International Conference on Hydroinformatics

2014

Analyzing And Modelling Sewage Discharge Process Of Typical Area Using Time Series Analysis Method

Sisi Zhou

Pengfei Du

Yong Pang

Shuyi Li

11th International Conference on Hydroinformatics
HIC 2014, New York City, USA

ANALYZING AND MODELING SEWAGE DISCHARGE PROCESS OF TYPICAL AREA USING TIME SERIES ANALYSIS METHOD

SISI ZHOU(1), PENGFEI DU(2), YONG PANG(1), SHUYI LI(2)
(1): College of Environmental Science and Engineering, Hohai University, China
(2): School of Environment, Tsinghua University, China

This study is conducted to develop a mathematical model for typical sewage discharge area like residential area, commercial district and institutional area. An approach of time series analysis is applied to build the model involving model selection, parameter estimation, simulation and prediction. The description of sewage discharge process is divided into two parts: Periodic change and stationary random process. Periodic change process is simulated by harmonic analysis which composites a number of trigonometric function together. Stationary random process is described using Stationary time series including six steps: stationary test of the series; calculation of autocorrelation function and partial autocorrelation function for the series; identification of model type; determination of the model order; estimation of model parameters; verification of the model. In this paper daily variation process models for Sewage discharge of residential areas are built using this method. The numerical results show that the present method is effective and produce good agreements with the measured curve. Sewage discharge simulation of other areas like commercial area or institutional area could take the same way. This model could be used as a tool for uncertainty analysis of sewage discharge predicting. And the model also could be coupled with pipe flow model like SWMM to build sewage discharge analysis system in urban scale.

1 INTRODUCTION

As the urbanization process is fast in China, sanitary sewage and industrial wastewater become the main pollution source of urban water body, which makes the river water quality deteriorate [1]. So sewage collection facilities are important to reduce the level of pollution. The path of waste water movement is mainly including discharge, transportation of pipes, control of facilities, treatments and receiving by water bodies [2]. Estimation flow of Sewage discharge of different landuse areas is important for waste load calculation and pollution source tracking. Meanwhile it can be the flow hydrograph input of drainage network model like SWMM [3] which used to simulate the moving process. Reasonable modeling of Sewage discharge is helpful to estimate overflow load [4] and also useful for designing of pipe net, pump stations and treatment plants.

In this paper, we use a common and convenient method to build the model of discharge for areas. The model building process includes three steps: firstly, data preprocessing; secondly, separating and expressing the Period Term; thirdly, analyzing the remainder using the theory of stationary random time series. Sewage discharge of residential area is discussed mainly using this method.

2 ANALYZING AND MODELING PROCESS OF SEWAGE DISCHARGE

2.1 Field Sampling and Data Preprocessing

(1) Field Sampling

The method for data collection used in this research is field sampling with equal time intervals. Flows are recorded by the flow gauge, with the time interval of 10 minutes. Samples of water quality are collected hourly (with interval of 1 hour) over a 24-hour period.

(2) Data Preprocessing

Misreading of probes or flow gauge may occur inevitably in the sampling process, which leads to abnormal data, so data preprocessing is necessary. Firstly, we can easily find some unreasonable data as those negative or zero flow measurements, and correct them using a common method like interpolation. There are also some mistakes which can be recognized by simply observation, for example, water depth greater than the diameter of the pipe on sunny days. Secondly, we need identify some anomalous measurements by mathematical statistics. The methods we use to deal with abnormal data nowadays are mainly based on the principles of probability and statistics. As the random errors obey the normal distribution, we take the theoretical distribution range of the random errors as reasonable distribution range. Data outside the range is regarded as abnormal. In this article, we choose the Chauvenet Criterion to identify anomalous flows.

2.2 Analysis and Modeling of Time Series

There are sorts of random processes. Up to now, we have not found a universal way to analysis all kinds of random processes. But there is a kind of random process called stationary time series which is relatively mature in theory and is widely used. Therefore time series are usually divided into two parts: stationary time series and non-stationary time series. In fact, flow data recorded by flow gauge are not stationary time series. So we use classical decomposition method to describe the series and convert the series to a combination of stationary components and non-stationary components. The series can be decomposed into three elements: the period term, the trend term and the stationary time series term.

2.2.1 Sewage Discharge Modelling of Typical Area

The trend term $\mathrm{Tr}(\mathrm{t})$ will not be considered as the analysis time range is one day that is not enough long for trend analysis like a year. The analysis time range is determined as 24 hours and the sampling interval is 10 minutes for flow, with the total sample number of 144 a day. We take the period term and the stationary time series term into consideration and discribe the time series of sewage discharge as follows.

$$\mathrm{Y} = \mathrm{Pe}(t) + \mathrm{St}(t) \qquad \text{Eq. (1)}$$

Where $\mathrm{Pe}(t)$ is the period term and $\mathrm{St}(t)$ is the stationary time series term.

2.2.2 Calculation and Expression of Period Term

Methods used to identify and calculate the period term mainly include analysis of variance analysis, correlation analysis, harmonic analysis and schuster periodgram analysis. In this research, we figure out the periodic pattern of pollution discharge based on the harmonic analysis.

According to Fourier Series Theory, a function whose period is T can be approximately expressed using the trigonometric system$\{\sin \omega_i, \cos \omega_j\}$, which is a finite trigonometric series. A finite series of the form,

$$\hat{y}_t = \hat{a}_0 + \sum_{k=1}^{p} (\hat{a}_k \cos \omega_k t + \hat{b}_k \sin \omega_k t) \qquad \text{Eq. (2)}$$

Where, $\hat{y}_t$ is estimation of $\{y_t\}$ expression by *p* harmonic waves, $\omega_k = \frac{2\pi k}{N}$ is frequency of *k*th harmonic wave; $\hat{a}_0$、$\hat{a}_k$ and $\hat{b}_k$ are called Fourier's Constants。

In order to test the significance of the *k*th harmonic, we use the F-test. First we construct the F statistics, which obeys the F distribution. Under the significance level given as $\alpha = 0.05$, if $F_k > F_\alpha$, then the harmonic is significant. Moreover, the more the variance of the harmonic contributes, the more significant the harmonic will be.

2.2.3 Analysis on Stationary Random Time Series

Autoregressive Moving Average Model is a common stationary time series, denoted as ARMA (*p, q*) model. In this paper, we use ARMA (*p, q*) to model the process of sewage flow. The steps of model building are as follow:

(1) The stationarity test of time series. We use Daniel test method to test the stationarity of the time series, a rank test method based on the Spearman correlation coefficient.

(2) Caculation of the autocovariance function (ACF) and the partial autocorrelation function (PACF) of Stationary time Series.

(3) Preliminary Identification of Model Type. AR(*p*)、MA(*q*) and ARMA(*p, q*) have their own statistical characteristics. We can preliminarily identify them according to ACF and PACF. The AR(*p*) process X*t* has $\emptyset_{k,k} = 0$ for all *k*, |*k*| > *p*. So a AR(*p*) process has negligible PACF after the *p*th term. The MA(*q*) process X*t* has $\rho_k = 0$ for all *k*, |*k*| > *q*. So a diagnostic for MA(*q*) is that the process has negligible ACF after the *q*th term. If two conditions above are not met, we may consider the process X*t* is ARMA(*p, q*) process.

(4) Determination of the order for the model. Common methods used to determine the order of models include the F-test criterion, residual variogram and the FPE/AIC/BIC criterion. Among these methods, the AIC criterion is quite widely used in determining not only the order of the ARMA model, but also other non-linear models as well. In this article, we choose AIC and BIC criterion to determine the order of the model.

(5) Estimation of model parameters. Maximum likelihood method is used for estimation of the parameters of the model.

(6) Model Verification. The Box-Pierce test is used to verificate the model. If the residual error

obeys normal distribution, it illustrates the model are accepted.

2.3 Case Study

In our study, three types of typical sewage discharge monitoring site are chosen, residential area, commercial district and institutional area. Here we take residential area for example. The monitoring site Jun'an Zutuan (JZ) is a residential area where rain water is separated systematically from sewage water, which means the sewage flow would not be influenced by precipitation.

2.3.1 Identification of Anomalous Measurements

Before analysis of the flow measurements, we should preprocess the data. Take ZJ for example. Figure 1 shows the different between raw data (red line) and corrected data (blue line). The interval of the series is 10 minutes.

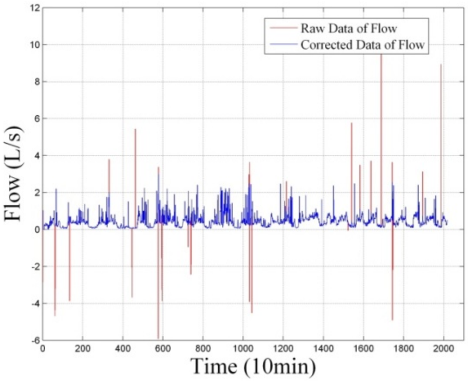

Figure 1. Identification of Anomalous Flow Measurements of JZ

2.3.2 Identification and Expression of the Period Term

As the flow measurements discussion is based on daily analysis, we break the modified flow time series into several days, starting from 0:00 to 24:00 with a 10 minute interval and 144 values in total a day. As result，we break 2016 flow measurements into 14 days. Then we get 144 mean flow $\{v_t\}$ of exactly the same time in 14 days using Eq. (3) (i=time, k=total days).

$$\bar{v}_i = \frac{1}{n}\sum_{k=1}^{n} v_{i,k} \qquad \text{Eq. (3)}$$

Table 1. Estimation of the Harmonic Parameters and Harmonic Contribution by Variance

k	T	a_k	b_k	C_k	φ_k	F_k	Contribution%
1	24	-0.0195	-0.1608	0.1619	3.2623	48.8707	0.4094
2	12	0.0923	-0.1003	0.1363	2.3975	28.7821	0.2899
3	8	0.0148	0.0258	0.0297	3.6631	0.9878	0.0138
4	6	0.0028	0.0123	0.0126	3.3629	0.1742	0.0025
5	4.8	0.0470	-0.0094	0.0479	1.7679	2.6180	0.0358
6	4	0.0184	0.0036	0.0188	4.5181	0.3902	0.0055

Through harmonic analysis, we separate the period term of $\{v_t\}$ and calculate the value of the parameters. The results are shown in the table below. Table 1 shows us, the first harmonic and the second harmonic are significant with the total contribution by the variance of 70.93%. Thus, let the first and the second harmonic be the composition of the harmonic function (with k=2). Then we get $a_0 = 0.4685$. The harmonic function Pe(t) of the flow series $\{v_t\}$ is,

$$\mathrm{Pe}(t) = 0.4685 + 0.1619\sin\left(\frac{2\pi}{24}t + 186.9°\right) + 0.1363\sin(\frac{2\pi}{12}t + 137.4°) \qquad \text{Eq. (4)}$$

The period term is plotted in Figure 2, which composited by those two main harmonic functions, and black solid line represent it. There are two wave crests and one trough of the line. The corresponding times of the two crests are round 12:00 and 8:30. They are mainly on the account of the habits of cooking and bathing in CHINA.

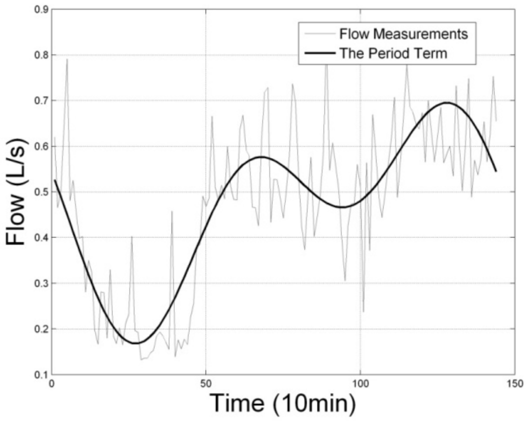

Figure 2. Extraction of the Flow Period Term of JZ

Results of periodic component analysis on other types of area are shown below. Figure 3 (a) shows periodic term of commercial district. And Figure 3 (b) shows periodic term of institutional area. It looks like the same as the periodic term of residential area.We can see institutional area flow drops to almost zero at 21:00~6:00. That confirms to our habits of working.

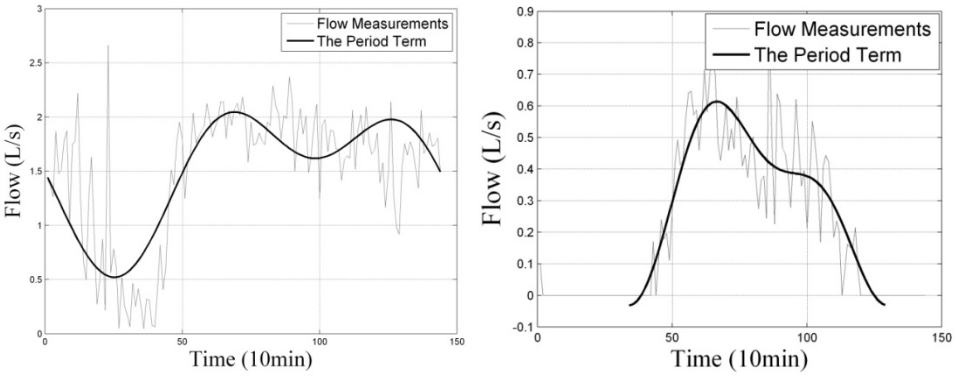

Figure 3. Extraction of the Flow Period Term of: (a) commercial district; (b) institutional area.

2.3.3 Analysis and Expression of Stationary Time Series

(1) The stationarity test of time series

We gain remainder series $\{f_t\}$ by minusing the Period term from flow series. Series $\{f_t\}$ has a constant mean of zero as Figure 4 shows to us. And the value fluctuates randomly above and below 0. Stationary of the series $\{f_t\}$ should be estimated, before we model it with Stationary Time Series Theory.

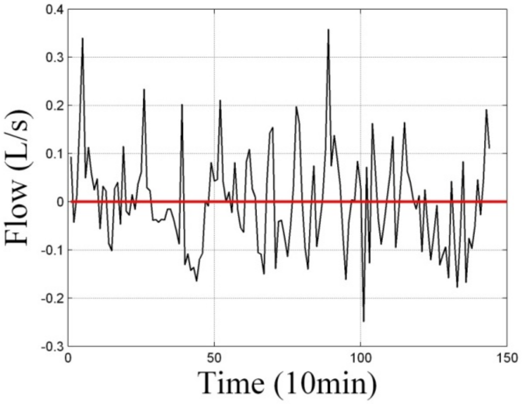

Figure 4. Stationary Time Series of JZ

By conducting a stationary test, we get H_0 is acceptable, which stands for a stationary sequence. Therefore the theory of stationary time series can be applied to build a model.

(2) Estimation of ACF and PACF for Preliminary Identification of Model Type

ACF and PACF are calculated separately. The results demonstrate that ACF (Figure 5 (a)) is trailing, and PACF (Figure 5 (b)) is convergent to 0 after the 2th term ($|k| > 2$) which can be seen as truncation. So the initial diagnostic for the process is AR(p) process.

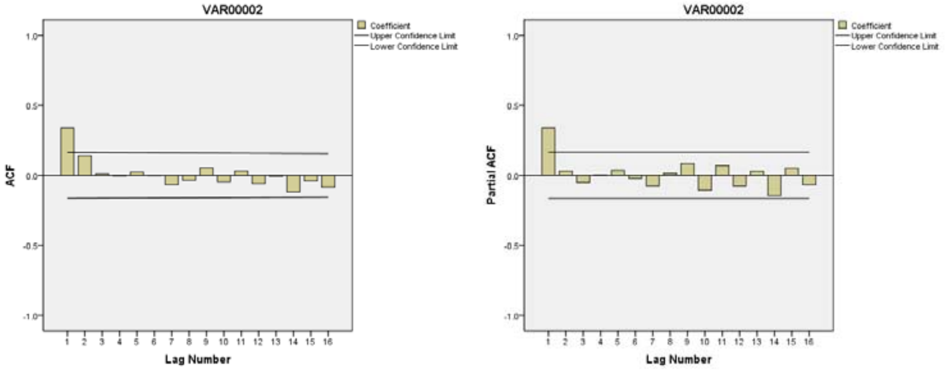

Figure 5. Caculation Result of: (a) ACF; and (b) PACF.

(3) Determination of the Order of the Model and Estimation of Model Parameters

ACF and PACF can provide us some information to estimate the model preliminary. But we should determine the order by accurate calculation. There are several methods to determine the order of ARMA model. However, the AIC and the BIC methods are most widely used at present. From Table 2, we can see AR (1) has the minimum value of AIC and BIC in the same time.

Table 2. Calculation Results of the Model Orders (JZ)

Type of Model	p	q	AIC	BIC
ε_t	0	0	-249.4	-243.5
MA(1)	0	1	-262.1	-253.2
MA(2)	0	2	-263.7	-251.8
MA(3)	0	3	-261.7	-246.9
AR(1)	1	0	-265.0	-256.1
AR(2)	2	0	-263.2	-251.4
AR(3)	3	0	-261.6	-246.8
ARMA(1,1)	1	1	-263.2	-251.3
ARMA(1,2)	1	2	-261.7	-246.9
ARMA(1,3)	1	3	-264.1	-246.3
ARMA(2,1)	2	1	-261.4	-246.6
ARMA(2,2)	2	2	-259.8	-242.0
ARMA(2,3)	2	3	-261.7	-240.9

So the final results of the model order is AR(1). After estimating the parameter of AR(1), we get $\hat{\varphi}_1 = 0.3406$ as the coefficient. Here is the model equation St(t),

$$y_t = 0.3406 y_{t-1} + \varepsilon_t \qquad \text{Eq. (5)}$$

(4) Model Verification

After the Box-Pierce test, we get the residual error obeys normal distribution and the model we build can be used to predict.

2.3.4 Modeling Results

According to Eq. (1), we get the discharge model. One of modeling results for JZ is shown below,

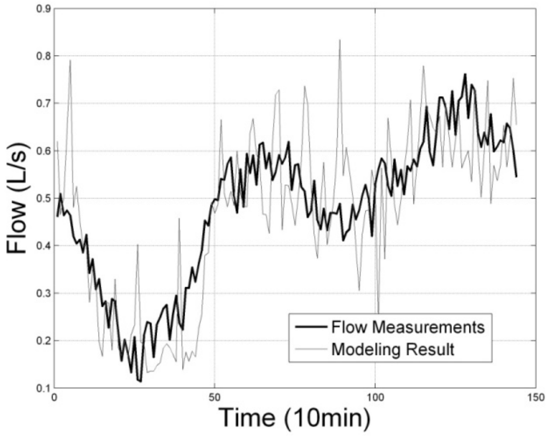

Figure 6. one of modeling results for JZ

3 CONCLUSIONS

This article adopts the method of periodic analysis and stationary time series to model sewage discharge. Simulations show good results. This model provides full consideration to the randomness of discharge. The dynamic modeling provides an analysis basis of uncertainty, which could be used as a tool for uncertainty analysis of sewage discharge predicting. The model also could be coupled with pipe flow model like SWMM to build sewage discharge analysis system in urban scale.

Acknowledgements

This study is supported by the Major Science and Technology Program for Water Pollution Control and Treatment (2012ZX07302002). The authors want to thank all fellows of the team.

REFERENCES

[1] Shen Y.Q., Wei Y. E., Li F. Q., et al. Influence of urbanization of small city on water environment: a case study of Changxing County in Zhejiang Province[J]. Water Resources Protection, 2005, 21(4)

[2] Xu S., Chen J. N., Zeng S. Y., et al. Study on integrated simulation of urban drainage system[J]. China Water & Wastewater, 2006, 22(15)

[3] Rossman L. A..Storm water management model user's manual, version 5.0[M]. National Risk Management Research Laboratory, Office of Research and Development, US Environmental Protection Agency, 2010.

[4] Ma J. H., Li J. F., Xu Y. J., et al. Solution to the problem on urban drainage system overflow in rainy season with storm water management model(SWMM)[J]. Water Purification Technology, 2012, 31(3)