

Pollutant load prediction for RTC by system identification method using time-series monitored data

Prédiction des flux de pollution pour la gestion en temps réel par un modèle d'identification utilisant des séries temporelles de données

Akihiro NAGAIWA¹, Hiroaki FURUMAI², Fumiyuki NAKAJIMA³ and Soichi TONARI⁴

¹ Department of Urban Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan (nagaiwa@env.t.u-tokyo.ac.jp)

² Research Center for Water Environment Technology, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan (furumai@env.t.u-tokyo.ac.jp)

³ Environmental Science Center, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan (nakajima@esc.u-tokyo.ac.jp)

⁴ Toshiba corporation, 1-1-1 Shibaura, Minato-ku, Tokyo 105-8001, Japan (soichi.tonari@toshiba.co.jp)

RESUME

Le Contrôle en Temps Réel (RTC) permet de réduire les rejets de charges polluantes dans les Surverses de Réseaux Combinés (CSO) grâce au contrôle d'installations telles que les bassins et les pompes d'eau de pluie. Des algorithmes de contrôle pourvus d'informations prédictives précises permettent l'utilisation efficace du RTC. Cette étude propose un modèle de prédiction de la charge polluante étalonné par méthode d'identification du système au moyen d'une série temporelle de données sur les charges polluantes en un point spécifique amont des conduites. Ce point amont sur la conduite a été choisi par analyse d'une série temporelle de charge polluante basée sur les résultats de la simulation d'un modèle correctement étalonné de qualité des ruissellements. Les résultats de ces prévisions ont été évalués en comparant les données de la série temporelle fournies par le modèle distribué étalonné.

ABSTRACT

RTC (Real Time Control) is one of the effective options to reduce CSO (Combined Sewer Overflow) pollutant discharge by control of sewage facility such as storage tanks and rainwater pumps. Control algorithms with accurate predicted information can let RTC act effectively. In this study, a pollutant load prediction model was proposed, which was calibrated by system identification method using time-series data on pollutant loads at a concerning point and its upstream pipe. The upstream pipe location was selected through time-series pollutant load analysis based on simulation results of a well calibrated distributed runoff quality model. The prediction results were evaluated by comparing time-series data given by the calibrated distributed model.

KEYWORDS

CSO; pollutant load prediction; RTC; system identification method; time-series data.

1 INTRODUCTION

Combined Sewer Overflow (CSO) is one of the most concerning pollution phenomena in urban area. RTC (Real Time Control) which controls the operation of sewage facility such as storage tanks and rainwater pumps is one of the effective options to reduce CSO pollutant discharge. Control algorithms with accurate predicted information by a simple (or quickly responding) and reliable model can let RTC act effectively. For example, when the model provides the prediction 10 minutes ahead, sewer operators can investigate several possible operational scenarios and make a decision on operating sewer facilities to reduce the pollutant discharge.

Ruan *et al.* (1997) utilized a simple model to predict the CSO discharge from time-series data of pollutant load and CSO discharge calculated with rainfall data. They described that the relationship between the pollutant load and the CSO discharge was non-linear and the real-time collection of the pollutant load for model input was difficult. Though it is still difficult to conduct an on-line and accurate monitoring of the pollutant load in a sewer system, sensing device has been developed for monitoring SS and COD concentrations by UV absorbance and visible light absorbance (MLIT, 2005).

The objective of this study is to propose a simple pollutant load prediction model which can be applied to RTC using time-series monitoring data for CSO pollutant reduction. For the purpose, time-series data of pollutant load were analyzed and the model structure and their parameter values were calibrated by system identification method. Input data for the model were rainfall, pollutant loads at the point in concern as well as at its upstream pipe. The prediction results were evaluated by comparing time-series data given by the well calibrated distributed model.

2 STUDY AREA AND THE DISTRIBUTED MODEL FOR ANALYSIS

Study area has a combined sewer system with catchment area of 168.8 ha and population of 14,600. The sewer network consisting of 934 pipes were modelled with a distributed model (InfoWorksCS Ver.6.5, Wallingford Software Lit.) (Figure 1). In this area, there were three subcatchments with main pipelines as shown in Figure 1.

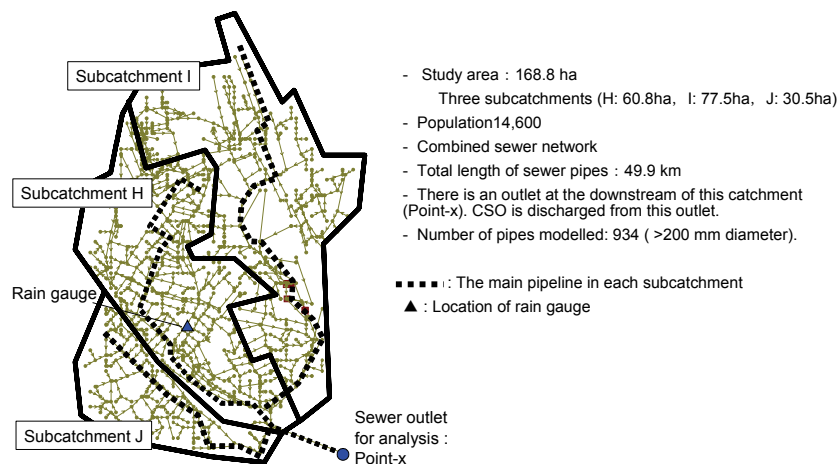


Figure 1 Study area and the sewer network modelled

Runoff monitoring was carried out in 2000 and 2001, and runoff and pollutant load parameters for a distributed model were calibrated in our previous study (Jinadasa *et al.*, 2004). Simulations with the calibrated parameters satisfactorily represented several rainfall events, but we found a considerable difference between the measured and the time-series data in long term (one-year of 2000) simulation including dry weather period. Therefore the parameter values of build-up factor and specific gravity were reconsidered in this study.

First the build-up factor was fixed to 1.85 kg/ha/day used in a simulation in Tokyo (Watanabe *et al.*, 1987), and then the specific gravity was calibrated by changing its value from 1.3 to 1.7 in increment of 0.1. As a result, the specific gravity of 1.4 minimized square error between monitored data and time-series data calculated with distributed model in three rainfall events. Table 1 shows the sub-models and parameter values used in this study.

Calculation was conducted by 30 seconds of time-step, and time-series data of flow rate and pollutant load were provided every 10 minutes by the well calibrated distributed model.

Sub-model	Parameter	Unit	Parameter value
Surface build-up model	Decay factor	1/day	0.08*
	Build-up factor	kg/ha/day	1.85**
Surface washoff model	Coefficients of the erosion equation	C_1	100,000,000*
		C_2	2.0*
		C_3	29.0*
Erosion/Deposition model (Ackers-White model)	Sediment particle size	d_{50}	mm
	Specific gravity	—	1.4

* Jinadasa *et al.*, 2004 . **Watanabe *et al.*, 1987

Table 1 Sub-models and parameter values used in this study.

3 MODEL DEVELOPMENT

Figure 2 shows a hydrograph ($Q_x(t)$) and a pollutograph ($L_x(t)$) in the heaviest rainfall event in 2000. Flow rate had the peak after 23:00, corresponding with increasing the rainfall intensity. However, pollutant load had a different pattern from flow rate, and the load peak appeared in an earlier timing (from 16:00 to 17:00). Then the pollutant load did not reach the highest rate level even if flow rate remarkably increased afterwards (after 22:00). This time-series data shows a phenomenon called first flash of pollutant load.

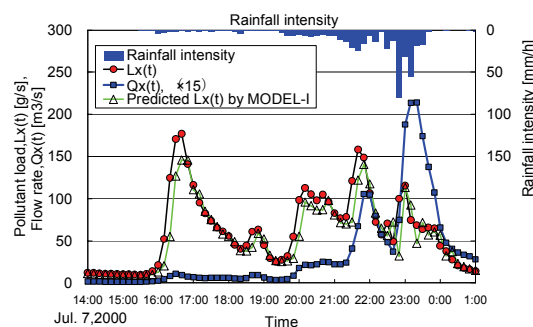


Figure 2 Hydrograph, pollutographs at Point-x calculated by the distributed model, prediction result by MODEL-I (AR model) and rainfall intensity on Jul. 7, 2000.

System identification method is applied to predict pollutant load in this study. The model is expressed by a simple linear equation. The selection of the equation components is essential in order to describe the variation of the pollutant load. Therefore time-series data were analyzed to select components of the pollutant load prediction model.

In the first step of the model development, we focussed on the time-series data of flow rate and pollutant load at the sewer outlet (Point-x). Table 2(A) shows the result of the correlation analysis. The correlation coefficient between pollutant load $L_x(t)$ and flow rate $Q_x(t)$ was not high (0.21). In contrast, correlation of $L_x(t)$ and $L_x(t-1)$ is high as 0.88. Therefore, it would be possible to predict pollutant load by using its time-series data. This prediction model type (called MODEL-I) is an auto-regressive (AR) model. When a prediction system using AR model is constructed, it needs only monitored pollutant load data at the point of sewer outlet.

Data		Time*					
		t	t-1	t-2	t-3	t-4	t-5
(A)	Flow rate at Point-x $Q_x(t)$	0.21	0.09	0.00	-0.07	-0.12	-0.16
	Pollutant load at Point-x $L_x(t)$	—	0.88	0.64	0.42	0.25	0.15
(B)	Pollutant load in Pipe-a $La(t)$	0.73	0.87	0.77	0.57	0.41	0.28
	Pollutant load in Pipe-b $Lb(t)$	0.38	0.54	0.57	0.49	0.39	0.34
	Pollutant load in Pipe-c $Lc(t)$	0.77	0.88	0.77	0.57	0.39	0.28
	Pollutant load in Pipe-d $Ld(t)$	0.23	0.09	0.02	0.00	-0.01	-0.03
	Pollutant load in Pipe-e $Le(t)$	0.44	0.47	0.37	0.20	0.09	0.00

*Time-step is 10 minutes

Table 2 Correlation coefficients between $L_x(t)$ and other time-series data on Jul. 7, 2000.

The MODEL-I was calibrated using time-series data of nine rainfall events from May to June in 2000 and then was validated using those of twelve rainfall events from March to April in 2000. This model is expressed by equation (1).

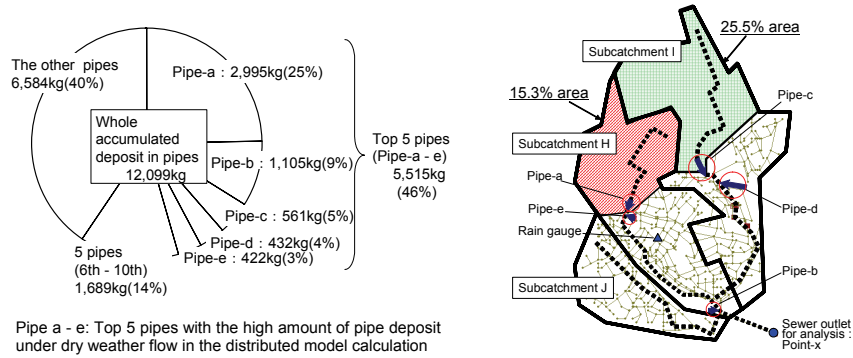
MODEL-I (AR model)

$$L_x(t) = 1.24 \cdot L_x(t-1) - 0.66 \cdot L_x(t-2) + 0.31 \cdot L_x(t-3) \quad (1)$$

L_x : pollutant load at Point-x [g/s]

An example of prediction result of the MODEL-I is shown in Figure 2. Although the MODEL-I explains the general trend of pollutant load variation, there is one step delay.

In the second step of the model development, time-series data of pollutant load in an upper stream pipe was given as additional equation component to improve MODEL-I having a problem of one step delay. It was shown that there were several pipes which were easy to accumulate deposit by the simulation analysis and the pipe observation at this area in our previous study (Jinadasa *et al.*, 2004). In addition, the simulation results by the calibrated distributed model revealed that the deposit in the pipes contributed significantly to CSO pollutant discharge (Nagaiwa *et al.*, 2006). Based on the previous knowledge, accumulated pipe deposits during two days of dry weather flow condition before the rainfall on July 7 were calculated. The top five pipes were Pipe-a, -b, -c, -d and -e as shown in Figure 3. The amount of accumulated deposit in these five pipes was estimated 46% of the whole accumulated deposit (12,099kg) in the sewer network as shown in Figure 3. The deposit in Pipe-a occupied 25% of the whole accumulated deposit.



Pipe a - e: Top 5 pipes with the high amount of pipe deposit under dry weather flow in the distributed model calculation

Figure 3 Fractional distribution of accumulated deposit in pipes (left) and upstream catchment area of Pipe-a and Pipe-c (right)

The correlative analysis were carried out to examine whether time-series data of pollutant load in these pipes become effective input data for the prediction model. As shown in Table 2(B), pollutant loads $L_a(t)$ and $L_c(t)$ directly correlated with $L_x(t)$, while the correlation coefficients of $L_b(t)$, $L_d(t)$ and $L_e(t)$ were not so high. Figure 3 shows that Pipe-a and Pipe-c were located in a halfway point of two main pipeline routes in subcatchments H and I, and their upstream catchment areas occupied 15.3% and 25.5% of the whole catchment, respectively. On the other hand, Pipe-b, -d and -e were located at the upper end in the network, having very narrow catchment areas. It seems important to select time-series data of pollutant load at a pipe having large area of the upstream catchment with high accumulation of pipe deposit. Additional analysis for selection of time-series data at a specific pipe is shown in the next section.

Bar graph in Figure 4 shows the cumulative pipe deposit from the upper end of the main pipeline in the subcatchment H. The cumulative pipe deposit is gradually increased from the upper end (H1) to downstream and sharply rise at H-18 (Pipe-a) by 3,789kg. This sudden increase is caused by pipe deposit in Pipe-a and a junction of branch catchment. Furthermore, at H-36 the cumulative pipe deposit is further increased by contribution of the cumulative pipe deposit in subcatchment I.

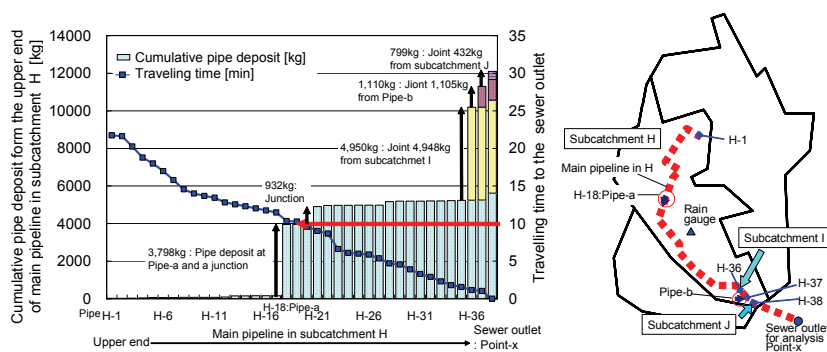


Figure 4 Cumulative pipe deposit and travelling time from upper end of the main pipeline in subcatchment H to the sewer outlet (left), and the route of main pipeline in subcatchment H (right)

To resolve the time delay problem which occurred in the prediction by MODEL-I, proactive input data are needed to predict pollutant load with a certain lead time. Travelling time along the pipeline was evaluated as the lead time for prediction. It was calculated using flow velocity of each pipe assuming a uniform flow and half water depth of pipe diameter.

In Figure 4, the pipes which have travelling time longer than 10 minutes are located in the upstream pipes of H-19. Therefore, Pipe-a is one of important pipes to get proactive input data of this prediction model because it can provide sufficient lead time and reflect upstream pollutant load runoff behaviour.

Pipe-c in Subcatchment I was also analyzed in the same procedure, but it was not such a pipe with sufficient travelling time as Pipe-a.

From these analysis results, MODEL-II was formulated with pollutant load data in an upstream pipe as an additional input. This model type is called ARX (Auto-Regressive with eXogenous) model. The prediction system using MODEL-II needs two sets of time-series data on pollutant load at the point of outlet and at a specific pipe in upstream.

The MODEL-II was also calibrated using nine rainfall events from May to June in 2000 and then was validated using twelve events from March to April in 2000. This model is described by equation (2).

MODEL-II (ARX model)

$$\begin{aligned} L_x(t) = & 0.65 \cdot L_x(t-1) - 0.03 \cdot L_x(t-2) \\ & + 0.08 \cdot L_x(t-3) + 0.01 \cdot L_x(t-4) - 0.01 \cdot L_x(t-5) \\ & + 2.98 \cdot L_a(t-1) - 1.35 \cdot L_a(t-2) - 0.19 \cdot L_a(t-3) - 0.24 \cdot L_a(t-4) \end{aligned} \quad (2)$$

L_x : pollutant load at Point-x [g/s], L_a : pollutant load in Pipe-a [g/s]

4 PREDICTION RESULTS AND EVALUATION OF THE MODELS

Figure 5 and Table 3 show comparison of prediction results by the two models. In MODEL-I, one step delay of prediction appeared conspicuously at 15:20 on Aug. 9, 2000. The error of total amount of pollutant load is -11 % in both Jul. 7, 2000 and Aug. 9, 2000, the predicted load is a little less than the time-series data. On the other hand, in MODEL-II, there was not a time delay as observed in MODEL-I. The errors of total amount of pollutant load in the two events were reduced from -11% to -1 and -4 %, respectively. It implies that the predicted load agreed better with the time-series data in MODEL-II than MODEL-I.

The root square mean error (RSME) of pollutant load was 14.1 g/s in MODEL-II and 18.8 g/s in MODEL-I in the evaluation period from July to December. This means that pollutant load in the specific pipe in upstream becomes effective and proactive information for prediction of the pollutant load at sewer outlet (Point-x).

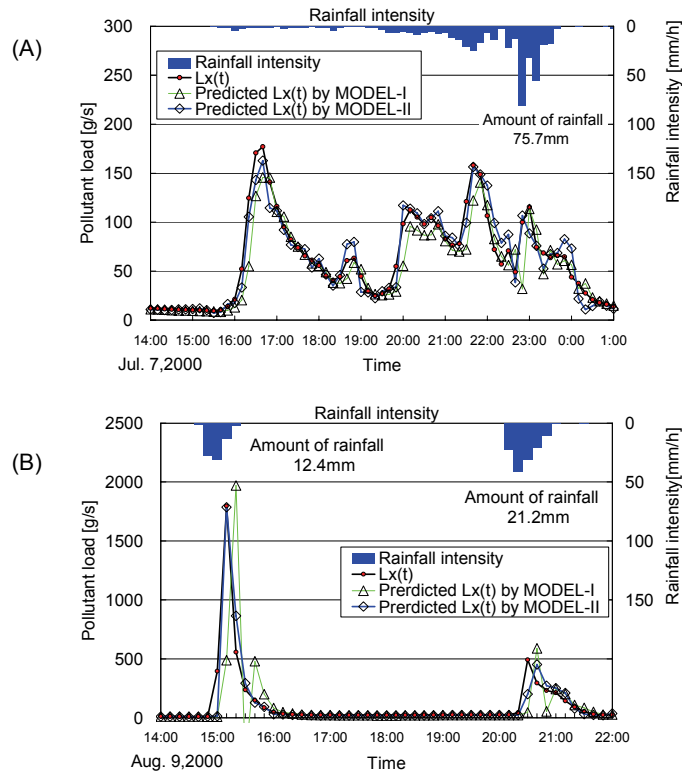


Figure 5 Comparison of pollutographs predicted by MODEL-I and II, (A) Rainfall on Jul. 7, 2000 (Maximum rainfall was monitored), (B) Rainfall on Aug. 9, 2000 (Maximum pollutant load was calculated).

Table 3 Evaluation of the prediction error

		Time-series data	Predicted pollutant load	
			MODEL-I	MODEL-II
Jul. 7,2000	Total amount of pollutant load [kg]	2523	2243	2497
	*Error [%]	-	-11	-1
	RSME [g/s]	-	18.8	12.5
Aug. 9,2000	Total amount of pollutant load [kg]	3423	3038	3291
	*Error [%]	-	-11	-4
	RSME [g/s]	-	309.0	86.1
Jul.- Dec., Evaluation period		-	18.8	14.1

*Error=(Predicted load - Time-series data)/(Time-series data)x100

5 CONCLUSIONS

In this study, an easy and fast prediction by ARX model using time-series monitoring data was proposed, which predicted the pollutant load 10 minutes ahead. It is expected that an RTC algorithm using this prediction system might be applied for operation of sewage facilities in order to reduce overflow pollution of CSO effectively. The following conclusions were drawn.

- Pollutant load was predicted by an AR (auto-regressive) model using time-series monitoring data. But one time step delay appeared in the prediction when the pollutant load fluctuated quickly.
- To eliminate the delay, ARX (auto-regressive with exogenous) model was proposed by adding the time-series data of pollutant load in an upstream pipe as a proactive input. The time-series data becomes effective information for prediction of the pollutant load at sewer outlet.
- Selection procedure of the upstream pipe for the proactive input was essential for the ARX model development. For this selection, graphical diagram of accumulative pipe deposit and travelling time was useful as well as correlative analysis.

LIST OF REFERENCES

- Jinadasa H. K. P. K., Itoi Y., Furumai H. and Nakajima F. (2004). Specification and categorization of combined sewer pipes with highly accumulated deposits using model simulations, Proceedings of 1st Asian junior scientist workshop on sewer network processes & urban drainage.
- MLIT (the Ministry of Land, Infrastructure and Transport, Japan). (2005). Organic pollutant monitor (UV meter), Japan Institute of Wastewater Engineering Technology (JIWET) ENGLISH Web-site, http://www.jiwet-spirit21.jp/confluence/eng_03.html
- Nagaiwa A., Settsu K., Nakajima F. and Furumai H. (2006). Identification of sewer pipes to be cleaned for reduction of CSO pollutant load, Water Science & Technology, (in press).
- Ruan M and Wiggers J. B. M. (1997). Application of time-series analysis to urban storm drainage, Water Science and Technology, Vol 36 No 5 pp 125-131.
- Watanabe K. and Kawamoto K. (1987). Research about an anti-improvement measure of combined sewer system, Journal of Japan Sewage Works Association, Vol.24, No.280, pp.35-53 (in Japanese).