

ORIGINAL ARTICLE

감천 유역의 TOC 농도 추정을 위한 회귀 모형 개발 및 평가

정강영 · 안정민 · 이경락 · 김신 · 유재정 · 천세억 · 이인정*

국립환경과학원 낙동강물환경연구소

Development and Evaluation of Regression Model for TOC Contentation Estimation in Gam Stream Watershed

Kang-Young Jung, Jung-Min Ahn, Kyung-Lak Lee, Shin-Kim, Jae-Jeong Yu,
Se-Uk Cheon, In Jung Lee*

Nakdong River Environment Research Center, National Institute of Environmental Research, NIER, Goryeong 717-873, Korea

Abstract

In this study, it is an object to develop a regression model for the estimation of TOC (total organic carbon) concentration using investigated data for three years from 2010 to 2012 in the Gam Stream unit watershed, and applied in 2009 to verify the applicability of the regression model. TOC and COD_{Mn} (chemical oxygen demand) were appeared to be derived the highest correlation. TOC was significantly correlated with 5 variables including BOD (biological oxygen demand), discharge, SS (suspended solids), Chl-a (chlorophyll a) and TP (total phosphorus) of $p < 0.01$. As a result of PCA (principal component analysis) and FA (factor analysis), COD, TOC, SS, discharge, BOD and TP have been classified as a first factor. TOC concentration was estimated using the model developed as an independent variable BOD₅ and COD_{Mn}. R squared value between TOC and measurement TOC is 0.745 and 0.822, respectively. The independent variable were added step by step while removing lower importance variable. Based on the developed optimal model, R squared value between measurement value and estimation value for TOC was 0.852. It was found that multiple independent variables might be a better the estimation of TOC concentration using the regression model equation(in a given sites).

Key words : TOC, Gam Stream, Regression analysis model, Correlation analysis

1. 서론

기존의 유기물질 관리는 생물학적으로 분해 가능한 유기물질에 의해서 발생하는 영향을 최소화하는 측면에 중점을 두고 이루어져왔으며 이를 행정목표로 정하여 공 공수역을 관리해왔다(Choi 와 Han, 2011). 그리고 지금 까지 환경부는 환경기초시설확대, 오염원 관리 등 지속

적인 유역중심의 예방적 수질관리 정책을 통해 유기물질 (BOD₅)에 대한 수질개선을 노력해 왔다(Kim 등, 2014). 그 결과로 BOD₅ 오염도는 지속적으로 개선되어 2011년 에는 전국하천의 81.6%가 BOD₅ 농도 3 mg/L이하인 2 급수 이상인 것으로 조사 보고되었다. 이처럼 유기물질 에 의한 오염정도를 평가하기 위한 지표로써 생물학적산 소요구량(BOD₅)과 화학적산소요구량(COD_{Mn})이 널리

Received 27 January, 2015; Revised 20 May, 2015;

Accepted 28 May, 2015

*Corresponding author : In Jung Lee, Nakdong River Environment Research Center, National Institute of Environmental Research, NIER, Goryeong 717-873, Korea
E-mail: ijlee@me.go.kr

© The Korean Environmental Sciences Society. All rights reserved.

© This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

이용되어 왔으며, 국내 공공수역의 수질관리 정책에서도 수질 오염도를 평가하는 중요한 지표로 활용되어 왔다(Choi 와 Han, 2011). 하지만 최근 급격한 산업화와 도시화로 인해 유역에서 발생하는 오염물질이 다양하고, 난분해성 유기물을 포함한 비점오염의 증가로 BOD₅와 COD_{Mn}만으로는 유기물 오염도를 제대로 평가 할 수 없다(Lyu 와 Lee, 2007; Choi 등, 2012). 또한, 하 · 폐수 내 유기물의 50% 이상이 분석 불가능한 복잡한 형태의 유기물로 구성되어 있다고 보고한 바 있다(Dignac 등, 2000). BOD₅는 미생물이 쉽게 분해 가능한 저분자 유기물질의 양을 나타내는 지표로, 전체 유기물의 20~40%를 측정 할 수 있지만, 질산화 등의 오차요인과 5일 간의 분석 소요기간의 한계가 있다(Lyu 와 Lee, 2007; Choi 와 Han, 2011). COD_{Mn}는 미생물의 분해가 쉽지 않고 분해하는데 오래 걸리는 고분자 유기물질까지 측정 할 수 있으나 이 역시 난분해성 유기물 관리에 한계가 있다(Park 등, 2013). 이와 같은 이유 때문에 BOD₅와 COD_{Mn}의 한계를 보완하는 일환으로 분석이 용이하고, 난분해성 유기물을 포함해서 전체유기물의 안정적이고 정확한 측정이 가능한 수단인 TOC를 고려해 왔다. 2006년 TOC 시험방법이 채택이 되었고, 2013년에는 공공수역의 유기물질 수질환경기준에 TOC가 도입되었다(Choi 등, 2012). 하지만 TOC의 경우 그 동안 축적된 자료가 많지 않고 과거와 현재의 자료 연계성 문제 때문에 유기물질 관리 항목 전환에 어려움이 있다(Choi 등, 2012; Park 등, 2014). 그간 선행된 연구에 의하면 TOC와 BOD₅, COD_{Mn}간의 상관성을 규명한 연구는 많은 연구자들을 통해 수행되어졌으며, BOD보다 COD_{Mn}가 상관성이 높다고 보고 하였다(Kim 등, 2007; Park 등, 2006). 최근에는 하수처리장 방류수의 TOC농도 추정을 위해 COD_{Mn}를 독립변수로 가진 단순회귀모형을 개발하여 하수처리장 방류수 유기물질 농도의 자료 보완에 활용하였다(Lee 등, 2010). Choi 등(2012)은 영농기간 동안 논 유출수의 BOD₅, COD_{Mn}자료를 이용하여 TOC농도를 추정하였으며, Jung 등(2013)은 영산강을 대상으로 하천수내 TOC농도 추정을 위한 회귀모형을 개발하여 적용 평가하였다. 선행된 연구에서는 회귀모형개발을 수행하기 위해 모집단의 변수들을 선정함에 있어 pearson 상관분석을 많이들 사용하였다. 하지만 본 연구에서는 pearson 상관분석과 함께 하천의 수질특성을 해석하기

위한 연구로 사용되는 다변량분석 기법인 주성분분석 및 요인분석을 수행하여 회귀모형을 개발하려 한다. 이 두 방법을 사용하여 추출한 요인들을 회귀분석에 이용하여 과거 TOC를 추정한다면 보다 효과적인 것으로 판단된다. 그리고 낙동강 수계에서는 결측 자료의 보완에 활용될 수 있는 TOC 농도추정을 위한 연구는 드문 실정이다. 따라서 앞서 언급한 방법으로 감천유역에 대해 TOC 농도추정을 위한 회귀모형을 개발하여 적용성을 검증 평가하였다.

2. 재료 및 방법

2.1. 대상유역 및 자료분석

감천은 경상북도 김천시 대덕면 태리와 경상남도 거창군 웅양면 산포리의 경계인 우두령재에서 발원하여 북동류한 후 낙동강 중류부 우안측(낙동강 하구로부터 223.4 km지점)으로 합류하는 하천으로 낙동강의 제1지류이며, 유역면적은 약 1,006 km²이다. 유로연장은 비교적 긴 편으로 69 km이며, 하상은 상류부는 잔자갈 및 모래로 형성되어 있고, 중 · 하류부는 양질의 모래로 이루어져 있다. 감천의 총 지류는 부향천, 무릉천, 대방천, 직지사천, 아천, 울곡천등 총 16개 정도가 합류한다. 행정구역은 김천시가 879.48 km²로 전체유역면적의 87%, 구미시가 126.40 km²로 13%를 차지하고 있으며, 감천유역의 연평균 강수량은 구미지역이 1,271.5 mm로 전국 평균인 1,302.1 mm보다 적은 편이다. 그리고 본 연구에서는 감천 단위유역발단(Fig. 1)에서 2010년~2012년까지 8일 간격으로 3년간 관측한 자료를 회귀모형 개발에 활용하였으며, 2009년에 8일 간격으로 관측한 자료는 회귀모형의 적용성을 검증 평가하는데 활용하였다. 수질자료는 국립환경과학원 낙동강물환경연구소 측정자료를 이용하였으며, 항목은 BOD₅, COD_{Mn}, TOC, TN, TP, Chl-a, Water temperature, pH, DO, EC, Discharge, SS 등 총 12개 항목을 분석에 사용하였다.

2.2. 상관분석

상관성 분석은 수질인자간 연관성이 있는 수질항목을 파악하기 위하여, 인자간의 상관관계를 분석하여 상호연관성이 높은 인자와 항목을 추출하는 방법이다. 상관분석은 두 변수간의 선형적인 상관관계를 파악해보는 것

으로 한 변수가 다른 변수와의 상관성이 있는지 여부와 상관성이 있다면 어느 정도의 상관성이 있는지를 알고자 할 때 유용한 분석방법이다. 보편적으로 Pearson 상관계수(Pearson Correlation Coefficient, r)가 가장 많이 사용되며, 상관계수는 -1부터 +1 사이의 값을 갖는다(Cho, 2011).

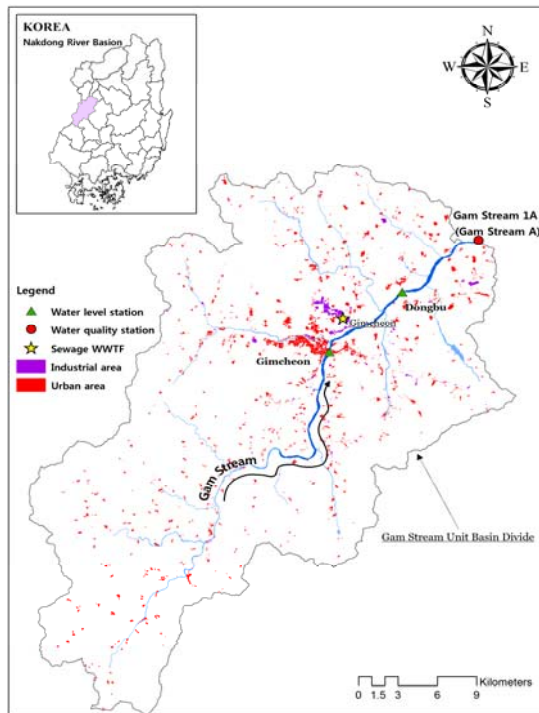


Fig. 1. Sampling locations in the Gam Stream watershed.

2.3. 회귀분석

회귀분석이란 변수들 중 한 개를 종속변수로 나머지를 독립변수로 하여 이들 변수들이 서로 상관관계를 가질 때 이러한 관계를 규명하고자 하는 방법으로 가장 많이 이용되는 통계기법이다. 회귀분석에는 단순회귀분석(Simple Regression Analysis)과 다중회귀분석(Multiple Regression Analysis)이 있으며, 회귀분석의 기본식은 다음과 같이 표현된다.

$$Y = \beta_1 + \beta_2 X \quad (1)$$

여기서, 기본적인 과업은 β_1 β_2 를 구하는 것인데 β_1 는

회귀계수, β_2 는 상수항으로 종속변수 Y 는 TOC농도이며 독립변수 X 는 BOD_5 또는 COD_{Mn} 의 농도이다. 또한, 종속변수와 여러 개의 독립변수들 사이의 관계를 규명하고자 할 때의 기본식은 다음과 같이 나타낸다.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_i X_i \quad (2)$$

여기서, 종속변수 Y 는 추정치 TOC농도이며 X 는 i 번째 독립변수, β 는 Y 에 대한 i 번째 독립변수 X_i 의 회귀계수이다. 회귀모형은 SPSS(ver. 17.0) 프로그램을 이용하였으며, 통계적 유의성을 검증하기 위해서 F 검정법에 의한 F 통계량을 이용하였다. 이 때 계산된 F 통계량의 값에 대한 유의확률이 유의수준보다 작으면 회귀계수가 유의함을 의미한다(Jung 등, 2013; Park 등, 2014). 또한, 회귀식의 적합성을 나타내는 것이 결정계수(R^2)이다. 결정계수(R^2)은 회귀식에 의하여 설명된 변동이 총 변동에서 차지하는 상대적 크기를 나타내며, 결정계수가 1에 가까울수록 잔차가 없는 회귀식으로 설명되고, 반대로 0에 가까울수록 회귀식이 변동을 설명하지 못함을 뜻한다.

2.4. 주성분 분석 및 요인분석

여러개의 변수들이 내포된 다변량 자료를 분석하는 일은 매우 복잡하고 힘든 경우가 많다. 이런 경우 변수들의 일차결합을 통해 간단한 구조를 갖도록 자료를 축약하고 공분산 행렬을 최대한 설명하는 몇 개의 일차결합을 산출하는 것을 주성분이라 하며, 이러한 주성분을 구하고 해석하여 자료를 축약하는 방법을 주성분분석이라 한다(Kim 등, 2007). 요인분석은 p (임의의 변수 수)차원 변수들간의 상호의존 구조를 나타내는 공분산이나 상관행렬에서 m 개(단, $m < p$)의 공통인자(common factor)를 추출하여 원래 변수들이 나타내는 복잡한 구조를 쉽게 파악하기 위한 자료축약(data reduction)기법이다(Park 등, 2001; Sakamoto 등, 1988). 또한, 수많은 변수들 중에 상관관계를 이용하여 여러 변수들로 측정된 자료를 보다 적은 수로 줄여 정보를 요약하고 변수들을 적절히 조합함으로써 중요도가 낮은 변수를 제거 축약하는 분석 방법이기도 하다. 요인 패턴의 구조를 단순화시켜서 요인에 대한 해석을 보다 쉽게 하기 위해서 요인축

회전시키며, 요인축을 회전하는 방법에는 크게 직교(orthogonal)회전과 사각(oblique) 회전이 있다. 직교회전으로는 Varimax, Quartimax, Eqimax 등이 있는데, 본 연구에서는 Varimax 직각회전방식을 사용하였다.

3. 결과 및 토의

3.1. 수질특성분석

모니터링지점에서 수질분석 결과는 Table 1(2010~2012)과 Table 2(2009)와 같다. Table에서 2010부터 2012까지 산술평균하여 요약한 수질 자료와 2009년의 요약한 수질 자료를 비교하여 살펴보면 BOD₅의 농도범위는 0.4~8.4 mg/L 정도이며, 평균값은 각각 1.5 mg/L와 1.6 mg/L으로 생활환경기준 I b(좋음)등급에 해당되며 차이가 없는 것으로 나타났다. COD_{Mn}의 농도범위는 2.5~13.7 mg/L 정도이며, 평균값은 각각 4.6 mg/L와 5.2 mg/L으로 생활환경기준 II(약간좋음)~III(보통)등급으로 2009년에 비해 낮지만 큰 차이는 없는 것으로 나타났다. TOC의 경우는 농도범위가 1.8~11.2 mg/L 정도이며, 평균값은 각각 3.1 mg/L와 3.7 mg/L으로 생활환경기준 II(약간좋음)등급으로 COD_{Mn}와 동일한 결과를 보였다. TP의 농도범위는 0.04~0.675 mg/L이고, 평균 농도는 각각 0.156 mg/L와 0.279 mg/L으로 생활환경기준 III(보통)~IV(약간나쁨)등급으로 2009년에 비해 농도가 많이 개선된 결과를 보였다. 그러나 비점오염원 대책과 하수처리장에서 인 처리시설 확충 강화 등의 영양염류 저감(삭감) 대책이 필요 할 것으로 판단된다. 감천의 유량은 0.314~703.482 m³/sec로 매우 큰 변동 폭을 보이고 있으며, 다소 차이는 있지만 평균 17.017 m³/sec~4.055 m³/sec 정도의 양이 흐르는 것으로 실측 조사되었다.

3.2. 상관분석

감천 단위유역발단에서 회귀식 개발의 독립변수를 결정하는 첫 번째 방법으로 SPSS(ver. 17.0) 통계프로그램을 이용하여 항목 간 상관관계를 분석하였다. 통계분석을 실시하기 전에 관측값이 정규분포형태를 가지는지 검토하였으며, 대부분의 관측값이 정규분포형태를 가지는 것으로 나타났다. 정규분포를 따르지 않는 관측값에 대해서는 적절한 자료의 변환을 검토하였다. 그러나 이러

한 정규성의 가정이 반드시 필요한 것은 아니어서 본 연구에서는 자료의 변환 없이 자료를 그대로 표준화 하여 분석하였다(Kim 등, 2007; Kim 등, 2013). 통계분석에 앞서 언급한대로 2009년 자료는 회귀모형의 적용성을 검토하기 위한 자료로 사용하였기 때문에 2010년~2012년까지 3년간 관측한 자료만을 사용하여 상관분석을 실시하였다.(Table 3.) Discharge이 SS와 0.784(p<0.01)로 유의성이 높게 나타났는데 이는 강우 시 토사 유출의 형태로 하천에서 유량이 증가 할 때 탁수(부유물이 함유된 물) 유발에 영향을 주고 있음을 제시한다. 또한 COD_{Mn}와 0.585(p<0.01), TOC와 0.684(p<0.01), 그리고 SS와 TOC가 0.640(p<0.01)로 비교적 높은 양의 상관관계를 보인 것은 비점유입 영향의 결과로 볼 수 있다(Yoon 등, 2006). EC는 BOD₅ 0.308(p<0.01)와 영양염류인 T-P 0.346(p<0.01), T-N 0.522(p<0.01)과도 양의 상관성을 보여 수중에서 유기물과 함께 전해질로 작용함을 의미한다. TOC와 T-P사이의 상관계수는 0.498 (p<0.01)로서 상관관계가 있는 것으로 나타났는데 이는 유기물과 영양염류(T-P)사이의 상관성을 의미하는 결과로 해석된다. Chl-a와 유기물 지표인 BOD₅, COD_{Mn}, TOC의 상관성은 0.469(p<0.01), 0.448(p<0.01), 0.390 (p<0.01)로 나타나 Chl-a이 유기물 증가에 기여할 가능성이 높다(Park 등, 2001). TOC와 BOD₅는 0.608(p<0.01)로 비교적 높은 양의 상관성으로 나타났다. TOC는 생분해성과 난분해성 유기물을 모두 포함한 것이고, BOD₅는 생분해성 유기물만 포함한 것으로 서로 의미가 다르다. 그러므로 유기물 성상에 따라서 시 · 공간적으로 상관성의 차이(높고, 낮음)가 있을 수 있다. 한편 TOC와 COD_{Mn} 사이의 상관계수는 0.926(p<0.01)로 나타났는데 이 값은 통계적으로 상관관계가 매우 높은 범위에 해당한다. 이처럼 상관관계가 높은 TOC와 COD_{Mn} 그리고 BOD₅는 모두 유기물에 기인하기 때문에 유기물 오염특성 해석 시 주요 지표로 활용 될 수 있다. 그 외에 상관분석에서 TOC와 p<0.01 유의 수준에서 상관관계가 있는 변수로는 Discharge와 SS, Chl-a, T-P 등으로 나타났다.

3.3. 주성분 분석 및 요인분석

회귀식 개발의 독립변수를 결정하는 두 번째 방법으로 대상지점에 대하여 먼저 주성분분석을 실시하였다. 주성분분석은 그 자체가 목적이 아니라기보다는 일단 자료를

Table 1. Summary of water quality parameters in the Gam stream during January, 2010 - December, 2012

| | Discharge (m ³ /sec) | Watertemp (°C) | DO (mg/L) | pH | EC (μs/cm) | BOD ₅ (mg/L) | COD _{Mn} (mg/L) | TOC (mg/L) | SS (mg/L) | TN (mg/L) | TP (mg/L) | Chl- <i>a</i> (mg/m ³) |
|-------|------------------------------------|-------------------|--------------|-------|---------------|----------------------------|-----------------------------|---------------|--------------|--------------|--------------|---------------------------------------|
| Max | 703.482 | 30.9 | 14.7 | 8.7 | 497.0 | 8.4 | 13.7 | 11.2 | 158.0 | 7.320 | 0.525 | 26.9 |
| Min | 0.625 | 0.0 | 7.0 | 6.8 | 91.0 | 0.4 | 2.5 | 1.8 | 1.6 | 1.884 | 0.043 | 0.7 |
| Mean | 17.017 | 17.0 | 9.9 | 7.6 | 290.6 | 1.5 | 4.6 | 3.1 | 13.3 | 4.160 | 0.156 | 5.5 |
| Count | 116 | 116 | 116 | 116 | 116 | 116 | 116 | 116 | 116 | 116 | 116 | 116 |
| SD | 66.866 | 8.233 | 1.592 | 0.380 | 85.932 | 1.008 | 1.501 | 1.124 | 18.406 | 1.075 | 0.097 | 4.612 |

Table 2. Summary of water quality parameters in the Gam stream during January, 2009 - December, 2009

| | Discharge (m ³ /sec) | Watertemp (°C) | DO (mg/L) | pH | EC (μs/cm) | BOD ₅ (mg/L) | COD _{Mn} (mg/L) | TOC (mg/L) | SS (mg/L) | TN (mg/L) | TP (mg/L) | Chl- <i>a</i> (mg/m ³) |
|-------|------------------------------------|-------------------|--------------|-------|---------------|----------------------------|-----------------------------|---------------|--------------|--------------|--------------|---------------------------------------|
| Max | 53.041 | 30.3 | 16.7 | 9.5 | 611.0 | 4.8 | 9.2 | 7.2 | 37.8 | 6.058 | 0.675 | 117.2 |
| Min | 0.314 | 2.1 | 7.1 | 6.4 | 171.0 | 0.5 | 2.8 | 1.9 | 2.2 | 2.146 | 0.127 | 2.9 |
| Mean | 4.055 | 19.0 | 10.8 | 8.1 | 440.6 | 1.6 | 5.2 | 3.7 | 8.4 | 4.213 | 0.279 | 14.2 |
| Count | 42 | 42 | 42 | 42 | 42 | 42 | 42 | 42 | 42 | 42 | 42 | 42 |
| SD | 9.779 | 8.015 | 2.128 | 0.621 | 99.393 | 0.956 | 1.354 | 1.064 | 6.982 | 0.985 | 0.127 | 18.615 |

Table 3. Pearson correlation coefficient among the water quality parameters

| | Discharge (m ³ /sec) | Watertemp (℃) | DO (mg/L) | pH | EC (μs/cm) | BOD ₅ (mg/L) | COD _{Mn} (mg/L) | TOC (mg/L) | SS (mg/L) | TN (mg/L) | TP (mg/L) | Chl- <i>a</i> (mg/m ³) |
|---------------|------------------------------------|------------------|--------------|----------|---------------|----------------------------|-----------------------------|---------------|--------------|--------------|--------------|---------------------------------------|
| Discharge | 1.000 | 0.131 | -0.192* | -0.163 | -0.374** | 0.128 | 0.585** | 0.684** | 0.784** | -0.199* | 0.224* | 0.025 |
| Watertemp | | 1.000 | -0.857** | 0.326** | -0.149 | -0.337** | -0.085 | -0.010 | 0.081 | -0.751** | -0.188* | 0.199* |
| DO | | | 1.000 | -0.249** | 0.245** | 0.264** | 0.048 | -0.014 | -0.162 | 0.635** | 0.172 | -0.151 |
| pH | | | | 1.000 | 0.331** | -0.007 | -0.077 | -0.086 | -0.065 | -0.077 | -0.093 | 0.207* |
| EC | | | | | 1.000 | 0.308** | 0.163 | 0.088 | -0.400** | 0.522** | 0.346** | 0.256** |
| BOD | | | | | | 1.000 | 0.688** | 0.608** | 0.352** | 0.520** | 0.385** | 0.469** |
| COD | | | | | | | 1.000 | 0.926** | 0.630** | 0.265** | 0.598** | 0.448** |
| TOC | | | | | | | | 1.000 | 0.640** | 0.195* | 0.498** | 0.390** |
| SS | | | | | | | | | 1.000 | -0.129 | 0.163 | 0.191* |
| TN | | | | | | | | | | 1.000 | 0.402** | -0.070 |
| TP | | | | | | | | | | | 1.000 | 0.205* |
| Chl- <i>a</i> | | | | | | | | | | | | 1.000 |

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Table 4. Initial eigenvalues and selected factor loading after varimax rotation

| Component | Initial Eigenvalues | | | Rotation Sums of Squared Loadings | | |
|-----------|---------------------|---------------|--------------|-----------------------------------|---------------|---------------|
| | Total | % of Variance | % Cumulative | Total | % of Variance | % Cumulative |
| 1 | 4.040 | 33.668 | 33.668 | 3.928 | 32.730 | 32.730 |
| (a) 2 | 3.107 | 25.893 | 59.561 | 3.090 | 25.747 | 58.478 |
| 3 | 1.903 | 15.860 | 75.421 | 2.033 | 16.943 | 75.421 |

간단히 하여 추후에 추가적인 다른 통계적 절차를 적용하기 위한 사전분석이라 볼 수 있다. 주성분 수를 결정하기 위해 고유치를 기준으로 결정하는 방법을 사용하여 고유치가 1 이상일 경우에 하나의 요인이 변수 1개 이상의 분산을 설명하며, 고유치가 1 이하인 경우 1개의 요인이 변수 1개의 분산을 설명할 수 없으므로 고유치가 1 이상이 되는 요인을 추출하는 방법으로 고유치가 1 이상을 만족하는 인자를 추출한 결과 3개의 요인이 추출되었다. 제 1요인은 32.730%, 제 2요인은 25.747%, 제 3요인은 16.943%를 기여하고 있으며, 주성분 전체 수질변동에 대하여 75.421%를 설명하고 있다(Table 4.).

요인분석 실행 전 분석에 대한 타당성을 알아보기 위해 Kaiser-Meyer-Olkin(KMO) test와 Bartlett's test를 실행하였다. KMO test는 변수들 간의 상관관계가 다른 변수에 의해 잘 설명되는 정도를 나타내는 척도로 1에 근

접한 값일수록 요인분석의 타당성이 높고 0.5이하일 경우 분석이 타당하지 않다. Bartlett's test의 구형 검증치는 각 변수가 서로 상관성이 존재함을 나타낼 때 유의적 관계라 하며 이때 0에 가까울수록 유의성이 높다(Bernard 등, 2004; Kim 등, 2007).

본 연구의 KMO test 결과 0.686로 분석되었고, Bartlett's test는 모두 0.000($p < 0.05$)을 나타내어 두 분석 모두 조건에 만족하여 요인분석이 가능함을 설명해 주고 있다. 요인분석은 관측된 변수들에 영향을 미치고 있는 숨어 있는 공통인자를 찾아내는 데 목적이 있으며 직교회전이 후 더 뚜렷하게 각 영향 요인을 규정한다. 제1요인은 COD_{Mn}, TOC, SS, Discharge, BOD₅, TP로 분류되었으며, 제2요인은 Watertemp, DO, TN이 제3요인은 EC, pH, Chl-a로 분류되었다(Table 5.).

Table 5. Rotated factor matrix extracted from principal component analysis

| Variable | Component | | |
|-------------------|--------------|--------|--------|
| | Com 1 | Com 2 | Com 3 |
| COD _{Mn} | 0.949 | 0.126 | 0.125 |
| TOC | 0.942 | 0.042 | 0.039 |
| SS | 0.782 | -0.129 | -0.370 |
| Discharge | 0.736 | -0.245 | -0.475 |
| BOD ₅ | 0.661 | 0.399 | 0.336 |
| TP | 0.551 | 0.335 | 0.248 |
| Watertemp | -0.007 | -0.928 | 0.167 |
| DO | -0.052 | 0.885 | -0.075 |
| TN | 0.146 | 0.867 | 0.254 |
| EC | -0.001 | 0.366 | 0.812 |
| pH | -0.090 | -0.348 | 0.630 |
| Chl-a | 0.461 | -0.205 | 0.569 |

Rotation Method : Varimax with Kaiser Normalization.

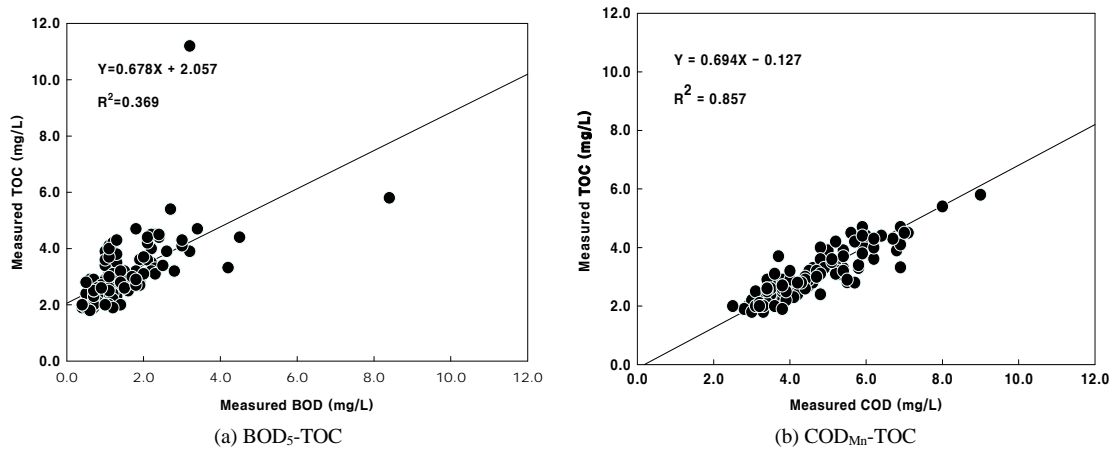


Fig. 2. Simple regression model for (a) and (b).

3.4. 회귀모형의 개발

먼저 관측한 BOD₅ 농도와 COD_{Mn} 농도를 독립변수로 가지고 TOC를 종속변수로 가지는 단순회귀모형을 개발하였다.(Fig. 2)

산포도(Scatter plot)상의 회귀선(Regression line)은 회귀식으로 나타낼 수 있는데 Fig. 2의 산포도(a)에서 구한 회귀식은 $TOC = 0.6782 \times BOD_5 + 2.0568$ 이다.(Table 6.) 회귀식의 기울기 0.6782는 BOD₅의 단위변화량에 대한 TOC의 변화량이며, BOD₅가 1 mg/L 증가할 때 TOC는 평균적으로 0.6782 mg/L 증가함을 의미한다.

(a)에서 TOC와 BOD₅의 적합성을 나타내는 결정계수 R^2 값은 0.369 낮게 나타났으며, (b)에서 TOC와 COD_{Mn}의 적합성을 나타내는 결정계수 R^2 값은 0.857로 높게 나타났다. 이와 같은 결과는 BOD₅ 산화율이 20~40% 정도이고, 그에 비해 COD_{Mn}과 COD_{Cr} 산화율은 60~80% 정도임을 감안할 때 상대적으로 BOD₅보다는 COD_{Mn}가 결정계수 R^2 값이 높게 나타날 수 있다. Kim

등(2007)은 산화되는 유기물의 양을 탄소의 양으로 환산하기 위해 산소와 탄소의 무게비를 BOD₅와 COD_{Mn} 농도에 곱한 후 산소와 탄소의 물비로 나누어서 BOD₅, COD_{Mn}의 산화율을 계산하였다. 이를 근거로 관측한 값을 가지고 계산한 평균 탄소 산화율이 BOD-C/TOC는 16.15%, COD-C/TOC 50.60%로 생분해성 유기물뿐만 아니라 난분해성 유기물까지 포함한 COD_{Mn}의 탄소 산화율이 BOD₅보다 높게 산정되어 선행된 연구 결과와도 유사하다(Choi 등, 2012). 한편, Kim 등(2013)은 낙동강 지류하천에서 TOC와 COD_{Mn}의 높은 선형 상관계수를 근거로 과거에 측정된 COD_{Mn} 자료를 TOC로 전환하여 장기 TOC를 구축 활용 안을 제안하였으며, Choi 등(2012)의 연구에서 BOD₅-TOC, COD_{Mn}-TOC의 선형 회귀식을 산정한 결과 결정계수가 각각 0.18, 0.53으로 COD_{Mn}자료를 이용하는 것이 BOD₅자료를 이용하는 것보다 유의성이 있음을 보고한 바 있다.

앞서 언급한바와 같이 pearson 상관분석과 함께 주성

Table 6. Results of the simple regression equations to estimate total organic carbon

| Model | Rgression model | r | R ² | F-value | P-value |
|-------|--|-------|----------------|---------|---------|
| a | $TOC = 2.057 + 0.678 \times BOD_5$ | 0.608 | 0.369 | 66.709 | 0.000 |
| b | $TOC = -0.127 + 0.694 \times COD_{Mn}$ | 0.926 | 0.857 | 686.244 | 0.000 |

a : BOD₅ , b : COD_{Mn}.

Table 7. Results of the multiple regression equations to estimate total organic carbon

| Model | Rgression model | r | R ² | F-value | P-value |
|-------|--|-------|----------------|---------|---------|
| c | $TOC = -0.133 + 0.704 \cdot COD_{Mn} + (-0.008 \cdot Chl-a)$ | 0.926 | 0.858 | 340.775 | 0.000 |
| d | $TOC = -0.166 + 0.726 \cdot COD_{Mn} + (-0.005 \cdot Chl-a) + (-0.005 \cdot BOD_5)$ | 0.927 | 0.859 | 227.437 | 0.000 |
| e | $TOC = -0.040 + 0.677 \cdot COD_{Mn} + (-0.003 \cdot Chl-a) + (-0.043 \cdot BOD_5) + 0.005 \cdot SS$ | 0.929 | 0.863 | 175.326 | 0.000 |
| f | $TOC = -0.099 + 0.727 \cdot COD_{Mn} + (-0.005 \cdot Chl-a) + (-0.049 \cdot BOD_5) + 0.004 \cdot SS + (-0.826 \cdot TP)$ | 0.931 | 0.866 | 142.341 | 0.000 |
| g | $TOC = 0.329 + 0.582 \cdot COD_{Mn} + 0.008 \cdot Chl-a + 0.101 \cdot BOD_5 + (-0.009 \cdot SS) + (-0.730 \cdot TP) + 0.006 \cdot Discharge$ | 0.948 | 0.899 | 161.419 | 0.000 |

c : (COD_{Mn}, Chl-a), d : (COD_{Mn}, Chl-a, BOD₅), e : (COD_{Mn}, Chl-a, BOD₅, SS), f : (COD_{Mn}, Chl-a, BOD₅, SS, TP), g : (COD_{Mn}, Chl-a, BOD₅, SS, TP, Discharge).

분분석 및 요인분석을 통해 결정되어진 독립변수는 COD_{Mn}, BOD₅, Discharge, SS, Chl-a, T-P 등 6개 항목이다. Park 등(2014)은 다른 유역에서 TOC농도를 추정함에 있어 해당 유역에서만 나타나는 특성에 주의하여 인자를 선별 할 필요가 있다고 제시하였다. 이와 같이 다중회귀모형을 개발하기 위하여 추출된 독립변수를 단계적으로 추가하면서 분석을 시행하여 결과를 Table 7에 나타내었다.

개발된 model은 모두 TOC농도를 종속변수로 가지며, model들의 F-value에 대한 P-value가 0.000으로 나타나 개발된 model이 종속변수를 설명하는데 유용한 것으로 나타났다. 따라서, 이들 5개의 다중회귀모형 중 6개의 독립변수(COD_{Mn}, BOD₅, Discharge, SS, Chl-a, T-P)를 가지며 결정계수 R²값이 0.899인 g model을 TOC농도 추정을 위한 최적의 다중회귀모형으로 선정하였다.

3.5. 회귀모형의 적용성 평가

이상의 논의를 토대로 개발된 회귀모형의 적용성을 검토 평가하기 위해 2009년 관측 자료를 활용하여 TOC 농도를 추정하였다. 개발한 단순회귀모형과 다중회귀모형을 적용한 실측치와 추정치에 대한 분석 결과는 Fig. 3과 같다. 단순회귀모형을 이용하여 산정한 TOC농도 추정치와 실측치간의 결정계수 R²값은 BOD₅=0.745, COD_{Mn}=0.822로 나타났고, 다중회귀모형을 이용하여 산정한 TOC농도 추정치와 실측치간의 결정계수 R²값은

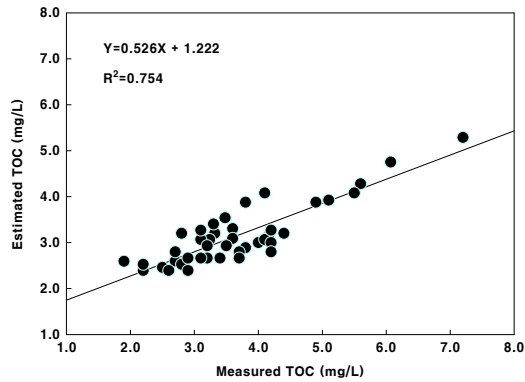
0.852로 나타났다. 이 결과는 단순회귀모형을 이용하여 추정한 TOC값 보다 다중회귀모형을 이용하여 추정한 TOC값이 좀 더 정확성이 높다는 의미로 해석 할 수 있다. 그러므로 단순히 하나의 독립변수만 가진 회귀식보다 여러개의 다양한 구성인자를 가진 회귀식을 이용하여 TOC값을 추정하는 것이 타당하다고 판단된다.

4. 결론

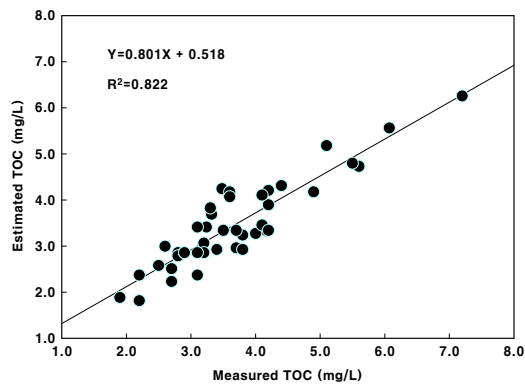
본 연구는 낙동강수계에 위치한 감천 단위유역 말단 지점의 2010년~2012년까지 3년간 관측한 수질자료로 상관분석을 통해 회귀모형을 개발하고 2009년에 관측한 자료는 회귀모형의 적용성을 검토 평가하는데 활용하였으며, 주요 결과는 다음과 같다.

1) 감천 단위유역 모니터링지점에서 관측한 수질자료를 분석한 결과 생활환경기준 I b(중음)등급의 BOD₅ 항목은 관리가 잘 되고 있는 것으로 나타났다. 하지만 COD_{Mn}, TOC의 평균농도가 생활환경기준 II(약간중음)~III(보통)등급으로 나타나 생물학적으로 분해 불가능한 유기물을 포함한 전체 유기물질 관리가 필요할 것으로 보인다. 또한, TP의 경우도 하수처리 강화와 토양 표면으로부터 비점오염원의 표면 유출에 따른 인의 발생 비율을 낮추는 대책을 추진하여 효과적으로 영양염류를 관리하는 방안을 모색할 필요가 있을 것으로 판단된다.

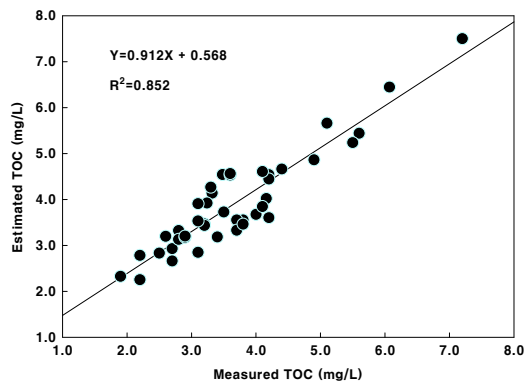
2) 항목 간 상관관계를 분석한 결과 TOC와 COD_{Mn}가 가장 큰 상관성을 보이는 것으로 나타났으며, 다음으



(I) Simple regression model using BOD



(II) Simple regression model using COD



(III) Multiple regression model

Fig. 3. Comparison of the measured and estimated total organic carbon using regression models.

로 Discharge, SS, BOD₅, T-P, Chl-a 순으로 나타났다. 또한 BOD₅-TOC, COD_{Mn}-TOC의 선형 회귀식을 산정한 결과 결정계수가 각각 0.369, 0.857으로 선행 연구들과 유사한 결과를 보여 TOC는 BOD₅보다는 COD_{Mn}와 더 높은 상관성이 있음을 알 수 있다.

3) BOD₅와 COD_{Mn}를 독립변수로 가진 단순회귀모형을 이용하여 추정된 TOC농도 실측치와 추정치간의 결정계수 R²값은 각각 0.745, 0.822로 나타났고, 요인분석에서 제1요인으로 추출된 독립변수들을 단계적으로 추가하면서 개발된 회귀모형 중 최적 모형을 이용하여 추정된 TOC농도 실측치와 추정치간의 결정계수 R²값은 0.852로 나타나 단순회귀모형을 이용하여 추정된 TOC값 보다 다중회귀모형을 이용하여 추정된 TOC값이 더 정확성이 높은 것을 알 수 있었다. 또한 유역 특성이 다른 지점에서 TOC농도를 추정함에 있어서 해당 유역에서만 나타나는 특성을 분석하여 인자를 선별 할 필요가 있음을 확인하였다. 따라서 본 연구결과들은 향후 특정유역에서의 효율적인 회귀모형개발을 위한 좋은 정보를 제공해 줄 것이다.

REFERENCE

- Bernard, P., Antoine, L., Bernard, L., 2004, Principal component analysis an appropriate tool for water quality evaluation and management application to a tropical lake system, *Ecological Modeling*, 178, 295-311.
- Cho, H. K., 2011, A study on the related characteristics of discharge-water quality in Nakdong river, *J. Environ. Sci.*, 20(3), 373-384.
- Choi, J. Y., Han, D. H., 2011, Development of water quality standard for TOC as organic matter index, *The Seoul Institute*, 12(3), 173-184.
- Choi, D. H., Jung, J. W., Yoon, K. S., Lee, K. S., Choi, W. J., Lim, S. S., Park, H. N., Yim, B. J., Hwang, T. H., 2012, Estimation of TOC concentration using BOD, COD in runoff from paddy fields, *J. Korean Soc. Water Environ.*, 28(6), 813-818.
- Dignac, M. F., Ginestet, P., Rybacki, D., Bruchet, A., Urbain, V., and Scribe, P., 2000, Fate of wastewater organic pollution during activated sludge treatment: nature of residual organic matter, *Water Res.*, 34(17),

- 4185-4194
- Jung, J. W., Cho, S. H., Choi, J. H., Kim, K. S., Jung, S. J., Lim, B. J., 2013, Development and evaluation of simple regression model and multiple regression model for TOC contentation estimation in stream flow, *J. Korean Soc. Water Environ.*, 29(5), 625-629.
- Kim, J. K., Shin, M. S., Jang, C. W., Jung, S. M., 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, *J. Korean Soc. Water Environ.*, 23(1), 72-80.
- Kim, B. C., Jung, S. M., Jang, C. W., Kim, J. K., 2007, Comparison of BOD, COD and TOC as the indicator of organic matter pollution in streams and reservoirs of Korea, *J. Korean. Soc. Environ. Eng.*, 29(6), 640-643.
- Kim, M. A., Lee, J. K., Zoh, K. D., 2007, Evaluation of the Geum river by multivariate analysis: principal component analysis and factor analysis, *J. Korean Soc. Water Environ.*, 23(1), 161-168.
- Kim, H. S., Hong, J. J., Seong, J. U., Choi, K. S., Park, J. C., 2013, Comparison of organic matter distribution in major tributaries of the Nakdong river, *J. Korean Soc. Water Environ.*, 29(5), 618-624.
- Kim, G. H., Jung, K. Y., Lee, I. J., Lee, K. L., Cheon, S. U., Im, T. H., Yoon, J. S., 2014, A study on instream flow for water quality improvement in lower watershed of Nam River dam, *J. Korean Soc. Water Environ.*, 30(1), 44-59.
- Kim, G. H., Jung, K. Y., Yoon, J. S., Cheon, S. U., 2013, Temporal and spatial analysis of water quality data observed in lower watershed of Nam river dam, *J-KOSHAM.*, 13(6), 429-437.
- Lyu, J. H., Lee, D. G., 2007, Inquiry of water environment in mihoccheon (stream) -water quality monitoring focused on TOC-, *J. Korean Soc. Water Environ.*, 23(5), 731-739.
- Lee, T. H., Lee, B. M., Hur, Jin., Jung, M. S., Kang, T. G., 2010, Conversion of COD_{Mn} into TOC and refractory organic matter concentrations for treated sewage using regression equations, *J. Korean Soc. Water Environ.*, 26(6), 969-975.
- Park, H. S., Park, C. K., Song, M. K., Baek, K. H., Shin, S. K., 2001, Evaluation of water quality characteristics using factor analysis in the Nakdong river, *J. Korean Soc. Water Environ.*, 17(6), 693-701.
- Park, H. Y., Lee, J. K., Ha, H., Lee, H. B., Kim, Y. K., Park, C. O., Park, S. I., 2006, A correlation study of organic matters by TOC, *J. Korean. Soc. Environ. Eng.*, 2006 Proceeding of Spring Conference, 834-842.
- Park, J. H., Park, B. K., Lee, I. J., Lee, J. K., Rhew, D. H., 2013, Necessity of refractory organic matters management in total maximum daily loads (TMDLs), *J. Korean Soc. Water Environ.*, 29(3), 393-399.
- Park, J. W., Moon, M. J., Han, S. W., Lee, H. J., Jung, S. J., Hwang, K. S., Kim, K. S., 2014, Application of regression analysis model to TOC concentration estimation -Osu stream watershed-, *J. Environ. Impact Assess.*, 23(3), 187-196.
- Sakamoto, H., Hamaguchi R., Iwase M., 1988, Evaluation of river water pollution in the chanai district of eastern hokkaido by the multivariate analysis method, *Water and Waste*, 30(3), 239-245.
- Yoon, Y. S., Yu, J. J., Kim, M. S., Lee, H. J., 2006, Computation and assessment of delivery pollutant loads for the streams in the Nakdong river basin, *J. Korean Soc. Water Environ.*, 22(2), 277-287.