



Tomasz CHRULSKI<sup>1</sup>

## Using the R language computing environment in forecasting natural gas consumption

**ABSTRACT:** There are many IT tools available on the market that carry out various types of forecasts in the gas industry. Programming evolves with the availability and capability of computers. IT tools support the user in engineering calculations, but also present the obtained results in an interesting visualization, e.g. in the form of interactive charts. The software can support making business decisions, which, in turn, can be used as business intelligence.

In the era of digitization, huge metadata of measurements are created, so conducting data analyzes in the energy sector is very common. Moreover, rapidly evolving artificial intelligence creates new opportunities.

The article presents a sample analysis of calculations using RStudio, an integrated development environment for the R language, a programming language for statistical calculations and graphics. The aim of the article is to present the possibility of using R language software to make a forecast and to determine the quality of forecasts. The article aims to present the possibility of making forecasts based on mathematical models available in R packages and the possibilities offered by the forecasting platform to readers.

The article presents the U.S. market and has a particular focus on Natural Gas Residential Consumption in Pennsylvania (publicly available data from the U.S. Energy Information Administra-

---

✉ Corresponding Author: Tomasz Chrulski; e-mail: [chrulski@agh.edu.pl](mailto:chrulski@agh.edu.pl)

<sup>1</sup> Faculty of Drilling, Oil and Gas, University of Science and Technology AGH, Kraków, Poland; ORCID iD: 0000-0002-8842-518X; e-mail: [chrulski@agh.edu.pl](mailto:chrulski@agh.edu.pl)



© 2021. The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-ShareAlike International License (CC BY-SA 4.0, <http://creativecommons.org/licenses/by-sa/4.0/>), which permits use, distribution, and reproduction in any medium, provided that the Article is properly cited.

tion). This dataset represents the monthly consumption of natural gas between 2015 and 2020. Forecasts were presented over a span of 12 months.

KEYWORDS: forecasting, gas consumption, RStudio

## Introduction

In the energy mix, natural gas plays a key role. Gas consumption is increasing worldwide in the energy market but it should be noted that as a result of the spread of the coronavirus pandemic, the demand for energy, including natural gas, has decreased (Strielkowski et al. 2021). However, the forecasts made by the U.S. Energy Information Administration show a positive trend for the coming years.

In United State of America natural gas can be used by not only large industrial consumers to product fabrics, steel, glass, paint, plastics and other products but also by households. Natural gas is used for the following domestic purposes: cooking, heating water, cooling, clothes drying, home heating (EIA 2020).

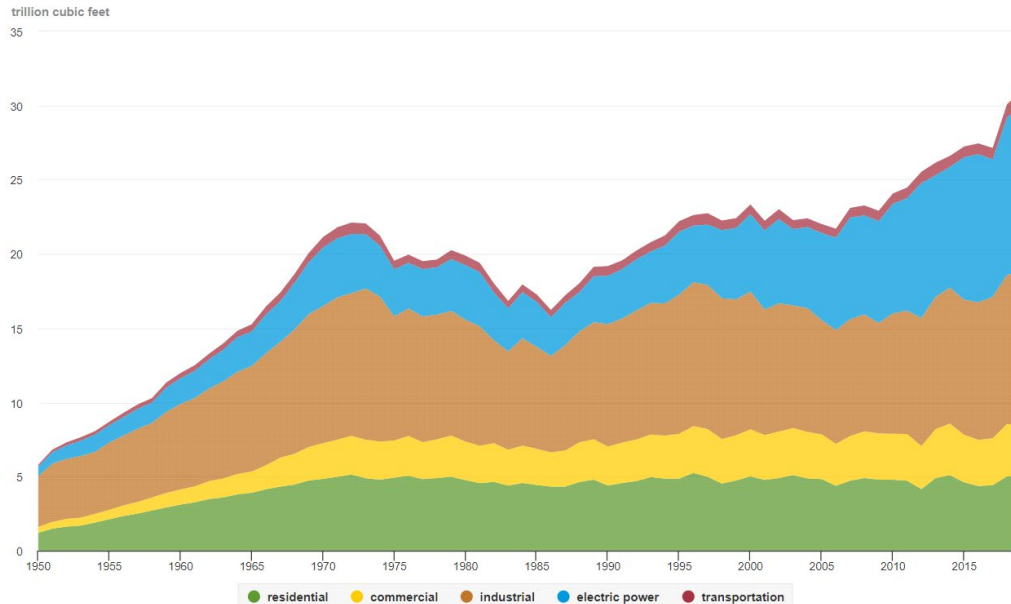


Fig. 1. U.S. Natural Gas consumption by sector, 1950–2020

Source: own study based on the U.S. Energy Information Administration (EIA)

Rys. 1. Zużycie gazu ziemnego w USA według sektorów, 1950–2020

The industrial sector is the one that consumes the largest amount of natural gas. This sector uses natural gas as a fuel for process heating and as a raw material for the production of chemicals, fertilizers and hydrogen. In 2020, the industrial sector accounted for approximately 33% of total U.S. natural gas consumption.

The next important sector that uses natural gas is the electricity sector, mainly to generate energy. In 2020, the electricity sector accounted for about 30% of the total consumption of natural gas in the U.S.

The residential sector is also an important sector that uses natural gas. Approximately 50% of the U.S. population uses this kind of fuel and in 2020, the residential sector accounted for approximately 17% of total U.S. natural gas consumption.

The next sector is commercial sector which uses natural gas to heat buildings and water, to operate refrigeration equipment, to cook, dry clothes and to light streets. In 2020, the commercial sector accounted for approximately 11% of total U.S. natural gas consumption.

The last transport sector uses natural gas as a fuel for vehicles. In 2020, this sector consumed about 4% of total U.S. natural gas consumption.

## 1. Dependence of gas consumption on external factors

The consumption of natural gas depends on weather conditions, natural gas prices, contracts, economic growth and the political situation. These are direct external factors affecting the consumption of natural gas. There are also random factors that also affect its consumption (Fig. 5). One of these external factors is temperature. It has been proven that there is a strong dependence of temperature on natural gas consumption (Timmer and Lamb 2007). Statistics shows that in the state of Pennsylvania, the summers are warm and humid, the winters are very cold, and it is wet and partly cloudy year round. Over the course of the year, the temperature typically varies from 26°F to 87°F and is rarely below 13°F or above 94°F.

The state of Pennsylvania is an important producer of natural gas. This state is famous mainly for Marcellus shale. This state is the second largest producer of natural gas in the country.

About half of Pennsylvania households use natural gas as their primary fuel to heat their homes. There are 49 underground gas storage facilities in the state. Due to the easy access to shale gas, gas storage infrastructure and an inexpensive energy source, the main factor of gas consumption by customers is the weather (Fig. 3). Pearson's Correlation Coefficient is  $-0.96$ .

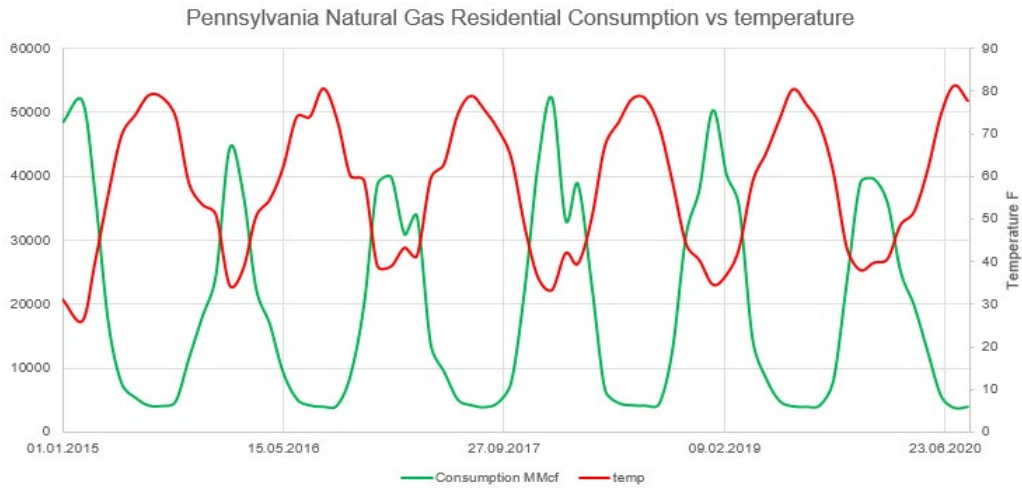


Fig. 2. Pennsylvania Natural Gas Residential Consumption vs temperature  
 Source: own study based on the U.S. Energy Administration

Rys. 2. Zużycie gazu ziemnego mieszkańców Pensylwanii do średniej temperatury

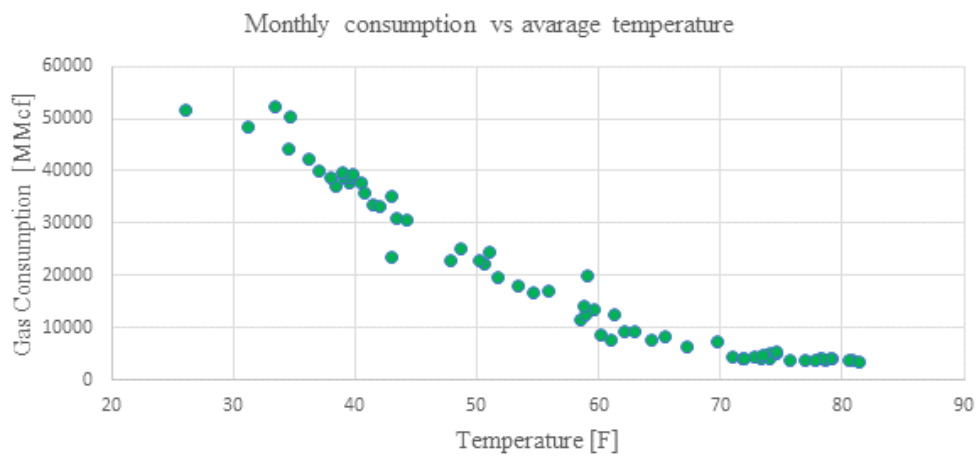


Fig. 3. Monthly consumption vs average temperature in Pennsylvania  
 Source: own study based on the U.S. Energy Administration

Rys. 3. Miesięczne zużycie gazu ziemnego do średniej temperatury w stanie Pensylwania

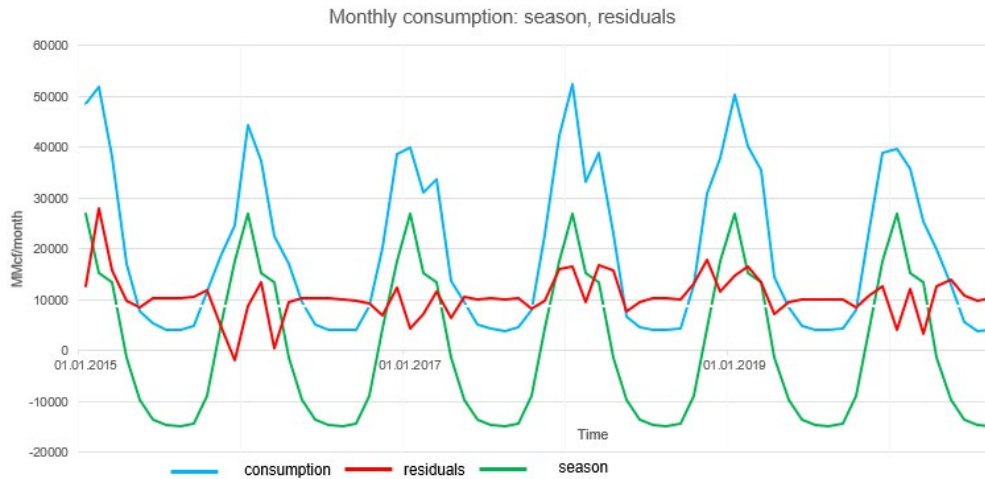


Fig. 4. Monthly natural gas consumption for the state of Pennsylvania in 2015–2020, broken down by season, residuals  
Source: own study based on the U.S. Energy Administration

Rys. 4. Miesięczne zużycie gazu ziemnego dla stanu Pensylwania w latach 2015–2020, z podziałem na sezonowość i czynniki losowe

## 2. Literature review

Many IT tools are available to support the user in engineering calculations in the mining, energy and renewable energy sectors. The software provides ready-made services that are the basis for making business decisions in enterprises. Forecasting brings targeted benefits for the company (Paliński 2019).

Forecasts of electricity consumption appear most often in the literature. Forecasting the demand for electricity is very important from the point of view of improving the efficiency and reliability of the power grid. In particular, short-term forecasting in this sector is the most important. One of the interesting article (Truong 2018) analysis on Electrical Demand in Australia. The author shows a variety of forecasts using RStudio and draws qualitative conclusions from forecasts.

In the mining industry, these may include forecasting the production volume, coal sales volume, coal market development and others (Manowska 2013). (Makridakis 2020) in an interesting way shows forecasting: Methods and Applications annual bituminous coal production in the US: 1920–1968 using self-made fma packages and bicoal functions.

A new trend in the literature is the publication of forecasts for the consumption of energy produced from renewable energy sources. Solar engineering has increased tremendously in the

past decade. The authors pay attention to the quality of the data and the right choice of external factors. Important factors to be considered as opposed to natural gas are wind power density, mean hourly wind speed, wind direction, solar power, solar generation, hourly solar irradiance. (Yang 2018) created a package named SolarData. This package was created for easy access to the publicly available solar datasets. The authors mainly used machine learning methods and neural networks.

In general, there are several good reviews on forecasting gas consumption using RStudio. An interesting example is the GAS package. Ardia et al. (2019) generalized the autoregressive score models. The GAS package allows for forecasting using univariate and multivariate descriptive statistics.

The most recent forecasting package available in the RStudio environment is the USgas package (Rami 2020). The package provides an overview of natural gas demand in the U.S. in a time-series format while the TSstudio package provides a set of descriptive tools and a predictive analysis of time series data. Another a new package is “fuel economy” (EPA 2020). Fuel economy data from the EPA, 1985–2015, conveniently packaged for consumption by RStudio users. The data frame containing Gasoline: Gasoline Consumption OECD, a panel of 18 observations from 1960 to 1978 is also an interesting issue in RStudio. The authors make predictions using Linear Models for Panel Data (Baltagi 2013).

An entertaining use of RStudio in forecasting natural gas consumption is the Smart City area (Makssod and Achuthan 2017). Authors prove the effectiveness of smart cities by using Data Mining Techniques (DMT) to predict energy consumption in Oman.

RStudio can also be used for forecasting natural gas with various mathematical models (Zhu et al. 2014) just like modern artificial intelligence technology that offers advantages in energy demand prediction (RBF neural network, BVAR model) and traditional forecasting methods.

In the available scientific articles, forecasting methods are divided into: classical methods and neural methods. In this article we presented forecasts with selected basic classical forecasting methods and one of the newer TBATS methods. Each RStudio package has its own package documentation. Therefore, only the functions that have been used in article will be recalled. However, a detailed description of the methodology is available on the RStudio platform.

### 3. Forecasting

Forecasting gas consumption is an inference about the development of the analyzed phenomenon. In the forecasting process, it is important to keep the current knowledge of the tools available on the market, one of such tools is RStudio. The further part of the article presents the possibilities of using the RStudio computing environment in the aspect of forecasting natural gas consumption. In the article we presented the implementation of the following time series forecasting techniques using the RStudio standard statistical programming language: Simple moving

average method, Naive method, Random walk with drift, Simple Exponential Smoothing method, ARIMA, TBATS,

## 4. Different methods for forecasting Time Series tutorial

Before building a forecasting model it is necessary to review some of the basics of forecasting theories. This should help focus attention on the most relevant issues and avoid some of the traps in forecasting (Bergheim 2008). To review the available forecasting tools in RStudio to predict gas consumption, the following packages should be installed: “tidyverse”, “forecast”, “TSstudio”, “reader” and it is possible to use “ggplot2” statistical graphing tools.

Occasionally, the downloaded telemetry gas flow may contain errors. This may be caused by the suspension of the measurement, error in the data. The user must analyze the gas consumption line and correct any data anomalies. This operation can be done with the “dplyr” package. This library will help clean and process the data.

With data previously cleared of inaccuracies, the user is obligated to use exploratory data. When the user prepares to forecast gas consumption, they must classify and group the data that will be used for the model. The simple forecasting model consists of: time, the amount of transmitted natural gas and possibly other additional parameters that could affect gas consumption. Following these steps, a model (Fig. 5) that shows the gas consumption of Pennsylvania residents in 2015–2020 (it indicates a time series with seasonality depending on the season of the calendar year) can be built.

## 5. Simple moving average

In this method, the arithmetic mean of the earlier values (Hayes 2021) of the time series is used as the forecast. In the R package, this function is available as a mean.

```
meanf(x,h = 12, level = c(80,95), fan = FALSE, lambda = NULL)
```

Parameters: x – time series, h – horizon, level – confidence interval, fan – TRUE/FALSE draw a fan chart, lambda – Box-Cox transformation.

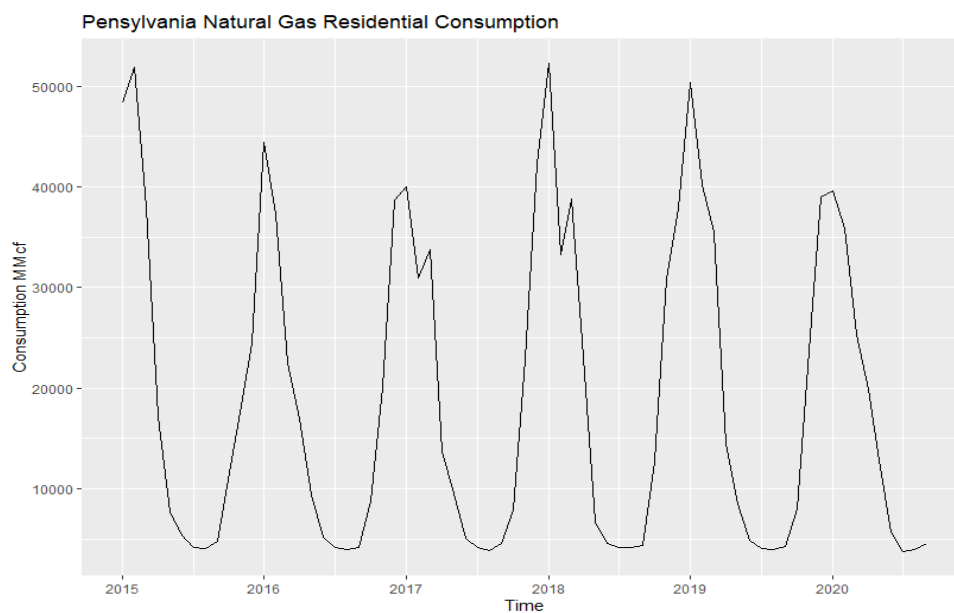


Fig. 5. Pennsylvania natural gas consumption forecast for 2015–2020  
Source: own study based on the U.S. Energy Administration

Rys. 5. Prognoza zużycia gazu ziemnego w stanie Pensylwania w latach 2015–2020

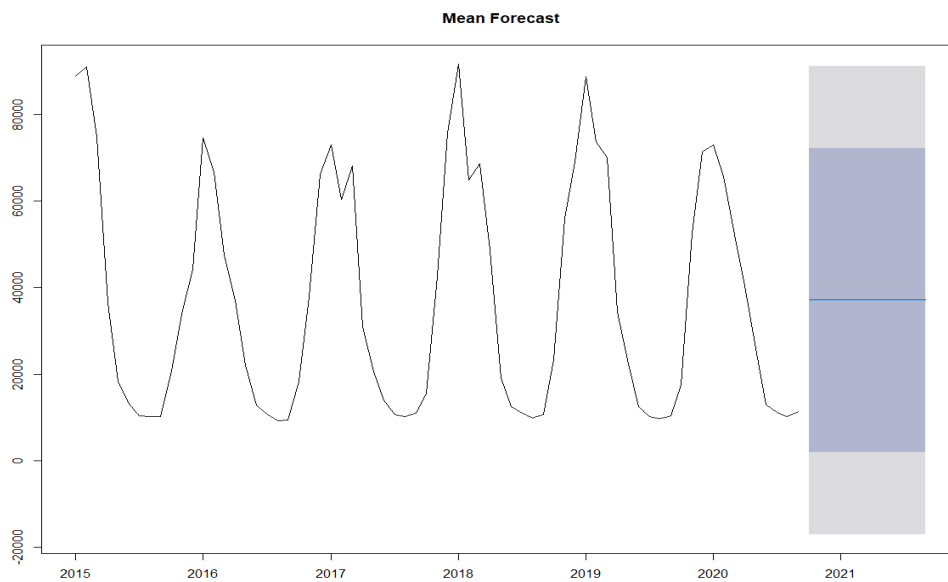


Fig. 6. Forecast using the Simple moving average method of natural gas consumption in Pennsylvania in 2015–2020  
Source: own study based on the U.S. Energy Administration

Rys. 6. Prognoza, oparta na średniej ruchomej, zużycia gazu ziemnego w stanie Pensylwania w latach 2015–2020



## 6. Naive methods

The main assumption of this method is the fact that consumption in the future period will be the same as in the preceding period. Forecasts using naive methods, such as: the level unchanged, the growth unchanged, the percentage increase unchanged.

There are packages available in RStudio based on naive methods. These are the `naive` and `snaive` functions from the `forecast` package.

```
naive(x,h = 12, level = c(80,95), fan = FALSE, lambda = NULL)
```

```
snaive(x,h = 12, level = c(80,95), fan = FALSE, lambda = NULL)
```

Parameters: `x` – time series, `h` – horizon, `level` – confidence interval, `fan` – TRUE/FALSE draw a fan chart, `lambda` – Box-Cox transformation.

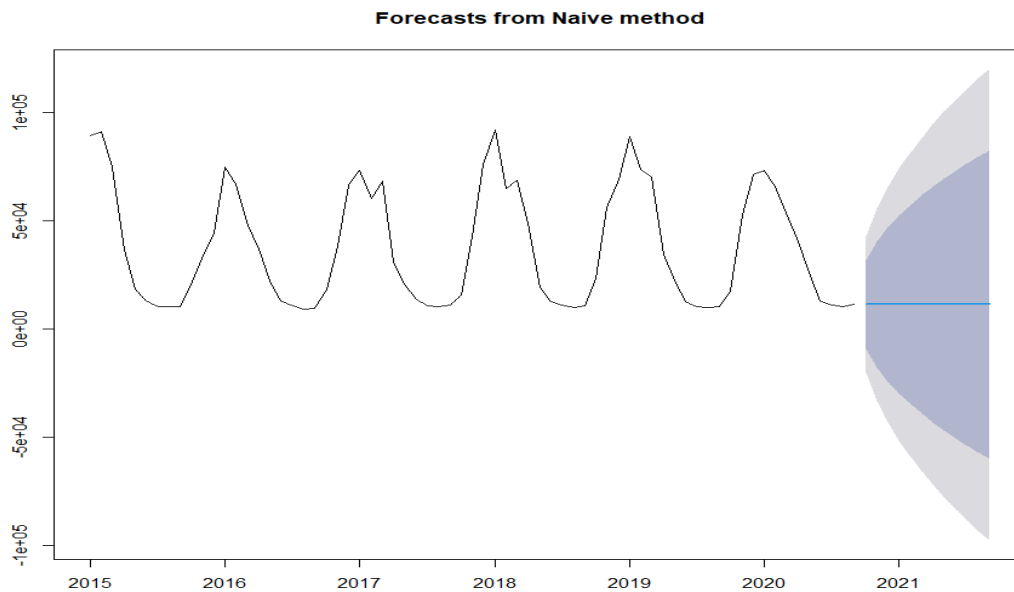


Fig. 7. Forecast using the naive method of natural gas consumption in Pennsylvania in 2015–2020  
Source: own study based on the U.S. Energy Administration

Rys. 7. Prognoza, oparta na metodzie naiwnej, zużycia gazu ziemnego w stanie Pensylwania w latach 2015–2020

## 7. Random walk with draft

A random walk with drift is defined as a process where the current value of a variable is composed of the past value plus an error term defined as white noise (a normal variable with zero mean and variance one).

```
rwf(x,h = 12, drift = FALSE, level = c(80,95), fan = FALSE, lambda = NULL, tidy = FALSE
```

Parameters: x – time series, h – horizon, drift – TRUE/FALSE use random walk model with drift, level – confidence interval, fan – TRUE/FALSE draw a fan chart, lambda – Box-Cox transformation.

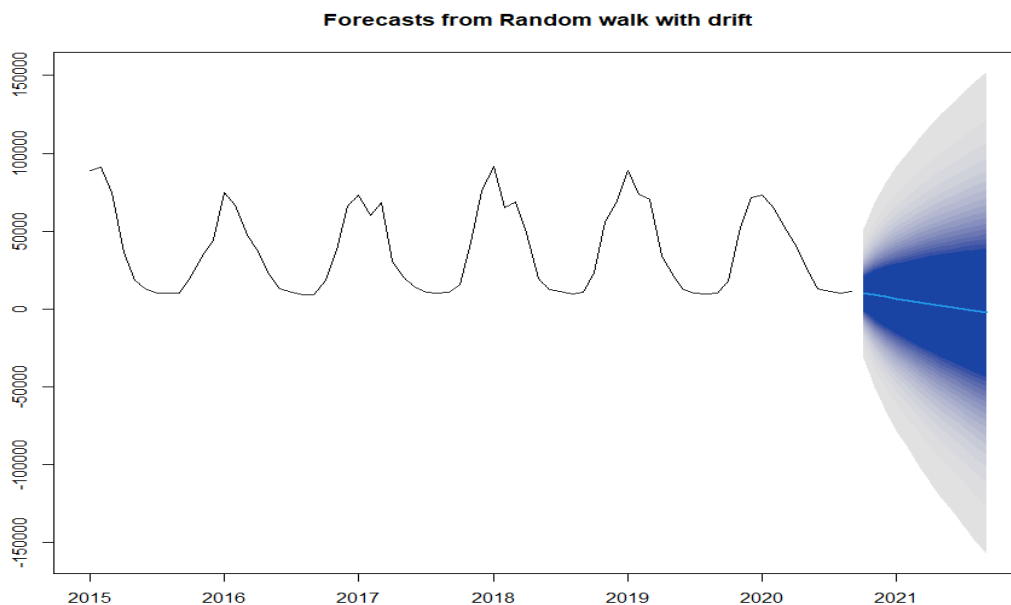


Fig. 8. Forecast using the random walk with draft method of natural gas consumption in Pennsylvania in 2015–2020  
Source: own study based on the U.S. Energy Administration

Rys. 8. Prognoza, oparta na metodzie uwzględniającej dryf, zużycia gazu ziemnego w stanie Pensylwania w latach 2015–2020

For the above forecasts, their accuracy was determined using the accuracy () function in RStudio.

Criteria based on prediction errors: ME (mean error), RMSE (root-mean-square error), MAE (mean absolute error), MPE (mean percentage error), MAPE (mean absolute percentage error),

MASE (mean absolute scaled error), ACF (estimates of the autocovariance or autocorrelation function).

TABLE 1. Compare forecast accuracy

TABELA 1. Porównanie dokładności prognoz

Training set	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
meanf()	-2.320746e-12	26751.63	23865.26	-81.3097	114.9119	4.140273	0.7892816
naive()	-1143.559	16018.08	11921.88	-12.87163	36.40534	2.068272	0.4693871
rwf()	8.428781e-13	15977.21	11825.72	-7.218666	35.30887	2.051589	0.4693871
rwf(), lambda=0	-20.61905	15727.65	11657.91	-9.48927	34.90955	2.022477	0.4680642

Source: based on the RStudio.

In order to effectively compare the results of different models, the time series should be divided into learning and testing parts. The window () package was used for this purpose.

In the following, the criteria for the prediction errors of the test set were determined.

TABLE 2. Compare the accuracy of the predictions for the training data

TABELA 2. Porównanie dokładności prognoz dla danych uczących

Training Set (test)	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
meanf()	-2.320746e-12	26751.63	23865.26	-81.3097	114.9119	4.140273	0.7892816
naive()	-1143.559	16018.08	11921.88	-12.87163	36.40534	2.068272	0.4693871
rwf()	8.428781e-13	15977.21	11825.72	-7.218666	35.30887	2.051589	0.4693871
rwf(), lambda=0	-20.61905	15727.65	11657.91	-9.489279	34.90955	2.022477	0.4680642

Source: based on the RStudio.

## 8. ARIMA

ARIMA, i.e. the autoregressive integrated moving average model is a generalization of the autoregressive moving average model. In RStudio the package is named as arima(). The most important parameters that define ARIMA in the RStudio package:

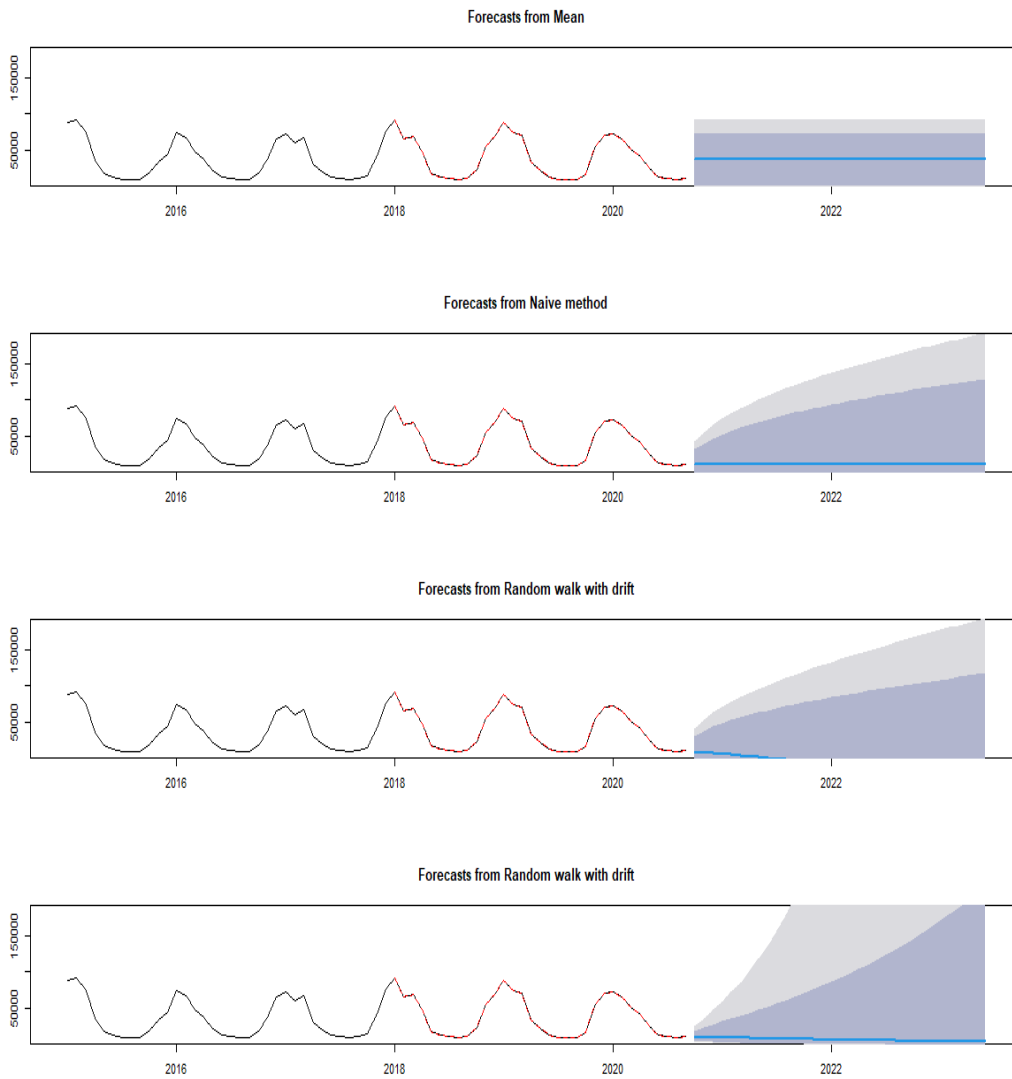


Fig. 9. Compare forecasts for the test set  
Source: own study based on the R

Rys. 9. Porównanie prognoz dla zbioru testowego

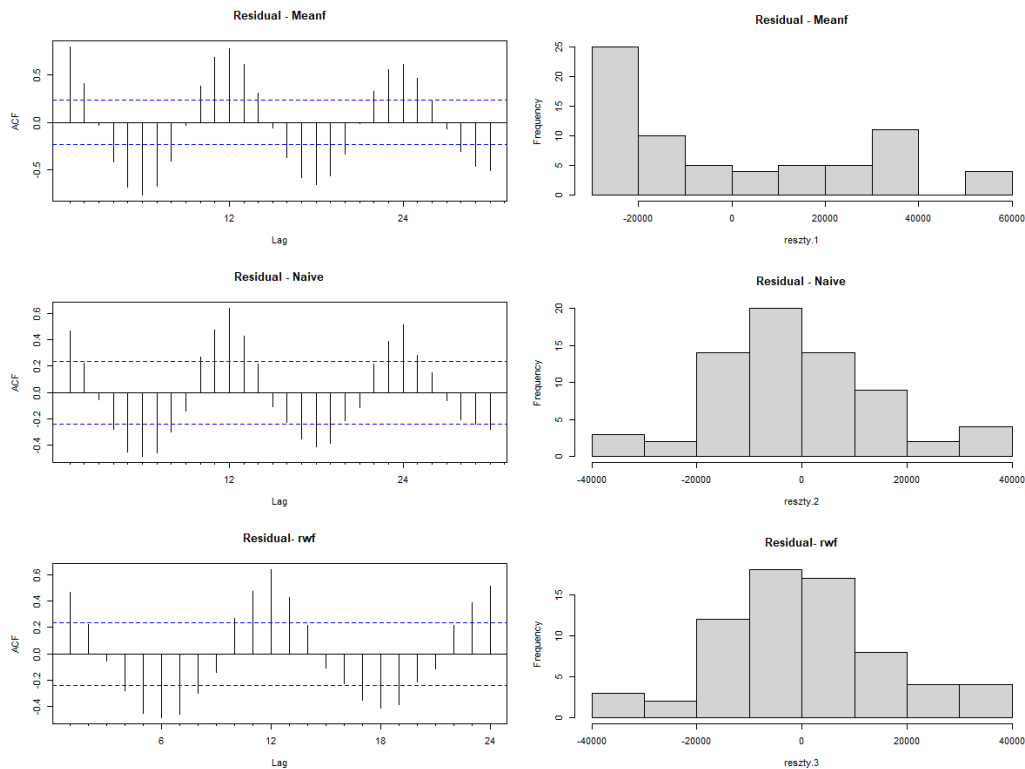


Fig. 10. Analysis of the properties of residuals  
Source: own study based on RStudio

Rys. 10. Analiza własności reszt

`Arima(x, order = c(0,0,0), seasonal = list(order = c(0,0,0), period = NA), include.mean = TRUE, include.drift = FALSE, include.constant = TRUE, lambda = NULL, ...)`

Parameters:  $x$  – time series,  $order = c(0,0,0)$  – define the ARIMA model  $(p, d, q)$ ,  $include.mean$  – forms the inclusion of the mean or not,  $include.drift$  – determines whether the drift should be taken into account,  $include.constant$  – whether a constant value is to be included in the model,  $lambda$  – takes into account a possible Box-Cox transformation.

The ACF and PACF plots indicate that the correct model for the data may be MA (12) or AR (13). Therefore, the models will be adjusted and compared (ARIMA (13,1,0) (0,1,0), ARIMA (0,1,12) (0,1,0), ARIMA (1,1,1) (0,1,0).

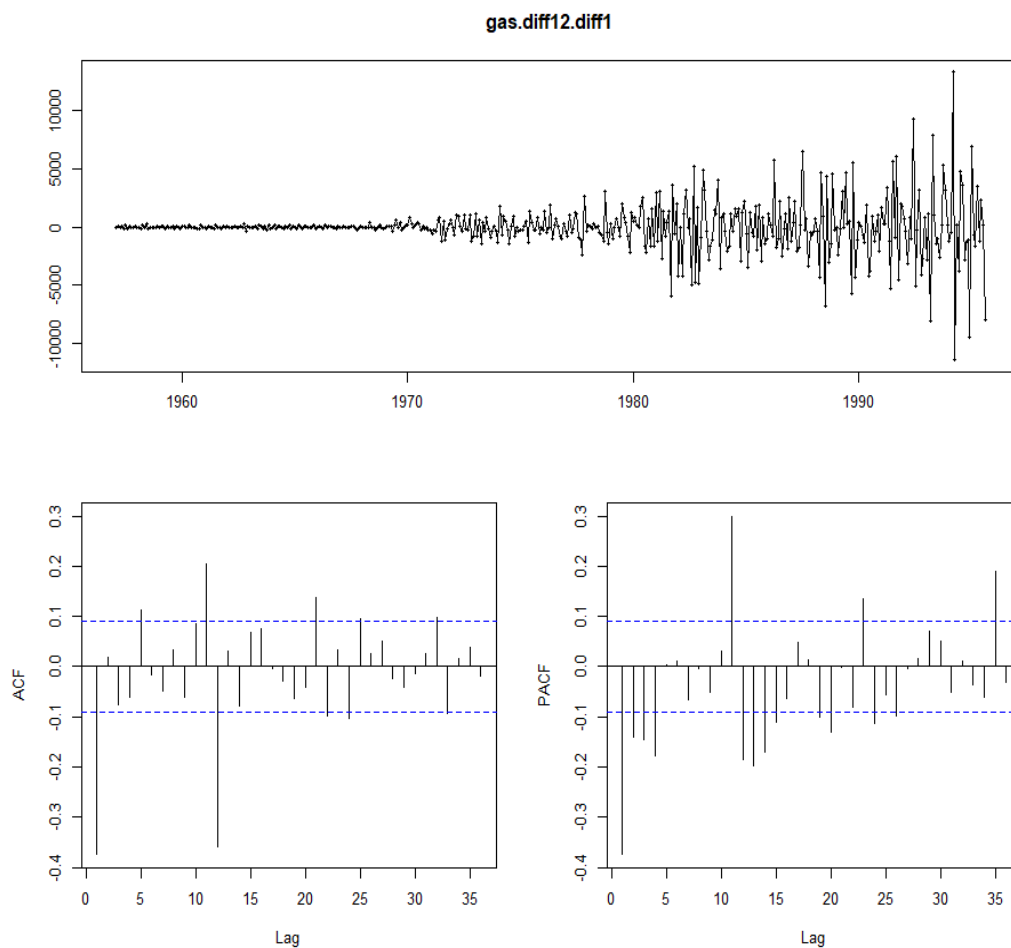


Fig. 11. ACF and PACF graph after applying the differentiation with a delay of 12 and 1  
 Source: own study based on the U.S. Energy Administration

Rys. 11. Wykres ACF i PACF po zastosowaniu różnicowania z opóźnieniem 12 oraz 1

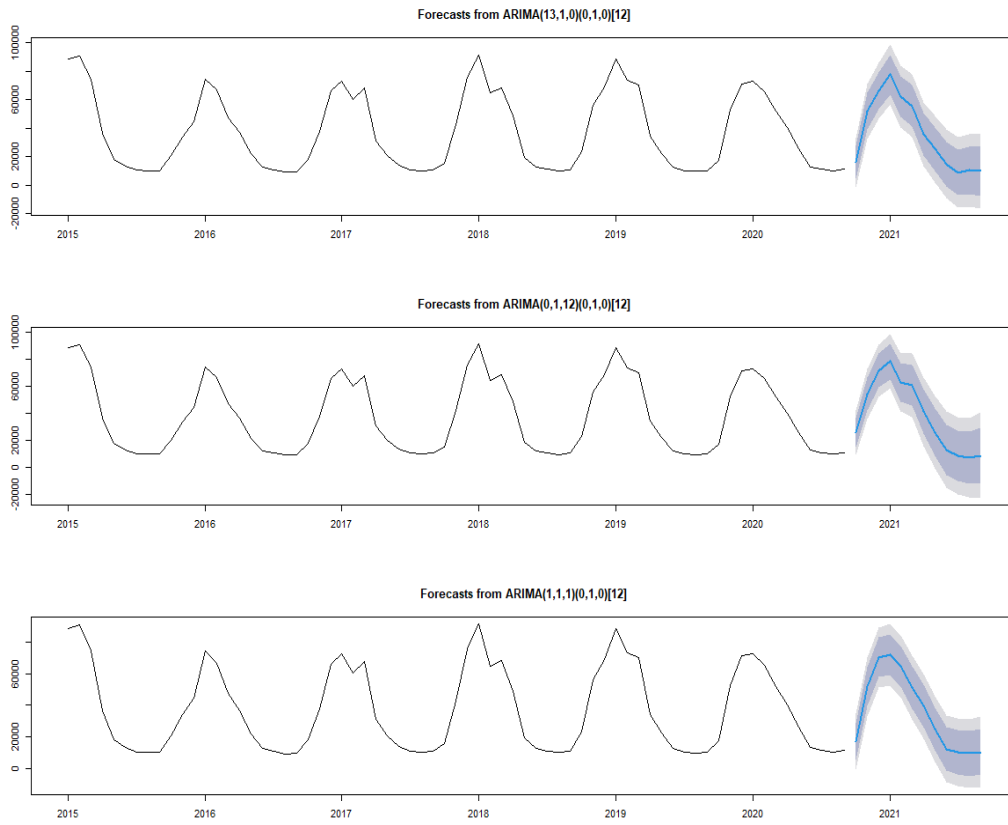


Fig. 12. Forecasts using the ARIMA method of natural gas consumption in Pennsylvania in 2015–2020  
Source: own study based on the U.S. Energy Administration

Rys. 12. Prognozy, oparte na modelu ARIMA, zużycia gazu ziemnego w stanie Pensylwania w latach 2015–2020

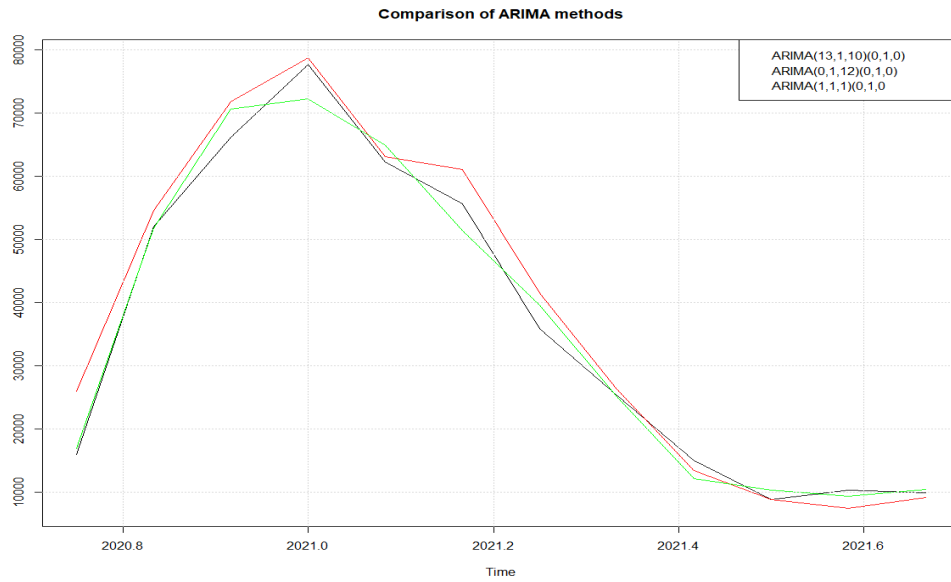


Fig. 13. Comparison of forecasts using ARIMA methods  
 Source: own study based on the U.S. Energy Administration

Rys. 13. Porównanie prognoz przy użyciu metod ARIMA

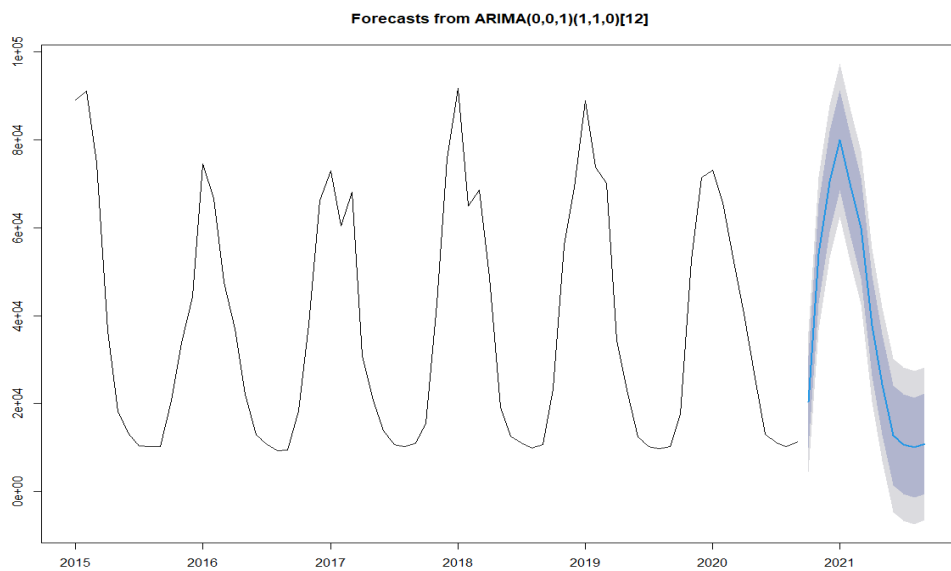


Fig. 14. Forecast using the auto.ARIMA method of natural gas consumption in Pennsylvania in 2015–2020  
 Source: own study based on the U.S. Energy Administration

Rys. 14. Prognoza, oparta na modelu auto.ARIMA, zużycia gazu ziemnego w stanie Pensylwania w latach 2015–2020



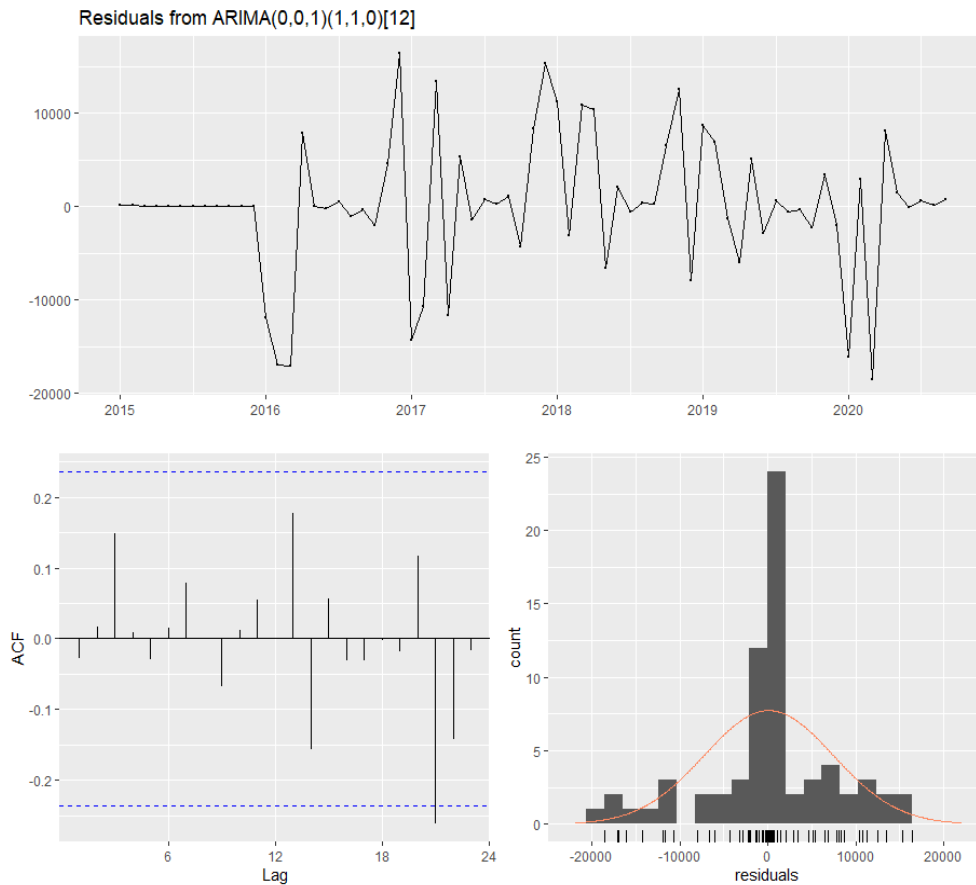


Fig. 15. ACF and PACF graph after applying the differentiation with a delay of 12  
Source: own study based on the U.S. Energy Administration

Rys. 15. Wykres ACF i PACF po zastosowaniu różnicowania z opóźnieniem 12

```

Model Information:
Series: gas
ARIMA(0,0,1)(1,1,0)[12]

Coefficients:
      ma1      sar1
      0.4397 -0.4333
s.e.  0.1366  0.1419

sigma^2 estimated as 66263798: log likelihood=-594.48
AIC=1194.96  AICc=1195.41  BIC=1201.08

Error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 99.18434 7267.666 4748.225 -0.1533682 11.42258 0.8237475 -0.02858196

```

Ljung-Box test  
data: Residuals from ARIMA(0,0,1)(1,1,0)[12]  
Q\* = 7.8078, df = 12, p-value = 0.8  
Model df: 2. Total lags used: 14

## 9. Simple Exponential Smoothing method

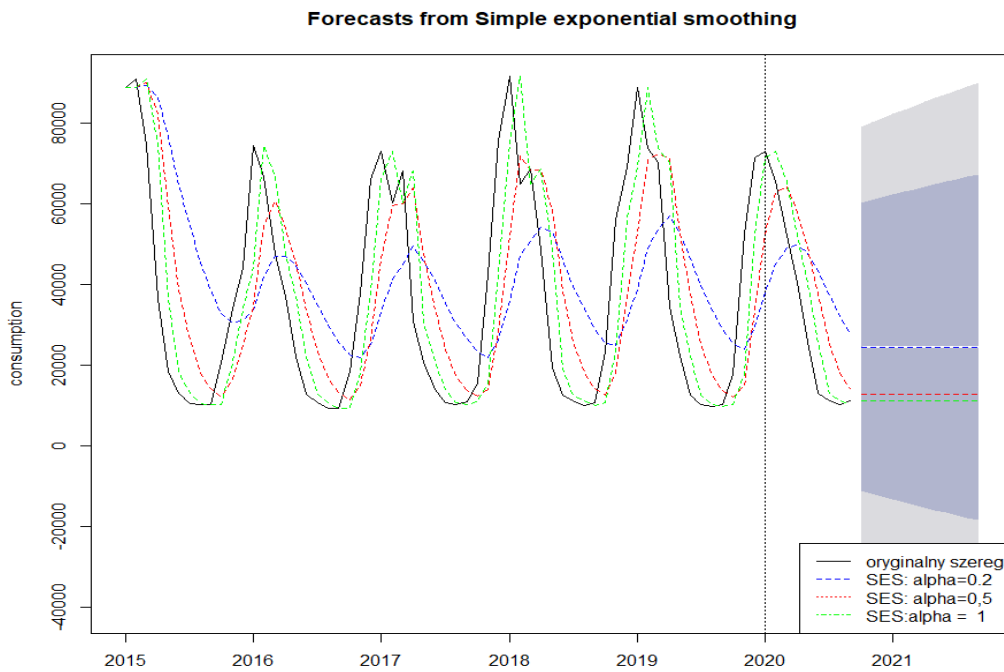


Fig. 16. Forecast using the SES method of natural gas consumption in Pennsylvania in 2015–2020  
Source: own study based on the U.S. Energy Administration

Rys. 16. Prognoza, oparta na modelu SES, zużycia gazu ziemnego w stanie Pensylwania w latach 2015–2020

TABLE 3. Compare the accuracy of the predictions for the training data

TABELA 3. Porównanie dokładności prognoz dla danych uczących

Training Set (test)	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
SES (0,2)	-4678.827	27913.69	24302.72	-90.99433	120.3631	4.216165	0.8160078
SES(0,5)	-2210.963	22385.49	18099.75	-43.78026	72.96971	3.140041	0.7428763
SES(1)	-1126.782	15902.33	11750.21	-12.68741	35.88172	2.038489	0.4696492

Source: based on the RStudio.

In Simple Exponential Smoothing, return forecasts and other information for exponential smoothing forecasts applied to y.

```
ses(x, h = 12, level = c(80, 95), fan = FALSE, initial = c("optimal", "simple"),
    alpha = NULL, ...)
```

Parameters: x – time series, h – horizon, fan – TRUE/FALSE fan chart, initial c() – method of initialization, alpha – smoothing parameter.

## 10. TBATS Model (Exponential Smoothing State Space Model with Box-Cox Transformation, ARMA Errors, Trend and Seasonal Components)

Fits a TBATS model applied to y, as described in (TBATS 2020). Parallel processing is used by default to speed up the computations.

Applying this method through code:

```
tbats(x, use.box.cox = ..., use.trend = ..., use.damped.trend = ..., seasonal.periods =
    = ..., use.arma.errors = ..., ...)
```

Parameters: x – time series, use.box.cox – the Box-Cox transformation or not, use.trend – include a trend or not, use.damped – to include a damping parameter in the trend or not, seasonal.periods – periods can be specified with this parameter, use.arma.errors – to include ARMA errors or not.

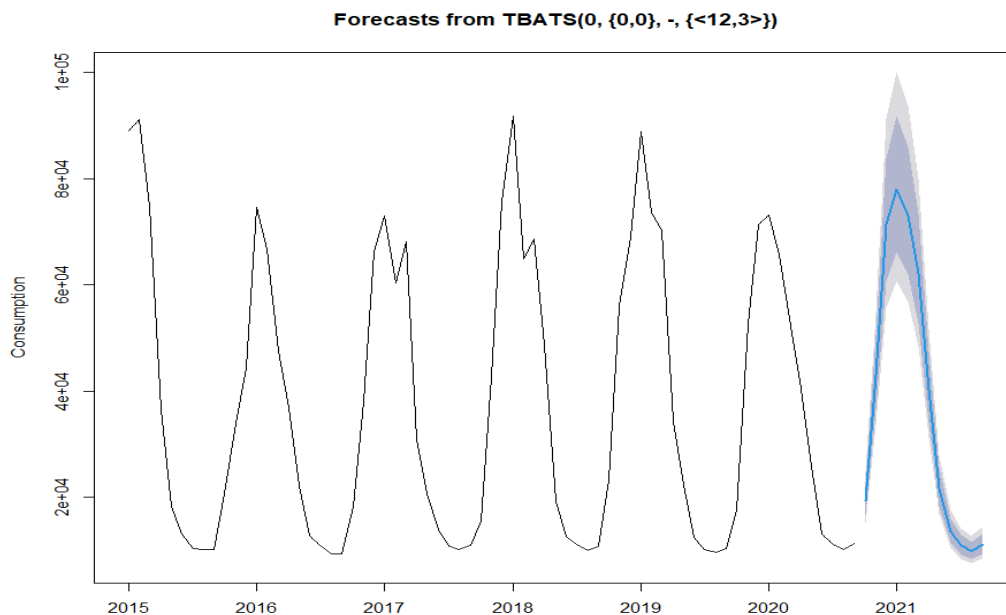


Fig. 17. Forecast using the TBATS method of natural gas consumption in Pennsylvania in 2015–2020  
Source: own study based on the U.S. Energy Administration

Rys. 17. Prognoza, oparte na modelu TBATS, zużycia gazu ziemnego w stanie Pensylwania w latach 2015–2020

## 11. Analysis results and discussion

The given time series, showing the consumption of natural gas, was analyzed by the classical methods used in forecasting and the new TBATS method. In order to select the best method, the libraries implemented in RStudio were used. In order to compare the methods, it is necessary to evaluate and compare the accuracy of the forecasts made. The main criterion that was chosen was the MAPE error. This error is the basic measure of the correctness of the calculations.

In the case of a forecast based on an average model, it is difficult to consider the forecasts as satisfactory. Based on this forecast, you can only use information about the presented prediction ranges of 80% and 95%. MAPE errors in Table 1 definitely show that this method is the worst among the other methods. All criteria indicate that of the three methods, the random walk drift method is the best. Additionally, the data division of the training and test sets was analyzed (Table 2). However, the errors for the test set are clearly larger compared to the forecast errors on the training set.

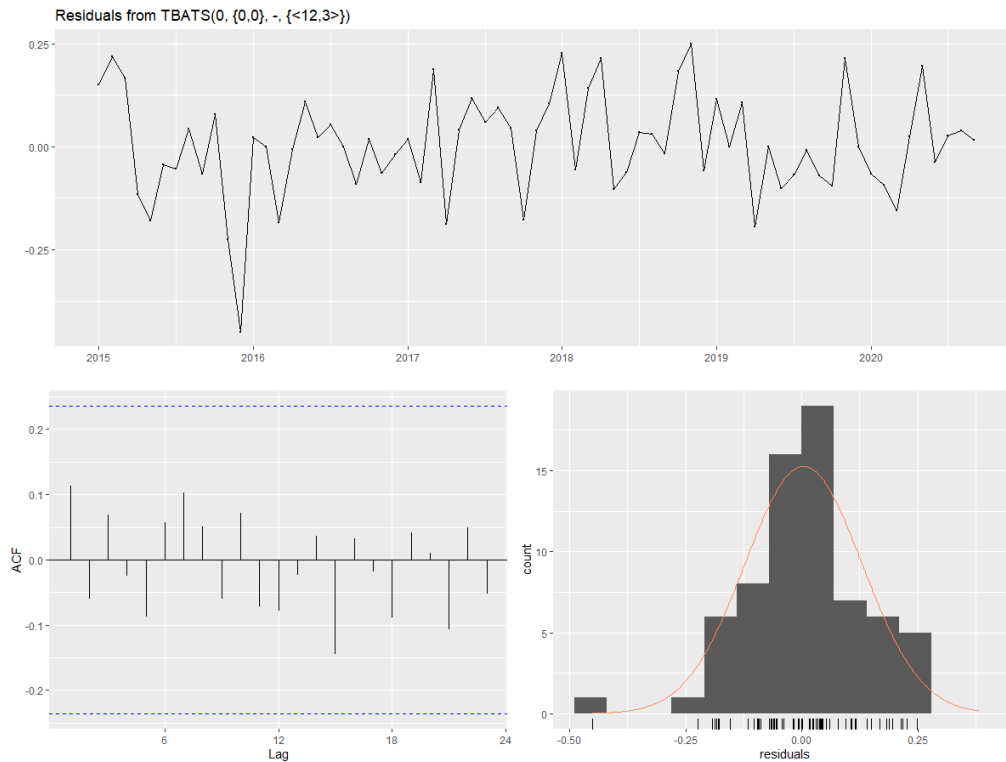


Fig. 18. Check the residual distribution from the model  
 Source: own study based on the U.S. Energy Administration

Rys. 18. Sprawdzenie rozkładu reszt z badanego modelu

#### Ljung-Box test

data: Residuals from TBATS(0, {0,0}, -, {<12,3>})  
 Q\* = 5.3289, df = 3, p-value = 0.1492

Model df: 11. Total lags used: 14

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	666.8058	6501.054	4037.017	-0.4134031	9.540862	0.7003634	0.1605979

However, the results are not that satisfying. Therefore, the analysis was carried out on the basis of the ARIMA, SES, TBATS methods. The forecasts based on ARIMA models are definitely better. For example, the accuracy of the forecast compiled with the auto.arima function was assessed. The MAPE error showed 11.4% in this case. Additionally, forecasts were constructed based on the SES methods. Prediction errors in Table 3 show that the forecasts are not satisfactory.

Finally, a forecast was made using the TBATS model, which turned out to be the most accurate. The MAPE error is 9.5%. The MAPE error states that the forecast may be considered good. Therefore, it can be assumed that TBATS is the best forecasting method for the presented methods, for the tested time series. Plots suggest that, as expected, the residuals are normally distributed (Fig. 18). The forecast confirms that the distribution of gas consumption depends on a given month of the year., The consumption of natural gas decreases with higher temperatures, and the demand for gas fuel increases with lower temperatures. The forecast indicates that the consumption trend will be constant.

As science develops, new methods are still appearing, for example neural networks, deep learning, machine learning. Therefore, it can be expected that in the future new methods will appear that will be even more accurate and will be able to adapt to exogenous factors.

## References

- ARDIA et al. 2020 – ARDIA, D., BOUDT, K. and CATANIA, L. 2019. *Generalized Autoregressive Score Models in R: The GAS Package*. [Online] <https://www.jstatsoft.org/article/view/v088i06> [Accessed: 2021-02-08].
- BALTAGI, B. 2013. *Gasoline: Gasoline Consumption*. [Online] <https://rdr.io/cran/plm/man/Gasoline.html> [Accessed: 2021-03-13].
- BERGHEIM, St. 2008. *Long – Run Growth Forecasting*. Springer, Berlin–Heidelberg.
- EIA 2020. *Natural gas explained*. [Online] <https://www.eia.gov/energyexplained/natural-gas/use-of-natural-gas.php> [Accessed: 2021-03-12].
- EPA 2020. *Package fueleconomy*. [Online] <https://cran.r-project.org/web/packages/fueleconomy/fueleconomy.pdf> [Accessed: 13-03-12].
- HAYES, A. 2021. *Simple Moving Average*. [Online] <https://www.investopedia.com/terms/s/sma.asp> [Accessed: 2020-03-13].
- MAKRIDAKIS, S. 2020. *Bicoal: Annual bituminous coal production*. [Online] <https://rdr.io/cran/fma/man/bicoal.html> [Accessed: 2021-02-15].
- MAKSSOD, Z. and ACHUTHAN, G. 2017. Sustainability in Oman: Energy Consumption Forecasting using R. *Indian Journal of Science and Technology* 10(10), DOI: 10.17485/ijst/2017/v10i/97008.
- MANOWSKA, A. 2013. Forecasting the quantities of coal sales by employing artificial neural networks (*Prognozowanie wielkości sprzedaży węgla kamiennego z wykorzystaniem sztucznej inteligencji*). *Zeszyty Naukowe Politechniki Śląskiej. Organizacja i Zarządzanie* 63 (in Polish).
- PALIŃSKI, A. 2019. Forecasting gas demand using artificial intelligence methods (*Prognozowanie zapotrzebowania na gaz metodami sztucznej inteligencji*). *Nafta-Gaz* 2, pp. 111–117 (in Polish).
- RAMI, K. 2020. *Introduction to USgas Package*. [Online] <https://ramikrispin.github.io/2020/12/intro-to-us-gas/> [Accessed: 2021-02-01].
- STRIELKOWSKI et al. 2021 – STRIELKOWSKI, W., FIRSOVA, I., LUKASHENKO, I., RAUDELUNIENE, J. and TVARONAVICIENE, M. 2021. Effective Management of Energy Consumption during the COVID-19 Pandemic: The Role of ICT Solutions. *Energies* 14, 893, DOI: 10.3390/en14040893.
- TBATS, BATS model (Exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components). [Online] <http://search.r-project.org/> [Accessed: 2020-01-12].
- TIMMER, P. and LAMB, J. 2007. Relations between Temperature and Residential Natural Gas Consumption in the Central and Eastern United States. *Journal of Applied Meteorology and Climatology* 46(11), DOI: 10.1175/2007JAMC1552.1.

- TRUONG, E. 2018. *Time series analysis on Electrical Demand in Australia*. [Online] [https://rstudio-pubs-static.s3.amazonaws.com/430567\\_e3e7a704a4dd45bbbb01abd7f03c7383.html](https://rstudio-pubs-static.s3.amazonaws.com/430567_e3e7a704a4dd45bbbb01abd7f03c7383.html) [Accessed: 2021-02-10].
- YANG, D. 2018. *SolarData: An R package for easy access of publicly available solar datasets*. [Online] [https://www.researchgate.net/publication/327313860\\_SolarData\\_An\\_R\\_package\\_for\\_easy\\_access\\_of\\_publicly\\_available\\_solar\\_datasets](https://www.researchgate.net/publication/327313860_SolarData_An_R_package_for_easy_access_of_publicly_available_solar_datasets) [Accessed: 2021-02-11].
- ZHU et al. 2014 – ZHU, Q., QUAN-YING, L., XIAO-YANG, Z. and KEUNG, L. 2014. A Driving Force Analysis and Forecast for Gas Consumption Demand in China. *Mathematical Problems in Engineering* 2014, DOI: 10.1155/2014/980410.

Tomasz CHRULSKI

## Wykorzystanie środowiska obliczeniowego języka R w prognozowaniu zużycia gazu ziemnego

### Streszczenie

Na rynku dostępnych jest wiele narzędzi informatycznych, które wykonują różnego rodzaju prognozy w gazownictwie. Programowanie ewoluuje wraz z dostępnością i możliwościami komputerów. Narzędzia informatyczne wspierają użytkownika w obliczeniach inżynierskich, ale także prezentują uzyskane wyniki w ciekawej wizualizacji, np. w postaci interaktywnych wykresów. Oprogramowanie może wspomagać podejmowanie decyzji biznesowych, które z kolei mogą służyć jako inteligencja biznesowa.

W dobie cyfryzacji powstają ogromne metadane pomiarów, prowadzenie analiz danych w energetyce jest bardzo powszechne. Ponadto szybko rozwijająca się sztuczna inteligencja stwarza nowe możliwości.

W artykule przedstawiono przykładową analizę obliczeń z wykorzystaniem RStudio, zintegrowanego środowiska programistycznego dla języka R, języka programowania do obliczeń statystycznych i grafiki. Ten język jest jednym z najbardziej rozpowszechnionych języków programowania używanych do tego typu obliczeń. Dodatkową zaletą jest to, że R to bezpłatne środowisko programowe do obliczeń statystycznych i grafiki. Kompiluje się i działa na wielu różnych platformach UNIX, Windows i MacOS.

Artykuł ma na celu przedstawienie czytelnikom możliwości tworzenia prognoz na podstawie modeli matematycznych dostępnych w pakietach R oraz możliwości, jakie daje platforma. Prezentuje funkcje zawarte w pakiecie, które z czasem są aktualizowane i pojawiają się nowe możliwości.

Wynikiem badań są analizy wybranych modeli dostarczonych przez R. W artykule przedstawiono zużycie gazu ziemnego w budynkach mieszkalnych stanu Pensylwania. Zbiór ten przedstawia miesięczne zużycie gazu ziemnego w latach 2015–2020. Prognozy przedstawiono w horyzoncie 12 miesięcy.

SŁOWA KLUCZOWE: prognozowanie, zużycie gazu, rstudio

