

## 하천수내 TOC 농도 추정을 위한 단순회귀모형과 다중회귀모형의 개발과 평가

정재운 · 조소현\* · 최진희\* · 김갑순\* · 정수정\* · 임병진\*\*

전남대학교 지역바이오시스템공학과, 국립환경과학원 영산강물환경연구소

\*국립환경과학원 영산강물환경연구소

### Development and Evaluation of Simple Regression Model and Multiple Regression Model for TOC Contentation Estimation in Stream Flow

Jaewoon Jung · Sohyun Cho\* · Jinhee Choi\* · Kapsoon Kim\* · Soojung Jung\* · Byungjin Lim\*\*

Department of Rural & Bio-systems Engineering, Chonnam National University, Yeongsan River Environment Research Center

\*Yeongsan River Environment Research Center

(Received 19 July 2013, Revised 26 August 2013, Accepted 10 September 2013)

#### Abstract

The objective of this study is to develop and evaluate simple and multiple regression models for Total Organic Carbon (TOC) concentration estimation in stream flow. For development (using water quality data in 2012) and evaluation (using water quality data in 2011) of regression models, we used water quality data from downstream of Yeongsan river basin during 2011 and 2012, and correlation analysis between TOC and water quality parameters was conducted. The concentrations of TOC were positively correlated with Chemical Oxygen Demand (COD), Biochemical Oxygen Demand (BOD), TN (Total Nitrogen), Water Temperature (WT) and Electric Conductivity (EC). From these results, simple and multiple regression models for TOC estimation were developed as follows :  $TOC=0.5809 \times BOD+3.1557$ ,  $TOC=0.4365 \times COD+1.3731$ . As a result of the application evaluation of the developed regression models, the multiple regression model was found to estimate TOC better than simple regression models.

**Key words** : BOD, COD, Simple and multiple regression model, TOC

#### 1. Introduction

그간 우리나라 수질관리 지표는 BOD와 COD를 중심으로 관리되어 왔으며, 현재 많은 양의 BOD와 COD자료가 축적되었다(Kim, Jung et al., 2007; Choi et al., 2012). 하지만, BOD는 긴 측정시간과 실험기간 큰 오차를 갖고 있고(Byoun et al., 2008a), COD는 산화제의 종류, 농도, 반응 온도, 시간 등에 따라서 크게 영향을 받는다. 이에 분석시간이 짧고 정확성이 높은 TOC에 대한 관심이 높아지고 있다(Byoun et al., 2008a; Choi et al., 2012). 또한, 현재 환경정책 기본법 시행령에 의해 2013년 1월 1일부터 하천 생활환경기준에 TOC가 추가되어 시행되고 있다. 하지만, TOC는 BOD와 COD에 비해 자료의 축적기간이 짧아 과거 수질관리 정책과 현재 수질관리 정책의 연계성 부분에 대한 문제가 발생될 수 있다. 이러한 문제를 해결하기 위해 많은 연구자들에 의해 다양한 지역의 하천 및 호소에 대한 BOD, COD, TOC간의 상관성에 관한 연구를 통해 이들 간의 상관성을 규명하고(Seo et al., 2010; Byoun et al., 2008a; Kim, Jung et al., 2007; Kim, Shin et al., 2007; Park et al., 2006) 회귀모형을

개발하여 BOD와 COD를 이용하여 TOC를 추정하는 연구가 수행되고 있다(Han et al., 2012; Min et al., 2012).

TOC와 BOD, COD와의 상관성에 관한 연구를 살펴보면, Choi et al. (2012)은 논으로부터 유출되는 유출수중 TOC는 BOD보다 COD가 더 유의한 상관성을 나타낸다고 보고 하였으며, Seo et al. (2010)은 하천수와 하수처리장 방류수를 대상으로 TOC, BOD, COD의 상관분석을 실시한 결과 하수처리장 방류수 보다 하천수의 TOC, BOD, COD의 상관성이 높고 BOD 보다 COD의 상관성이 더 높다고 보고 하였다. TOC, BOD, COD와의 상관분석에 관한 대부분의 연구결과들이 BOD보다 COD와 TOC의 상관성이 더 높은 것으로 나타났다(Park et al., 2006; Ji et al., 2010). 한편, BOD, COD, TOC를 이용하여 회귀모형을 도출한 결과 Choi et al. (2012)은 논 유출수의 경우 COD를 독립변수로 가진 단순회귀모형은 BOD를 독립변수로 가진 단순회귀모형보다 TOC 추정의 정확성이 높다고 보고하였고, 이러한 결과는 Park et al. (2006)의 연구결과와 유사하다. 또한 Lee et al. (2010)은 하수처리장 방류수의 TOC를 추정하기 위해 COD를 독립변수로 가진 단순회귀모형을 개발하여 하수처리장 방류수 유기물질 농도의 결측치 보완에 활용될 수 있다고 주장하였다. 이처럼 다양한 연구자들에 의해 다양한 지역에서 TOC 추정을 위한 회귀모형을 개발하여 적용성을 검토

\* To whom correspondence should be addressed.  
imnolim@korea.kr

하였지만 대부분의 회귀모형의 형태가 하나의 독립변수를 가진 단순회귀모형의 형태로 개발되었다. 따라서 본 연구에서는 여러 개의 독립변수를 가진 다중회귀모형과 단순회귀모형을 개발하여 TOC 추정의 적용성을 평가하였다.

## 2. Materials and Methods

### 2.1. 연구대상지구

본 연구는 영산강 수계 최하류 지점(Fig. 1)의 수질자료를 활용하여 단순회귀모형과 다중회귀모형을 개발하였다. 수질자료는 물환경정보시스템(NIER, 2013)의 약 8일간 측정자료를 이용하였으며 2012년 1월부터 2012년 12월의 측정자료는 회귀모델 개발에 활용하였으며, 2011년 1월부터 2011년 12월의 자료는 회귀모델의 적용성을 평가하는데 활용하였다. 수질항목은 수온, pH, DO(Dissolved Oxygen), EC(Electric Conductivity), BOD(Biochemical Oxygen Demand), COD(Chemical Oxygen Demand), TN(Total Nitrogen), TP(Total Phosphorus), TOC(Total Organic Carbon), SS(Suspended Solids)이다.

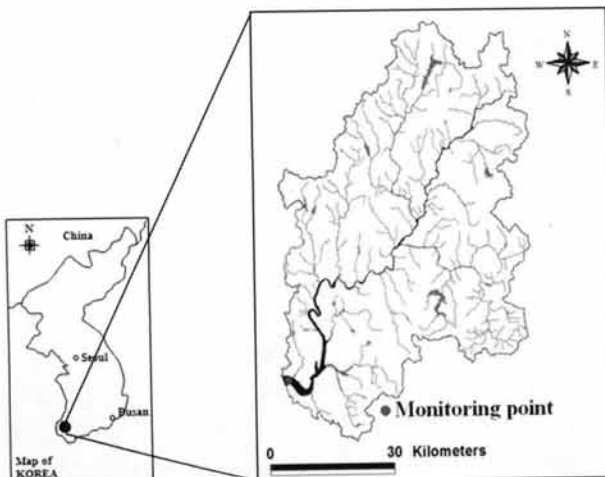


Fig. 1. Location of study site where simple and multiple regression models developed.

### 2.2. 상관분석

상관분석은 변수간의 관련성의 정도를 계량화하여 나타내는 척도로 피어슨의 상관계수( $r$ )를 보편적으로 사용하고 있다(Jung et al., 2012). 상관계수는 -1에서 1까지의 값을 가지는데 절대값이 1에 가까울수록 상관성이 강하다는 것을 의미하며 상관성이 없는 경우에는 0에 가깝다. 본 연구에서는 피어슨의 상관계수를 이용하여 수질 항목간 상관분석을 실시하여 회귀식 개발의 독립변수를 결정하는데 활용하였다.

### 2.3. 회귀모형개발

회귀모형은 단순회귀모형과 다중회귀모형의 형태로 개발하였고 2012년 수질자료는 회귀모형의 개발에 활용하였으며, 2011년 수질자료는 회귀모형의 적용성 검토에 활용

였다. 단순회귀모형의 형태는 BOD와 COD를 각각의 독립변수로 가진 식 (1)과 식 (2)와 같은 두가지 형태로 개발하였고 다중회귀모형은 형태는 식 (3)과 같다.

$$Y = a_1 \times X_1 + b_1 \quad (1)$$

$$Y = a_2 \times X_2 + b_2 \quad (2)$$

여기서, 종속변수  $Y$ 는 TOC 농도,  $a_1$ ,  $a_2$ 는 회귀계수,  $b_1$ ,  $b_2$ 는 각각의 상수항, 독립변수  $X_1$ 과  $X_2$ 는 각각 BOD와 COD의 농도이다.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + \varepsilon \quad (3)$$

여기서, 종속변수  $Y$ 는 TOC 농도,  $X$ 는  $i$ 번째 독립변수,  $\beta$ 는  $Y$ 에 대한  $i$ 번째 독립변수  $X_i$ 의 회귀계수,  $\varepsilon$ 는 오차항이다.

다중회귀모형은 SPSS(ver. 18.0) 프로그램을 이용하였으며, 다중회귀분석에서 변수선택법은 단계선택법(Stepwise)을 적용하였다. 다중회귀모형에 대한 통계적인 유의성을 검증하기 위해서 F 검정법에 의한 F 통계량을 이용하였고, 이때 계산된 F 통계량의 값에 대한 유의확률이 유의수준보다 작으면 적어도 하나의 회귀계수가 유의하다는 것을 의미한다(Han et al., 2012). 또한, 각 독립변수의 유의성을 검정하기 위해서 t 검정법에 의한 t 통계량을 이용하였으며, t 통계량값에 대한 유의확률이 유의수준보다 작으면 추정된 회귀계수가 통계적으로 유의하다는 결과로 해석할 수 있다(Han et al., 2012).

## 3. Results and Discussion

### 3.1. 수질특성

조사지점에서 2011년부터 2012년까지 수질분석 결과는 Table 1과 2와 같다. 2011년의 경우 유기물 오염도를 나타내는 간접지표인 BOD, COD, TOC의 범위는 각각 0.4~6.4 mg/L (평균 2.3 mg/L), 4.0~12.5 mg/L (5.9 mg/L), 2.5~6.7 mg/L (4.1 mg/L)로 나타났고 BOD, COD, TOC의 크기는 COD>TOC>BOD 순으로 나타났다. 또한, TN과 TP의 범위는 각각 1.8~8.9 mg/L (3.4 mg/L)와 0.02~0.23 mg/L (0.10 mg/L)로 나타났다. 한편, 2012년의 경우 BOD는 0.7~4.7 mg/L (1.7 mg/L), COD는 4.7~12.4 mg/L (6.3 mg/L), TOC는 2.2~7.1 mg/L (4.1 mg/L)로 조사되었다. BOD와 TOC의 연평균 농도는 2011년이 다소 높게 관측되었고 COD의 경우 2012년이 다소 높게 관측되었지만 큰 차이는 나타나지 않았다. 2012년 TN과 TP의 범위는 각각 1.6~5.4 mg/L (3.1 mg/L), 0.01~0.20 mg/L (0.09 mg/L)로 나타나 2011년에 비해 다소 낮게 관측되었다.

### 3.2. 상관분석

조사지점에서 수질항목별 상관분석은 2012년 자료만을

**Table 1.** Summary of water quality variations in the study site (2011)

	WT (°C)	pH	DO (mg/L)	EC (μm/cm)	BOD (mg/L)	COD (mg/L)	TN (mg/L)	TP (mg/L)	TOC (mg/L)	SS (mg/L)
Min	1.0	6.3	5.7	130.0	0.4	4.0	1.8	0.02	2.5	5.3
Max	29.0	8.4	13.9	1867.0	6.4	12.5	8.9	0.23	6.7	44.8
Median	18.0	7.6	9.6	990.0	2.1	5.5	3.2	0.10	4.0	16.7
Mean	16.1	7.6	10.0	911.1	2.3	5.9	3.4	0.10	4.1	17.8

**Table 2.** Summary of water quality variations in the study site (2012)

	WT (°C)	pH	DO (mg/L)	EC (μm/cm)	BOD (mg/L)	COD (mg/L)	TN (mg/L)	TP (mg/L)	TOC (mg/L)	SS (mg/L)
Min	2.1	6.5	5.8	118.0	0.7	4.7	1.6	0.01	2.2	2.5
Max	30.4	8.6	14.8	1144.0	4.7	12.4	5.4	0.20	7.1	152.5
Median	17.1	7.5	10.1	642.0	1.4	5.8	3.1	0.09	4.3	10.4
Mean	15.7	7.5	10.4	666.1	1.7	6.3	3.1	0.09	4.1	17.9

이용하여 분석하였다. 앞서 언급했듯이 2012년 수질자료를 이용하여 회귀모형을 개발하고 2011년 자료는 개발된 회귀모형의 적용성을 검토하기 위해 사용되었기 때문에 2012년 수질자료만을 이용하여 수질항목간 상관분석을 실시하였다. Table 3은 2012년 수질항목간 상관분석 결과를 나타내고 있다. 조사지점의 경우 TOC와 COD의 상관계수가 0.570으로 나타나 조사된 수질항목중 가장 높은 상관성을 보였고, 다음은 BOD로 상관계수가 0.471, TN은 0.411로 나타났다. 또한 TOC와 수온, EC와의 상관계수는 각각 0.330과 0.238로 약한 상관성을 보였지만 유의한 상관성(p-value<0.01)을 보였다.

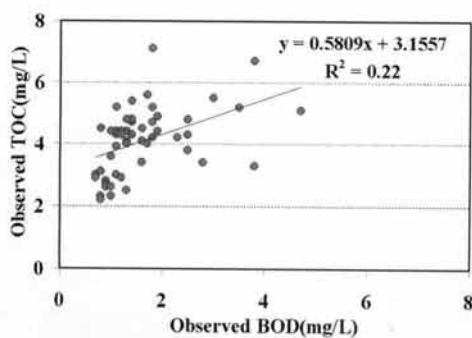
### 3.3. 회귀모형의 개발

실측 BOD농도와 COD농도를 각각의 독립변수로 가지고 TOC농도를 종속변수로 가진 단순회귀모형의 개발 결과는 Fig. 2와 같다. Model 1은 BOD를 이용하여 개발한 단순회귀모형으로  $R^2$ 가 0.22로 나타났고 Model 2는 COD를 이용한 단순회귀모형으로  $R^2$ 가 0.33으로 나타났다. COD를 이용한 단순회귀모형이 BOD를 이용한 단순회귀모형보다  $R^2$ 이 높게 나타났다. 한편, TOC 추정을 위한 다중회귀모형을 개발하기 위해 단계적으로 독립변수를 추가하면서 분석을 실시하였다(Table 4). Model 3은 상수항과 COD, TN을 독립변수를 가진 다중회귀모형이고, Model 4는 COD, TN, TP,

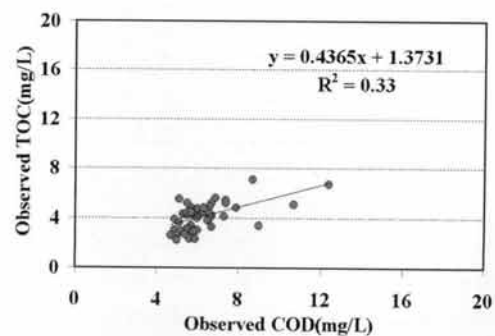
**Table 3.** Correlation coefficients among 10 water quality parameters measured

Items	WT	pH	DO	EC	BOD	COD	TN	TP	TOC	SS
WT	1	.001	-.844**	-.692**	.377**	.516**	-.381**	-.036	.330**	.177
pH		1	.246	.202	.161	.179	.025	-.319**	.204	-.075
DO			1	.744**	-.229	-.442**	.462**	-.147	-.120	-.287**
EC				1	-.361**	-.389**	.691**	-.357**	.238**	-.458**
BOD					1	.680**	-.162	.257	.471**	.425**
COD						1	-.260	.198	.570**	.454**
TN							1	-.076	.411**	-.446**
TP								1	-.161	.533**
TOC									1	-.097
SS										1

(\*\* p value < 0.01)



(a) Model 1



(b) Model 2

**Fig. 2.** Simple regression model for BOD-TOC (left) and COD-TOC (right).

**Table 4.** Results of the multiple regression model analysis

Model	r	R <sup>2</sup>	F-value	p-value
3 <sup>a</sup>	.814	.662	47.955	.000
4 <sup>b</sup>	.856	.732	43.703	.000
5 <sup>c</sup>	.874	.765	38.165	.000
6 <sup>d</sup>	.917	.841	48.517	.000
7 <sup>e</sup>	.917	.841	61.961	.000
8 <sup>f</sup>	.933	.871	61.938	.000

a: COD, TN, b: COD, TN, TP c: COD, TN, TP, Water Temp., d: COD, TN, TP, Water Temp, EC, e: COD, TN, Water Temp., EC, f: COD, TN, Water Temp., EC, BOD

Model 5는 COD, TN, TP, 수온, Model 6은 COD, TN, TP, 수온, EC, Model 7은 COD, TN, 수온, EC, Model 8은 COD, TN, 수온, EC, BOD를 각각의 독립변수로 가진 다중회귀모형이며, 이들 모형은 모두 TOC농도를 종속변수로 가진다. 개발된 모형들의 F-value에 대한 p-value(유의확률)는 모두 0.000으로 나타나 개발된 회귀모형이 종속변수를 설명하는데 유용한 것으로 나타났다. 이들 6개의 다중회귀모형 중 5개의 독립변수(COD, TN, 수온, EC, BOD)를 가진 Model 8의 결정계수(R<sup>2</sup>)가 0.871로 가장 높아 Model 8을 TOC 추정을 위한 최적 다중회귀모형으로 선정하였다.

### 3.4. 단순회귀모형과 다중회귀모형의 적응성 평가

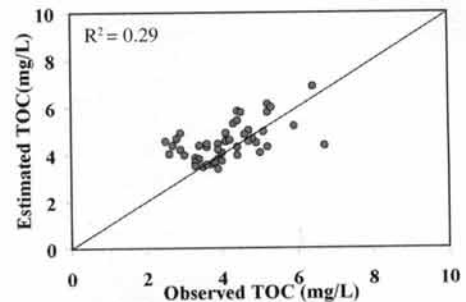
단순회귀분석과 다중회귀분석을 통해 선정된 최적 회귀모형을 Table 5와 같이 정리하였다. 개발된 회귀모형의 적응성을 검토하기 위해 2011년 수질자료를 활용하였으며 단순회귀모형과 다중회귀모형을 적용한 결과는 Fig. 3과 같다. BOD와 COD를 이용하여 개발된 단순회귀모형의 TOC 실측치와 추정치의 R<sup>2</sup>는 각각 0.29와 0.36으로 나타났고 다중회귀모형을 이용하여 추정된 TOC와 실측 TOC의 R<sup>2</sup>은 0.48로 나타났다. 이는 다중회귀모형을 이용하여 추정된 TOC값이 단순회귀모형을 이용하여 추정된 TOC값보다 정확성이 더 높다는 것을 의미한다. 따라서, 향후 TOC를 추정할 경우 BOD와 COD만을 독립변수로 가진 회귀식보다 다양한 수질인자로 구성된 다중회귀식을 이용하여 TOC를 추정하는 것이 더 정확한 TOC값을 추정할 것으로 판단된다.

**Table 5.** Selected the simple and multiple regression model for TOC estimation

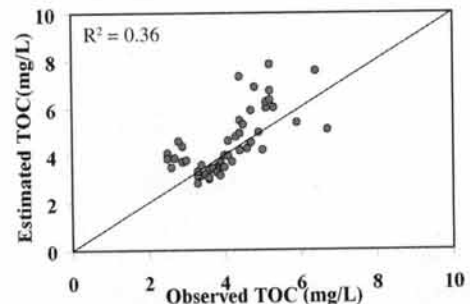
Type	Regression models	R <sup>2</sup>
Model 1	TOC=0.5809×BOD+3.1557	0.220
Model 2	TOC=0.4365×COD+1.3731	0.330
Model 8	TOC=0.339×COD+0.350×TN+0.072×수온+0.002×EC+0.300×BOD-2.335	0.871

## 4. Conclusion

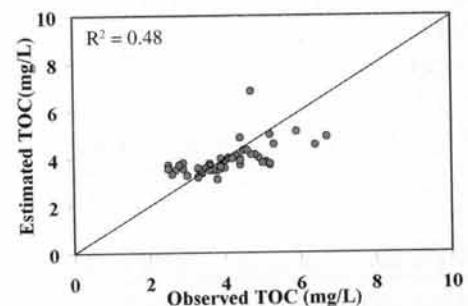
본 연구는 영산강 최하류 지점의 2011년과 2012년 수질자료와 TOC 상관분석을 통해 TOC 추정을 위한 단순회귀모형과 다중회귀모형을 개발하고 적응성을 평가하였으며, 도출된 주요 결론은 다음과 같다.



(a) Simple regression model using BOD



(b) Simple regression model using COD



(c) Multiple regression model

**Fig. 3.** Comparison of the observed and estimated TOC using regression models.

- 1) 조사된 수질항목중 TOC와 COD가 가장 큰 상관성을 보였고 다음은 BOD, TN, 수온, EC 순으로 TOC와 상관성이 있는 것으로 나타났다.
- 2) 상관분석결과를 이용하여 BOD와 COD를 독립변수로 가지고 TOC를 종속변수로 가진 단순회귀모형은 각각  $TOC = 0.5809 \times BOD + 3.1557$ ,  $TOC = 0.4365 \times COD + 1.3731$  형태로 개발되었고 다중회귀모형은  $TOC = 0.339 \times COD + 0.350 \times TN + 0.072 \times \text{수온} + 0.002 \times EC + 0.300 \times BOD - 2.335$  형태로 개발되었다.
- 3) BOD와 COD를 독립변수로 가진 단순회귀모형을 통해 추정된 TOC 실측치와 추정치의 R<sup>2</sup>는 각각 0.29와 0.36으로 나타났고 다중회귀모형을 통해 추정된 TOC 실측치와 추정치의 R<sup>2</sup>는 0.48로 나타나 단순회귀모형보다 다중회귀모형을 통해 TOC를 추정하는 것이 더 정확한 것으로 나타났다. 따라서, 미래측 유역이나 과거 TOC를 추정할 경우 다양한 수질인자를 독립변수로 가진 다중회귀모형을 통해 TOC를 추정하는 것이 바람직할 것으로 판단된다.

## References

- Byoun, J., Kim, T., Lee, S., Hong, T., and Kim, H. (2008a). A Study on Correlation between TOC and COD<sub>Mn</sub> in Environmental Water Samples, *Journal of Korean Society of Environmental Engineers*, 2008(1), pp. 340-343. [Korean Literature]
- Byoun, J., Kim, T., Lee, S., Hong, T., and Kim, H. (2008b). Correlation between TOC, UVA and COD<sub>Mn</sub> of Environmental Water Samples, *Journal of the Korean Society for Environmental Analysis*, 11(4), pp. 286-291. [Korean Literature]
- Choi, D. H., Jung, J. W., Yoon, K. S., Choi, W. J., Lim, S. S., Park, H. N., Yim, B. J., and Hwang, T. H. (2012). Estimation of TOC Concentration using BOD, COD in Runoff from Paddy Fields, *Journal of Korean Society on Water Environment*, 28(6), pp. 813-818. [Korean Literature]
- Han, S. H., Lee, E. H., and Kim, H. (2012). Development of Software Sensor for Total Phosphorus and Nitrogen in Various Waterbody, *Journal of the Korean Society for Environmental Analysis*, 15(3), pp. 215-223. [Korean Literature]
- Ji, J. Y., Tae, E. N., Choi, J. Y., Kim, S. U., and Jeon, T. W. (2010). Comparison of TOC and the Correlation of BOD and COD in the Geum River Basin, *Journal of Korean Society on Water Environment*, 23(1), pp. 36-44. [Korean Literature]
- Jung, J. W., Park, J. H., and Lim, B. J. (2012). The Analysis of Water Quality Characteristics in Lake Juam using Multivariate Analysis, *Journal of Korean National Committee on Irrigation and Drainage*, 19(2), pp. 61-70. [Korean Literature]
- Kim, B. C., Jung, S. M., Jang, C. W., and Kim, J. K. (2007). Comparison of BOD, COD and TOC as the Indicator of Organic Matter Pollution in Streams and Reservoirs of Korea, *Journal of Korean Society of Environmental Engineers*, 29(6), pp. 640-643. [Korean Literature]
- Kim, J. K., Shin, M. S., Jang, C. W., Jung, S. M., and Kim, B. C. (2007). Comparison of TOC and DOC Distribution and the Oxidation Efficiency of BOD and COD in Several Reservoirs and Rivers in the Han River System, *Journal of Korean Society on Water Environment*, 23(1), pp. 72-80. [Korean Literature]
- Lee, T. H., Lee, B., Hur, J., Jung, M. S., and Kang, T. G. (2010). Conversion of COD<sub>Mn</sub> into TOC and Refractory Organic Matter Concentrations for Treated Sewage using Regression Equations, *Journal of Korean Society on Water Environment*, 26(6), pp. 969-975. [Korean Literature]
- Min, S. Y., Lee, S. P., Kim, J. S., Park, J. U., and Kim, M. S. (2012). Development and Validation of Multiple Regression Models for the Prediction of Effluent Concentration in a Sewage Treatment Process, *Journal of Korean Society of Environmental Engineers*, 34(5), pp. 312-315. [Korean Literature]
- National Institute of Environment Research (NIER). (2013). *Water Information System*, National Institute of Environment Research, <http://water.nier.go.kr/>.
- Park, H. Y., Lee, J. K., Ha, H., Lee, H. B., Kim, Y. G., Park, C. O., and Park, S. I. (2006). A Correlation Study of Organic Matters by TOC, *Proceedings of the 2006 Spring Conference of the Korean Society of Environmental Engineers*, pp. 834-842. [Korean Literature]
- Seo, H. J., Kang, Y. J., Min, K. W., Lee, K. S., Seo, G. Y., Kim, S. H., Paik, K. J., and Kim, S. J. (2010). Characteristics of Distribution and Decomposition of Organic Matters in Stream Water and Sewage Effluent, *Analytical Science and Technology*, 23(1), pp. 36-44. [Korean Literature]