DE GRUYTER OPEN

Environmental & Socio-economic Studies

DOI: 10.1515/environ-2015-0074

Environ. Socio.-econ. Stud., **2015, 3, 4: 57-65**

Short-term forecasting of the chloride content in the mineral waters of the Ustroń Health Resort using SARIMA and Holt-Winters models

Dominika Dąbrowska[1], Marek Sołtysiak[1], Jan Waligóra[2]

[1]Department of Hydrogeology and Engineering Geology, Faculty of Earth Sciences, University of Silesia, 41-200 Sosnowiec, Będzińska Str. 60, Poland,
[2]Health Resort Ustroń, Ustroń, Sanatoryjna Str. 1, Poland
E–mail address (corresponding author): d.dabrowska@us.edu.pl

ABSTRACT

The Ustroń S.A. Health Resort (southern Poland) uses iodide-bromide mineral waters taken from Middle and Upper Devonian limestones and dolomites with a mineralisation range of 110-130 g/dm^3 for curative purposes. Two boreholes - U-3 and U3-A drilled in the early 1970s were exploited. The aim of this paper is to estimate changes in mineral water quality of the Ustroń Health Resort by taking into consideration chloride content in the water from the U-3 borehole. The data has included the results of monthly analyses of chlorides from 2005 to 2015 during the tests carried out by the Mining Department of the Health Resort. The triple exponential smoothing (ETS) function and the Seasonal Autoregressive Integrated Moving Average (SARIMA) method of modelling time series were used for the calculations. The ability to properly forecast mineral water quality can result in a good status of the exploitation borehole and a limited number of failures in the exploitation system. Because of the good management of health resorts, it is possible to acquire more satisfied customers. The main goal of the article involves the real-time forecast accuracy, obtained results show that the proposed methods are effective for such situations. Presented methods made it possible to obtain a 24-month point and interval forecast. The results of these analyses indicate that the chloride content is forecast to be in the range of 72 to 83 g/l from 2015 to 2017. While comparing the two methods of analysis, a narrower range of forecast values and, therefore, greater accuracy were obtained for the ETS function. The good performance of the ETS model highlights its utility compared with complicated physically based numerical models.

KEY WORDS: brines, water quality, time series modelling

1. Introduction

Precise methods of forecasting data allow for a better assessment of physical phenomena changing. It has been shown through the forecasting of groundwater levels (MOHANTY ET AL., 2010) or seasonal runoff forecasting systems constructed from a statistical relationship between the model – assimilated precipitation and subsequent runoff (RAHIMI ET AL., 2014; VALIPOUR, 2015) which show that data-driven methods ensure the robustness of the model.

There are many methods intended for forecasting, such as physically based numerical models, stochastic models, artificial neural network (MENHAJ, 2012), and fuzzy or rough sets theory (DUBOIS & PRADE, 1990). Several studies focusing on data – driven models have shown their accuracy and reliability in forecasting is based on Autoregressive Integrated Moving Average (ARIMA), seasonal autoregressive integrated moving average (SARIMA) and exponential smoothing models (TRATAR, 2013).

The reliable assessment and forecasting of groundwater chemical status is extremely important in determining the physico-chemical properties of water used in health resorts for curative purposes. Among the many methods available for investigating environmental forecasts, integrated moving average and exponential smoothing have been used in groundwater chemical status assessments at the Ustroń Health Resort.

Autoregressive integrated moving average and exponential smoothing are very useful models for environmental forecasting (KARAMOUZ & ARAGHINEJAD, 2012; MOHAMMADI ET AL., 2005). In this

study we have examined the sensitivity of SARIMA and Holt-Winters models to forecast the chloride content.

Several studies have highlighted the limitations of these methods connected with analyzing non-linear data patterns (ADAMOWSKI & CHAN, 2011). Having a full database, however, allows the use of these methods for forecasting.

The time series analysis generally assumes that models such as ARIMA, consisting of three elements such as the autoregressive process, moving average process and the degree of integration, are more general than exponential smoothing which is de facto a special variant of the ARIMA model (HYNDMAN & ATHANASOPOULOS, 2014). Data driven models have not been examined to date for the forecasting of chemical status.

2. Study area

The study area is the Ustroń Health Resort located in the southern part of Poland (Fig. 1) within the Silesian Beskids, which are part of the Carpathians (KONDRACKI, 2011). Although Ustroń is still considered to be one of the youngest health resorts in Poland, its balneological traditions date back to the early nineteenth century (RAJCHEL ET AL., 2007). The resort has had a license to capture curative waters since 1994. Water exploitation is carried out there from two U-3 and U3-A boreholes drilled in 1971 and 1978 respectively (WALIGÓRA & SOŁTYSIAK, 2011). The drilling of the latter completed the exploration stage of curative waters in the region. The C-1 absorptive borehole, the first hole for injection post-treatment waters into rocks, was drilled in 1992. It has been assumed that the resource area equals the "Ustroń" mining area in the hydrogeological documentation drawn up so far (WALIGÓRA, 2012). It has an area of 5.4 km^2 (Fig. 1).

The geological profile in the region of the Ustroń contains such forms as Precambrian mica gneiss, Devonian karst limestone and dolomite with a thickness between 460 to 550 m, Carboniferous complexes of claystones, slates, mudstones and sandstones with thicknesses in the range of 53-298 m, flysch forms (alternate layers of slates, shales, limestones, marls, sandstones and mudstones) with thicknesses of more than 1300 m, and quaternary sediments resulting from river accumulation or due to weathering (WALIGÓRA & SOŁTYSIAK, 2011).

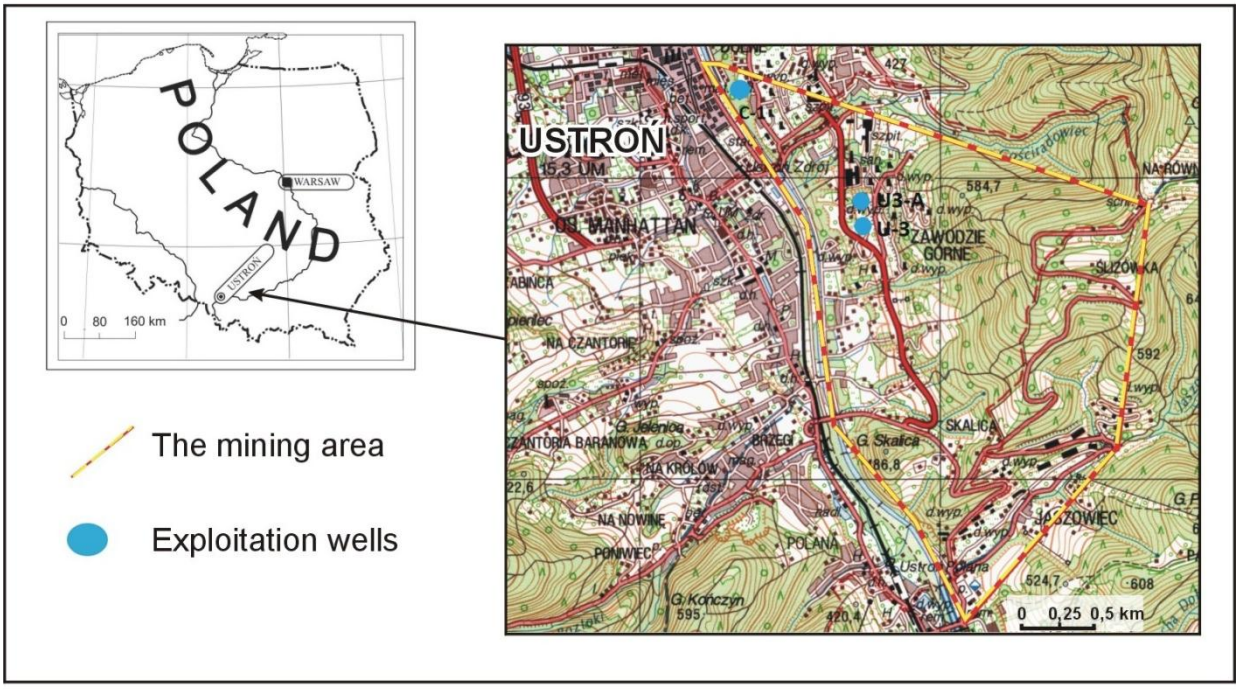

Fig. 1. Location of the study area

In the region of the resort, there are four layers of aquifers: Neogene, Cretaceous-Paleogene, Carboniferous and Devonian (SOLECKI, 2007). The Neogene layer is associated with coarse sandstones and conglomerates belonging to the Dębowiec layers (CHOWANIEC, 1993). The Cretaceous-Paleogene layer is full of crevasses and void fractions. Because of the great lithological diversity of the Silesian and Sub Silesian Nappes, the mineral waters have a mineralisation equal to 2 and 3.2 g/l respectively. The Carboniferous water-bearing layer is poorly known and is not used due to the predominance of clay forms.

The crevasse-karst Devonian aquifer is the most important in this region. It is associated with Upper and Middle Devonian carbonates such as grey and dark- grey limestones with the depositions of dolomites, marls or dark shales. Within this layer, there is a level of mineral waters used in balneology with a mineralisation ranging from 110 g/dm^3 (the U-3) to 130 mg/dm^3 (the U3-A) and the Cl Na-Ca, Fe, I type exploited from a depth of 1318 to 1728 m. The direction of the Devonian water migration is difficult to determine. According to WALIGÓRA (2012), the slow movement of these waters is through privileged zones, i.e. dislocation zones and karst zones, from south to the north.

3. Methodology

Maintaining good groundwater status is very important for environmental reasons. It is necessary to maintain stable physico-chemical conditions of curative waters for the purposes of balneotherapy. The performed physical and chemical analyses offer the possibility of forecasting changes in the concentration of the individual components and thus the possibility of preventing the potential negative consequences of the deterioration in water quality. Time series forecasting functions allow for the prediction of the future values based on historical data. For this purpose, various forecasting methods such as the Holt-Winters model or the ARMA/ARIMA model are used.

Logarithm and then additive decomposition should be applied to the data, which is to be used in forecasting. This will allow for the determination of individual components in the series – the definition of seasonality and the determination of a trend line.

Holt. There are two variations to the Holt Winters model – the additive and the multiplicative method. The additive method is preferred when the seasonal variations are roughly constant through the series and the multiplicative method is used when the seasonal variations are changing proportional to the level of the time series. The ETS function which is a special case of the Holt-Winters model uses advanced machine learning algorithms such as triple exponential smoothing. The ETS function in the R package allows you to choose the correct model from the 20 available.

The selected model will allow for the determination of the monthly point and interval forecasts. The interval forecasts can be determined at different confidence levels (usually 0.95 and 0.8). Such forecast values are verified using 6 indicators such as: mean error, mean absolute error, root mean squared error, percentage error, mean absolute percentage error, mean absolute scaled error.

The second method used for the forecasting purposes is the SARIMA (Seasonal Autoregressive Integrated Moving Average) model composed of three parameters: autoregression, degree of series integration and moving average (ASTERIOU & HALL, 2011), which evaluate the impact of historical values and the volume of historical and current disturbances on current values.

The autoregressive process (p) is a process in which each value is a linear combination of the preceding values. In the autoregressive process the autoregressive order is calculated to determine how many previous values have impacted on the current analysed value. The moving average process (d) is an average of recent anomalies and random components of the model. Reducing a non-stationary series to a stationary one is an integration of the process (q). The multiplicative SARIMA model (p, d, q) (P, D, Q)$_\omega$ is defined as follows:

$$(1-\Phi_1 B^{\omega}-\Phi_2 B^{2\omega}-\ldots-\Phi_p B^{P\omega})(1-\varphi_1 B-\varphi_2 B^2-\ldots-\varphi_p B^{p})(1-B^{\omega})^{D}(1-B)^{d} Z_t = (1-\Theta_1 B^{\omega}-\Theta_2 B^{2\omega}-\ldots-\Theta_Q B^{Q\omega})(1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^{q})\varepsilon_t \quad (1)$$

where:
ϕ - parameter of seasonal autoregressive model
B - difference operator
B(Z_t) - the Dth seasonal difference measure ω
$(1-B^{\omega})^{D}$ - the dth non- seasonal difference
⊙ - seasonal moving average model
φ - parameter of non- seasonal autoregressive model
θ - parameter of non- seasonal moving average model
p - order of non- seasonal autoregressive model
q - order of non- seasonal moving average model
P - order of seasonal autoregressive model
Q - order of seasonal moving average model
ω - periodic term
$d=(1-B)^{d}$ – difference order

Choosing the most accurate SARIMA model is performed by using the procedure developed in 1976 by Box and Jenkins. The procedure for selection of the optimal model includes identification, forecast and diagnosis. For this purpose, an autocorrelation (ACF) diagram and partial autocorrelation (PACF) graph are drawn (BALAGUER ET AL., 2008). The horizontal axis shows the delay time and the vertical one presents the amounts of ACF and PACF, respectively (EDIGER & AKAR, 2007).

As a result of the forecast procedure two estimated parameters (MA1 and SMA1, where MA1 = moving average parameter of order 1, SMA1 = seasonal MA1) are obtained, so are the corresponding estimate errors (s.e.), which are used to verify the hypothesis about the importance of these estimates. Also, the values of information criteria such as AIC (Akaike Information Criterion) and BIC (Bayes Information Criterion) are obtained. These criteria are based on the log likelihood function. Both criteria are used to select the most accurate and easiest model.

The analysis of residuals obtained from AIC and BIC should be performed for the model built in such a way. When a model has been selected correctly, standardised residuals behave as independent random variables with a mean of 0 and variance equal to 1. The residuals and time ratio graph and the autocorrelation graph are performed, as well as the Ljung-Box test (LJUNG & BOX, 1978) which is carried out to check the validity of the model selection.

Developing such a model allows for the execution of point and interval forecasts for several different levels of confidence. After this operation, the results are verified by calculating the errors, as is done after the ETS function.

4. Results

What is presented is the forecasted changes in chloride content in the curative waters of the Ustroń Health Resort from the U-3 borehole. The study period included the mean value of chloride content for each month for the period from January 2005 to January 2015 (Fig. 2). To assess changes in chloride concentration and to give a 24-month forecast, two statistical methods – the Holt-Winters model and the SARIMA model were used. All calculations were made in the R programming environment (R Core Team 2015). The data was subjected to logarithm and additive decomposition. The individual components of the series: seasonality and trend line were determined in that way (Fig. 3).

As observed between 2005 and 2009, there is a decreasing trend of chloride concentration in the U-3 borehole. Most changes in the whole study period took place between 2009 and 2013, and there was a growing trend since 2013. It is worth noting that slightly higher values of chloride concentrations occur in spring and summer months, and lower ones from November to January, and are connected with the plan of the deposit exploitation. The efficiency of the pumping system is lower in the winter period because of less demand for mineral water. Thus the mineralization will be lower as the pump efficiency decreases (WALIGÓRA, 2012).

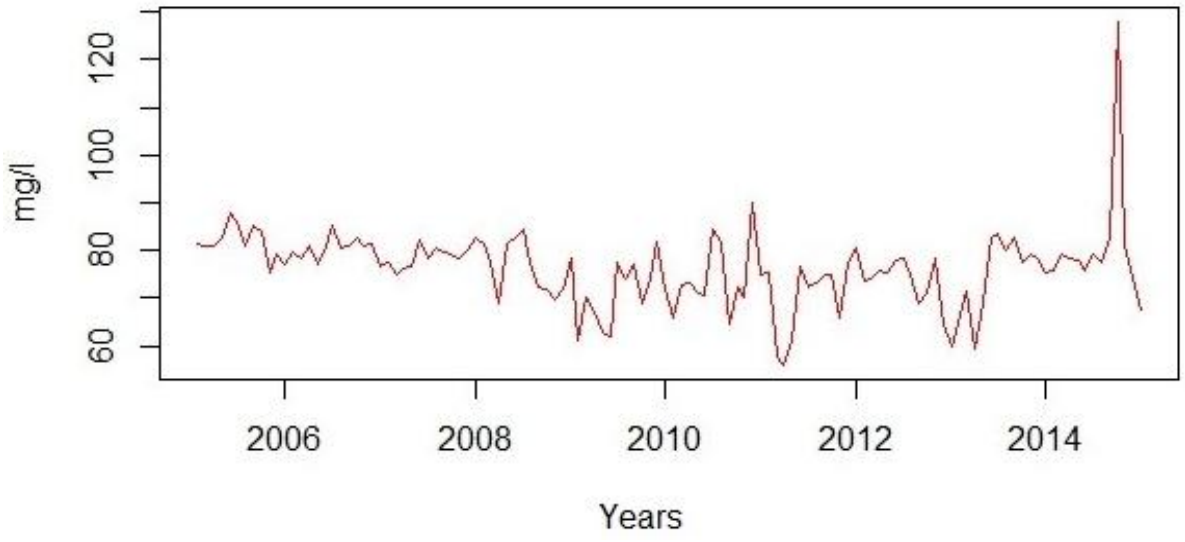

Fig. 2. Concentration of chlorides in the U-3 borehole in the period of January 2005 to January 2015

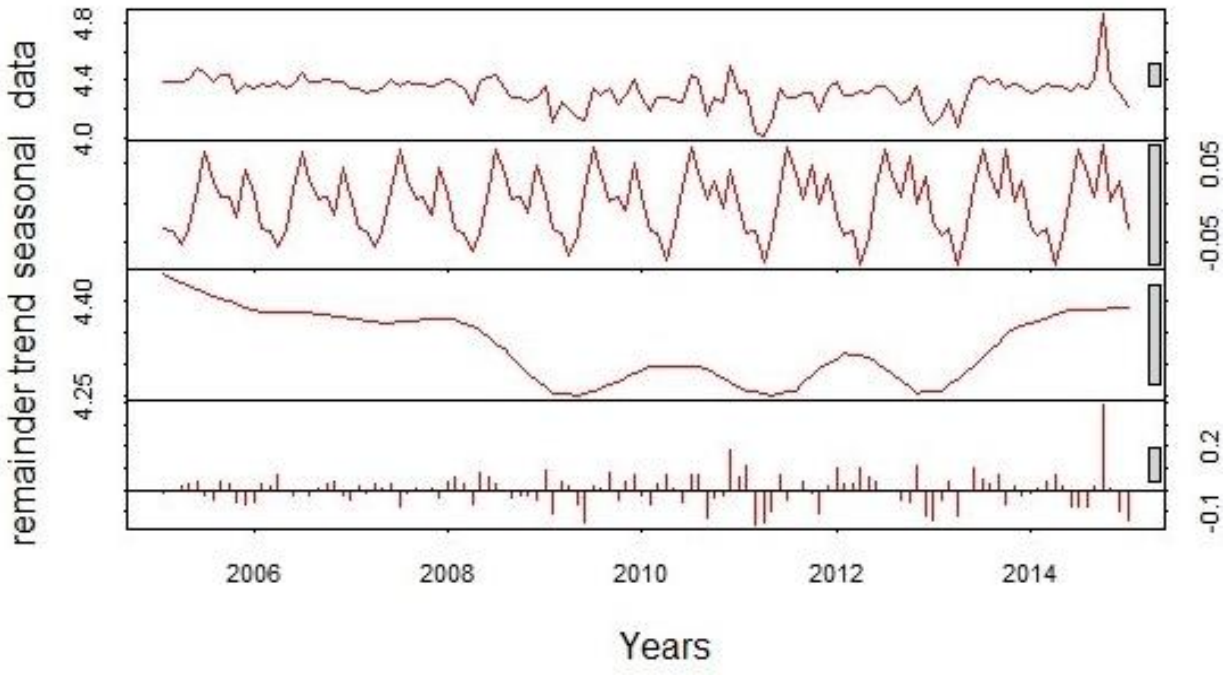

Fig. 3. The additive decomposition of the time series studied

When choosing the ETS function in the R package, four models of exponential smoothing were tested, i.e. a simple exponential smoothing model, Holt's model, Winters' additive and multiplicative model. Finally, the last model was selected, as the one with the smallest number of mistakes and the best match, was chosen to be compared with the SARIMA model. The Winters' multiplicative model consists of three equations:

(1) $$L_{t-1} = \alpha \frac{y_{t-1}}{C_{t-1-r}} + (1-\alpha)(L_{t-2} + B_{t-2}),$$

(2) $$B_{t-1} = \beta(F_{t-1} - F_{t-2}) + (1-\beta)B_{t-2},$$

(3) $$S_{t-1} = \gamma \frac{y_{t-1}}{F_{t-1}} + (1-\gamma)S_{t-1-r},$$

where:
L_{t-1} - the equivalent of the smoothed value obtained from a simple exponential smoothing model (the weighted arithmetic mean)
B_{t-1} - the evaluation of a growth trend of the moment or period t-1,
S_{t-1} - the assessment of the seasonal indicator for a moment or period t-1,
r - the length of the seasonal cycle - the number of phases
α, β, γ - smoothing parameters of the model with the values in the range of [0, 1].

Initial values were determined for each of the equations in the Winters' model: for the average level of a series (L), for the trend (B), and for the seasonal component (S) (L = 84.8648, B = -0.2074, and S = 0.9838, 1.0359, 0.9901, 1.0716, 0.98, 1.0153, 1.073, 1.0282, 0.9621, 0.9316, 0.9699, 0.9585) and smoothing parameters (α = 0.5112, β = 3e-04, γ = 1e-04). The values obtained were the lowest among the smoothing models tested.

After selecting the model, it was used to forecast changes in chloride concentration. A period of 24 months was assumed for the period of the forecast. Point and interval forecasts were made. The latter was made at two levels of confidence: 0.95 and 0.8. Forecasts of the changes in chloride content using the ETS function is shown in Fig. 4.

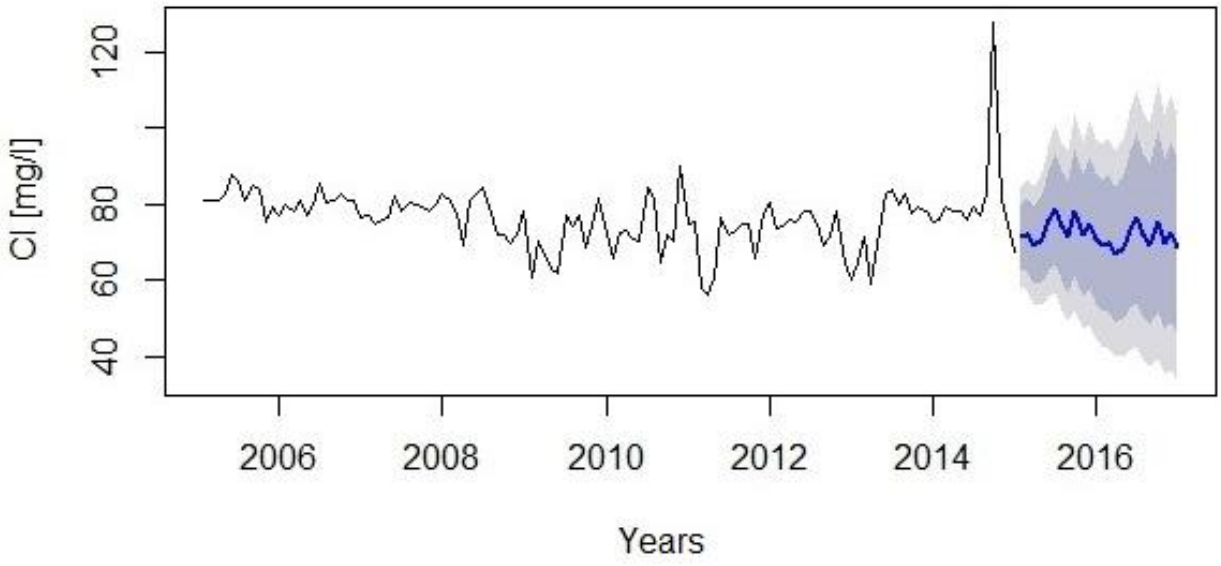

Fig. 4. Forecast of changes in chloride content in the period between February 2015 and February 2017 using the ETS function

Also, the accuracy of the estimation was verified by determining the following values: Mean error 0.2145021, Mean absolute error 5.127926, Root mean squared error 7.316485, Percentage error -0.285565, Mean absolute percentage error 6.71916, Mean absolute scaled error 0.7205705.

The value of the average error should be close to zero. In this case, there is a slight underestimation. The MAE informs about the average aberrations of the actual realisations from the forecast ones for the forecasted variable (the absolute magnitude) in the period of forecast. The PE indicates what percentage of the actual realisation of the forecast variable is the forecast error. The MAPE error should rather be used to compare models and should not be used to determine a single forecast error (SZMUKTA-ZAWADZKA & ZAWADZKI, 2012). Secondly, the SARIMA model was used. The autocorrelation diagram (ACF) were produced (Fig. 5) and the partial autocorrelation graph – PACF (Fig. 6) (BALAGUER ET AL., 2008).

The SARIMA model (1,1,1), (1,1,1) was used to project the chloride concentration forecast. As a result of the estimation procedure, there was obtained a forecast of two parameters ma1 = 0.3695 and sma1 = -0.9241 and the corresponding estimate errors (s.e. = 0.1023; 0.0419). The value of information criteria AIC, BIC equal respectively: -226.39 and -218.05.

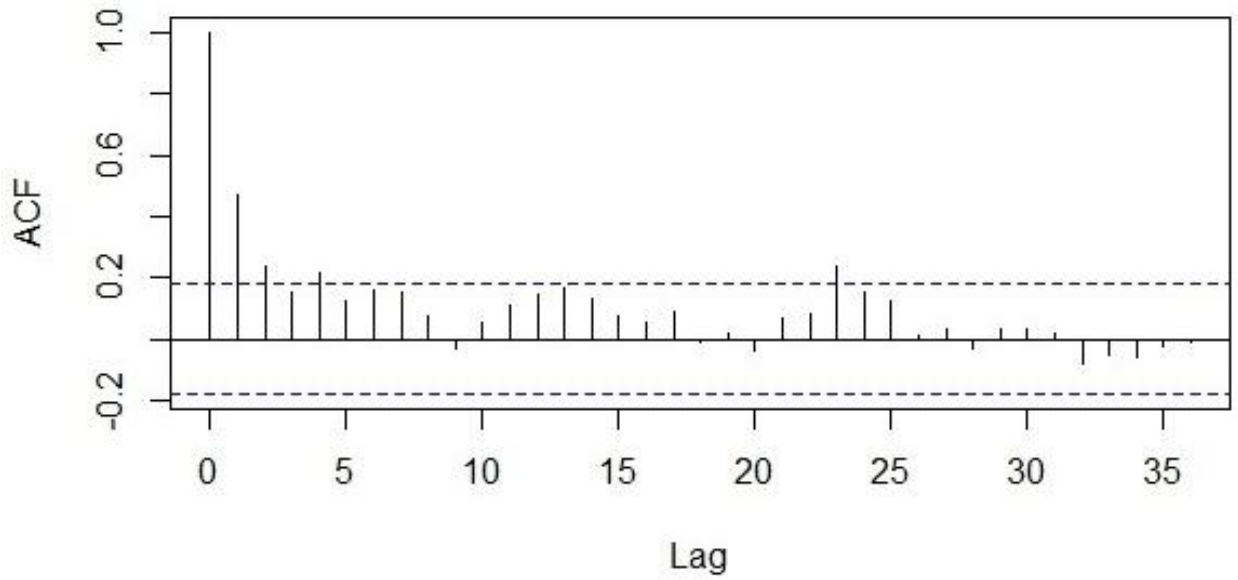

Fig. 5. The autocorrelation chart of the ARIMA model

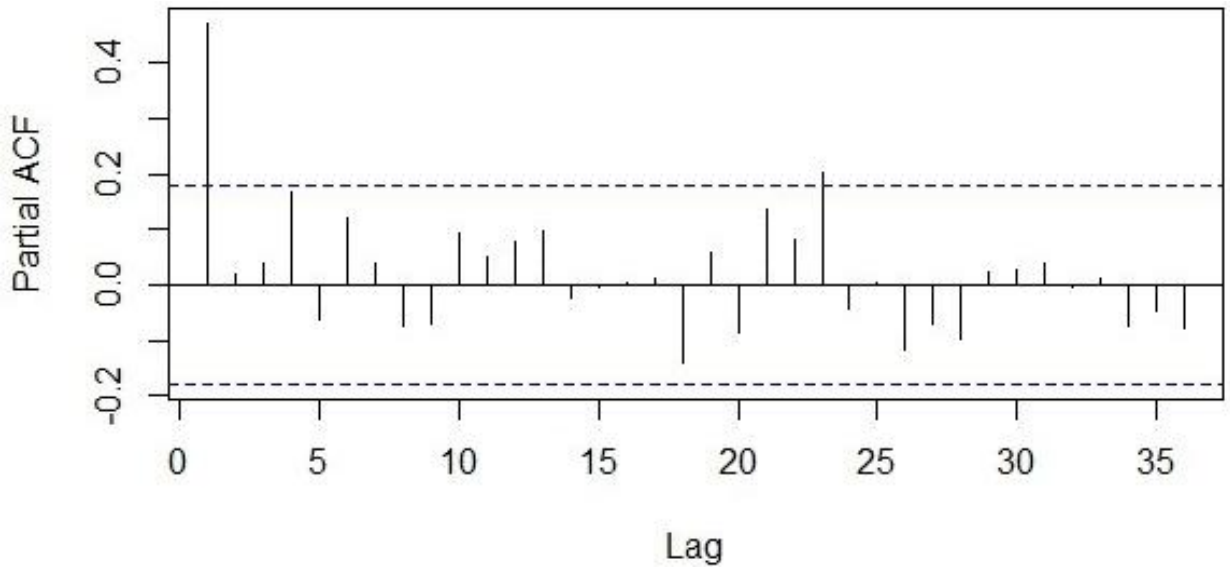

Fig. 6. The partial correlation chart of the ARIMA model

To check if the model was matched accurately, a diagram of the values of residuals and time ratio, and autocorrelation function diagram were done. Thirdly, the Ljung-Box test was performed (Fig. 7) P-value of the Ljung-Box statistics is so high that the hypotheses of zero autocorrelations cannot be denied. 24 monthly point and interval forecasts at the same level of confidence as for the ETS function were made for the model constructed in such a way. The results are shown in Fig. 8.

As can be seen when comparing the change models of chloride concentration in the U-3 borehole using the ETS function and the SARIMA, better forecast and their narrower range were achieved for the first one. Higher forecast values of chloride concentration in the U-3 borehole were received by the SARIMA method.

The resulting forecast values are not significantly different from the previous ones. This indicates the stability of the chemical composition of therapeutic waters from the boreholes in the Ustroń Health Resort. Both forecasting models indicate a slightly higher concentration of chlorides in the summer months and slightly lower in the winter, which is also noticeable now. Both full physical and chemical analyses as well as individual marks of chloride ions confirm the stable conditions of the chemical composition of water in the exploited holes. Both methods can be used to analyse changes in the chemical composition of groundwater.

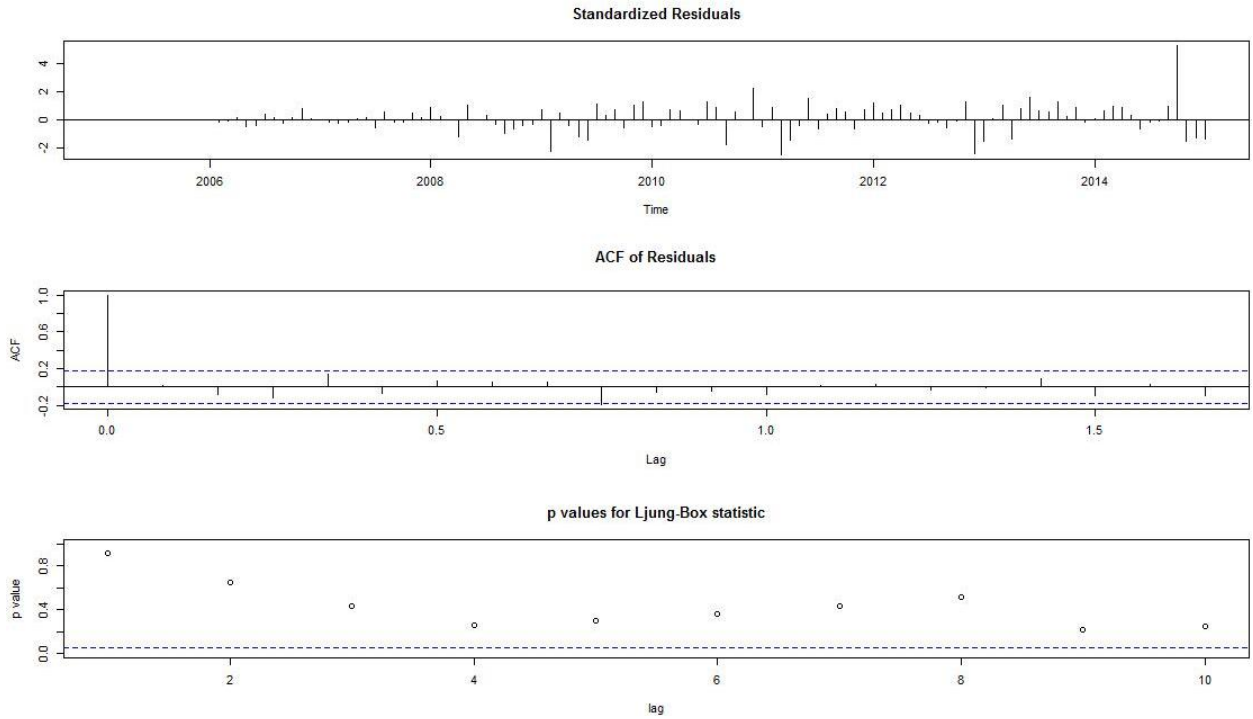

Fig. 7. Analysis of residuals of the model tested

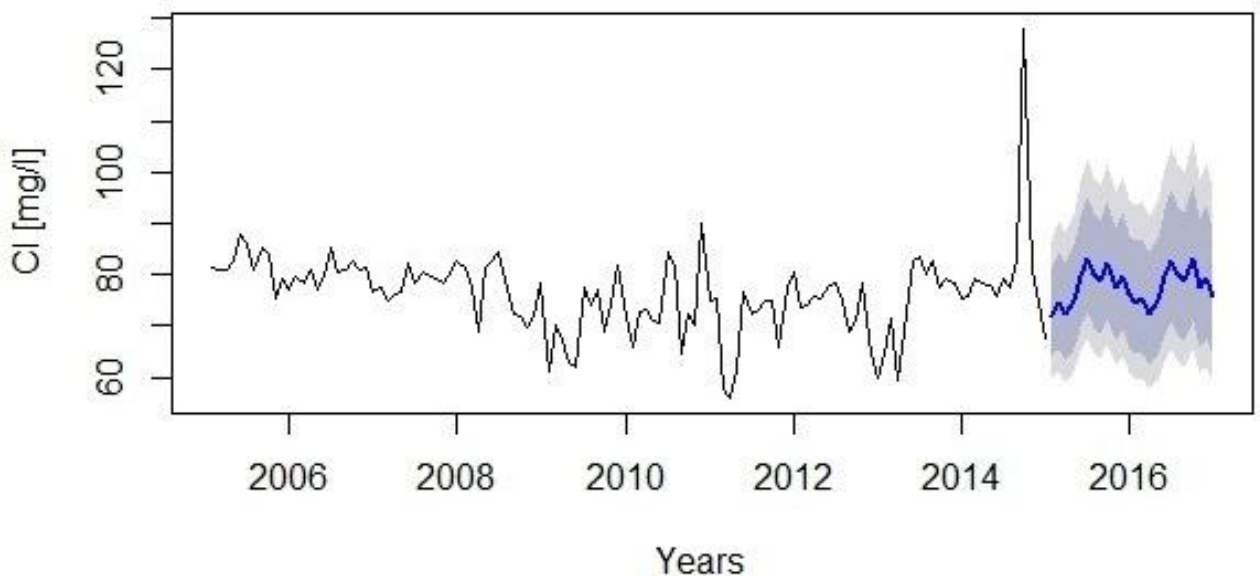

Fig. 8. Forecast of the changes in chloride content in the period from February 2015 to February 2017 using the SARIMA method

5. Summary and conclusions

The Ustroń Health Resort uses Devonian brines for therapeutic purposes with the mineralization of 130 g/l from the U-3 and U-3A boreholes. The brines are characterised by a high stability of physico-chemical composition.

Conducted measurements of the chemical composition of the therapeutic waters of the Ustroń Health Resort show that higher concentrations of certain components, e.g. chlorides, are slightly higher in summer than in winter which is associated with pump efficiency.

In the case of water used in health treatment, it is important to assess the current changes in their chemical composition and performance of a forecast. For weekly data on the concentration of chlorides in the U-3 borehole, time series analyses were performed using the ETS function and the ARIMA method. The data covered the period from January 2005 to January 2015. The analysis shows that it is possible to use both methods in order to perform the forecast changes in the content of individual components of groundwater based on current monitoring data.

The resulting models suggest a slight variation in the concentration of chlorides in the entire study period, and also in the forecast. Taking chlorides as an example and referring the results of both methods to general mineralisation of the waters tested, the stability and high quality of the chemical composition of the waters in the Ustroń Health Resort were confirmed.

References

Adamowski J., Chan H.F. 2011. A wavelet neural network conjunction model for groundwater level forecasting. *J. Hydrol.* 407: 28-40.

Asteriou D., Hall S. 2011. ARIMA Models and the Box–Jenkins Methodology. *Applied Econometrics*: 265-286.

Balaguer E., Palomares A., Sorie E., Martin- Guerrero J.D. 2008. Predicting service request in support centers based on nonlinear dynamics, ARMA modeling and neural networks. *Expert Syst. App*. 34: 665-672.

Box G.E.P., Jenkins G.M. 1976. *Series Analysis Forecasting and Control*. 1st ed. Holden-Day, San Francisco.

Chowaniec J. 1993. Budowa geologiczna i warunki hydrogeologiczne okolic Ustronia z uwzględnieniem wyników otworu chłonnego Ustroń C-1. Ustroń Health Resort Archive, unpublished.

Dubois D., Prade H. 1990. Rough fuzzy sets and fuzzy rough sets. *Int. J. General Systems,* 17 (2–3): 191–209.

Ediger V., Akar S. 2007. ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy*, 35: 1701-1708.

Karamouz M., Araghinejad Sh. 2012. *Advanced Hydrology*. Amirkabir Univ. of Tech. Press.

Kondracki J. 2011. Geografia regionalna Polski. PWN, Warszawa.

Ljung G., Box G. 1978. On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 66: 67–72.

Malina A. 1994. The Forecasting of Economic Phenomena on the Basis of the Methods of Exponential Smoothing of Time Series. *Cracow Rev. Econ. Manage.,* 440: 15-29.

Menhaj M.B. 2012. Artificial Neural Network. Amirkabir Univ. of Tech. Press.

Mohammadi K., Eslami H.R., Dayyani Sh. 2005. Comparison of regression ARIMA and ANN models for reservoir inflow forecasting using snowmelt equivalent. *J. Agric. Sci. Tech.,* 7: 17-30.

Piłatowska M. 2011. Porównanie kryteriów informacyjnych i predykcyjnych w wyborze modelu. *J. Manage. Finance,* 4: 499-512.

Piontek K. 2002. *Modeling and forecasting financial instruments variability* (PhD thesis). Wroc. Uniw. Econ., Wrocław.

Rajchel L., Śliwa T., Waligóra J. 2007. Uwagi o wodach leczniczych Ustronia. Współczesne problemy hydrogeologii, 13, Krynica-Kraków.

R Core Team 2015. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna URL http://www.R-project.org/

Solecki T. 2007. Zastosowanie odwiertów chłonnych w ochronie środowiska na przykładzie uzdrowiska Ustroń. *Wiertnictwo, nafta, gaz*, 24: 465-473.

Szmukta-Zawadzka M., Zawadzki J. 2012. O metodzie prognozowania brakujących danych w szeregach czasowych o wysokiej częstotliwości z lukami. Metody ilościowe w ekonomii, 13: 212-223.

Tratar L.F. 2013. Improved Holt-Winters method: A case of overnight stays of tourists in Republic of Slovenia. *Econ. bus. rev.,* 16: 5-17.

Valipour M., Banihabib M., Behbahani S. 2013. Comparison of the ARMA, ARIMA and the autoregressive artificial neural networks model in forecasting the monthly inflow of Dez dam reservoir. *J. Hydrology,* 476: 433-441.

Waligóra J. 2012. Projekt zagospodarowania złoża wody leczniczej „Ustroń" z utworów dewonu, w granicach obszaru górniczego „Ustroń". Ustroń Health Resort Archive, unpublished.

Waligóra J., Sołtysiak M. 2011. Zatłaczanie wód pozabiegowych w utwory serii węglanowej dewonu w uzdrowisku Ustroń. *Biul. Państ. Inst. Geol.,* 445: 701-708.