abstract: According to the Green Deal, the carbon neutrality of the European Union (EU) should be reached partly by the transition from fossil fuels to alternative renewable sources. However, fossil fuels still play an essential role in energy production, and are widely used in the world with no alternative to be completely replaced with, so far. In recent years, we have observed the rapidly growing prices of commodities such as oil or gas. The analysis of past fossil fuels consumption might contribute significantly to the responsible formulation of the energy policy of each country, reflected in policies of related organisations and the industrial sector. Over the years, a number of papers have been published on modelling production and consumption of fossil and renewable energy sources on the level of national economics, industrial sectors and households, exploiting and comparing a variety of approaches. In this paper, we model the consumption of fossil fuels (gas and coal) in Slovakia based on the annual data during the years 1965–2020. To our knowledge, no such model, which analyses historical data and provides forecasts for future consumption of gas and coal, respectively, in Slovakia, is currently available in the literature. For building the model, we have used the Box–Jenkins methodology. Because of the presence of trend in the data, we have considered the autoregressive integrated moving average (ARIMA (p,d,q)) model. By fitting models with various combinations of parameters p, d, q, the best fitting model has been chosen based on the value of Akaike's information criterion. According to this, the model for coal consumption is ARIMA(0, 2, 1) and for gas consumption it is ARIMA(2, 2, 2).

Key words: ARIMA model, coal, gas, consumption, Slovakia, prediction

1. INTRODUCTION

Passing the Green Deal in 2019, the European Union (EU) committed to turn Europe into the first climate-neutral continent. According to this, each EU member has set itself goals to be achieved by 2030. Slovakia has aimed towards the following goals:

- to reduce greenhouse gas emissions by −20% until 2030 compared with 2005;
- to increase the share of energy from renewable sources in gross final consumption of energy to 19.2%;
- to increase the energy efficiency by energy savings that will lead to 15.7 [Mtoe] for primary energy consumption and 10.3 [Mtoe] for final energy consumption [1, 2].

Fulfilling these targets requires a reconsideration of the national strategy for energies. Fossil fuels are still an important part of the Slovak energy mix with a share of nearly 28% [3]. With respect to the EU targets, the use of coal should end entirely by 2030. Due to reduction of coal use, there has been an obvious switch from coal to gas in electricity production recently. In 2019, the volume of electricity produced in gas-fired power plants increased by about 11% in Europe, whereas the production of coal-fired power plants decreased by 24% [4]. In addition to power generation, thermal coal is used for operations, such as cement production and industrial and household heat applications, where alternatives are also being sought. Despite gas being considered the "greener" among fossil fuels, such status is only temporary, and in the future, part of the gas consumption will be replaced with renewable sources (see, for example, Jandačka et al. [5] and Nandimandalam et al. [6]).

In this paper, we model and forecast coal and gas consumption, respectively, in Slovakia by applying autoregressive integrated moving average (ARIMA) models. The ARIMA model is commonly used for modelling production and/or consumption of fuels. Dritsaki et al. [7] built the ARIMA model to forecast oil consumption in Greece. The time series covering the period 1960–2020 was modelled by the ARIMA(1,1,1) model and the forecasts for years 2021–2023 were calculated. Ozturk and Ozturk [8] forecast consumption of coal, oil, natural gas, renewable energy sources as well as of total energy, respectively, by modelling the historical data with the ARIMA models. Based on the predictions until the year 2040, they estimated the rate of increase to be between 4% and 5% for all the sources except that of the renewable energy sources, which have been expected to increase by about 1.6%. Akpinar and Yumusak [9] predicted a year-ahead demand for natural gas for household and low consumption consumers in Turkey. They applied three models – time series decomposition, ARIMA model and Holt–Winters exponential smoothing – for monthly consumption between the years 2011 and 2014. Among these considered models, the ARIMA model performed the best. Chaturvedi et al. [10] discussed and compared the performances of three models – the seasonal ARIMA (SARIMA) model, the Long Short-Term Memory Recurrent Neural Network (LSTM RNN) model, the Facebook (Fb) Prophet model and the Indian Central

DOI 10.2478/ama-2023-0004

Energy Authority (CEA) model as a reference model – for fitting the monthly total and peak energy demand in India.

Recently, the ARIMA models have been combined with other methods for creating hybrid models in order to obtain more precise results. It is common to incorporate the artificial neural network to model the non-linearity in the data, as the ARIMA model is able to describe only the linear relationship between the inputs and the output. Manowska et al. [11] forecast the natural gas consumption in Poland using the ARIMA–LSTM hybrid model. The residuals of the ARIMA model were further modelled by the LSTM neural network taking historical consumption and prices of energy resources, that is, crude oil, natural gas and thermal coal, as predictors of the model. Using this approach, the authors achieved the average percentage error of 2%. Based on the model, the predictions of natural gas consumption in Poland up to the year 2040 were constructed. Wang [12] predicted per capita coal consumption in China using the ARIMA-BP combined model. Proceeding from the hybrid model that combines the ARIMA model and the back-propagation neural network and simply sums these two models' results, Wang improved the model accuracy by obtaining the predictions via multiple linear regression with the linear fitting of the ARIMA model and the non-linear fitting of the BP model as independent variables and the consumption as the dependent variable. The energy demand in China and India were forecast by Wang et al. [13], applying the rolling metabolic grey (MGM) model, the rolling metabolic grey – ARIMA (MGM-ARIMA) model and the non-linear metabolic grey (NMGM) model. Here, the MGM-ARIMA model combined the grey model and the ARIMA model in a different way as in the hybrid ARIMA artificial neural network models, mentioned previously. The ARIMA model was used to model the MGM model's residuals to minimise their volatility. According to the comparison of three considered models, the MGM-ARIMA model fitted the energy demand of India as the best one, and the energy demand of China as the second best, outperformed by the NMGM model. South Africa's energy consumption was analysed and predicted by Ma and Wang [14]. They considered the ARIMA model, the nonlinear grey model (NGM) and the nonlinear grey – ARIMA model. Achieving the value of the mean average percentage error less than 3%, all three models made highly reliable predictions.

The number of papers providing the prediction models regarding the energy sector in Slovakia is scarce. Recently, Pavlicko et al. [15] forecast the electricity consumption in Slovakia by applying and comparing two approaches – grey models and multi-layer feed-forward back-propagation network. They also proposed a new model combining both approaches. Based on this model, the authors obtained more accurate maximum hourly electricity consumption per day forecasts, compared with the official load predictions. Brabec et al. [16] presented a non-linear mixed effect model that was able to predict the daily consumption of natural gas for an individual consumer. This model was applied on daily-recorded consumptions of 62 larger commercial entities in Slovakia and its performance was compared with performances of the ARIMAX and ARX model, respectively. Hošovský et al. [17] modelled the daily gas consumption considering three particular types of buildings, each in a different town in Slovakia. In the paper, they compared the performance of reg(S)ARMA (the regression model with ARMA-modelled time series error terms) with regWANN (the regression wavelet neural network) model and the SARMA model as a reference model. However, as far as we have found out, no prediction model of gas or coal consumption for Slovakia as a country has been published. To fill this gap, we propose such models applying the Box–Jenkins methodology. These models can serve as reference models and the bases for further research.

This paper is organised as follows: In Section 2, the time series used for study are characterised and their descriptive statistics are given. In Section 3, we provide the methodology for building ARIMA models. In Section 4, the results of modelling process for coal and gas time series, respectively, are summarised. In addition, the forecasts for the next 10 years are made. The last section "Discussion and Conclusions" summarises and interprets the obtained results, and provides the proposals for future research. All the calculations in the paper were conducted using MATLAB software, version R2020b.

2. CHARACTERISTICS OF DATA

Data modelled in the paper represent annual values of consumption of fossil fuels in Slovakia. Namely, we analyse gas and coal consumption, respectively. The records cover the period of years 1965–2020, that is, the data count 56 observations for each commodity. The gas data are given in milliards of cubic metres; the coal data are given in exajoules. The data are obtained from the literature [18].

The descriptive statistics of datasets are summarised in Tab. 1. The value of skewness of the coal data close to zero implies that the consumptions in the considered years are distributed almost symmetrically around the mean. On the other hand, the negative value of skewness for the gas consumption shows that higher consumptions dominate. The kurtosis values in both cases are greater than 1, indicating too peaked (leptokurtic) distributions.

Tab. 1. The descriptive statistics for coal and gas dataset, respectively

Statistic	Coal [EJ]	Gas [10^9 m^3]
Minimum	0.08296	0.29853
Maximum	0.35881	7.17380
Mean	0.23709	4.28430
Variance	0.00572	3.78942
Skewness	−0.03980	−0.48146
Kurtosis	1.79441	2.17448
Lower quantile	0.17385	2.89133
Median	0.24985	4.63595
Upper quantile	0.29202	5.91770

The presence of outliers, which indicate anomalies in the data, is checked by the boxplot. As shown in Fig. 1, there are no outliers in the respective datasets of the commodities.

3. METHODOLOGY

In general, the process of modelling and analysing the time series contains several steps, namely:
- graphical analysis of data, identification of components;
- selection of the model, estimation of parameters;
- checking the adequacy of the model in relation to the data;
- forecast of future values.

3.1. Pre-analysis of data

The first step in modelling the time series is its visualisation that enables us to identify the presence of particular components.

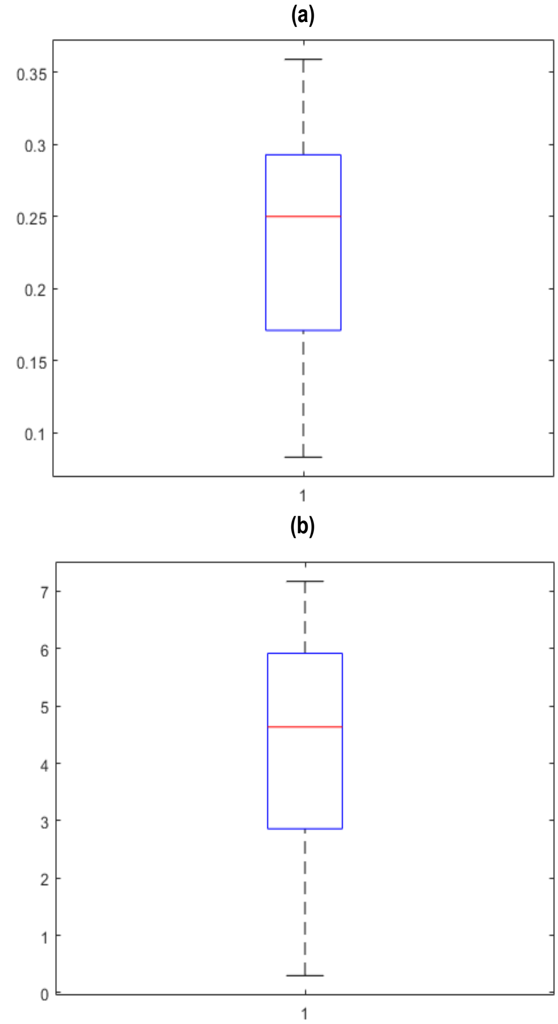

Fig. 1. Boxplot of dataset (a) coal and (b) gas

Generally, the time series includes the following components:
- trend;
- seasonal;
- cyclic;
- residual.

A trend occurs in the data when there can be observed a long-term change in the mean either as an increase, or as a decrease. However, there are cases when the trend is not monotonic. Seasonality is represented by fluctuations with a fixed frequency. These fluctuations are related to the seasonal aspects, such as a season of the year, a day in the week, etc. Similar to seasonality, cycle is also represented by altering the increase and the decrease in the data. Contrary to the seasonal component, these fluctuations do not have a fixed frequency. Usually, they are explained as a consequence of business cycles in economics. The residual component represents random changes in the data. Time series of residuals should be a white noise.

A stationary time series does not have a predictable pattern in the long-term. Therefore, the time series with trend or seasonality is not stationary. The verification of the presence of these components in the series can be done by testing the series for stationarity. There exist several stationarity tests. In this paper, we have selected the unit root tests, namely the augmented Dickey–Fuller (ADF) test and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test. The ADF test tests the null hypothesis H_0: there is a unit root in an autoregressive AR model (the data series is not stationary) against the alternative that the data series is stationary. The KPSS test works in a reverse manner to the ADF test since it tests the null hypothesis H_0: there is no unit root in an AR model (the data series is stationary) against the alternative that there is a unit root in an AR model, thus the data series is not stationary.

The results of these two tests should be interpreted as follows:
- the time series is stationary when H_0 in the ADF test is rejected and H_0 in the KPSS test cannot be rejected;
- the time series is non-stationary when H_0 in the ADF test cannot be rejected and H_0 in the KPSS test is rejected [19, 20].

The non-stationarity in the time series is eliminated by differencing.

3.2. ARIMA model

The ARIMA models are based on regression models built on the observations themselves and on the residual component of the time series. The ARIMA(p, d, q) model is given as follows:

$$y'_t = c + \alpha_1 y'_{t-1} + \alpha_2 y'_{t-2} + \ldots + \alpha_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (1)$$

where y'_t is the differenced series, c is constant, $\alpha_1, \ldots, \alpha_p$ are the coefficients of AR(p) process, $y'_{t-1}, \ldots, y'_{t-p}$ are lagged values of the differenced series, $\theta_1, \ldots, \theta_q$ are coefficients of the MA(q) process and $\varepsilon_t, \ldots, \varepsilon_{t-q}$ are independent identically distributed error terms with zero mean.

The parameters of the ARIMA(p, d, q) model are as follows:
- p is the order of the autoregressive part (the AR process);
- d is the degree of differencing involved;
- q is the order of the moving average part (MA process) [21].

The degree of differencing depends on the stationarity/non-stationarity of the time series. A stationary time series has $d = 0$. The values of parameter $d > 2$ seldom occur. The order of the AR and the MA processes, respectively, can be estimated from the correlogram – a plot of autocorrelation coefficients (ACF), and from a plot of partial ACF (PACF). The estimates of the ACF are given as follows [20]:

$$r_k = \frac{\sum_{t=k+1}^{n} (y_t - \bar{y}).(y_{t-k} - \bar{y})}{\sum_{t=k+1}^{n} (y_t - \bar{y})^2}, k = 0, 1, \ldots, n-1 \quad (2)$$

where y_t are the observations, $\bar{y}$ is the average of the observations; and the estimates of the PACF are given as follows [20]:

$$\begin{aligned} & r_{11} = r_1, \\ & r_{kk} = \frac{r_k - \sum_{j=1}^{k-1} (r_{k-1,j} \cdot r_{k-j})}{1 - \sum_{j=1}^{k-1} (r_{k-1,j} \cdot r_j)}, k > 1, \\ & r_{k,j} = r_{k-1,j} - r_{kk} . r_{k-1,k-j}, j = 1, 2, \ldots k-1. \end{aligned} \quad (3)$$

The methodology for estimating the parameters of the ARIMA model from its ACF and PACF can be found in the literature [20].

DOI 10.2478/ama-2023-0004

However, such estimation of the parameter orders is subjective; therefore, it is more convenient to use it as a supporting information.

In this paper, several ARIMA models with different combinations of parameters are fitted to the time series. The coefficients of each model are found as maximum likelihood estimates. The best fitting model is chosen according to the smallest value of Akaike's information criterion (AIC) [21]

$$\text{AIC}= -2\log(L) + 2(p+q+k+1), \tag{4}$$

where $\log(L)$ denotes the maximised value of log likelihood function, p, q are the parameters of ARIMA model, $k=1$ if constant $c \neq 0$ and $k=0$ if $c=0$.

3.3. Verification of the ARIMA model

When the model is selected and the coefficients are estimated, we need to verify the model by checking whether the residuals, given as

$$e_t = y_t - \hat{y}_t, \tag{5}$$

are a white noise. Here $\hat{y}_t$ are the modelled values. A sequence of random variables ε_t is said to be a white noise under these conditions:

- the mean is zero, $E(\varepsilon_t)=0$;
- the variance is constant, $D(\varepsilon_t)=\sigma^2$;
- random variables are not correlated,

$$cov(\varepsilon_t, \varepsilon_{t-k}) = cov(\varepsilon_t, \varepsilon_{t+k}).$$

Furthermore, if the random variables ε_t are drawn from the standard normal distribution ($\varepsilon_t \sim N(0, \sigma^2)$), they are called Gaussian white noise.

The absence of correlation among the residuals is tested by the Ljung–Box Q test that tests the null hypothesis H_0: the residuals are not correlated, against the alternative that the residuals are correlated. When the H_0 is rejected, the considered model of the time series is not adequate and it needs to be changed.

Zero mean of residuals is tested using the t-test, when we test the hypothesis H_0: the data come from a normal distribution with a mean equal to zero and unknown variance, against the alternative hypothesis H_A: the population distribution does not have a mean equal to zero.

The homoscedasticity (constant variance) of residuals is tested by the two-sample F-test for equal variance. The normality of residuals can be tested by the Kolmogorov–Smirnov (KS) test or the Anderson–Darling (AD) test.

The performance of the model for fitting the data may be also considered by the following measures:

- the root mean square error (RMSE)

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2} \tag{6}$$

- the mean absolute percentage error (MAPE)

$$\text{MAPE}=\frac{1}{n}\cdot\sum_{t=1}^{n}\frac{|e_t|}{y_t}\cdot 100\% \tag{7}$$

- the mean percentage error (MPE)

$$\text{MPE}=\frac{1}{n}\sum_{t=1}^{n}\frac{e_t}{y_t} \tag{8}$$

Here n is sample size.

4. RESULTS

In this section, we summarise the results obtained when modelling the coal and the gas time series, respectively, in accordance with the procedure described in the Methodology section.

4.1. Coal time series

The visualisation of the time series is presented in Fig. 2. As we can see, the series is obviously decreasing with no fluctuations of a fixed frequency.

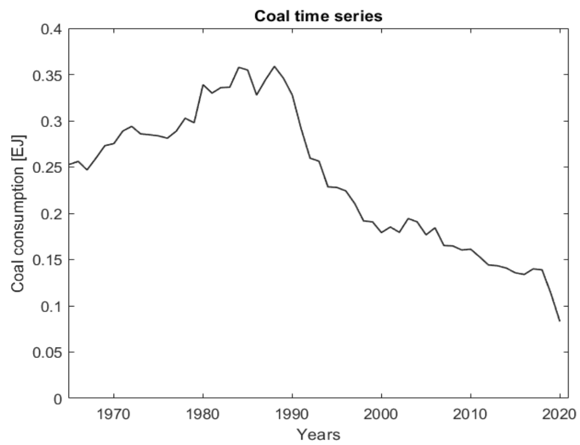

Fig. 2. The time series of coal consumption in Slovakia during the years 1965–2020

The presence of trend is indicated also by the correlogram of the time series (Fig. 3). The slow decrease of the ACF is caused by a strong correlation between the consecutive observations.

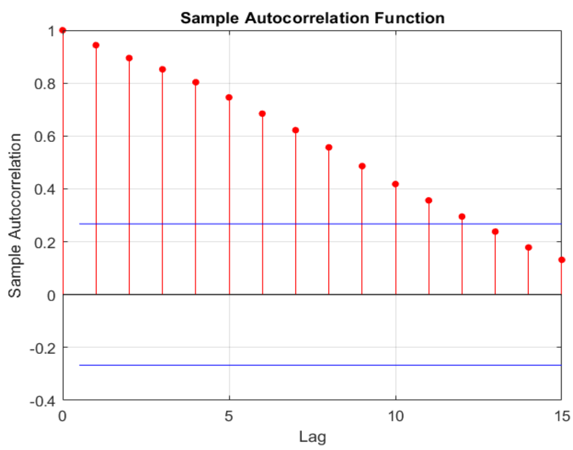

Fig. 3. The correlogram of the coal time series

We conduct the unit root tests on the significance level $\alpha=0.05$. The *p*-value of the ADF test and the KPSS test, respectively, are summarised in Tab. 2.

Tab. 2. Unit root tests for the coal time series

Original time series	**ADF test**	**KPSS test**
p-value	0.2845	0.0100

According to the *p*-value of the ADF test, the null hypothesis cannot be rejected on the significance level α=0.05; according to the *p*-value of the KPSS test, we reject the null hypothesis on the significance level α=0.05. The time series is non-stationary.

To eliminate the trend in the data, we replace the original series with the series of differences between the consecutive observations. This time series of first-order differences is also tested for stationarity. The *p*-value of the ADF test and the KPSS test, respectively, are presented in Tab. 3.

Tab. 3. Unit root tests for the first-order difference series (coal)

Series of first-order differences	**ADF test**	**KPSS test**
p-value	0.0159	0.0449

According to the *p*-value of the ADF test, we reject the null hypothesis on the significance level α=0.05; according to the *p*-value of the KPSS test, we also reject the null hypothesis on the significance level α=0.05. Because of the conflicting results of both tests for the series of differenced data, we cannot make any conclusions whether this series is stationary or not. Therefore, we calculate the differences of the consecutive observations of the differenced time series and test the stationarity of the time series of second-order differences. We summarise the *p*-value of the ADF test and the KPSS test, respectively, in Tab. 4.

Tab. 4. Unit root tests for the second-order difference series (coal)

Series of second-order differences	**ADF test**	**KPSS test**
p-value	0.0010	0.1000

On the significance level $\alpha = 0.05$, we reject the null hypothesis of the ADF test, while we do not reject the null hypothesis of the KPSS test. We may conclude that the series of second-order differences is stationary.

We fit the ARIMA models with parameters considered as follows:

$$d=\{1,2\};\; p=\{0,1,2\};\; q=\{0,1,2\}. \tag{9}$$

The values of parameter d are determined by results of the unit root tests; the values of the other two parameters are considered to not exceed 2 because higher values occur only seldom. The best fitting model is chosen according to the AIC value. The ARIMA models along with their AIC values are in Tab. 5.

Tab. 5. The fitted ARIMA models

ARIMA model	$\log(L)$	**AIC**
$(1,1,0)$	159.299	−314.597
$(0,1,1)$	159.011	−314.022
$(1,1,1)$	160.896	−315.792
$(2,1,0)$	160.158	−314.315
$(0,1,2)$	159.637	−313.274
$(1,1,2)$	160.937	−313.874
$(2,1,1)$	160.925	−313.849
$(2,1,2)$	162.585	−315.171
$(1,2,0)$	154.658	−305.316
$(0,2,1)$	160.246	−316.492
$(1,2,1)$	160.275	−314.550
$(2,2,0)$	157.471	−308.942
$(0,2,2)$	160.273	−314.546
$(1,2,2)$	160.275	−312.551
$(2,2,1)$	160.465	−312.929
$(2,2,2)$	162.163	−314.327

The smallest value of the AIC is achieved by the MA$(0,2,1)$ model, which means that the second-order differences of time series follow the MA(1) model in the form

$$y_t''=\varepsilon_t - 0.859\varepsilon_{t-1} \tag{10}$$

where y_t'' is a series of second-order differences, $\varepsilon_t, \varepsilon_{t-1}$ are the independent identically distributed error terms with zero mean. The model fitted to the time series is depicted in Fig. 4.

We verify the model by checking the residuals. The results of the tests are summarised in Tab. 6.

Tab. 6. The *p*-values of the tests for verification of the model

Ljung-Box Q test (*p*-value)	**t-test (*p*-value)**	**Two-sample F-test (*p*-value)**	**AD test (*p*-value)**
0.8534	0.4028	0.0626	0.2188

According to the *p*-values of all the tests, we may conclude that on the significance level $\alpha = 0.05$, the residuals of the model are not autocorrelated and are normally distributed with constant variance. Thus, the residual time series is a white noise.

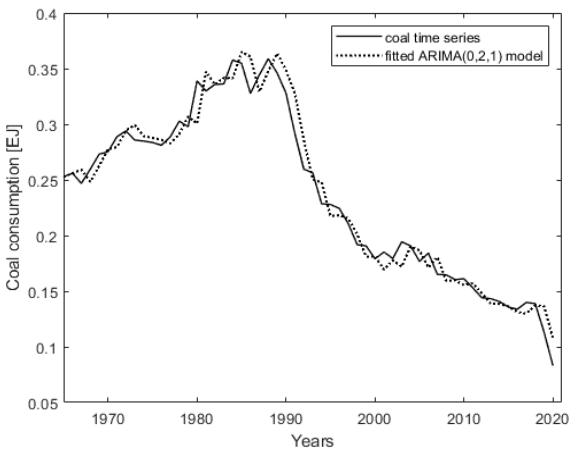

Fig. 4. Model ARIMA$(0,2,1)$ fitted to the coal time series

The performance of the model is assessed by the measures RMSE, MAPE and MPE, respectively; the results are given in Tab. 7. The measure MPE indicates that the majority of errors is negative, which means that the model systematically overestimates the reality. According to MAPE, the mean absolute percentage error between the consumption of coal predicted by the model and the actual consumption is 4.75%.

DOI 10.2478/ama-2023-0004

Tab. 7. Performance of the ARIMA(0,2,1) model for the coal time series

RMSE	MAPE	MPE
0.0138	4.7506	–0.9684

Based on the fitted model, we may forecast future coal consumption. The forecasts for years 2021–2030 are presented in Tab. 8 and visualised in Fig. 5.

Tab. 8. The forecast of coal consumption for years 2021–2030

Year	Point forecast	Lower 95% confidence level	Upper 95% confidence level
2021	0.072525	0.04462	0.10042
2022	0.06207	0.01896	0.10519
2023	0.05163	−0.00573	0.10900
2024	0.04119	−0.03039	0.11277
2025	0.03075	−0.05531	0.11681
2026	0.02031	−0.08063	0.12125
2027	0.00987	−0.10641	0.12614
2028	−0.00057	−0.13267	0.13152
2029	−0.01102	−0.15943	0.13740
2030	−0.02146	−0.18669	0.14378

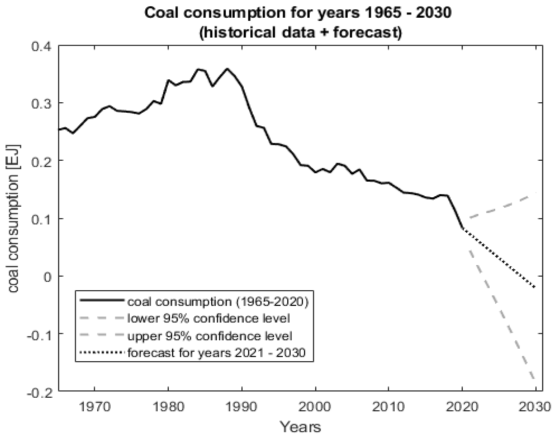

Fig. 5. Coal consumption in years 1965–2030 (actual values and forecast)

4.2. Gas time series

The visualisation of the gas time series is in Fig. 6. As we can see, the series is increasing with no obvious seasonality.

Similarly as in the coal time series, the presence of a trend is indicated also by the correlogram of the time series (Fig. 7) where the ACF only slowly decrease.

We conduct the unit root tests on the significance level $\alpha = 0.05$. The *p*-values of the ADF test and the KPSS test are presented in Tab. 9.

According to the *p*-value of the ADF test, the null hypothesis cannot be rejected on the significance level $\alpha = 0.05$; according to the *p*-value of the KPSS test, we reject the null hypothesis on the significance level $\alpha = 0.05$. This proves our assumption that the time series is non-stationary.

To eliminate the trend in the data, we transform the series by differencing. Then we test the stationarity of the first-order difference series and draw a conclusion from the *p*-values of the ADF test and the KPSS test (Tab. 10).

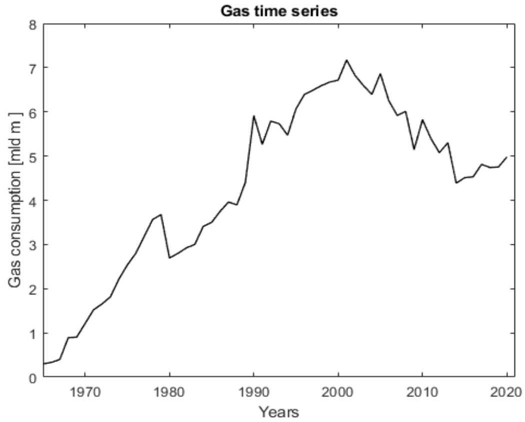

Fig. 6. The time series of gas consumption in Slovakia during the years 1965–2020

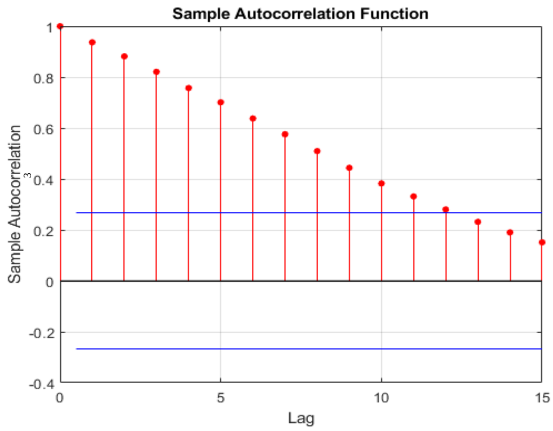

Fig. 7. The correlogram of the gas time series

Tab. 9. Unit root tests for the gas time series

Original time series	ADF test	KPSS test
p-value	0.8200	0.0100

Tab. 10. Unit root tests for the first-order difference series (gas)

First-order difference series	ADF test	KPSS test
p-value	0.0046	0.0327

Just as for the coal time series, according to the *p*-value of the ADF test, we reject the null hypothesis on the significance level $\alpha = 0.05$. According to the *p*-value of the KPSS test we also reject the null hypothesis on the significance level $\alpha = 0.05$. Because of the conflicting results of both tests for the first-order difference series we cannot make any conclusions whether the series is stationary or not. We replace this series with the series of the differences of its consecutive observations and test the stationarity of such second-order difference series. Tab. 11 presents the *p*-values of the ADF test and the KPSS test.

Tab. 11. Unit root tests for the second-order difference series (gas)

Second-order difference series	ADF test	KPSS test
p-value	0.0010	0.1000

On the significance level $\alpha = 0.05$ we reject the null hypothesis of the ADF test, while we do not reject the null hypothesis of the KPSS test. We may conclude that the second-order difference series is stationary.

We fit the ARIMA models with parameters considered as follows:

$$d=\{1,2\};\ p=\{0,1,2\};\ q=\{0,1,2\}. \tag{11}$$

Again, the values of parameter d are determined by results of the unit root tests. The best fitting model is chosen according to the smallest AIC value. The ARIMA models along with their AIC are in Tab. 12.

Tab. 12. The fitted ARIMA models

ARIMA model	$\log(L)$	AIC
(1,1,0)	−30.391	64.782
(0,1,1)	−30.495	64.990
(1,1,1)	−30.324	66.647
(2,1,0)	−29.997	65.995
(0,1,2)	−29.216	64.432
(1,1,2)	−27.788	63.577
(2,1,1)	−27.995	63.991
(2,1,2)	−27.788	65.576
(1,2,0)	−40.226	84.452
(0,2,1)	−30.115	64.229
(1,2,1)	−29.342	64.685
(2,2,0)	−31.455	68.910
(0,2,2)	−28.829	63.658
(1,2,2)	−28.708	65.415
(2,2,1)	−29.285	66.569
(2,2,2)	−25.210	60.419

The smallest value of the AIC is obtained by the MA(2,2,2) model, which means that the series of second-order differences follow the ARMA(2,2) model in the form

$$y''_t = -1.1375y''_{t-1} - 0.3978y''_{t-2} + \varepsilon_t - 0.8086\varepsilon_{t-2} \tag{12}$$

where y''_t is the second-order difference series, $\varepsilon_t, \varepsilon_{t-2}$ are the independent identically distributed error terms with zero mean. The model fitted to the time series is shown in Fig. 8.

We verify the model by checking the residuals. The results of the tests are summarised in Tab. 13. According to the *p*-values of the tests, we may conclude that on the significance level $\alpha = 0.05$, the residuals of the model are not autocorrelated, they have constant variance but do not come from the normal distribution $N(0, \sigma^2)$, that is, the residual time series is a white noise, not Gaussian white noise.

The performance of the model is assessed by the measures RMSE, MAPE and MPE; the results are given in Tab. 14.
The value of MPE indicates that the majority of errors is positive, which means that the model systematically underestimates the reality. According to the value of MAPE, the mean absolute percentage error between the consumption of gas predicted by the model and the actual consumption is 7.88%.

Based on the fitted model, we predict the annual gas consumption. The forecasts for years 2021–2030 are given in Tab. 15 and visualised in Fig. 9.

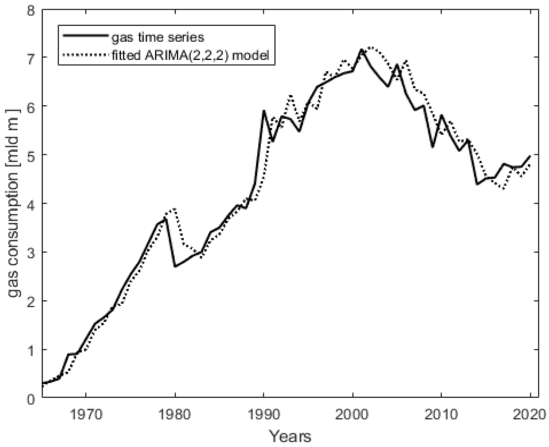

Fig. 8. The model ARIMA(2,2,2) fitted to the gas time series

Tab. 13. The *p*-values of tests for verification of the model

Ljung-Box Q test (*p*-value)	t-test (*p*-value)	Two sample F-test (*p*-value)	AD test (*p*-value)
0.8489	0.5287	0.4288	0.0258

Tab. 14. Performance of the ARIMA(0,2,1) model for the gas time series

RMSE	MAPE	MPE
0.3824	7.8804	0.3135

Tab. 15. The forecast of gas consumption for years 2021–2030

Year	Point forecast	Lower 95% confidence level	Upper 95% confidence level
2021	4.8126	4.0378	5.5875
2022	4.8620	3.7920	5.9320
2023	4.8091	3.5133	6.1048
2024	4.7933	3.1506	6.4360
2025	4.7736	2.8838	6.6634
2026	4.7439	2.5213	6.9665
2027	4.7277	2.2143	7.2411
2028	4.6993	1.8550	7.5436
2029	4.6801	1.5114	7.8489
2030	4.6547	1.1438	8.1656

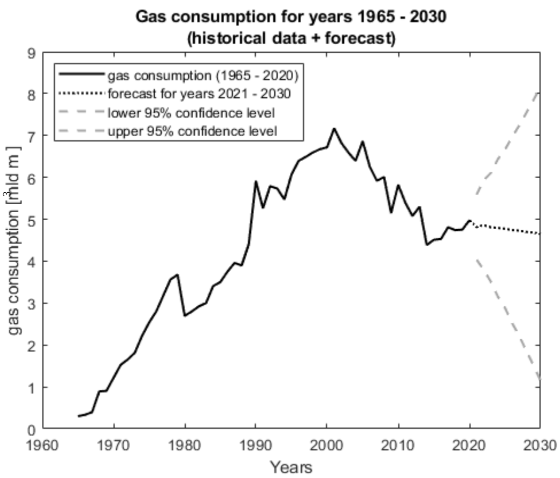

Fig. 9. Gas consumption in years 1965–2030 (actual values and forecast)

DOI 10.2478/ama-2023-0004

5. DISCUSSION AND CONCLUSIONS

In the paper, we modelled the time series of coal and gas consumption, respectively, in Slovakia during the years 1965–2020 by applying the ARIMA(p, d, q) models. Because of the trend in each time series, parameter $d > 0$. After fitting several models with various combinations of parameters p, d, q, we have chosen the ARIMA$(0,2,1)$ model for the coal consumption and the ARIMA$(2,2,2)$ model for the gas consumption, respectively, as they achieved the smallest values of AIC. The results of the Ljung–Box test verified that for each time series, the chosen ARIMA model explains all the autocorrelation in the series, therefore it is adequate for modelling the actual time series and can be used for predicting future values. The values of MAPE less than 10% (4.75% for coal and 7.88% for gas) indicate that the fitted ARIMA models provide reliable predictions.

Based on the constructed forecasts, we can formulate the following conclusions:

- coal consumption shall follow the decrease that has been observed in recent years. According to the forecast, close to zero coal consumption will be achieved between years 2027 and 2028. Such a scenario is in agreement with the obligations of Slovakia to finish the electricity production from coal by 2030 (as electricity production alongside households heating is one of main coal consumers);
- the gas consumption in the next decade exhibits a very mild decrease, almost a stagnation.

Although there is no other model that can be used for comparison, we can compare our predictions of the gas consumption with official predictions reported by the Ministry of Economy of Slovak Republic. The ministry issues annually a Report on the results of gas supply security monitoring (Správa o výsledkoch monitorovania bezpečnosti dodávok plynu), where it is summarised the consumption, the production and the import during the year. In addition, the ministry provides the prognosis of development in the consumption and the production for the following period, including the predictions for the next 5 years. The predictions for years 2021–2025 are presented in Tab. 16. The methodology for obtaining the forecasts declared by the ministry is not available.

Tab. 16. Predictions of gas consumption for years 2021–2025

Year	Gas consumption [10^9 m^3]	
	Predictions from the Ministry of Economy [22]	Point forecasts from the ARIMA(2,2,2) model
2021	5.1	4.8126
2022	5.0	4.8620
2023	5.0	4.8091
2024	5.0	4.7933
2025	5.0	4.7736

According to the report [22], the Ministry anticipates the stagnation of consumption on the level of approx. 5.0 [10^9 m^3] in the upcoming years. In our predictions, we observe that after the increase in previous years, the consumption should start to slowly decrease. This predicted decrease of consumption can be caused by the situation in 2020. The COVID-19 pandemic brought restrictions to our everyday lives, influencing everything, including the industrial sector as the main gas consumer in Slovakia. To avoid the influence of unpredictable changes in 2020, Wang et al. [23, 24] suggested to consider the consumption from a COVID-free scenario simulation instead of the real consumption in 2020. It is of further research to estimate the influence of the pandemic on the future gas consumption, comparing the predictions for the original time series and the time series adjusted by the simulation.

To sum it up, the target to build a prediction model for the coal and the gas consumption, respectively, in Slovakia has been achieved. Both ARIMA models provide very good fit to the time series. Taking these results as reference models, we can carry on with the research in the future by applying other approaches, such as artificial neural networks, grey models, and by building hybrid models in order to improve the fit of the model to the data, and to obtain more precise predictions.

Making responsible energy policy requires trustworthy predictions. Therefore, we hope that the proposed models will be of help to the authorities when preparing future strategies.

REFERENCES

1. Slovak Ministry of Economy. Integrated National Energy and Climate Plan for 2021 to 2030. [Internet] 2019 December [cited 2022 May 31] Available from: https://energy.ec.europa.eu/system/files/202003/sk_final_necp_main_en_0.pdf
2. Brożyna J, Strielkowski W, Fomina A, Nikitina N. Renewable Energy and EU 2020 Target for Energy Efficiency in the Czech Republic and Slovakia. Energies. 2020; 13(4): 965.
3. OKTE, a.s. National Energy Mix. [Internet] 2022 May [cited 2022 May 31] Available from: https://www.okte.sk/en/guarantees-of-origin/statistics/national-energy-mix/
4. IEA. Gas 2020. [Internet] 2020 June [cited 2022 May 31]. Available from: https://iea.blob.core.windows.net/assets/555b268e-5dff-4471-ac1d-9d6bfc71a9dd/Gas_2020.pdf
5. Jandačka J, Holubčík M, Trnka J. Utilization of solid fuels with regard to the transport distances of the raw material. TRANSCOM 2021, Transp Res Proc 2021; 55: 829-836.
6. Nandimandalam H, Gude VG, Marufuzzaman M. Enviromental impact assessment of biomass supported electricity generation for sustainable rural energy systems – A case study of Grenada County, Mississippi, USA. Sci Total Environ. 2022; 802: 149713.
7. Dritsaki C, Niklis D, Stamatiou P. Oil Consumption Forecasting using ARIMA Models: An Empirical Study for Greece. Int J Energy Econ Policy. 2021; 11(4):214-224.
8. Ozturk, S, Ozturk F. Forecasting Energy Consumption of Turkey by ARIMA Model. J Asian Sci Res. 2018; 8(2): 52-60.
9. Akpinar M, Yumusak N. Year ahead demand forecast of city natural gas using seasonal time series methods. Energies. 2016; 9: 727.
10. Chaturvedi S, Rajasekar E, Natarajan S, McCullen N. A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for India. Energy Policy. 2022; 168: 113097.
11. Manowska A, Rybak A, Dylong A, Pielot J. Forecasting of Natural Gas Consumption in Poland Based on ARIMA-LSTM Hybrid Model. Energies. 2021; 14(24): 8597.
12. Wang X. Research on the prediction of per capita coal consumption based on the ARIMA-BP combined model. Energy Rep. 2022: 8(4): 285-294.
13. Wang Q, Li S, Li R. Forecasting energy demand in China and India: Using single-linear, hybrid-linear, and non-linear time series forecast techniques. Energy. 2018; 161: 821-831.
14. Ma M, Wang Z. Prediction of the energy consumption variation trend in South Africa based on ARIMA, NGM and NGM-ARIMA models. Energies. 2020; 13(1): 10.
15. Pavlicko M, Vojteková M, Blažeková O. Forecasting of electrical energy consumption in Slovakia. Mathematics. 2022; 10: 577.

16. Brabec M, Konár O, Pelikán E, Malý M. A nonlinear mixed effects model for the prediction of natural gas consumption by individual customers. Int J Forecast. 2008; 24: 659-678.
17. Hošovský A, Piteľ J, Adámek M, Mižáková J, Židek K. Comparative study of week-ahead forecasting of daily gas consumption in buildings using regression ARMA/SARMA and genetic-algorithm-optimized regression wavelet neural network models. J Build Eng. 2021; 34: 101955.
18. BP p.l.c. Statistical Review of World Energy – all data, 1965-2020. [Internet] 2021 July [cited 2022 May 29]. Available from: https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/xlsx/energy-economics/statistical-review/bp-stats-review-2021-all-data.xlsx
19. Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y. Testing the null hypothesis of stationarity against the alternative of unit root: How sure are we that economic time series have a unit root? J Econom. 1992; 54(1-3): 159-178.
20. Cipra T. Time Series in Economics and Finance. Cham: Springer Nature Switzerland; 2020. 410p.
21. Hyndman RJ, Athanasopoulos G. Forecasting: Principles and Practice. 2nd ed. Heathmont: OTexts; 2018. 382p.
22. Slovak Ministry of Economy. Report on the results of gas supply security monitoring. [Internet] 2021 July [cited 2022 October 21] Available from: https://www.mhsr.sk/uploads/files/WdL723Kw.pdf?csrt=6552196416131516580
23. Wang Q, Li S, Jiang F. Uncovering the impact of the COVID-19 pandemic on energy consumption: New insight from difference between pandemic-free scenario and actual electricity consumption in China. J Clean Prod. 2021; 313: 127897.
24. Wang Q, Li S, Zhang M, Li R. Impact of COVID-19 pandemic on oil consumption in the United States: A new estimation approach. Energy. 2022; 239: 122280.

This work has been supported by the Slovak Grant Agency KEGA through the project No. 027ŽU-4/2020 Innovation of learning text and implementation of new didactical means to improve the quality of Mathematics II at the university of the technical type and through the project No. 029ŽU-4/2022 Implementation of the principles of blended learning into the teaching of the subject Numerical Methods and Statistics.

Mária Michalková: https://orcid.org/0000-0001-7488-8514

Ivana Pobočíková: https://orcid.org/0000-0003-4357-260X