AMAZONIA Investiga

Vol. 7 Núm. 17 /Noviembre - diciembre 2018/

Artículo de investigación

Statistical analysis of the urban water consumption for administrative-business sector

Análisis estadístico del consumo de agua urbana para el sector empresarial administrativo
Análise estatística do consumo de água urbana para o setor administrativo-empresarial

Recibido: 20 de septiembre de 2018. Aceptado: 11 de octubre de 2018

Written by:
Gul'naz I. Galimova (Corresponding Author)[127]
Dinar T. Yakupov[128]

Abstract

In this work statistical forecasting of daily water consumption of the administrative and business sector of the city is considered for what the technique including 5 stages is offered. According to this technique, at the first stage the choice of set of the statistics influencing water consumption volume is derivative, the basic statistical data (BSD) are collected. Further, the main statistical characteristics of ISD are calculated: estimates of population mean, average quadratic deviations, medians, asymmetries, excesses, errors of calculation of averages, asymmetries, excesses. At the following stage the regression equation is received, assessment of extent of influence of factors on water consumption on their specific weight and coefficients of elasticity is made, forecasting on the basis of the received regression equation is executed, assessment of the forecast is made by criteria of R2, MAE, MAPE, MSE, $S/\bar{y}$. At the fifth stage the additive model of forecasting on the basis of autocorrelation of basic data of a productive factor is received, forecasting on the basis of the received additive model is executed, forecast assessment is made. At the final stage the regression and auto correlated model is received, forecasting on the basis of the received additive model is executed, forecast assessment is made. On the basis of the executed research the conclusion is drawn that the regression and auto correlated model has the best values of indicators of accuracy, the regression model is a little less

Resumen

En este trabajo, se considera el pronóstico estadístico del consumo diario de agua del sector administrativo y comercial de la ciudad para lo que se ofrece la técnica que incluye 5 etapas. De acuerdo con esta técnica, en la primera etapa la elección del conjunto de estadísticas que influyen en el volumen de consumo de agua es derivada, se recopilan los datos estadísticos básicos (BSD). Además, se calculan las principales características estadísticas de la DSI: estimaciones de la media poblacional, desviaciones cuadráticas promedio, medianas, asimetrías, excesos, errores de cálculo de promedios, asimetrías, excesos. En la siguiente etapa se recibe la ecuación de regresión, se realiza la evaluación del grado de influencia de los factores sobre el consumo de agua en su peso específico y los coeficientes de elasticidad, se realiza la previsión sobre la base de la ecuación de regresión recibida, la evaluación de la previsión se realiza mediante criterios de R2, MAE, MAPE, MSE, $S / \bar{y}$. En la quinta etapa, se recibe el modelo aditivo de pronóstico basado en la autocorrelación de los datos básicos de un factor productivo, se ejecuta el pronóstico sobre la base del modelo aditivo recibido y se realiza la evaluación del pronóstico. En la etapa final, se recibe la regresión y el modelo de correlación automática, se ejecuta el pronóstico sobre la base del modelo aditivo recibido y se realiza la evaluación del pronóstico. Sobre la base de la investigación realizada, se llega a la conclusión de que la regresión y el modelo auto correlacionado tienen los mejores

[127] Kazan Federal University, Naberezhnye Chelny Institute
Email: plosatik1208@mail.ru
[128] Kazan National Research Technical University named after A.N. Tupolev – KAI

exact, the worst indicators of accuracy correspond to additive model.

Keywords: daily water consumption, time dynamic row, regression analysis, autocorrelation, forecasting.

valores de los indicadores de precisión, el modelo de regresión es un poco menos exacto, los peores indicadores de precisión corresponden al modelo aditivo.

Palabras claves: Consumo diario de agua, hilera dinámica de tiempo, análisis de regresión, autocorrelación, pronóstico.

Resumo

Neste trabalho a previsão estatística do consumo de água diário do setor administrativo e empresarial da cidade é considerada para o que a técnica que inclui 5 estágios é oferecida. Segundo esta técnica, na primeira etapa a escolha do jogo da estatística que influi no volume de consumo de água é derivada, os dados estatísticos básicos (BSD) são coletados. Além disso, são calculadas as principais características estatísticas da DSI: estimativas da média populacional, desvios quadráticos médios, medianas, assimetrias, excessos, erros de cálculo de médias, assimetrias, excessos. No seguinte passo a equação de regressão recebe-se, a avaliação do ponto da influência de fatores no consumo de água no seu peso específico e coeficientes da elasticidade faz-se, a previsão com base na equação de regressão recebida executa-se, a avaliação da previsão faz-se critérios de R2, MAE, MAPE, MSE, $S / \bar{y}$. Na quinta etapa o modelo aditivo da previsão com base na autocorrelação de dados básicos de um fator produtivo recebe-se, a previsão com base no modelo aditivo recebido executa-se, a avaliação de previsão faz-se. No estágio final, o modelo de regressão e auto-correlacionado é recebido, a previsão com base no modelo aditivo recebido é executada, a avaliação de previsão é feita. Com base na pesquisa executada conclui-se que o modelo de regressão e autocorrelacionado possui os melhores valores de indicadores de precisão, o modelo de regressão é um pouco menos exato, os piores indicadores de exatidão correspondem ao modelo aditivo.

Palavras-chave: Consumo diário de água, linha dinâmica de tempo, análise de regressão, autocorrelação, previsão.

Introduction

For the solution of the tasks connected with operational management of system of giving and distribution of water it is expedient to use short-term expected loadings of network of water supply. At the same time the character of water consumption at different categories of consumers can be a miscellaneous. Depending on economic activity, consumers are divided into sectors, one of which is the administrative and business sector.

There is a set of examples of short-term forecasting of volumes of water consumption on the basis of statistical methods. For example, in works (Howe & Linaweaver, 1967; Oh & Yamauchi, 1974; Hughes, 1980; Anderson et al, 1980; Maidment & Parzen, 1984; Maidment et al, 1985; Miaou, 1990; Zhou et al, 2000) models of multiple regression where such factors as temperature, humidity, quantity or existence of rainfall, the price for water, growth rates of the population, the average income of inhabitants and so forth are considered are considered. The main advantage of regression models - comparative simplicity. But at the same time they are a little limited in the presence in basic data of the nonlinear relations and noise. In works (Bougadis et al, 2005; Adamowski, 2008) forecasting models on a temporary dynamic row, including on the basis of autocorrelation are considered. These models do not consider objective factors and work on the found dependence of the current values of a row on previous, but are useful where the statistics on objective factors is not available. Generally, statistical methods in relation to forecasting of short-term water consumption yield quite good results.

In this work forecasting of water consumption of the administrative and business sector with application of statistical methods of the analysis is considered: regression, auto correlated, regression and auto correlated.

Methods

There is a time dynamic number of daily water consumption one of office centers of the large city from 01.01.2016 till 31.03.2017. Values from 01.01.2016 (figure 1) are used till 31.12.2016 for the analysis and creation of models of forecasting which quality is estimated by means of values of water consumption from 01.01.2017 till 31.03.2017.

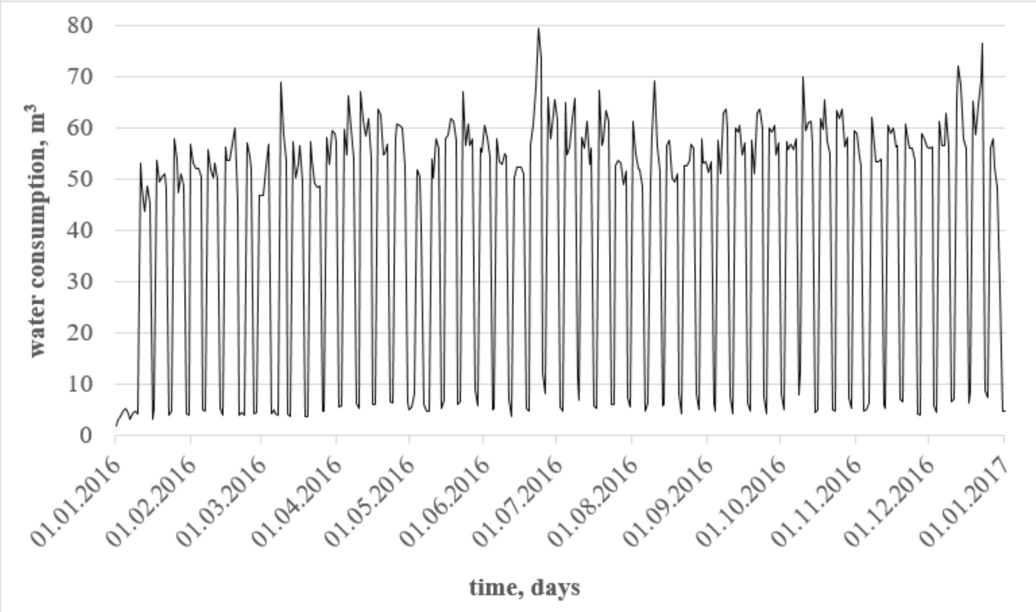

Fig. 1. Water consumption for the period 01.01.2016 - 31.12.2016

The task of development of models of forecasting of daily water consumption is set.

The first model is based on the regression equation:

$$y = f\left(x_1, x_2, \ldots x_M\right)$$

(1)

where y - the daily water consumption (a productive factor);
xi - i-y the objective factor influencing water consumption level;
M - quantity of objective factors.
The second model is described by the additive equation:

$$y = t + s + e$$

(2)

where t - trend a component;
s - seasonal component;
e - casual component.
The third model is based on association of two previous equations in expression:

$$y = \frac{f\left(x_1, x_2, \ldots x_M\right) + t + s}{2} \qquad (3)$$

For transformation of functional dependences (1-3) to an analytical look and ensuring correctness of the received results the technique which has to include the following stages is offered:

1) The choice of set of the statistics influencing the water consumption volume, collecting the basic statistical data (BSD) on them.

2) Calculation of the main statistical characteristics of ISD: estimates of population mean (average); average quadratic (standard) deviations; medians; asymmetries; excesses; errors of calculation of averages, asymmetries, excesses.

3) Receiving the equation of regression, assessment of extent of influence of factors on water consumption on their specific weight and coefficients of elasticity. Forecasting on the basis of the received regression equation, forecast assessment.

4) Receiving the VDR additive model on the basis of autocorrelation of basic data of a productive factor. Forecasting on the basis of the received additive model, forecast assessment.

5) Receiving regression and auto correlated model. Forecasting on the basis of the received model, forecast assessment.

Results and Discussion

- The choice of set of the statistics influencing water consumption volume. For the solution of a problem of daily forecasting of water consumption for the administrative and business sector of the city as the statistics influencing water consumption volume during a day of y (m3), the following factors are chosen:

x1 - average value of air temperature, To;
x2 - average value of atmospheric pressure, mm Hg;
x3 - average value of relative humidity of air, %;
x4 - average value of speed of wind, m/s;
x5 - the level of overcast, %;
x6 - the minimum value of air temperature per day, To;
x7 - the maximum value of air temperature per day, To;
x8 - an amount of precipitation per day, mm;
x9 - duration of the working day, h;

Factors h1-h8 is meteorological, and points values of an indicator x9 to a number of hours, carried out by workers (which are the main consumers of water) in a workplace.

- Calculation of the main statistical characteristics of ISD. The calculated statistical characteristics on all variables are provided in table 1.

Table 1. Basic statistical characteristics of the initial statistical data

Characteristic		Variable									
		x1	x2	x3	x4	x5	x6	x7	x8	x9	y
Mean	$\overline{v}_j$	279,678	747,426	74,508	1,549	0,789	276,566	282,954	4,805	6,096	40,231
Median	Me_j	278,344	746,850	77,563	1,000	0,950	275,375	282,000	0,600	9,000	53,122
Standard deviation	σ_j^*	10,652	6,818	13,705	0,781	0,304	9,584	12,140	12,270	4,211	24,410

Standard error of mean	S_j	0,557	0,356	0,716	0,041	0,016	0,501	0,635	0,641	0,220	1,276
Coefficient of variation	$\frac{\sigma^*_j}{\overline{v}_j}$	0,038	0,009	0,184	0,504	0,385	0,035	0,043	2,554	0,691	0,607

According to the table it is visible that in 6 cases from 10, the attitude of a standard mistake towards average value exceeds 0,05 that makes 60%. And for the x8 variable (rainfall) it is value makes 2,554. Points out great values of coefficient of a variation wide spacing of statistical data.

Let's accept that the available basic data on the attitude of a standard mistake towards average value are suitable for statistical researches. It is possible to hope that at increase in quantity of basic data of the attitude of standard mistakes towards average value will decrease.

- Regression model of forecasting. As all variables used for statistical researches are random quantitative and continuous variables, in this case application of the regression analysis is expedient.

Let's carry out the nonlinear regression analysis. As the independent variables influencing a productive indicator of y factors of xi, i=1 … 9 are chosen.

The nonlinear equation of regression connecting a productive indicator with the factors influencing it is received by means of the procedure of multiple regression of a package of application programs MATLAB.
The received equation of multiple regressions for a variable y has an appearance:

$$
\begin{aligned}
y = {} & 232{,}15 - 10{,}79 \cdot x_1 - 0{,}45 \cdot x_2 - 0{,}03 \cdot x_3 + 0{,}38 \cdot x_4 + 1{,}02 \cdot x_5 + 1{,}80 \cdot x_6 + 8{,}11 \cdot x_7 \\
& + 0{,}01 \cdot x_8 - 0{,}51 \cdot x_9 + 0{,}02 \cdot x_1^2 + 0{,}0004 \cdot x_2^2 - 0{,}23 \cdot x_4^2 - 0{,}76 \cdot x_5^2 - 0{,}0032 \cdot x_6^2 \\
& - 0{,}02 \cdot x_7^2 - 0{,}0001 \cdot x_8^2 + 0{,}69 \cdot x_9^2
\end{aligned} \tag{4}
$$

Results of assessment of influence of factors on productive indicators on coefficients of elasticity and specific weight are reduced in table 2.

Table 2. Specific weights and coefficients of elasticity of factors

Variable	d_i	e_i
x1	0,1932	0,4073
x2	0,0609	0,1283
x3	0,0198	-0,0417
x4	0,0349	-0,0735
x5	0,0024	0,0049
x6	0,0189	0,0399
x7	0,2019	-0,4256

x8	0,0062	-0,0129
x9	0,4618	0,9734

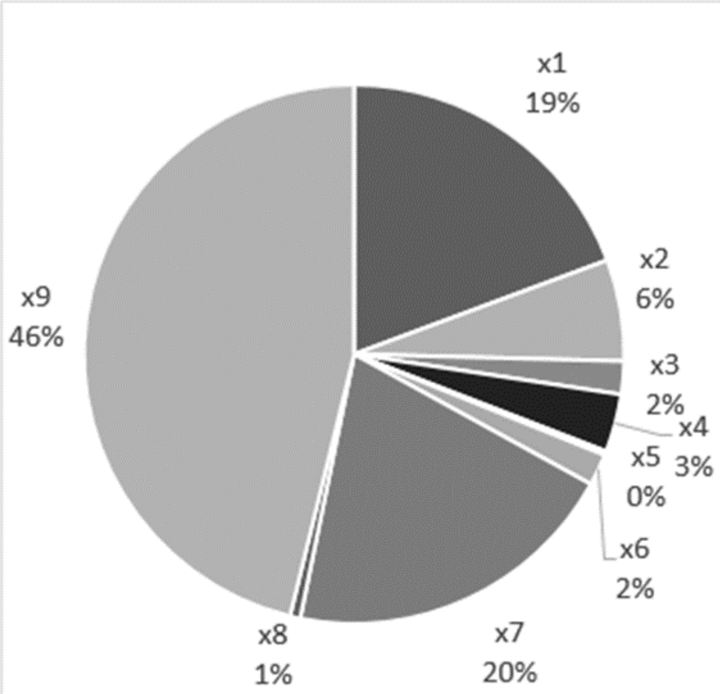

Fig. 2. The diagram of influence of factors on y by specific weights

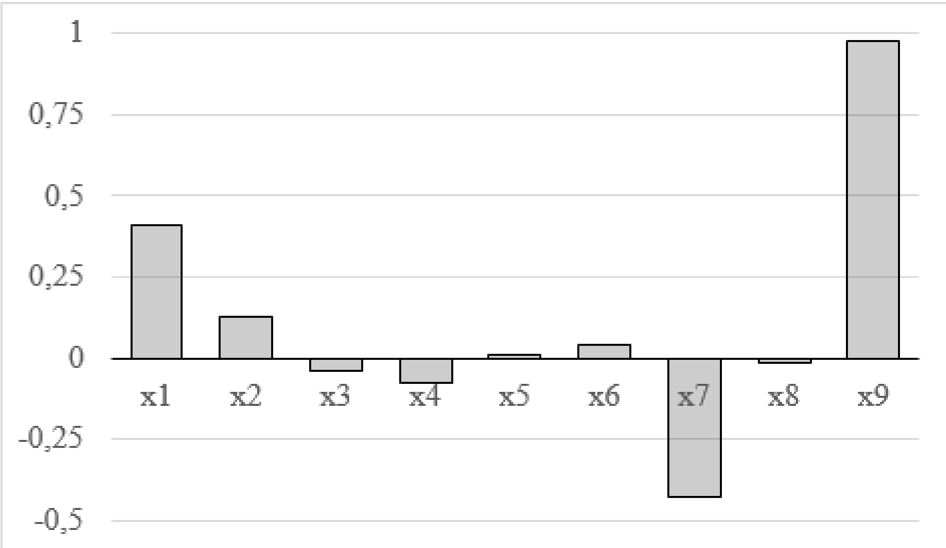

Fig. 3. The diagram of influence of factors on y by elasticity coefficients

The charts provided on figures 2, 3 allow to draw a conclusion that positive impact on water consumption in the administrative and business sector of the city is exerted: the x1st average value of air temperature, x2 - average value of atmospheric pressure, x5 - the level of overcast, x6 - the minimum value of air temperature per day, x9 - duration of the working day; negative impact on water consumption is exerted: x3 - average value of relative humidity of air, x4 - average value of speed of wind, x7 - the maximum value of air temperature per day, x8 - an amount of precipitation per day. At the same time the greatest impact is exerted by factors: x9 - duration of the working day, x7 - the maximum value of air temperature per day, x1 - average value of air temperature, x2 - average value of atmospheric pressure. Influence of other factors on a productive indicator insignificant (less than 5%).

The equation of regression (4) is used for obtaining expected values on water consumption for the period from 01.01.2017 till 31.03.2017. Schedules of the actual and predicted by regression model values of water consumption are provided on the figure 4.

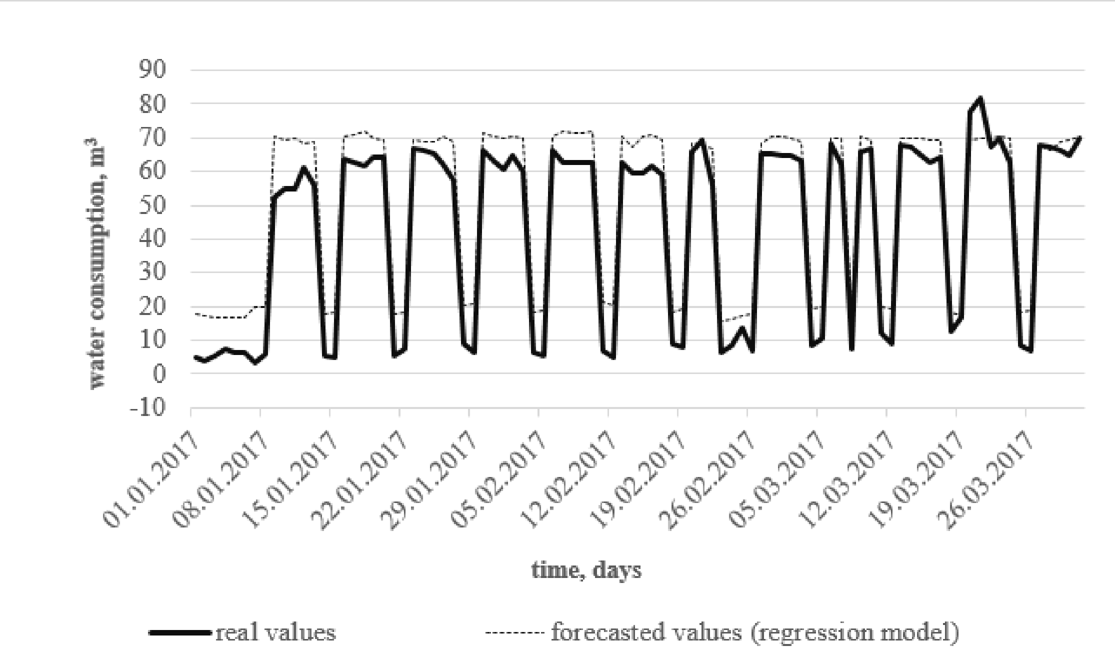

Fig. 4. Results of prediction by the regression model

Apparently from schedules, the regression model catches the jumps of values of a temporary row connected with existence of the days off. At the same time the predicted values both for workers, and for the days off exceed the actual values of water consumption.

The criterion of determination was 0,89 - it means that the additive model explains 89% of the general variation of levels of a temporary row. The average relative error of the forecast received by this model makes 70,55%, an average absolute mistake - 8,10, and the value of a mean square mistake is equal to 83,22. The attitude of a standard mistake towards average value made 2,23% that gives the grounds to consider model adequate.

- Additive model of forecasting. From the schedule of the water consumption (figure 1) it is possible to see that water consumption in the working days is much more, than during the week-end. It speaks haraktery for the administrative and business sector five-day single-shift working week. Therefore there is a sense to investigate function of water consumption on existence of statistical interrelation between the sequences of sizes of a temporary dynamic row taken with shift i.e. autocorrelations.

Coefficients of autocorrelated function are received by means of the procedure of autocorrelation of a package of application programs MATLAB.

In the figure 5 the korrelogramma on which it is visible that the nonlinear tendency and seasonal fluctuations by frequency in 7 days are characteristic of the studied temporary row is presented.

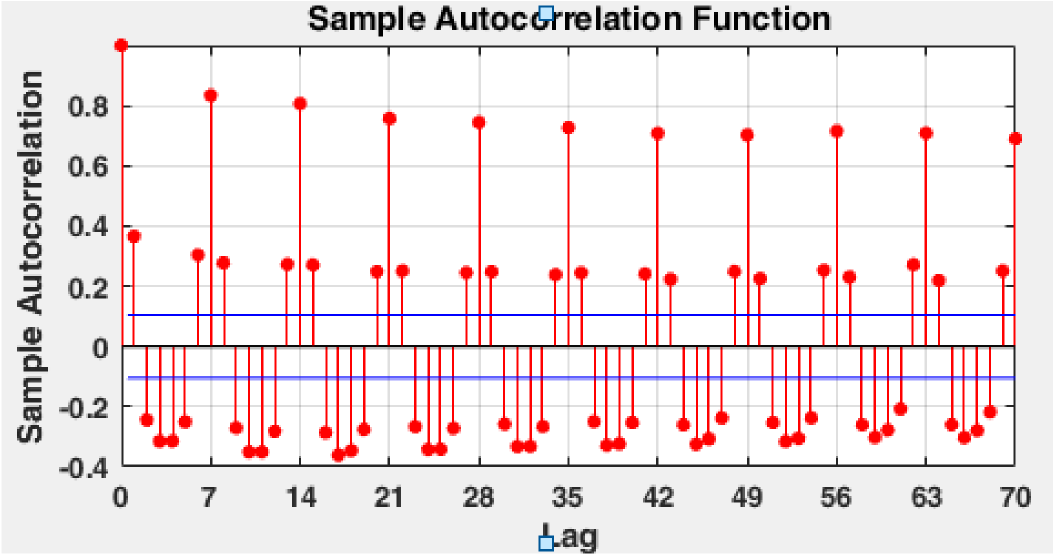

Fig. 5. Correlogram

Values of coefficients of autocorrelated function of the first 7 logs are presented in table 3.

Table 3. The coefficients of the autocorrelation function of the first 7 lags

Lag	Autocorrelation coefficient
1	0,3661
2	-0,2440
3	-0,3149
4	-0,3142
5	-0,2506
6	0,3040
7	0,8339

The additive model is described by expression $y = t + s + e$, i.e. it is supposed that a time dynamic row can be presented in the form of the sum (t)), seasonal (s) and casual (e) a component.

The frequency of seasonal fluctuations equal 7 is determined by the calculated coefficients of autocorrelated function. Alignment of initial levels of a row is carried out by a moving average method. Estimates and values seasonal components s are calculated.

Influence seasonal components is excluded. For definition trend components polynomial alignment of a row (t+e) (figure 6) therefore the trend equation is received is carried out:

$$t = 2\cdot 10\text{- } 6 \cdot j3\text{-}0{,}0012 \cdot j2+0{,}2603 \cdot j+24{,}743 \quad (5)$$

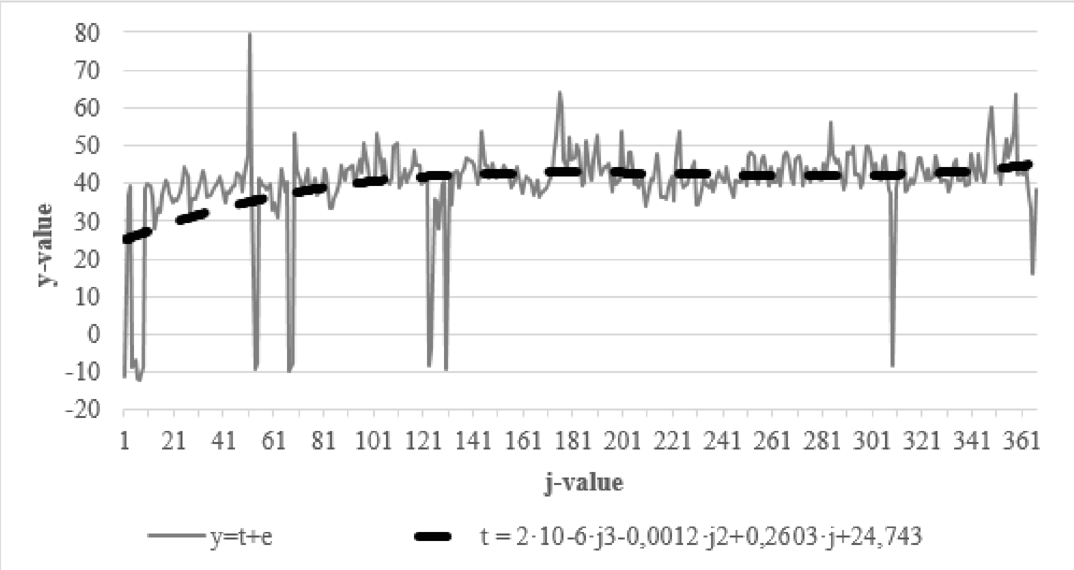

Fig. 6. The polynomial alignment of the series (t + e)

According to (5) and taking into account seasonal components are found values of levels of a row. Check of adequacy of model to data of observation is executed by Fischer's criterion.

As Frasch (1336,13)> Ftabl (3,87), additive model is statistically significant.

The additive model is used for obtaining expected values of water consumption for the period from 01.01.2017 till 31.03.2017. Schedules of the actual and predicted by additive model values of water consumption are provided on the figure 7.

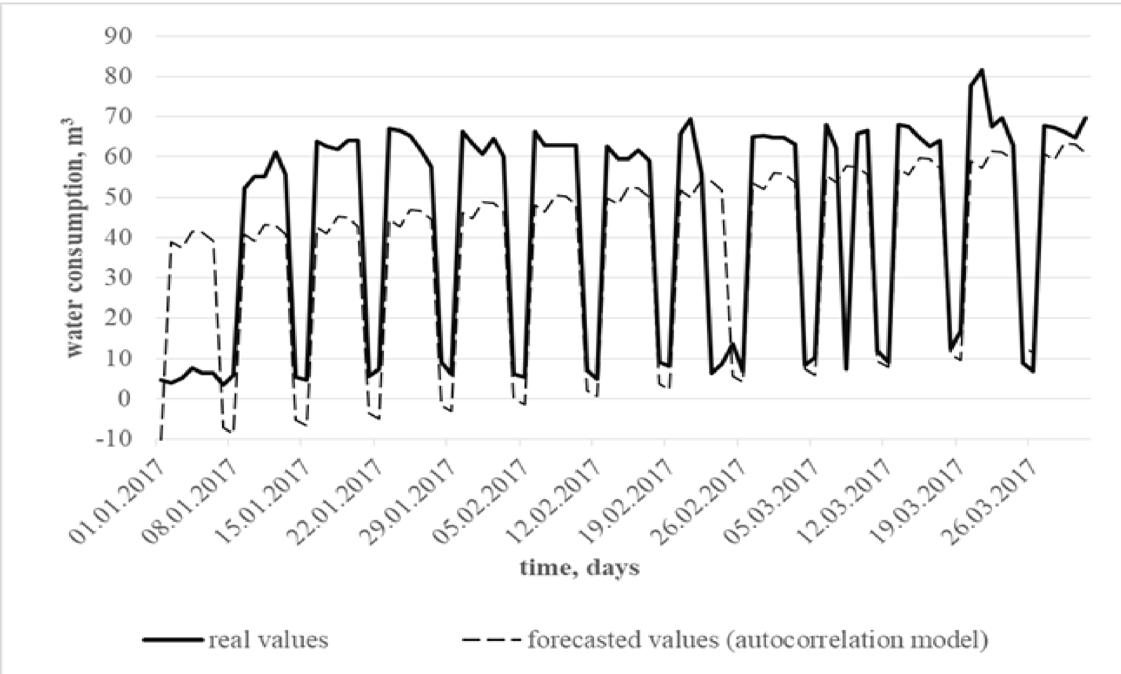

Fig. 7. Results of prediction by the additive (autocorrelation) model

Apparently from schedules, the additive model catches the jumps of values of a temporary row connected with existence of the regular days off, but does not react to long New Year's days off. At the same time the predicted values both for workers, and for the days off it is less, than the actual values of water consumption.

The criterion of determination was 0,63 - it means that the additive model explains 63% of the general variation of levels of a temporary row. The average relative error of the forecast received by this model makes 98,53%, an average absolute mistake - 13,44, and the value of a mean square mistake is equal to 278,38. The attitude of a standard mistake towards average value made 4,08% that gives the grounds to consider model adequate.

- Regression and auto correlated model of forecasting. This model of forecasting is based on finding of an arithmetic average between the expected values received on regression and additive models.
The regression and auto correlated model is used for obtaining expected values of water consumption for the period from 01.01.2017 till 31.03.2017. Schedules of the actual and predicted by regression and auto correlated model values of water consumption are provided on the figure 8.

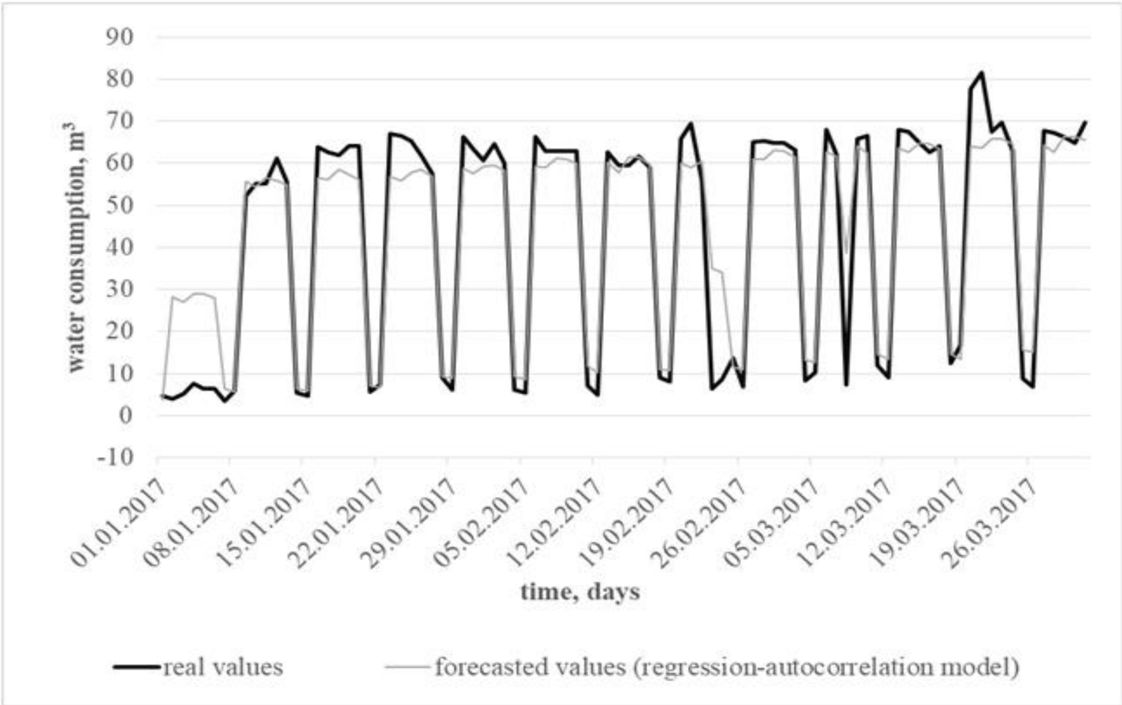

Fig. 8. Results of prediction by the regression-autocorrelation model

Apparently from schedules, the regression and auto correlated model also catches the jumps of values of a temporary row connected with existence of the days off. At the same time the predicted values for long New Year's days off have big deviations from values of the actual water consumption.

The criterion of determination was 0,90 - it means that the regression and auto correlated model explains 90% of the general variation of levels of a temporary row. The average relative error of the forecast received by this model makes 76,51%, an average absolute mistake - 5,60, and the value of a mean square mistake is equal to 76,51. The attitude of a standard mistake towards average value made 2,14% that gives the grounds to consider model adequate.

Summary

All 3 models of forecasting on the attitude of a standard mistake towards average value are adequate. Values of evaluation criteria on all models are given in table 4.

Table 4. Values of evaluation criteria of models

Forecasting model	R2	MAE	MAPE	MSE	$S_{cm}/\bar{y}$
Regression model	0,89	8,10	70,55	83,22	2,23
Additive (autocorrelation) model	0,63	13,44	98,53	278,38	4,08
Regression-autocorrelation model	0,90	5,60	50,62	76,51	2,14

Apparently from the table, the best values of indicators of accuracy at regression and auto correlated model. The regression model is a little less exact. The worst indicators of accuracy correspond to additive model. Schedules of the actual and predicted by all models values of water consumption are provided on the figure 9.

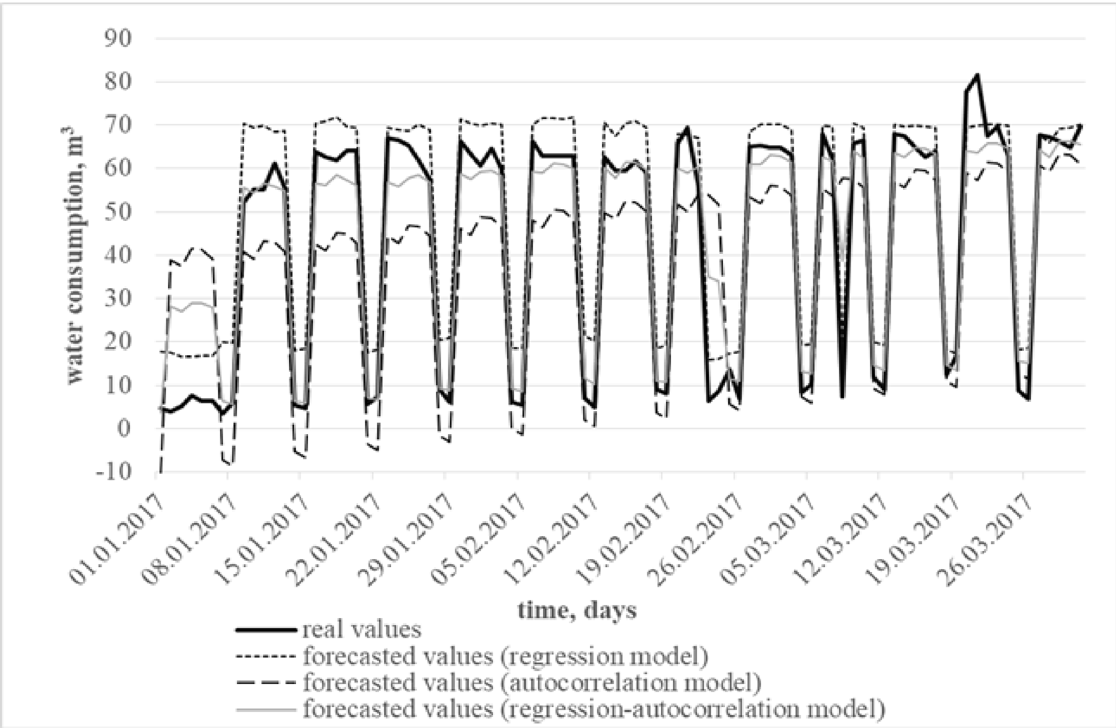

Fig. 9. Results of prediction by the regression, autocorrelation and regression-autocorrelation models

On the basis of the received results the research of opportunities ARIMA models and artificial neural networks is supposed further when forecasting daily water consumption of the administrative and business sector of the city.

Conclusions

In this work statistical forecasting of daily water consumption of the administrative and business sector of the city is considered by 3 methods: regression, auto correlated and regression and auto correlated for what the technique consisting of 5 stages is offered. 3 models of forecasting which adequacy is estimated on the attitude of a standard mistake towards average value are received. On the basis of the executed research the conclusion is drawn that the regression and auto correlated model has the best values of indicators of accuracy, the regression model is a little less exact, the worst indicators of accuracy correspond to additive model.

Acknowledgement

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

Reference

Adamowski, J. F. (2008). Peak daily water demand forecast modeling using artificial neural networks. Journal of Water Resources Planning and Management, 134(2), 119-128.

Anderson, R. L., Miller, T. A., & Washburn, M. C. (1980). Water savings from lawn watering restrictions during a drought year, Fort Collins, Colorado 1. JAWRA Journal of the American Water Resources Association, 16(4), 642-645.

Bougadis, J., Adamowski, K., & Diduch, R. (2005). Short-term municipal water demand forecasting. Hydrological Processes: An International Journal, 19(1), 137-148.

Howe, C. W., & Linaweaver, F. P. (1967). The impact of price on residential water demand and its relation to system design and price structure. Water Resources Research, 3(1), 13-32.

Hughes, T. (1980). Peak period design standards for small western US water supply systems 1. Journal of the American Water Resources Association, 16(4), 661-667.

Maidment, D. R., Miaou, S. P., & Crawford, M. M. (1985). Transfer function models of daily urban water use. Water Resources Research, 21(4), 425-432.

Maidment, D., & Parzen, E. (1984). Monthly water use and its relationship to climatic variables in Texas. Water Resour. Bull, 19(8), 409-418.

Miaou, S. P. (1990). A class of time series urban water demand models with nonlinear climatic effects. Water Resources Research, 26(2), 169-178.

Oh, H. S., & Yamauchi, H. (1974). An economic analysis of the patterns and trends in water consumption within the service area of the Honolulu Board of Water Supply. Univ. of Honolulu Rep, 84.

Zhou, S. L., McMahon, T. A., Walton, A., & Lewis, J. (2000). Forecasting daily urban water demand: a case study of Melbourne. Journal of hydrology, 236(3-4), 153-164.