Analysis of the time series of waste water quality at the inflow of the wastewater treatment plant and transfer functions

Ivan Nesmerak[1], Sarka D. Blazkova[2*]

[1] Former researcher at T. G. Masaryk Water Research Institute, Podbabska 30, 160 00, Prague, Czech Republic.
E-mail: ivan.nesmerak@seznam.cz
[2] T. G. Masaryk Water Research Institute, Podbabska 30, 160 00, Prague, Czech Republic.
[*] Corresponding author. Tel.: +420 724 602 956. E-mail: sarka_blazkova@vuv.cz

Abstract: Time series of the daily total precipitation, daily wastewater discharges and daily concentrations and pollution loads of BOD_5, COD, SS, $N-NH_4$, N_{tot} and P_{tot} were analyzed at the inflow to the wastewater treatment plant (WWTP) for a larger city in 2004–2009 (WWTP is loaded by pollution from 435,000 equivalent inhabitants). The time series of the outflow from a WWTP was also available for 2007. The time series of daily total precipitation, daily wastewater discharges, concentrations and pollution loads at the inflow and outflow from the WWTP were standardized year by year to exclude a long-term trend, and periodic components with a period of 7 days and 365 days (and potentially also 186.5 days) were excluded from the standardized series. However, these two operations eliminated only a small part of the variance; there was a substantial reduction in the variance only for ammonium nitrogen and total nitrogen at the inflow and outflow from a WWTP. The relationship between the inflow into a WWTP and the outflow from a WWTP for the concentrations and pollution loads was described by simple transfer functions (SISO models) and more complicated transfer functions (MISO models). A simple transfer function (SISO model) was employed to describe the relationship between the daily total precipitation and the wastewater discharge.

Keywords: Analysis of time series; Wastewater treatment plant; Inflow and outflow; Concentrations and pollution loads; Transfer functions.

INTRODUCTION

The modeling of water quality in water courses, in contrast to the past, is no longer confined to modeling the steady-state annual conditions, but rather entails either models based on a monthly step (where the input into the model consists in the monthly average of wastewater discharges and concentrations of water quality indicators or the monthly averages of the discharged amounts of substances in g/s or kg/day) or on a daily step (where the input into the model consists in daily values). However, the daily values are not available for most WWTP in the Czech Republic and thus it is necessary to model these daily values and it is simultaneously necessary to base the calculation on the statistical distribution of the wastewater discharges and the concentrations of substances or pollution loads.

A number of authors employ methods of time series analysis of indicators of wastewater quality at a WWTP to create a control chart for managing the treatment process. Alwan (1992), Alwan and Roberts (1988), Berthouex (1989), Van der Wiel (1996), Wardell et al. (1992) and others dealt with this aspect.

Berthouex et al. (1976) employed the method of time series analysis to analyze 1339 daily measurements at a WWTP in Madison for the January 1971–August 1974 period. The evaluation included the water temperature, wastewater discharge (and the retention time calculated from it) and BOD_5 at the inflow and outflow from the WWTP.

Tiao and Box (1981) theoretically analyzed the problem of modeling multi-parameter time series. In the general model of multi-parameter time series, the j-th time series is affected not only by previous values in the j-th series, but also by previous values in other series. They were looking for a solution for the ARMA multi-parameter model.

Umashankar and Ledolter (1983) resolved the problem of modeling multi-parameter time series using the simplifying assumption that the autocorrelation matrices and parameter MA matrices are diagonal. In the diagonal model of multi-parameter time series, the j-th time series is affected only by previous values in the j-th series, but not by previous values in other series; however mutual correlation of random members of the time series is assumed.

Ellis et al. (1990) analyzed the time series of wastewater quality at the inflow into a WWTP with a length of 92 daily values. The wastewater discharge, BOD_5, SS, NH_3 and total organic nitrogen were analyzed and the ARMA model and multi-parameter ARMA model were employed. The wastewater discharge at time t was correlated with the daily total precipitation at time t and the wastewater discharge at time $t-1$.

Novotny et al. (1991) employed the time series analysis to describe the process occurring at a WWTP. They employed both the ARMA transfer function which can be used to describe a system of linear or nonlinear equations (e.g., such as the inflow-outflow relationship for BOD_5 or suspended solids) and also neuron grids/neural networks.

Zheng and Novotny (1991) modeled the relationship between precipitation and the wastewater discharge at the inflow into a WWTP. The ARTF (7,7) and SISO (7,7,2) models were employed for data from the Fusina WWTP for Venice, Italy.

Capodaglio et al. (1992) evaluated annual daily measurements at a WWTP for 155,000 equivalent inhabitants in Green Bay, Wisconsin. The evaluation was related primarily to the wastewater discharge and the concentrations of suspended solids at the outflow, concentrations of sludge in activation, the sludge index and the concentration of returned sludge. The daily total precipitation was considered as an independent variable (for calculation of the transfer function model).

Ellis et al. (1993) evaluated the time series of the concentrations of BOD_5, NH_3, TON and pollution loads of BOD_5 at the inflow and the concentration of BOD_5 at the outflow from the WWTP for Albany and described them using the linear uncorrelated model, the ARMA model [or ARMA (1,0) model] for each indicator separately and the multivariate ARMA model with diagonal matrices. The data were taken for the June 1, 1988–August 31, 1988 period.

Berthouex and Box (1996) were concerned with prediction of the water quality at the outflow from a WWTP 1–5 days in advance. The time series were modeled using the EWMA (exponential weighted moving average) model. The models were used for data from 3 WWTPs. The output from the model, y, was considered to be BOD_5 or suspended solids at the outflow from the WWTP and the input into the model, x, consisted of a total of 14 parameters (e.g. water temperature, BOD_5 or suspended solids at the inflow into the WWTP).

Van Dongen and Geuens (1998) also employed the ARMA multi-parameter model and compared the results with the ARMA single-parameter model. They demonstrated the method on COD values.

Huo et al. (2005) employed time series analysis to evaluate the amount and quality of wastewater at the inflow into the WWTP for Oak Ridge, Tennessee. The measurement was performed for the wastewater discharge, water temperature, BOD_5, suspended solids and ammonium nitrogen from November 1, 1999 to July 31, 2002. The dependence of BOD_5 on the wastewater discharge was described by a transfer function, consisting only of first-order members.

Escalas-Cañellas et al. (2008) evaluated the time series of temperatures at the inflow into a lagoon at a municipal WWTP as the input parameter for calculation of the water temperature in the lagoon. A model including previous values of the water temperature, air temperature and daily precipitation and their previous values was created for the input temperature.

Computing methods

Computations were carried out in MATLAB (MathWorks, 2000), with the first author's own programs and with procedures of the CAPTAIN Toolbox (Pedregal et al., 2004).

Adjustment/Modification of the time series measured at the wastewater treatment plant

The analyses were performed for the time series of daily total precipitation, amounts of waste waters and concentrations as follows:

- series after supplementing the missing concentration values;
- the supplemented series centered year by year to exclude a long-term trend;
- the supplemented series standardized year by year to exclude a long-term trend;
- the standardized series after eliminating periodic components.

In some years, the time series were incomplete (especially values for Saturday were missing) and the missing values were supplemented by interpolation. The time series for the $N\text{-}NH_4$ concentrations was available only for 2005–2009 and the N_{tot} time series was available only for 2007–2009.

The basic statistical characteristics of the completed time series at the inflow into the wastewater treatment plant

Table 1 gives the basic statistical characteristics of the wastewater discharge Q (in m^3/day) and the concentrations at the inflow into the WWTP (in mg/l) for the 2004–2009 period; period average of daily precipitation was 1.496 mm/day. As an example, Fig. 1 depicts the inflow and outflow of the $N\text{-}NH_4$ concentrations in 2007.

Table 1. The basic statistical characteristics of the wastewater discharge Q (in m^3/day) and the concentrations at the inflow into the WWTP (in mg/l) for the 2004–2009 period.

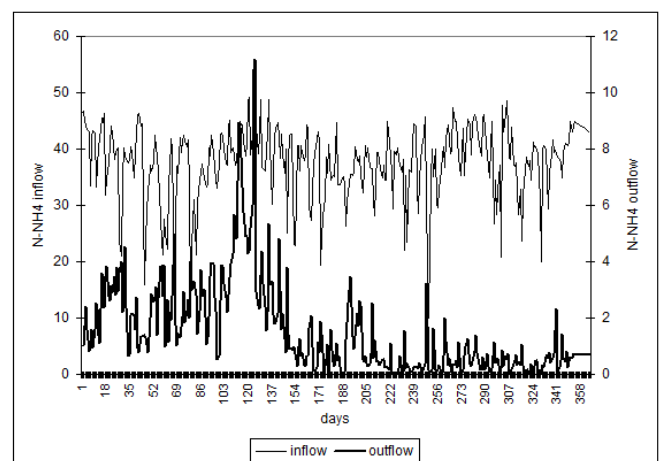|	Q	BOD_5	COD	SS	$N\text{-}NH_4$	N_{tot}	P_{tot}
number	2192	2192	2192	2192	1826	1096	2192
average	98021	264	2192	328	36.3	56.7	7.70
median	88949	265	595	320	37.6	57.7	7.72
variance	1.08E+09	4712.7	18882.1	6962.3	62.86	149.81	2.863
std. dev.	33007	69	137	83	8.1	12.4	1.69
var.coeff.	0.337	0.260	0.232	0.254	0.222	0.219	0.220

Fig. 1. Concentrations of the $N\text{-}NH_4$ inflow and outflow in 2007.

Elimination of periodic components of the time series

Because primarily the time series at the inflow into the WWTP were evaluated, it is necessary to anticipate certain periodic influences in the life of the city (e.g. weekly life cycle, summer vacations, etc.). The basic and secondary frequencies (individual components) can be found by harmonic analysis or using a periodogram.

While all the periodograms were calculated for the supplemented and standardized time series (i.e. for Q and the concentrations for 6 indicators) using the first author's procedure PERIOD, practically only three periods were finally employed (T=365 days and T=7 days, or T=182.5 days).

The selected periods (T_1=365 days and T_2=7 days, or T_3=182.5 days) correspond to frequencies f_1=0.0027, f_2=0.1429 and f_3=0.0055. Short periods (high frequencies) can be read off the complete periodogram and long periods (low frequencies) from the section of the periodogram.

Standard deviations of variously adjusted time series in 2007

Table 2 gives the standard deviations of variously adjusted time series for wastewater discharge Q and the concentrations at the inflow and outflow of the WWTP in 2007 (in m^3/day and mg/l, resp.).

Table 2. The standard deviations of variously adjusted time series for the concentrations at the inflow and outflow of the WWTP in 2007 (in m^3/day and mg/l, resp.).

	standard deviation							
	inflow				outflow			
	[1]	d[2]	n[3]	nc[4]	[1]	d[2]	n[3]	nc[4]
Q	27422	27422	1.0000	0.9270				
BOD_5	68.4	68.4	1.0000	0.9900	1.99	1.99	1.0000	0.9258
COD	125.3	125.3	1.0000	0.9324	9.7	9.7	1.0000	0.8751
SS	75.4	75.4	1.0000	0.9843	4.97	4.97	1.0000	0.9092
N-NH_4	6.36	6.36	1.0000	0.9649	1.655	1.655	1.0000	0.7156
N_{tot}	10.28	10.28	1.0000	0.9419	2.840	2.840	1.0000	0.7428
P_{tot}	1.54	1.54	1.0000	0.9799	0.198	0.198	1.0000	0.9249

[1] original time series after supplementing missing values;
[2] centered time series;
[3] standardized time series;
[4] standardized time series after elimination periodic components.

It can be seen from Table 2 that the standard deviations (and thus the variance too) of the concentrations at the inflow into the WWTP decrease only slightly after elimination the periodic components of the time series; the reduction in the standard deviations (and thus the variance too) is significant at the outflow from the WWTP for N-NH_4 and N_{tot}. The relatively substantial fluctuations of the concentrations of ammonium and total nitrogen during the year are also affected by the temperature of the wastewater; however, these values were not available.

Simple transfer functions

The sewer system + WWTP can be considered to constitute two simple linear SISO (single-input single-output) models or one more complicated linear MISO (multiple-input single output) model. Simple linear SISO models can be described by Eq. (1) or by transfer function model (Box and Jenkins, 1976) according to Eq. (2).

$$y_t + a_1 * y_{t-1} + ... + a_n * y_{t-n} = b_0 * x_t + \\ + b_1 * x_{t-1-\delta} + ... + b_m * x_{t-m-\delta} + N_t \qquad (1)$$

$$y_t = \frac{B(L)}{A(L)} x_{t-\delta} + \frac{1}{C(L)} e_t \quad , \qquad (2)$$

where y_t is the output from the linear model at time t; x_t the input into the model at time t (i=1, 2,…, k); δ the value of the time shift (delay) of the input compared to time t; N_t the value of a correlated random variable at time t; e_t the value of a random variable with standard distribution with zero average and constant scatter at time t.

$A(L)$, $B_i(L)$ and $C(L)$ are polynomials of order n, m and p and L is the backward shift operator ($L^j y_t = y_{t-j}$). Then $A(L)$, $B(L)$ and $C(L)$ are polynomials of type

$$A(L) = 1 + a_1 L + a_2 L^2 + ... + a_n L^n$$
$$B(L) = b_0 + b_1 L + b_2 L^2 + ... + b_m L^m$$
$$C(L) = 1 + c_1 L + c_2 L^2 + ... + c_p L^p .$$

In the first SISO model, the input is the time series of daily total precipitation values and the output is the time series of the daily wastewater discharge; in the second SISO model, the input is the time series of concentrations at the inflow into the WWTP and the output is the time series of concentrations at the outflow from the WWTP. The RIVID procedure was used to identify the structure of the model and the RIV procedure was employed to determine the parameters of the model; both procedures are part of MATLAB. A suitable model was selected according to the YIC identification criterion (Pedregal et al., 2004).

The values of calculated coefficients a, b, c and the time shift δ of the transfer functions of the first SISO model (daily rainfall precipitation RP – amount of waste waters Q) and the second SISO model (concentrations at the inflow and outflow) for the standardized series after elimination of the periodic components are given in Tables 3.

Table 3. Coefficients a, b, c in the TF selected according to the YIC criterion for Q and the concentrations.

	$A(L)$	$B(L)$		$C(L)$
	a_1	b_0	δ	c_1
$Q = f(RP)$		0.6754		
BOD_5		0.2041		
COD		0.2885		
SS		0.1909		–0.5520
N-NH_4		0.1442	1	
N_{tot}	–0.9178	0.0725		
P_{tot}		0.2138		

The determination coefficient was the highest for the relation between the daily precipitation RP and the amount of waste water Q ($R^2=0.5310$) and for the relation between SS in the inflow and the outflow from a WWTP ($R^2=0.3330$) and the lowest for the relation for N-NH_4 in the inflow and outflow from a WWTP ($R^2=0.0379$).

More complicated transfer functions

A WWTP can be considered to be a more complicated linear MISO (multiple inputs, single output) model with two inputs. More complicated linear MISO models can best be described by transfer functions according to Eq. (3). The inputs into the model consist in the concentrations at the inflow into the WWTP and the wastewater discharge and the output from the model is the concentration at the outflow from the WWTP; here, the wastewater discharge has a similar effect as the parameter of retention time in the WWTP.

$$C_{out,t} = \frac{B_1(L)}{A(L)} Q_{t-\delta_Q} + . \frac{B_2(L)}{A(L)} C_{in,t-\delta_C} + \frac{1}{C(L)} e_t , \qquad (3)$$

where $C_{out,t}$ is the output from the linear model (concentration at the outflow from the WWTP) at time t; Q_t is 1st input into the model (wastewater discharge at the inflow into the WWTP) at time t; δ_Q is the value of the time shift (delay) of the 2nd input compared to time t; $C_{in,t}$ is 2nd input into the model (concentration at the inflow into the WWTP) at time t; δ_C is the value of the time shift (delay) of the 1st input compared to time t; e_t the value of a random variable with standard distribution with zero average and constant variance at time t.

The values of calculated coefficients a, b, c and the time shift δ of the transfer functions of the MISO model (concentrations and wastewater discharge Q) for the standardized series after elimination of the periodic components are given in Tables 4.

Table 4. Coefficients a, b, c in the TF selected according to the YIC criterion for the concentrations according to Eq. (3).

	$A(L)$	$B_1(L)$		$B_2(L)$		$C(L)$		
	a_1	b_0	δ_Q	b_0	δ_{LO}	c_1	c_2	c_3
BOD_5	–0.7602	–0.0794		0.1537	1	–0.3402		
COD	–0.8284	–0.1274	1	0.0902				
SS		–0.1766	1	0.1461				
$N\text{-}NH_4$		0.2617		0.2280				
N_{tot}		0.1433		0.0837		–0.7495	–0.0868	–0.0265
P_{tot}		–0.1603	1	0.1497				

The determination coefficient was the highest for the relation between $N_{tot}+Q$ in the inflow and N_{celk} in the outflow from a WWTP (R^2=0.7076) and the lowest for the relation $N\text{-}NH_4+Q$ in the inflow and outflow from a WWTP (R^2=0.0623).

Because the determination coefficients for the more complicated transfer functions are highest than the those for the simple transfer functions for all the wastewater quality indicators (with the exception of concentration of suspended solids), it is advantageous to consider more complicated transfer functions, i.e. transfer functions where one input is the concentration in the inflow into the WWTP and the other input is the wastewater discharge. The output is then the concentration in the outflow from the WWTP.

Conclusion

The wastewater discharge and concentrations at the inflow into the WWTP undergo quite substantial fluctuations during the year (the coefficient of variation in the amount of waste waters is 0.337 and for the concentration equals about 0.230) with slight changes between the years. Periodic changes occur during the year and week. The most marked changes during the year occur in ammonium nitrogen. These periodic changes can be described and eliminated from the time series.

Long-term trends were eliminated using annual centering (calculation of deviations from the annual averages) or annual standardization. The elimination of periodic components leads to a relatively small reduction in the variance of the time series; a more substantial reduction was found for ammonium nitrogen and for total nitrogen.

When the Young identification criterion YIC is used to determine simple transfer functions for the relationship of the daily rainfall precipitation values and the wastewater discharge and also for the relationship of the concentrations at the inflow and outflow of the WWTP, simple forms of the transfer functions were obtained (for the RP–Q relationship and for most indicators were $A(L)$=0 and $C(L)$=0).

When the Young identification criterion YIC was used to determine more complicated transfer functions, where the input into the model is the concentration and the wastewater discharge and the output is the concentration, more complicated forms of the transfer functions were obtained, in which significant values of $A(L)$ and $C(L)$ were obtained.

List of notations

ARMA = Autoregressive moving average model
ARTF = Autoregressive transfer function model
BOD_5 = Biological oxygen demand (5 days)
C = Concentration
COD = Chemical oxygen demand
EWMA = Exponential weighted moving average
L = Backward shift operator
MA = Moving average model
MISO = Multiple input single output model
N_t = Coloured noise at time t (correlated random variable at time t)
$N\text{-}NH_4$ = Ammonium nitrogen
N_{tot} = Total nitrogen
P_{tot} = Total phosphorus
Q = Wastewater discharge
RP = Rainfall precipitation
R^2 = Determination coefficient
SISO = Single input single output model
SS = Suspended solids
T = Time in days
WWTP = Wastewater treatment plant
YIC = Young identification criterion
a = Coefficient in A(L) polynomial
b = Coefficient in B(L) polynomial
c = Coefficient in C(L) polynomial
e_t = White noise at time t (uncorrelated random variable at time t)
f = Frequency
t = Time
x = Measured input in the model
y = Measured output form the model
δ = Time shift (delay)

REFERENCES

Alwan, L.C., Roberts, H.V., 1988. Time-series modelling for statistical process control. Journal of Business & Economic Statistics, 6, 1, 87–95.

Alwan, L.C., 1992. Effect of autocorrelation on control chart performance. Commun. Statist.-Theory Meth., 21, 4, 1025–1049.

Berthouex, P.M., Hunter, W.G., Pallesen, L., Shih, C.Y., 1976. The use of stochastic models in the interpretation of historical data from sewage treatment plants. Water Research, 10, 8, 689–698.

Berthouex, P.M., 1989. Constructing control charts for wastewater treatment plant operation. Journal WPCF, 61, 9, 1534–1551.

Berthouex, P.M., Box, G.E., 1996. Time series models for forecasting wastewater treatment plant performance. Water Research, 30, 8, 1865–1875.

Box, G.E.O., Jenkins, G.M., 1976. Time series analysis forecasting and control. Revised Edition, Holden Day, San Francisco.

Capodaglio, A.G., Novotny, V., Fortina, L., 1992. Modelling wastewater treatment plants through time series analysis. Environmetrics, 3, 1, 99–120.

Ellis, G.W., Ge, X., Grasso, D., 1990. Time series analysis of wastewater quality. In: Instrumentation, Control and Automation of Water and WasteWater Treatment and Transport Systems, 5th Workshop of the International Association on Water Pollution Research and Control, Yokohama, Japan Jul 26–Aug 03, 441–448.

Ellis, G.W., Ge, X., Grasso, D., 1993. ARMA processes and reliability-based design of wastewater-treatment facilities. Journal of Environmental Engineering, 119, 3, 463–477.

Escalas-Caňellas, A., Ábrego-Góngora, C.J., Barajas-López, M.G., Houweling, D., Comeau, Y., 2008. A time series model for influent temperature estimation: application to dynamic temperature modelling of an aerated lagoon. Water Research, 42, 10–11, 2551–2562.

Huo, J.H., Seaver, W.L., Robinson, R.B., Cox, Ch.D., 2005. Application of time series models to analyse and forecast the influent components of wastewater treatment plants (WWTPs). In: Impact of Global Climate Change, Proceeding of World Water and Environmental Resources Congress 2005.

MathWorks, 2000. MATLAB, The language of technical computing, Using MATLAB, Version 6, MathWorks, Inc., Natick, MA, USA, p. 830.

Novotny, V., Jones, H., Feng, X., Capodaglio, A., 1991. Time series analysis of activated sludge plants. Water Science & Technology, 23, 4–6, 1107–1116.

Pedregal, D.J., Taylor, C.J., Young, P.C., 2004. System Identification, Time Series Analysis and Forecasting, The Captain Toolbox, CRES, Lancaster University, Lancaster, UK, pp. 226.

Tiao, G.C., Box, G.E.P., 1981. Modelling multiple time series with applications. Journal of the American Statistical Association, 76, 376, 802–816.

Umashankar, S., Ledolter, J., 1983. Forecasting with diagonal multiple time series models: an extension of univariate models. Journal of Marketing Research, XX, 58–63.

Van der Wiel, S.A., 1996. Monitoring processes that wander using integrated moving average models. Technometrics, 38, 2, 139–151.

Van Dongen, G., Geuens, L., 1998. Multivariate time series analysis for design and operation of biological wastewater treatment plant. Water Research, 32, 3, 691–700.

Wardell, D.G., Moskwitz, H., Plante, R.D., 1992. Control charts in the presence of data correlation. Management Science, 38, 8, 1084–1105.

Zheng, S., Novotny, V., 1991. Stochastic modelling of combined sewer flows. Water Science & Technology, 24, 6, 35–40.