APPLICATION OF HOLT-WINTERS METHOD IN WATER CONSUMPTION PREDICTION

**Edward KOZŁOWSKI, Dariusz MAZURKIEWICZ,
Beata KOWALSKA, Dariusz KOWALSKI**

Summary: One of the most strategic management tasks of water supplying companies is water consumption prediction**.** This consumers' demand forecast is done using several, specially designed models which generate data necessary for planning forthcoming operational activities. However, still one can observe a lack of universal modeling method. For this reason, the authors have proposed previously a solution, wherein hourly water consumption is determined by trend analysis and harmonic analysis and now continue the research on prediction methods enabling effective forecasts with the applications of Holt-Winters method.

Key words: water supply system, water demand, time series, Holt-Winters method.

1. Introduction

Water demand forecasting in city water supply system is one of the key tasks of municipal companies. This is because they have to ensure proper standards of service with reduced operational costs, taking into consideration the most important in economic calculations - cost of energy required for pumping. Water consumption depends on a number of factors, including periodicity, time, temperature and randomness. This is why water demand prediction is done using water several consumption models specially developed for this purpose, what was in detailed analyzed in [1]. Forecasting data obtained with such models are among others used to manage, control and modernize the existing infrastructure, as well as to design new elements of the system. However, still one can observe a lack of universal modeling method. For this reason, the authors have proposed previously a solution, wherein hourly water consumption is determined by trend analysis and harmonic analysis [1], presenting two methods for modeling hourly water consumption in a water supply system. The first method involved the use of linear phase trends dependent on ambient temperature. The other was based on the Fourier transform, wherein the harmonics are linearly dependent on ambient temperature. Structural parameters of the above models were estimated based on the date regarding two lines, "Puławy East" and "Puławy West". The application of the two water demand forecasting models demonstrated that the real part of theoretical values for the Fourier transform and the values of the expected phase trends are similar. Both the matching of these models to the empirical data and the forecasts obtained with these models reflect relatively accurately the variations in daily water consumption. Some differences between the theoretical and empirical data on water consumption indicated that there exist some additional factors affecting water consumption. This is why now we continue the research on prediction methods enabling effective forecasts with the applications of Holt-Winters method, what in details is discussed in this paper. In [1] it was found that both ambient temperature and working/free days affect the water intake. Below, the Holt-Winters method will be presented to identify

the trend and seasonality in the time series. Investigations results show that this is a good tool for short-term water demand forecasts.

2. Holt - Winters method in time series analysis

Let (Ω, F, P) be a complete probability space and the time series $\{x_t\}_{1 \le t \le n}$ be defined as follows:

$$x_t = a + bt + s_t + \varepsilon_t \qquad (1)$$

where the value a presents a level around which the elements of the series $\{x_t\}_{1 \le t \le n}$ evolve, the value b corresponds to the slope factor and series $\{s_t\}_{1 \le t \le n}$ describe the seasonal effect in time series $\{x_t\}_{1 \le t \le n}$, $\{\varepsilon_t\}_{1 \le t \le n}$ is a sequence of independent identically distributed random variables (a sequence of external disturbances) with normal distribution $N(0, \sigma^2)$.

Let us assume, that the series $\{x_t\}_{1 \le t \le n}$ have a frequency $p > 1$ (the number of seasonality components per unit of time). In the presented formula (1) the elements of time series $\{x_t\}_{1 \le t \le n}$ are presented as a sum of trend, seasonal effect and random disturbances. On (Ω, F, P) we define a family of sub-σ-fields $F_t = \sigma\{x_i : i = 1, 2, .., t\}$ for $1 \le t \le n$. In order to predict the behavior of a time series, first we need to determine the level a, the slope b and the seasonal effect $\{s_t\}_{1 \le t \le n}$. The task consists in the construction of estimators of unknown parameters a, b and seasonal effect $\{s_t\}_{1 \le t \le n}$ as $\hat{a}_t = E(a|F_t) = \phi(X_t)$, $\hat{b}_t = E(b|F_t) = \varphi(X_t)$, $\hat{s}_t = E(s_t|F_t) = \psi(X_t)$, where $\phi(\ )$, $\varphi(\ )$, $\psi(\ )$ denote the statistics and the sample $X_t = \{x_1, x_2, ..., x_t\}$. The prediction of behavior of series x_{t+k} based on σ-field F_t (observation X_t) is defined as:

$$x_{t+k|t} = E(x_{t+k}|F_t) = \hat{a}_t + \hat{b}_t k + \hat{s}_{t+k'-p}, \qquad (2)$$

where $k' = \left[\dfrac{k}{p}\right]$ means the remainder of the division k by p.

The Holt-Winters method (see e.q. [2,3]) has been known for a long time. This method is based on exponential smoothing and allows us to estimate the level, slope and seasonality component in a simple way. The unknown parameters a, b, $\{s_t\}_{1 \le t \le n}$ are determined as follows:

$$\hat{a}_t = \alpha(x_t - \hat{s}_{t-p}) + (1 - \alpha)(\hat{a}_{t-1} + \hat{b}_{t-1}), \qquad (3)$$

$$\hat{b}_t = \beta(\hat{a}_t - \hat{a}_{t-1}) + (1 - \beta)\hat{b}_{t-1}, \qquad (4)$$

$$\hat{s}_t = \gamma(x_t - \hat{a}_t) + (1 - \gamma)\hat{s}_{t-p} \qquad (5)$$

with initial values $\hat{a}_1 = x_1$, $\hat{b}_1 = 0$ and $\hat{s}_j = 0$ for $1 \le j \le p$ and $0 < \alpha < 1$, $0 < \beta < 1$ and $0 < \gamma < 1$ denote the smoothing parameters. Sometimes the observation

of time series $\{x_t\}_{1\le t\le n}$ is divided into two sets $\{x_t\}_{1\le t\le mp}$ and $\{x_t\}_{mp+1\le t\le n}$. The first of them contains m full seasons (periods) of historical data and is used to estimate the initial values of level, slope and seasonal effect based on $\{x_t\}_{1\le t\le mp}$. The initial values are determined as follows:

$$\hat{a}_{mp} = \frac{1}{mp}\sum_{i=1}^{mp} x_i \ , \ \hat{b}_{mp} = \frac{x_{mp} - x_1}{mp - 1} \ \text{and} \ \hat{s}_{(m-1)p+k} = \frac{1}{m}\sum_{i=0}^{m-1} x_{k+im} - \hat{a}_{mp} \qquad (6)$$

Next, for the times $mp < t \le n$ the estimators of level, slope and seasonal effect are calculated by usage the formulas (3) – (5). From (3) – (5) we see, that the at the moment t the estimators $\hat{a}_t = \hat{a}_t(\alpha,\beta,\gamma)$, $\hat{b}_t = \hat{b}_t(\alpha,\beta,\gamma)$, $\hat{s}_t = \hat{s}_t(\alpha,\beta,\gamma)$ depend on the values of smoothing parameters α,β,γ. In accordance with above to determine the unknown parameters (level, slope and seasonal effect) we must solve the task:

$$\min_{0<\alpha,\beta,\gamma<1} J_k(\alpha,\beta,\gamma) \qquad (7)$$

where the objective function is defined as a sum of squared errors of prediction:

$$J_k(\alpha,\beta,\gamma) = \sum_{t=p+1}^{n-k}\left(x_{t+k} - x_{t+k|t}\right)^2 \qquad (8)$$

for $1 \le k \le p$. The value $J_k(\alpha,\beta,\gamma)$ denotes a sum of squared k-step-ahead prediction errors (SS^{k}PE).

3. Water demand analysis as time series data model

The data of water consumption for region *"Puławy East"* from 1/07/2015 to 31/08/2015 were used to analyze and forecast the water demand. The water consumption is different during the day, whereas if we look at the water demand during the week or month, it is not difficult to notice that the water demand has a seasonal components. For example, the Figure 1a shows the water demand from 08/07/2015 - 00:00 to 16/07/2015 - 00:00 in region *"Puławy East"*.
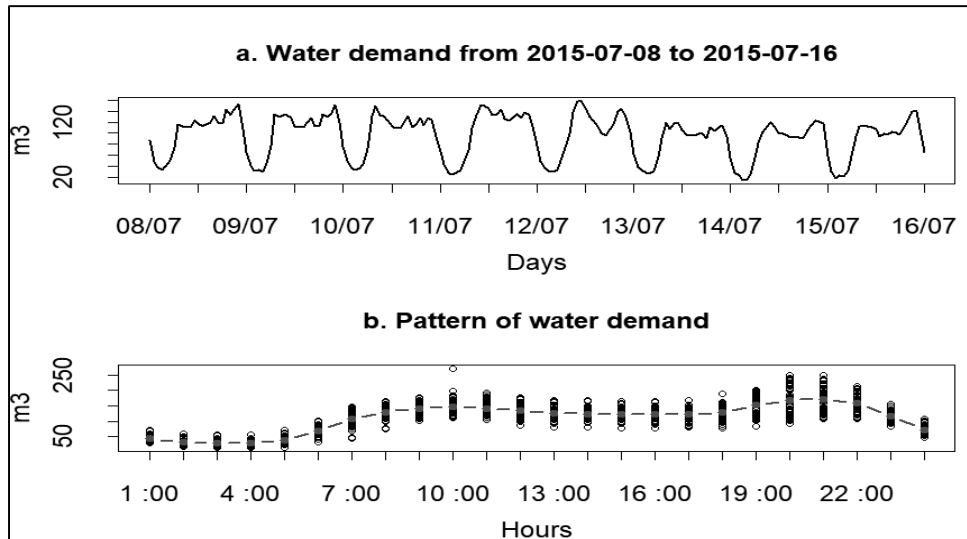
Fig. 1. Water consumption for Puławy East

The Figure 1b represents a pattern of water demand, which is based on hourly average water consumption (the dotted line).

To determine the trend component and seasonal effect in equation (1) the Holt-Winters method was applied. The initial values of level, slope and seasonal effect were estimated based on the water consumption during the first five days of July, and next the estimators of level, slope and seasonal effect were calculated with the use of the formulas (3)-(5). Taking into consideration the task (7) for performance criterion $J_1(\alpha, \beta, \gamma)$, which is defined by formula (8), it was determined the set of smoothing parameters. The optimal values of these parameters are:

$$\alpha = 0.54, \quad \beta = 0.64, \quad \gamma = 0.47.$$

This way the water demand prediction with the use of Holt-Winters method was achieved. The final result was presented on Figure 2a and Figure 2b. The black solid line (Fig. 2a) illustrates the real values of water consumption during the time period from 13/07/2015-00:00 to 19/07/2015-00:00 (as elements of the time series $\{x_t\}_{1 \le t \le n}$) whereas the red dotted line presents the expected values for 1-step-ahead prediction (as elements of time series $\{x_{t+1|t}\}_{1 \le t \le n}$).
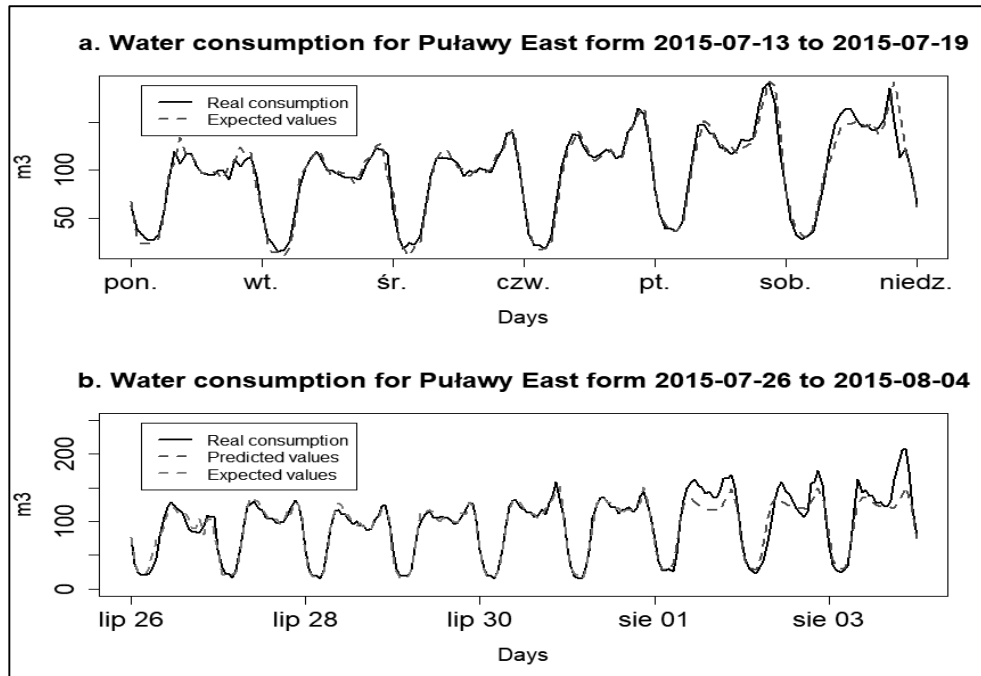
630

Fig. 2. Real and expected values of water consumption for Puławy East

Applying the Holt-Winters method we have successfully calculated the predicted values of water demand for the first three days of August. In Figure 2b the black solid line illustrate the real water consumption for the time period from 26/07/2015- 00:00   to 04/08/2015- 00:00. The green dotted line show the expected values of water consumption (as elements of time series $\left\{x_{t+1|t}\right\}_{1\leq t\leq n}$) for the time period from 26/07/2015- 00:00   to 31/07/2015- 00:00, while the red dashed curve presents the forecast for the next three days (72 hours) as elements of the series $\left\{x_{t+k|t}\right\}_{1\leq k\leq 72}$.

Analyzing the Figure 2b we can notice, that the short-time prediction with the use of Holt-Winters method is quite good, whereas the long-time prediction does not give expected effects. Analysis of estimated components sequences in the equation (1) gives a promising answer to the question why the long-term forecasts are less accurate than the short-term prediction? Figure 3 shows the realizations of sequences of values of level component $\left\{\hat{a}_t\right\}_{mp+1\leq t\leq n}$, slope component $\left\{\hat{b}_t\right\}_{mp+1\leq t\leq n}$, seasonal components $\left\{\hat{s}_t\right\}_{mp+1\leq t\leq n}$ and residuals $\left\{\varepsilon_t\right\}_{mp+1\leq t\leq n}$. On the Figure 3a and Figure 3 we can observe the realizations of sequences $\left\{\hat{a}_t\right\}_{mp+1\leq t\leq n}$ and $\left\{\hat{b}_t\right\}_{mp+1\leq t\leq n}$ (they describe the behaviors of intercept and slope estimators respectively) which have a fluctuations. Next the tests were carried out to analyze the significance and stability of the intercept estimators and slope estimators.
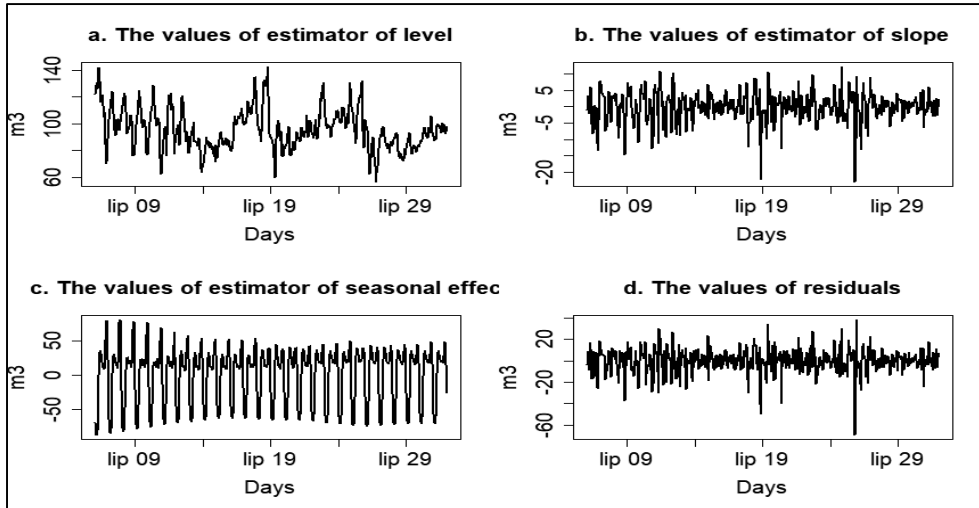
Fig. 3. The values of estimators of level, slope, seasonal and the time series of residuals

As a result it was considered the influence of the time factor on the values of estimators $\hat{a}_t$ and $\hat{b}_t$. For the sequences of estimators $\{\tilde{a}_t\}_{1\leq t\leq d}$ and $\{\tilde{b}_t\}_{1\leq t\leq d}$, where $\tilde{a}_t = \hat{a}_{t+mp}$, $\tilde{b}_t = \hat{b}_{t+mp}$ and $d = n - mp$ we have calculated and analyzed the dependences:

$$\tilde{a}_t = \lambda_0^a + \lambda_1^a t + \zeta_t^a, \tag{9}$$

$$\tilde{b}_t = \lambda_0^b + \lambda_1^b t + \zeta_t^b, \tag{10}$$

where $\{\zeta_t^a\}_{1\leq t\leq d}$ and $\{\zeta_t^b\}_{1\leq t\leq d}$ represent the external disturbances in linear formula (9) and (10) respectively. To test the significance and stability of the intercept estimators and slope we analyze and explain the existence of linear dependences in models (9) and (10) (see e.g. [4-7]). The table 1 presents the results of these investigations.

Tab. 1. Analysis of linear regression of values of intercepts and slope estimators

	i	λ_i^j	$S_{\lambda_i^j}$	$t_{\lambda_i^j}$	$P\left(|T| > \left|t_{\lambda_i^j}\right|\right)$	F_{val}	$P(F > F_{val})$	R^2
$j = a$	0	102.21	1.163	87.874	$<2*10^{-16}$	37.71	$1.466*10^{-9}$	0.057
	1	-0.0198	0.003	-6.141	$1.47*10^{-9}$			
$j = b$	0	-0.179	0.3378	-0.531	0.596	0.2149	0.6431	0.0003
	1	0.0004	0.0009	0.464	0.643			

We can see, that for the formula (9) the hypothesis that the multiple correlation coefficient is equal zero (or not significantly differs from zero) should be rejected, therefore the value of the estimator of intercept depends on the time factor. For the formula (10) we have no basis to reject the hypothesis, that the multiple correlation coefficient is equal zero (or not significantly differs from zero), therefore the value of the estimator of intercept does not depend on the time factor. From the other side we have no basis to reject the

hypothesis, that the parameter λ_0^b is equal zero.

Remark 1. Thus, using the Holt-Winters filter (3) - (5) for model (1) we have noticed that the estimator of parameter a is not stationary (the values of estimator depend on the time) and the estimators of parameter b not significantly differs from zero. Thus, the long time prediction of water demand based on the Holt-Winters method does not give a good effect.

4. Final conclusions

It is highly required to continue research on water demand prediction methods, a particular challenge for mathematical forecasting models being here long-term forecasts based on correctly selected sets of input data. This stems from the fact that aside from the random nature of water consumption, these methods should also take account of many additional factors. The results presented in this article confirm that the Holt-Winters method can be effectively used for water demand forecasting and designing controllers for water supply pumps in order to optimize the costs of maintaining water surface at the desired level. Current analysis results show that further investigation are necessary to improve to prediction performance in long-time period. Solution is expected to be found when modifying the performance criterion (8) to determine more stable estimators of intercept and slope in formula (1).

Literatura

1. Kozłowski E., Kowalska B., Kowalski D., Mazurkiewicz D.: Water demand forecasting by trend and harmonic analysis, Archives of Civil and Mechanical Engineering, 18, 2018, pp. 140-148.
2. Holt C.C.: Forecasting trends and seasonals by exponentially weighted moving averages, ONR Research Memorandum, Carnegie Institute of Technology, 52, 1957.
3. Winters P.R.: Forecasting sales by exponentially weighted moving averages, Management Science, 6, 1960, pp. 324–342.
4. Kozłowski E.: Analiza i identyfikacja szeregów czasowych. Wyd. Politechnika Lubelska, Lublin 2015.
5. Zeliaś A., Pawełek B., Wanat S.: Prognozowanie ekonomiczne. Teoria, przykłady, zadania. PWN, Warszawa, 2004.
6. Chow G.C.: Ekonometria. PWN, Warszawa, 1995.
7. Hamilton J.D.: Time Series Analysis. Princeton University Press, Princeton 1994.
8. Bakker M., Vreeburg J.H.G., Palmen L.J., Sperber V., Bakker G., Rietveld L.C.: Better water quality and higher energy efficiency by using model predictive flow control at water supply system. Journal of Water Supply; Research and Technology – AQUA, 62, 2013, pp.: 1-13.
9. Loska A., Moczulski W., Wyczółkowski R., Dąbrowski A.: Integrated system of control and management of exploitation of water supply system. Diagnostyka, 17(1), 2016, pp.: 65-74.
10. Loska, A.: Scenario modeling exploitation decision-making process in technical network systems. Eksploatacja i Niezawodnosc – Maintenance and Reliability 19 (2), 2017, pp.: 268–278, http://dx.doi.org/10.17531/ein.2017.2.15.

Dr Edward KOZŁOWSKI
Wydział Zarządzania/ Katedra Metod Ilościowych w Zarządzaniu
Politechnika Lubelska
20-618 Lublin, ul. Nadbystrzycka 38
Tel.: (0-81) 53-84-628
e-mail: e.kozlovski@pollub.pl

dr hab. inż. Dariusz MAZURKIEWICZ, prof. PL
Wydział Mechaniczny, Katedra Podstaw Inżynierii Produkcji
Politechnika Lubelska
20-618 Lublin, ul. Nadbystrzycka 36
Tel.: (0-81) 53-84-229
e-mail: d.mazurkiewicz@pollub.pl

dr hab. inż. Beata KOWALSKA, prof. PL
dr hab. inż. Dariusz KOWALSKI, prof. PL
Wydział Inżynierii Środowiska/Katedra Zaopatrzenia w Wodę i Usuwania Ścieków
Politechnika Lubelska
20-618 Lublin, ul. Nadbystrzycka 40B
Tel.: (0-81) 53-84-430
e-mail: b.kowalska@pollub.pl,
        d.kowalski@pollub.pl