Nonlinear regression approach to evaluate nutrient delivery coefficient

M.S. Bae* and S.R. Ha**

*Watershed Management Research Center, National Institute of Environmental Research, Kyongseo-dong, Seo-gu, Incheon, South Korea (E-mail: *mysoba@me.go.kr*)

** Department of Urban Engineering, Chungbuk National University, 48 Gaeshin-dong, Heungduk-gu, Cheongju, Chungbuk, South Korea (E-mail: *simplet@chungbuk.ac.kr*)

Water Science & Technology **Vol 53 No 2 pp 271–279** © IWA Publishing 2006

Abstract Implementation of the Korean Total Maximum Daily Load Act calls for new tools to quantify nutrient losses from diffuse sources at a river basin district scale. In this study, it was elucidated that the nonlinear regression model (NRM) reduces the uncertainty of the boundary conditions of the water quality model. The NRM was proposed to analyse the delivery coefficients of surface waters and retention coefficients of pollutants. Delivery coefficient of pollution load was considered as a function of two variables: the watershed form ratio, S_f, which is a measurable geomorphologic variable and the retention coefficient, ϕ, which is an empirical constant representing the basin-wide retarding capacity of pollutant wash-off. This model was applied on the Geum River, one of the major basins in South Korea. The QUAL2E was used to simulate stream water quality using NRM. In this paper, we elucidate the possibility to use a nonlinear regression model for delivery and retention of nutrients in a drainage basin characterized as both data-rich and data-poor, and the magnitude of the nutrient loads and sources has been uncertain for a long time.

Keywords Delivery coefficient; diffuse pollution; pollution load runoff; retention coefficient

Introduction

The total maximum daily load (TMDL) Act has been established for watershed-based water quality management in Korea since 1998. To meet water quality standards mandated by the TMDL Act at a water quality monitoring station (WQMS), a comprehensive and rational pollution runoff analysis must be carried out. Prerequisites for sustainable environmental management of watersheds require basic environmental statistics gathering and quantitative assessments of the riverine loads, in particular, including estimation of the pollution delivery and retention in the drainage basin. The nutrient level and fluxes at a specific location in a public river network depend on the pollution sources in the upstream area, and transfer, retention, and loss of nutrients in the soil, ground water, and surface water network. This is a complex function of biological, physical, and chemical processes. Therefore, it is needed to analyze how these processes influence nutrient fluxes from pollution sources to river outlets over large spatial and temporal scales in modeling (Vassilijev and Stalnacke, 2005). Several models for so-called source delivery and retention have already been developed worldwide (Ha, 1989; Grimvall and Stalnacke, 1996; Ha *et al.*, 1998; Liden et al., 1998). However, many watersheds in developing countries are often regarded as "data-poor" and characterized by high varying quantity and quality of input data. This is a typical problem in trans-boundary waters in which the richness and details of information about the watersheds' specified data may differ between provinces or countries both in terms of quantity and quality. Unfortunately, because most of the existing models require very detailed and spatially consistent input data, their applicability may be limited. Thus, more simple models are needed to address the limitations in these basins. A simple delivery coefficient (SDC), calculated by a conventional simple

doi: 10.2166/wst.2006.061

rate method, is called to either an assimilative capacity or a purification coefficient (Lee, 2000). Actually, these kinds of SDC account for the difference between the total pollution loads generated from pollution sources and total pollution loads discharged at the outlet of a stream in a specific watershed. The SDC, however, can't be used for the estimation of pollution load delivered in a watershed that lacked measurements because it is based on water quality data observed in the field. In this paper we elucidate the possibility to use a statistical model for delivery and retention of nutrients in a drainage basin characterized as both data-rich and data-poor, and the magnitude of the nutrient loads and sources has been uncertain for a long time.

Methods

Nonlinear statistical model

The SDC has been calculated by simply dividing the pollution load monitored with one discharged from a specific watershed. This relationship could be defined as:

$$P_M = P_O \times K \tag{1}$$

where P_M is the pollution load monitored at a specific water pollution monitoring site and P_O is the total pollution load discharged, which is an assumed amount of pollutants discharged from a specific watershed based on the unit loading factor method. K is a simple delivery coefficient. Because the calculation of the K in Equation (1) is impossible unless to get the data of water quality at a WQMS in each watershed, as an alternative way, it was usual to assign a mean value of the K coefficients, which have been determined from other watersheds with WQMS to the specific watershed without observed water quantity and quality data. This means that the delivery coefficient for the watershed without observed data can be estimated using an average of K values calculated from the watersheds with data observed by Equation (1). To overcome this weakness of SDC, the non-linear function defined with two variables, the watershed form ratio, S_f, which is a deterministic variable obtainable from the digital terrain map of a watershed as well as a quantitative index representing the geo-characteristic of the watershed and the retention coefficient, ϕ, which is an empirical variable was developed (Ha and Bae, 2001; Bae and Ha, 2003; Bae, 2003). Pollution loads monitored at WQMS, P_M, make a balance with the result of multiplying the pollution loads discharged from a watershed, P_O, and the innovated delivery coefficient, K. In the Equation (3), all variables are known values except the retention coefficient, ϕ. Consequently an empirical variable ϕ can be determined using the data set of known variables.

$$K = e^{-\phi \cdot S_f} \tag{2}$$

$$P_M = P_O \times e^{-\phi \cdot S_f} \tag{3}$$

where ϕ is an empirical variable and denotes the retardation effect of pollution loads in a watershed. And S_f denotes the Horton's watershed form ratio and quantifies the portion of channels in a watershed and is defined as follows:

$$S_f = \left(\sum L\right)^2 \Big/ A \tag{4}$$

where L is a sum of stream length and A is area of a specific watershed. Digital elevation model (DEM) data can be used to determine a length of waterway, the area, and average slope of a watershed.

Establishment of a reference flow rate and a standard water quality

The reference year was 1998 because the TMDL Act was established in this year in Korea. Using flow rate data for the last 10 years, a non-excess probability distribution of flow rate was analyzed and the mean of the data of Q_{275} was determined since the water quality at WQMS in the study watershed showed the worst state at the flow rate Q_{275}. As the flow rate in December 1998 was most similar to the average Q_{275}, the monthly average flow rate and BOD concentration obtained in December 1998 were used to calculate a basin-wide delivery coefficient. In addition, the total daily pollution loads discharged from all tributaries were referenced to the data released by the National Institute of Environment Research in Korea. The water quality data observed at WQMS in the Geum River were cited from the official homepage of the Ministry of the Environment, Korea.

Water quality simulation model

The feasibility analysis of the methodology innovated to determine a basin-wide delivery coefficient is carried out through a comparison of BOD concentration calculated by the innovated method with the water quality observed at WQMS. The EPA QUAL2E model was used for simulation of stream water quality. The effectiveness of introducing the innovated delivery coefficient of pollution loads, which had been taking into account the influence of stream watershed geomorphologic properties on a pollutant wash-off behavior, into the stream water quality simulation was evaluated using the root mean square errors (RMSE) method.

Results and discussion

Study watershed

The Geum River basin is located in the central part of South Korea and includes the Dacheong Reservoir, of which the total storage capacity is 1,490 million tons. It is a water supply resource for about three million people. The watershed area is about 9,910 km^2 and the annual precipitation is about 1,400 mm/year but more than half of it concentrates on a rainfall season from July to September. The delivery time of the storm peak is comparatively short and less than 2 days. This study watershed is divided into 121 sub-watersheds and the mean area of the watersheds is about 82 km^2. The mean

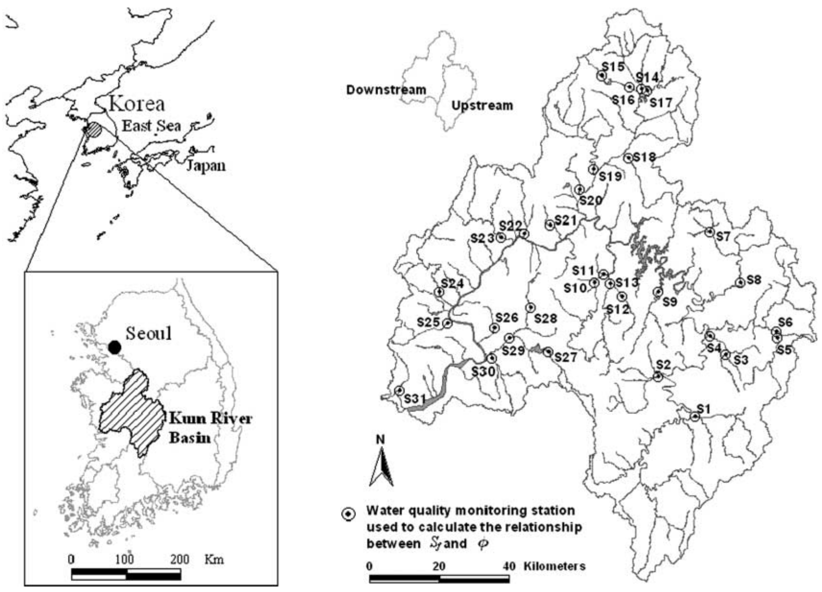

Figure 1 Study area and water quality monitoring stations

length of the waterways and the average slope of the watersheds is 28 km and 11 °, respectively. A spatial distribution of WQMS used to estimate the relationship between S_f and ϕ is shown in Figure 1.

Non-linear regression model on S_f and ϕ

Non-linear regression between S_f and ϕ was derived from the data set in terms of water quality data obtained at water quality monitoring stations and the estimated values of watershed pollution loads discharged in Dec. 1998. In this analysis, taking into consideration the differences of water use in the stream reach, the Geum River was divided into the upstream reach and the downstream one on the basis of the location of the Daecheong

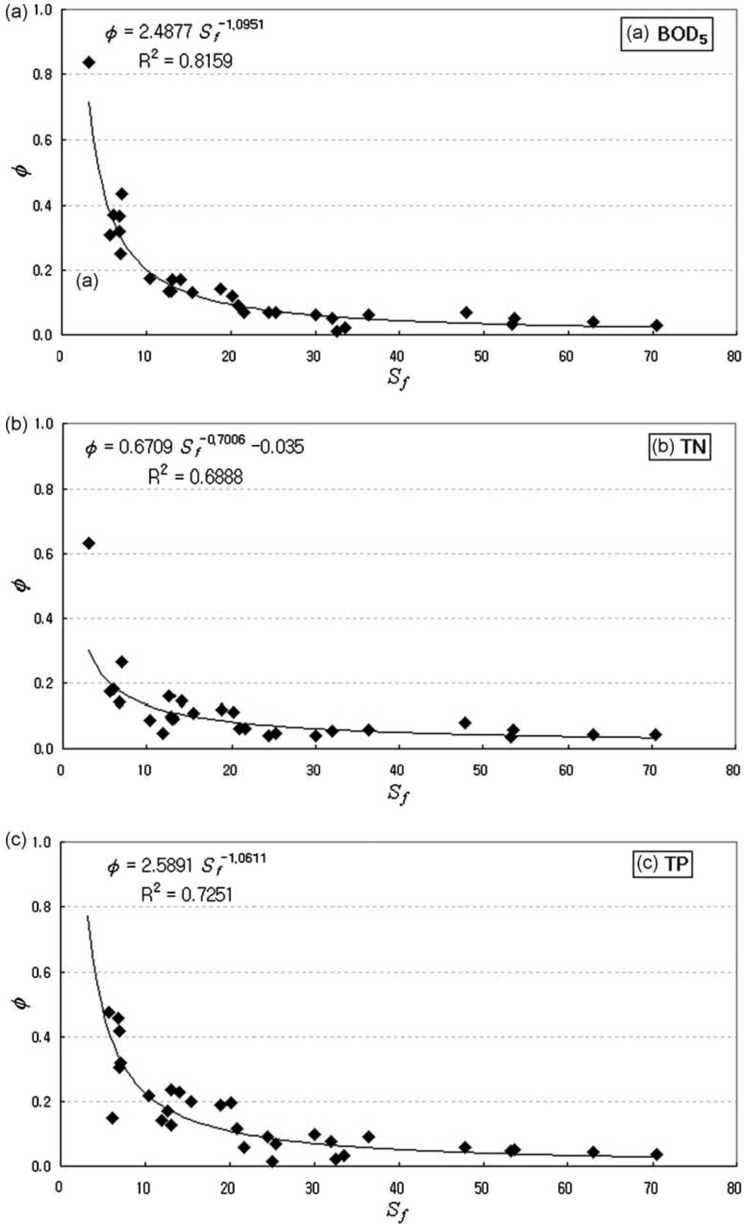

Figure 2 Relationship between S_f and ϕ on BOD, TN and TP

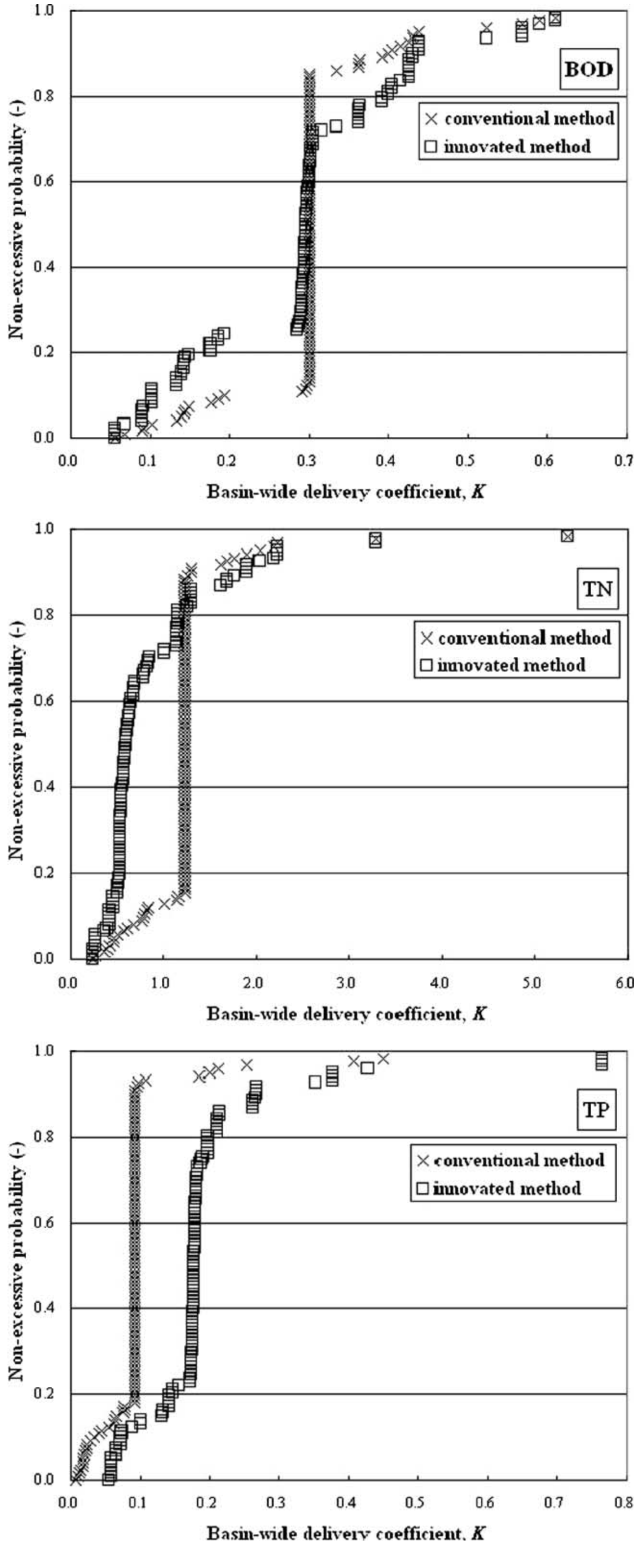

Figure 3 Comparison of non-excess probability distributions of the conventional delivery coefficient (X) and the innovated one, K (□), determined by the regression on S_f *vs.* ϕ

Reservoir. Figure 2 shows scatter plots of the relationship between S_f and ϕ. This regression analysis was done using 31 data sets available from all WQMSs including the upstream and the downstream of Geum River. Figure 2(a) shows the scattering of BOD_5 and the correlation coefficient is about 0.82. On the other hand, Figure 2(b) and Figure 2(c) are the plots on total nitrogen and total phosphorus, respectively. And correlation coefficients on total nitrogen and total phosphorus are 0.69 and 0.73, respectively.

Determination of the innovated delivery coefficients on BOD, TN, and TP

Because it is possible to calculate S_f for any watersheds, the retention coefficient, ϕ, in the basins that lack measurement of water quantity and quality can be estimated by substituting the S_f value for the specific sub-watershed into the regression equation. Then the innovated delivery coefficient, *K*, in the sub-watershed without WQMS is determined by substituting two values of S_f and ϕ to Equation (2). While difference between non-excessive probability distributions of two different delivery coefficients determined using the regression model and the average *K* values is shown in Figure 3. This has come from the assignation of the *K* value averaged from all values of *K* determined from 31 sub-watersheds with WQMS to the other sub-watershed without WQMS in the study watershed.

Impact of the innovated basin-wide delivery coefficient on the water quality simulation

The application results of the innovated delivery coefficient to the water quality simulation using QUAL2E model showed high consistency with observed water quality concentrations.

Tables 1 and 2 show the RMSE errors between the delivered pollution loads calculated using the regression (with innovated delivery coefficient) and the pollution loads monitored at 31 WQMS sites. RMSE of pollution load delivered on BOD, TN and TP using innovated delivery coefficients are 4.27 kg/day, 17.91 kg/day and 2.53 kg/day in upstream, respectively. While RMSE using conventional delivery coefficients are 105.76 kg/day, 251.58 kg/day and 2.40 kg/day in upstream, respectively. RMSE of BOD, TN and TP concentrations with innovated delivery coefficient are 0.001 mg/L, 0.080 mg/L and 0.002 mg/L in upstream, respectively. While RMSE using conventional delivery coefficients are

Table 1 RMSE of water quality simulation using QUAL2E (upstream)

	Pollution load (kg/day)		Concentration (mg/L)	
	Innovated *K* *1	**Conventional *K* *2**	**Innovated *K***	**Conventional *K***
BOD	4.27 (0.04)*3	105.76 (1.00)	0.001 (0.001)	0.826 (1.00)
TN	17.91 (0.07)	251.58 (1.00)	0.080 (0.012)	6.676 (1.00)
TP	2.53 (1.05)	2.40 (1.00)	0.002 (1.00)	0.002 (1.00)

*1 resulted from the application of the innovated basin-wide delivery coefficient proposed in this study
*2 resulted from the application of the conventional basin-wide delivery coefficient by simple rate method
*3 numeric values in parenthesis are relative to conventional delivery coefficient

Table 2 RMSE of water quality simulation using QUAL2E (downstream)

	Pollution load (kg/day)		Concentration (mg/L)	
	Innovated *K*	**Conventional *K***	**Innovated *K***	**Conventional *K***
BOD	38.55 (0.29)	132.98 (1.00)	0.039 (1.11)	0.035 (1.00)
TN	54.34 (0.16)	329.70 (1.00)	0.303 (0.16)	1.844 (1.00)
TP	6.44 (0.34)	19.11 (1.00)	0.001 (0.004)	0.262 (1.00)

0.826 mg/L, 6.676 mg/L and 0.002 mg/L in upstream, respectively. The simulation results from QUAL2E runs on BOD, TN and TP in upstream and downstream are shown in Figure 4 (a) and (b), respectively. As a result of water quality simulation, it was revealed that the innovated delivery coefficient resulting from the nonlinear regression model has more consistent than the conventional delivery coefficient. Following is the summary of the new findings from simulation analysis. The root mean square error for BOD (mean of observed concentrations is 1.1 mg BOD/L) was reduced from 0.82 mg BOD/L in the run case with the conventional delivery coefficient to 0.001 mg BOD/L in the run with the innovated delivery coefficient. For nutrient cases, RMSE for total nitrogen (mean is 2.57 mg TN/L) and total phosphorus (mean is 0.04 mg TP/L) are improved from 6.68 mg TN/L to 0.08 mg TN/L and from 0.0023 mg TP/L to 0.0016 mg TP/L by applying the innovated delivery coefficient to the simulation model.

Flow rate and retention coefficient

Hydrological conditions, such as rainfall intensity/duration, soil infiltration, retention time of reservoir etc., are crucial factors in diffuse pollution wash-off. The rainfall, especially, acts as a transmission mediator of pollutant from source to receiving water. In order to reveal the effects of rainfall on pollutant runoff, the relationship between S_f and ϕ was

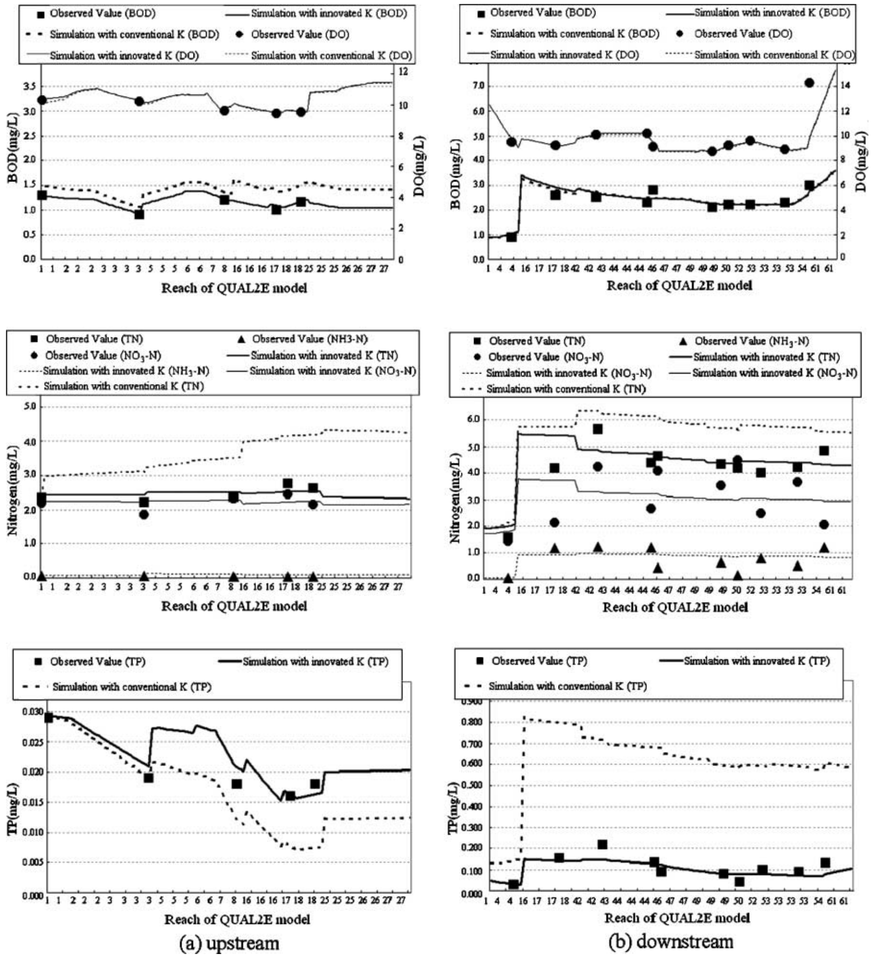

Figure 4 Consistency of simulation results on BOD and nutrients

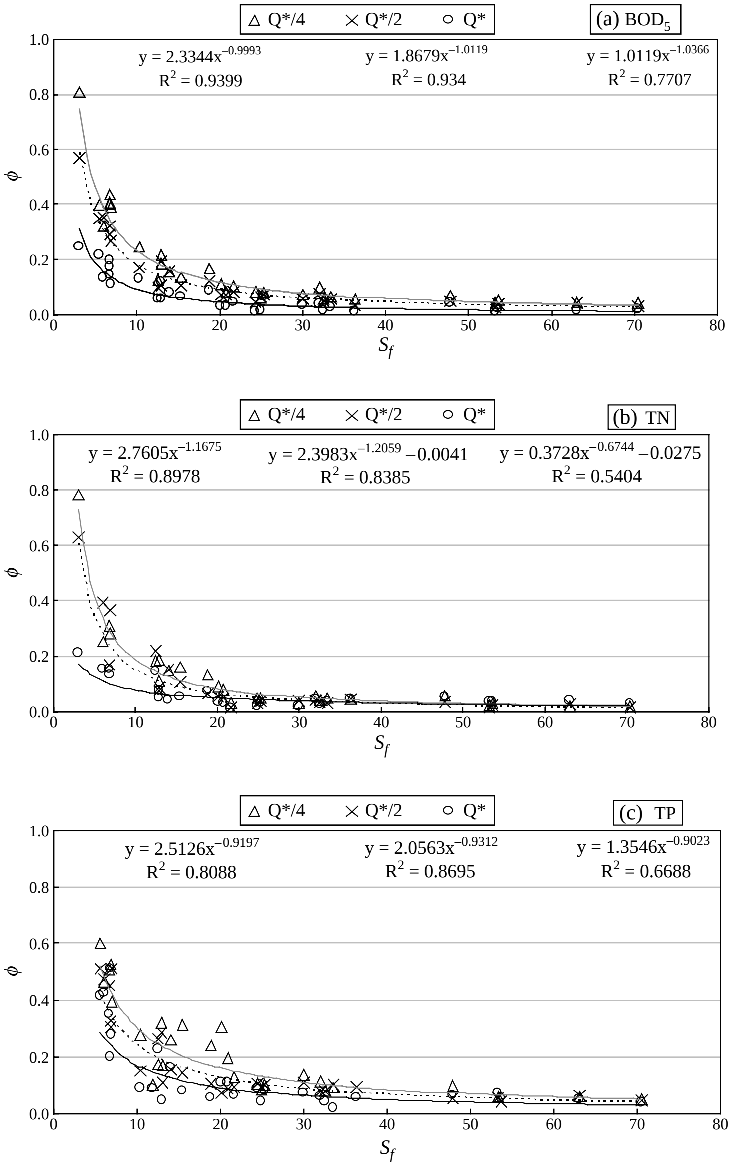

Figure 5 Variation of the flow rate and the relationship between S_f and ϕ on BOD, TN and TP

analyzed in three different periods (June, May and March, 1998) with different flow rates. Figure 5 shows the variation of the regression curve between S_f and ϕ on BOD, TN and TP based on different flow rates; Q^* is mean flow rate in June, $Q^*/2$ is mean flow rate in May when we had the most approximate flow rate to half of Q^*, and $Q^*/4$ is mean flow rate in March when we had the most approximate flow rate to a quarter of Q^*. As shown in Figure 5, it was revealed that the retention coefficient, ϕ, reduced as the flow rate increased, especially where watershed has a low S_f. This means that the variation of delivery coefficient, K, with rainfall variation can be quantified on a specific watershed.

Conclusions

In this study, the possibility to use the nonlinear regression model to calculate the delivery coefficients and the retention coefficients in terms of BOD and nutrients (TN and TP) was elucidated. The model takes into account the influence of stream watershed geomorphologic properties such as a watershed form ratio, S_f, and a basin-wide delivery retardation coefficient, ϕ, on pollution wash-off behavior. The application results of the innovated model to the stream water quality simulation done by QUAL2E model showed high consistency with the observed results. The root mean square error for BOD (mean of observed concentrations is 1.1 mg BOD/L) was reduced from 0.82 mg BOD/L in the run case with the conventional delivery coefficient to 0.001 mg BOD/L in the run with the innovated delivery coefficient. For nutrient cases, RMSE for total nitrogen (mean of observed concentrations is 2.57 mg TN/L) and total phosphorus (mean of observed concentrations is 0.04 mg TP/L) are improved from 6.68 mg TN/L to 0.08 mg TN/L and from 0.0023 mg TP/L to 0.0016 mg TP/L by applying the innovated delivery coefficient to the simulation model. As a result of this study, it was revealed that the innovated method to evaluate the delivery coefficient could cope with the limitations of the conventional method, especially in agricultural and forest areas. The accuracy improvement using this innovated model on water quality simulation in the study watershed was about 62% on average. Interestingly, it was revealed that the retention coefficient, ϕ, reduced as the flow rate increased, especially where the watershed has a low S_f. This means that the variation of delivery coefficient, K, with rainfall variation can be quantified on a specific watershed.

References

Bae, M.S. (2003). *Estimating of delivery coefficient of pollution load using geo-informatics*. PhD thesis, Department of Urban Engineering, Chungbuk National University, Korea.

Bae, M.S. and Ha, S.R. (2003). GIS-based influence analysis of geomorphological properties on pollutant wash-off in agricultural area. *Wat. Sci. Tech.*, **51**(3–4), 301–307.

Grimvall, A. and Stalnacke, P. (1996). Statistical methods for source apportionment of riverine loads of pollutants. *Environmetric*, **7**, 201–213.

Ha, S.R. (1989). *Kinematic modeling for wash-off on urban non-point source pollutants*. PhD thesis, Department of Environmental Engineering, Osaka University, Japan.

Ha, S.R. and Bae, M.S. (2001). Effects of land use and municipal wastewater treatment changes on stream water quality. *Environ. Monitor. Assess.*, **70**, 135–151.

Ha, S.R., Jung, D.I. and Yoon, C.H. (1998). A renovated model for spatial analysis of pollutant runoff loads in agricultural watershed. *Wat. Sci. Tech.*, **38**(10), 207–214.

Lee, J.H. (2000). Management of nonpoint sources in watershed with reference to Daechong Reservoir in Korea. *Journal of Environmental Impact Assessment*, **9**(3), 163–176.

Liden, R., Vassilijev, A., Loigu, E., Stalnacke, P. and Wittgen, H.B. (1998). Nitrogen source apportionment using a physically and statistically based model. *Ecological modeling*, **114**, 235–250.

Vassilijev, A. and Stalnacke, P. (2005). Statistical modeling of riverine nutrient sources and retention in the lake Peipsi drainage basin. *Wat. Sci. Tech.*, **51**(3–4), 309–317.