

Ewa Wąsik*, Krzysztof Chmielowski*

The use of Holt–Winters method for forecasting the amount of sewage inflowing into the wastewater treatment plant in Nowy Sącz

Wykorzystanie metody sezonowej Holta-Wintersa do prognozowania dopływu ścieków do oczyszczalni na przykładzie obiektu w Nowym Sączu

*Dr inż. Ewa Wąsik, dr hab. inż. Krzysztof Chmielowski - Department of Sanitary Engineering and Water Management, University of Agriculture in Krakow, Mickiewicza 24/28 Ave., 30-059 Krakow e-mail: e.wasik@ur.krakow.pl, k.chmielowski@ur.krakow.pl

Keywords: amount of inflow sewage, time series, Holt–Winters model, wastewater treatment plant

Słowa kluczowe: natężenie dopływu ścieków, szeregi czasowe, model Holta-Wintersa, oczyszczalnia ścieków

Abstract

The aim of the study was to determine changes of daily amount of sewage inflowing into a wastewater treatment plant in Nowy Sącz in the years 2008–2014. To this end, the data in the form of time series corresponding to the investigated multi-year period were analysed. Daily volume of sewage for annual periods was forecast using a seasonal method of Holt and Winters based on the exponential smoothing algorithms. The model fit to actual daily amount of sewage for 2014 was assessed using linear regression. The results of fit for the additive Holt–Winters model confirmed the usefulness of this tool for forecasting the amount of sewage inflowing into the wastewater treatment plant.

Streszczenie

Celem pracy była ocena zmian dobowego natężenia ścieków, które w latach 2008–2014 dopływały do oczyszczalni w Nowym Sączu. Dokonano analizy danych dopływu z wielolecia w postaci szeregu czasowego. Do prognozowania dobowej ilości ścieków dla rocznego okresu wykorzystano opartą na algorytmach wyrównania wykładniczego metodę sezonową Holta-Wintersa. Ocenę trafności dopasowania modelu do rzeczywistych wartości dobowych ilości ścieków z 2014 r. przeprowadzono z wykorzystaniem regresji liniowej. Uzyskane wyniki dopasowania addytywnego modelu Holta-Wintersa potwierdzają jego przydatność jako narzędzia do prognozowania natężenia ścieków dopływających do oczyszczalni.

© IOŚ-PIB

1. INTRODUCTION

Natural phenomena are often described using time series [Box, Jenkins 1983]. A time series consists of consecutive elements that are linearly combined with previous observations and a random component. Time series analysis may help not only in understanding the nature of a phenomenon represented by a sequence of observations but also in estimating future values of the time series (forecasting).

Popular methods for time series forecasting involve simple models of exponential smoothing. They assume that the impact of all previous elements of the series on the predicted value decreases over time. Assuming an exponential nature of this decrease, the following general form of the exponential smoothing model is given by the formula:

$$F_t = \alpha \times X_t + (1-\alpha) \times F_{t-1} \quad (1)$$

where:

F_t - time series values at t time that were exponentially smoothed;

α - smoothing factor;

X_t - observed time series values;

F_{t-1} - previous values of the series within $t-1$ period.

This model includes α factor that may range from 0 to 1. When $\alpha = 1$, the previous observations are completely ignored. When $\alpha = 0$, all predicted values are equal to the first observation. Therefore, the lower α , the better smoothing of the series but also 'weaker memory'.

Structure of a random phenomenon is often difficult to identify due to a systematic element and so-called noise (random noise). The components of time series for a given time point t include trend (T_t), seasonal (S_t), cyclical (C_t), and accidental (I_t) irregularities. A method of simple exponential smoothing is used to remove noise or to forecast the time series that do not show a clear trend and seasonality. In 1957, Charles Holt proposed a generalisation of the above method for forecasting time series characterised by a growth tendency, that is, trend [Box, Jenkins 1983].

Apart from a trend, numerous random phenomena include also seasonality of weekly, monthly, quarterly or annual character. The smoothing model based on Holt's algorithm and supplemented with seasonality smoothing factor d was introduced in 1960 by Peter Winters [Box, Jenkins 1983]. The Holt–Winters model is used when the seasonality factor cannot be determined directly from the source data and the time series must be smoothed to

estimate the values of d . Seasonality factor may be determined by means of simple exponential smoothing in two different ways:

- for an additive model (the time series shows regular seasonality irrespective of a trend):

$$I_t = I_{t-p} + d \times (1-\alpha) \times e_t \quad (2)$$

- for a multiplicative model (seasonality changes follow those of a trend):

$$I_t = I_{t-p} + d \times (1-\alpha) \times e_t / S_t \quad (3)$$

where:

I_{t-p} - smoothed seasonality factor at t minus p (length of periodic irregularity);

d - smoothing factor;

e_t - observed value minus the value predicted at time t .

d may range from 0 to 1. When $d = 0$, predictions for one step ahead are generated using a constant seasonality factor. When $d = 1$, the seasonality factor is modified by a corresponding forecast error (times $1-\alpha$).

Another component of time series that should be analysed is a growth trend. The trend may be continuously increasing or decreasing (monotonic trend). When a time series contains an essential random component (noise), smoothing with the moving average or median [www.statsoft.pl] is required to identify g (for the models of linear or exponential trends) or j (for the models of damped trends) parameters. When $g = 0$, the trend component is constant for all time series values (and all forecasts). When $g = 1$, the trend component is completely modified by the appropriate forecast error. j parameter shows how strongly the trend changes affect the trend assessment for future forecasts [www.statsoft.pl]. Therefore, it is important to find an appropriate way of noise filtering to reveal the systematic component. This is achieved by decomposition of a time series and focussing on the analysis of key data for the investigated phenomenon. In seasonal decomposition based on Census I method, where the trend component and cyclic component are combined into trend-cycle component (TC_t), functional relationship between the components may have a form of an additive model [www.statsoft.pl]:

$$X_t = TC_t + S_t + I_t \quad (4)$$

or a multiplicative model:

$$X_t = T_t \times C_t \times S_t \times I_t \quad (5)$$

where X_t is the observed value of time series at time t .

Forecasting methods using exponential smoothing models are not complex and do not assume time series stationarity.

Chotkowski and Lis [2006] and Cupak *et al.* [2011] used the additive model of Winters to predict the amount of sewage inflowing into a wastewater treatment plant (WTP), while Wałęga

and Bergel [2009] used it to achieve a reliable forecast of daily water consumption for domestic purposes.

The aim of the study was to evaluate the possibility of using time series for forecasting the amount of municipal sewage inflowing into the mechanical and biological WTP in Nowy Sącz within the multi-annual period 2008–2014.

2. MATERIALS AND METHODS

The investigated WTP is located in Subcarpathian region in the village of Wielopole. Minimum design capacity of the facility is 120 000 PE. Average design daily sewage amount Q_{davg} is 30 000 $m^3 \cdot d^{-1}$, and maximum daily sewage amount $Q_{dmax} = 42.200 m^3 \cdot d^{-1}$ [Water Permit 1999].

Municipal sewage from the city of Nowy Sącz and neighbouring communities of Stary Sącz, Nawojowa and Kamionka Wielka is supplied to the WTP via a sewerage system and by septic tankers that transport sewage from unconnected smaller facilities and households. Receiving water is the Dunajec river at 103 + 600 km.

The sewerage network is 201 km long, including 76.1 km of combined sewage system and 124.9 km of sanitary sewage system. The sewerage network serves about 86% of Nowy Sącz inhabitants [www.swns.pl].

Source data provided by Water and Sewerage Company in Nowy Sącz included daily amounts of raw sewage entering the investigated WTP. The study lasted from 1 January 2008 until 31 December 2014. Basic descriptive statistics, i.e. arithmetic average, median, standard deviation, minimum, maximum, coefficient of variation were calculated and the number of exceedance events with reference to a specification limit provided in a water permit [Water Permit 1999] was analysed.

The first step in assessing the sequence of the source data was identification and description of the time series structure. A graphical representation of daily flow intensity was analysed for outliers, trend and seasonal irregularities.

The next step was introducing seasonal irregularities into a model and determining their frequency by means of autocorrelation function. The results were then used to perform seasonal decomposition of the time series for trend effect, seasonality effect and other variations of sewage flow. Following classical decomposition of the time series with Census I method and the time series observation, the additive model was adopted.

An analysis of the time series properties was used to define a model based on exponential smoothing as the one most accurately describing the variability of sewage flow to the Nowy Sącz WTP in the investigated multi-year period.

This theoretical model was then used for forecasting future values of the time series. Accuracy of the forecasts generated on the basis of some measurements was verified by their comparison with source data from 2014. Determination and linear correlation coefficients were used to evaluate the fit of the curve for the analyzed data arranged by years.

The model for forecasting daily amount of sewage was selected using STATISTICA 10.

Table 1. Descriptive characteristics of the sewage inflowing into the wastewater treatment plant in Nowy Sącz in the consecutive months of multi-year period 2008–2014.

Month/Parameter	Average [m ³ ·d ⁻¹]	Median [m ³ ·d ⁻¹]	Minimum [m ³ ·d ⁻¹]	Maximum [m ³ ·d ⁻¹]	Standard deviation [m ³ ·d ⁻¹]	Coefficient of variation [%]
January	21 387	21 096	11 784	34 392	4 703	21.99
February	23 345	23 153	14 040	38 496	4 416	18.91
March	25 399	24 720	13 704	48 336	5 156	20.30
April	26 024	25 104	16 608	46 536	4 911	18.87
May	27 953	26 447	19 104	56 640	5 215	18.66
June	28 745	27 012	12 216	55 512	6 831	23.76
July	30 831	30 648	16 298	48 504	6 004	19.47
August	28 601	28 307	15 840	45 936	7 302	25.53
September	24 892	24 744	10 848	41 280	6 027	24.21
October	23 762	22 272	11 136	43 200	6 067	25.53
November	21 726	20 629	11 328	46 368	5 777	26.59
December	20 781	19 920	12 288	42 264	4 635	22.30
Multi-year period 2008–2014	25 301	24 504	10 848	56 640	6 451	25.50

3. RESULTS AND DISCUSSION

Statistical analysis of the source data for the study period (1 January 2008–31 December 2014) is presented in Table 1. Average daily amount of sewage inflowing into the WTP in Nowy Sącz in the study period was 25.301 m³·d⁻¹ and was by 16% lower than the amount allowed by the water permit [Water Permit 1999]. Maximum capacity of the facility $Q_{dmax} = 42.200$ m³·d⁻¹ was exceeded usually in the spring due to snow melt and in the summer due to heavy rains. The highest number of exceedance events was observed in 2009 (12 events) and 2010 (10 events). The greatest daily flows were recorded in May 2010 (56 640 m³·d⁻¹) due to heavy rains that caused a flood. Minimum flow intensity of 10 848 m³·d⁻¹ was recorded in September 2011. High standard deviations were typical for June and August, that is, the months characterised by intense rainfalls. The highest average and median values were reported for the summer months of June, July and August, and the lowest for the winter months, that is, December and January. Maximum coefficient of variation for the investigated period (26.6%) was calculated for November and minimum (19.5%) was reported for July.

Figure 1 presents the time series reflecting the changes in daily flow of sewage inflowing into the WTP in Nowy Sącz in the years 2008–2014. This time series is characterised by significant fluctuations of the monitored parameters, slightly decreasing linear trend and a lack of clear seasonality.

In the next step, a structure of the time series was analysed for the seasonality factor. To this end, serial correlation was performed and the resulting correlogram was used to work out the relationship between individual values of the time series (Figure 2).

Analysis of the serial correlation graph revealed that the second maximum happened on 325th day of the year and the third after the next 398 days. Therefore, the analysed time series values

correlated with observations delayed by, on average, 365 days and the observed deviations were caused by mismatch of the seasons and astronomical calendar.

One year (365 days) long time delay was then used for seasonal decomposition of the analysed time series by means of Census I method. As the analysis of actual series course showed that the magnitude of seasonal irregularities did not depend on general trend level, additivity of the model was assumed. The seasonality factor was identified after subtracting the series smoothed with moving average with average period of 365 days from the actual series (Figure 3). The course of this factor for annual intervals showed that sewage flow was by 10 140 m³·d⁻¹ higher and 6 146 m³·d⁻¹ lower than the values determined by the trend.

Furthermore, seasonal irregularities were constant over time, which is why actual and smoothed (not allowing for seasonality) series had a similar course.

Next steps were aimed at identifying an optimal model of exponential smoothing, that is, a model with parameters for which the theoretical values would be as similar as possible to the observed ones.

Figure 4 shows the results of exponential smoothing for sewage flow and an annual forecast for 2014. The fit of the assumed model to actual data was verified using mean error, mean absolute error, summed square of residuals, mean percentage error and mean absolute percentage error (Table 2).

Optimisation allowed for estimating the model parameters with the following values: $\alpha = 0.522$; $\delta = 0.000$ and $g = 0.000$. Here, the value of α meant that half of the previous observations were ignored. δ equal to zero indicated high stability of the seasonality factor. Finally, g equal to zero indicated the trend component constant for all values of the time series and their forecasts.

According to the forecast, sewage flow in the first 180 days would increase ranging from 13 860 to 31 611 m³·d⁻¹. After reaching this maximum, the flow would gradually decrease to its minimum

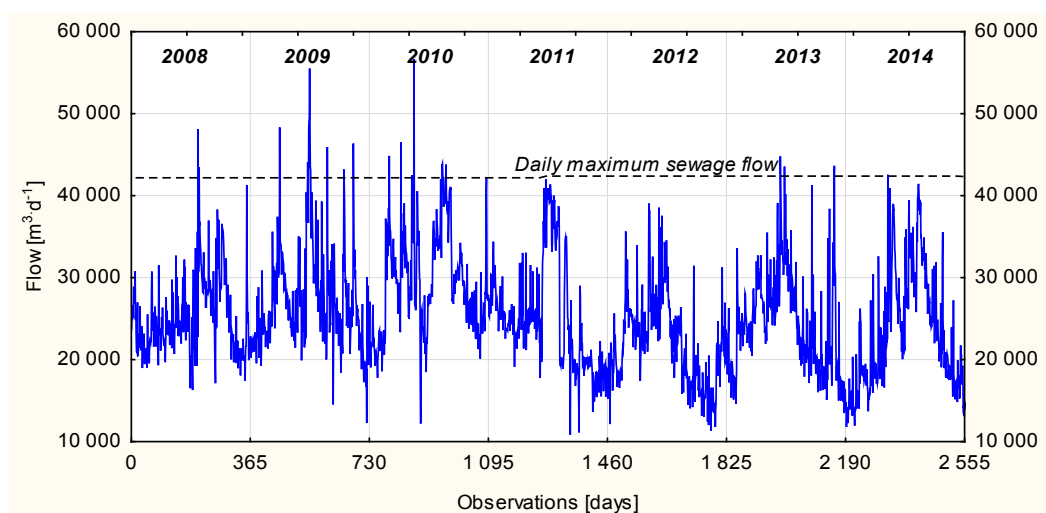


Figure 1. Time series of daily sewage amount inflowing into the wastewater treatment plant in Nowy Sącz in the years 2008–2014.

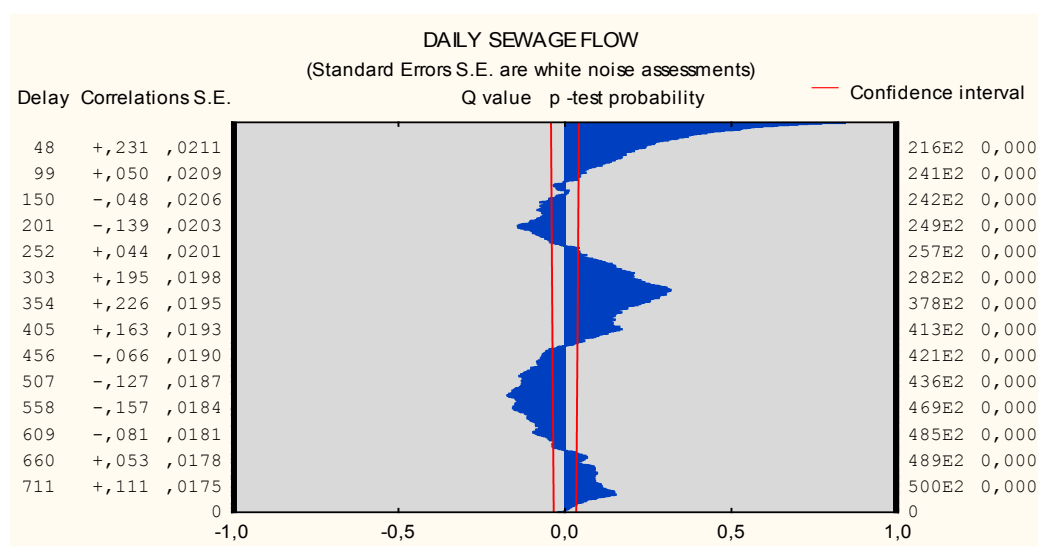


Figure 2. Autocorrelation of the series of daily sewage flow.

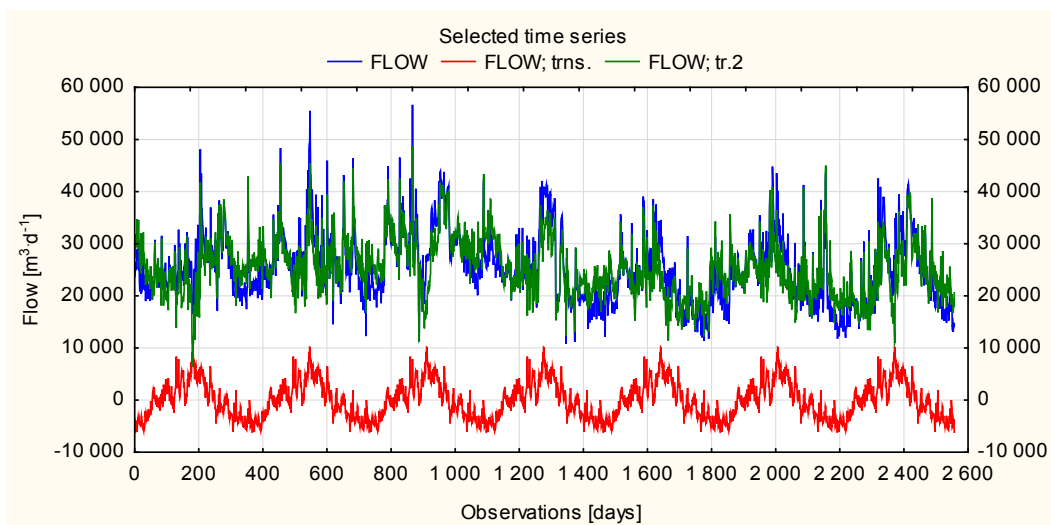


Figure 3. Seasonal decomposition results.

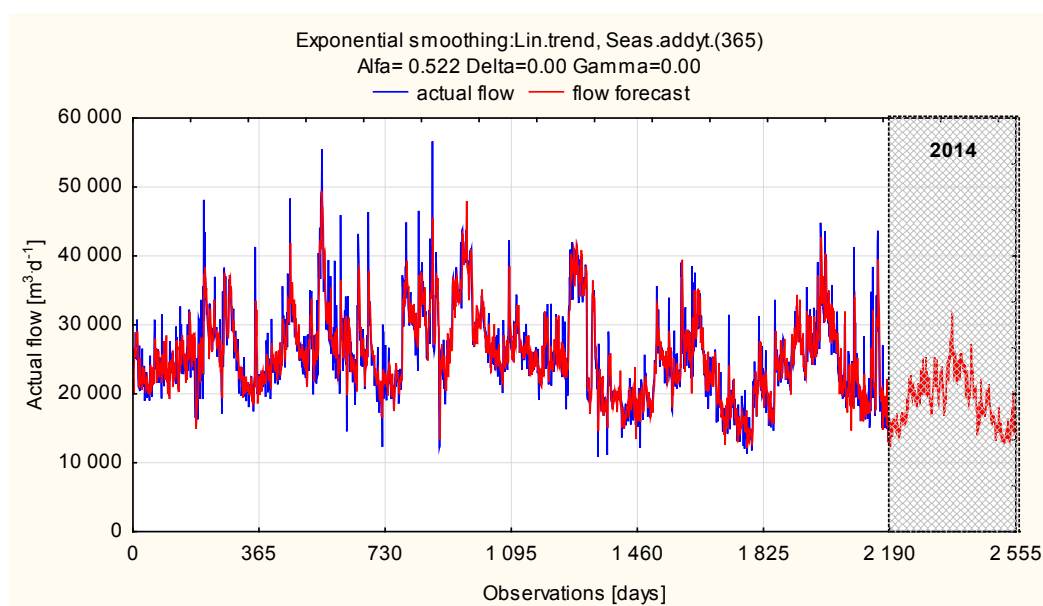


Figure 4. Exponential smoothing for daily sewage flow with 365-day forecast (forecast period for 2014 marked with grid).

Table 2. Accuracy of the model fit to the actual time series.

Goodness of fit - statistical measures of error	Error value [%]
Mean error ME	-4.522
Mean absolute error MAE	2162.1
Summed square of residual SSE	978,8438
Mean percentage error MPE	-0.983
Mean absolute percentage error MAPE	8.536

of 12 891 $\text{m}^3 \cdot \text{d}^{-1}$ on day 329 of the forecast and 13 113 $\text{m}^3 \cdot \text{d}^{-1}$ on day 365. Average daily flow intensity for an annual forecast would be 19 197 $\text{m}^3 \cdot \text{d}^{-1}$ with standard deviation amounting to 3 716 $\text{m}^3 \cdot \text{d}^{-1}$.

Despite many outliers and fluctuations in sewage flow, good values of model fit were achieved. Mean error was small, and mean percentage error of 1% was low enough to indicate usefulness of this model in predicting actual values of the series. The model quality was significantly improved by low mean absolute percentage error, which means an average forecast differs from actual values by $\pm 8.5\%$.

The final stage was a verification of the forecast for the generated model of 6-year prediction against actual values of flow intensity for 2014. Correlation coefficient between the forecast and actual data was 0.54 indicating moderate linear correlation. It is worth mentioning that correlation coefficient is shaped by extreme values and in the cases with any outliers, the correlation may be not very strong.

Coefficient of determination equal to 21% meant that for an annual period, 73 observations in a sample of total variation of flow forecast were explained by variation in the actual flow.

4. CONCLUSIONS

The study yielded the following conclusions:

1. Daily flow of sewage inflowing into the WTP in Nowy Sącz in the years 2008–2014 was characterised by high variability and annual seasonality.
2. Mean daily sewage amount was 25 301 $\text{m}^3 \cdot \text{d}^{-1}$ and was by 16% lower than the design capacity. The highest number of maximum capacity ($Q_{\text{dmax}} = 42\,200 \text{ m}^3 \cdot \text{d}^{-1}$) exceedance events in the investigated years occurred in 2009 (12 events) and 2010 (10 events) at the time of intense rainfall often resulting in a flood.
3. Holt–Winters seasonal method, based on exponential smoothing and not requiring time series stationarity, was especially useful in forecasting daily amount of sewage for annual periods.
4. Correlation coefficient for the forecast and actual data for 2014 was moderate. This was largely influenced by high number of extreme values, that is, exceedance of daily sewage flow.
5. Despite many outliers and fluctuations in sewage flow, good values of the additive Holt–Winters model fit were achieved.
6. Easily applicable method based of exponential smoothing algorithms proved useful in forecasting the amount of sewage inflowing into the Nowy Sącz WTP. This is particularly important as due to investments in water and sewage management network, municipal WTPs in Stary Sącz will be closed and from 1 July 2015, the city inhabitants will be served by the facility in Nowy Sącz [www.sadeczanin.info].

REFERENCES AND LEGAL ACTS

- BOX G., JENKINS G. 1983. Analiza szeregów czasowych. [Time series analysis] PWN, Warszawa.
- CHOTKOWSKI W., LIS G. 2006. Krótkoterminowe prognozowanie dopływu ścieków do oczyszczalni na podstawie ilości zużywanej wody. [Short-term prognosis of sewage flow to be treated based on the amount of consumed water] Gas, Water and Sanitary Technology [Gaz, Woda i Technika Sanitarna]. 5: 13-18.
- CUPAK A., KRZANOWSKI S., WAŁĘGA A. 2011. Dynamika ilości ścieków dopływających do oczyszczalni zbiorczej. [Dynamics of sewage inflow into the collective sewage treatment plant] Zeszyty Problemowe Postępów Nauk Rolniczych [Advances of Agricultural Sciences Problem Issues] 560: 55-61.
- Water Permit 1999 - Pozwolenie wodno-prawne wydane przez Prezydenta miasta Nowego Sącza** [Water permit issued by the President of Nowy Sącz].
- WAŁĘGA A., BERGEL T. 2009. Zastosowanie metod data mining w prognozowaniu zużycia wody na cele bytowe w gospodarstwach wiejskich. [Data mining implementation in household water usage forecasting in the farmhouses] Infrastruktura i Ekologia Terenów Wiejskich [Infrastructure and Ecology of Rural Areas] PAN, Oddział w Krakowie, Kraków branch of Polish Academy of Sciences. 5: 183-195.
- www.sadeczanin.info/wiadomosci,5/stary-sacz-zamykaja-oczyszczalnie-i-zwalniaja-pracownikow,73429
- www.statsoft.pl
- www.swns.pl