J. Korea Water Resour. Assoc. Vol. 57, No. 1 (2024), pp. 9-19
doi: 10.3741/JKWRA.2024.57.1.9
pISSN 2799-8746
eISSN 2799-8754

A study of statistical analysis method of monitoring data for freshwater lake water quality management

Chegal, Sundong[a]* · Kim, Jin[b]

[a]Director, Company-affiliated research institute, Hycerg, Gwacheon, Korea
[b]Assistant Manager, Company-affiliated research institute, Hycerg, Gwacheon, Korea

Paper number: 23-079
Received: 16 October 2023; Revised: 15 December 2023; Accepted: 19 December 2023

Abstract

As using public monitoring data, analysing a trends of water quality change, establishing a criteria to determine abnormal status and constructing a regression model that can predict Chlorophyll-a, an indicator of eutrophication, was studied. Accordingly, the three freshwater lakes were selected, approximately 20 years of water quality monitoring data were analyzed for periodic changes in water quality each year using regression analysis, and a method for determining abnormalities was presented by the standard deviation at confidence level 95%. By calculating the temporal change rate of Chlorophyll-a from irregular observed data, analyzing correlations between the rate and other water quality items, and constructing regression models, a method to predict changes in Chlorophyll-a was presented. The results of this study are expected to contribute to freshwater lake water quality management as an approximate water quality prediction method using the statistical model.

Keywords: Freshwater lake, Chlorophyll-a, Correlation, Multiple regression, Polynomial regression

담수호 수질관리를 위한 측정자료의 통계적 분석방법 연구

제갈선동[a]* · 김진[b]

[a](주)하이써그 기업부설연구소 이사, [b](주)하이써그 기업부설연구소 대리

요 지

본 연구는 공개된 수질측정 자료를 이용하여 담수호의 수질변화추이를 분석하고 수질항목의 이상여부의 판단기준을 마련하며, 자료로부터 부영양화의 지표인 Chlorophyll-a를 예측할 수 있는 회귀모형을 구성하여 담수호 관리에 이용할 수 있는 방안을 검토하고자 하였다. 이에 따라 서해안 담수호 3개소를 선정하여 약 20년간의 수질항목자료를 회귀분석 방법으로 분석하고, 각 수질항목의 연중 주기적인 변화를 나타내는 회귀식과 신뢰도 95%에서의 표준편차를 산정함으로서 이상 여부의 판단방법을 제시하였다. 또한 불규칙한 관측일로부터 Chlorophyll-a의 시간적 변화율을 산정하고, 다른 수질항목간의 상관관계 분석 및 회귀모형을 구성하여 분석함으로서 수질측정 자료만을 이용하여 Chlorophyll-a의 변화를 예상할 수 있는 방법을 제시하였다. 본 연구결과는 통계학적 모형에 의한 근사적인 수질예측방법으로서 향후 수질측정 자료의 양적·질적 개선이 이루어진다면 담수호 수질관리에 기여할 것으로 기대된다.

핵심용어: 담수호, Chlorophyll-a, 상관분석, 다중회귀, 다항회귀

***Corresponding Author.** Tel: +82-31-346-1941
E-mail: chegalsd@hanmail.net (Chegal, Sundong)

© 2024 Korea Water Resources Association. All rights reserved.

1. 서 론

국내 담수호는 유역의 말단에 해당되고 해안 인근에 인위적인 방조제 건설을 통해서 형성되었는데, 국토의 약 70%가 산지로 구성되어 있는 우리나라의 경우 강우의 유출이 빨라 생활용수 및 농업용수의 확보 목적으로서 담수호의 중요성이 강조되고 있다(Lee *et al*., 2022).

상공업의 고도화와 다양화, 농축산업의 고밀도화 및 주거지 확대로 인하여 담수호의 유역내에서 발생하는 점 및 비점오염(non-point pollution)이 증가하고 있으며, 특히 농업용 비료 보급이 원활해짐에 따라서 담수호로 유입되는 영양물질(nutrients)도 증가하였다. 하천과 달리 호소는 정체수역의 특성을 가지며 유입된 영양물질의 배출이 원활하지 않음으로 인해 수질악화의 가능성이 높음에도 불구하고 관심도 낮은 상태였다. 그러나 생활수준의 향상과 수자원부족에 당면하게 되면서 최근에는 담수호의 수질개선 및 수자원으로의 활용을 위한 다양한 시도와 연구가 이루어지고 있다. 담수호로 유입되는 영양물질이 증가하면 부영양 상태에서 녹조현상(water bloom)을 유발하고 부영양화(eutrophication)를 가속시키며, 수질악화가 반복되는 악순환이 일어난다.

우리나라에서는 1978년부터 하구담수호 수질현황조사를 시작하여 1990년에 본격적으로 주요 농업용 저수지를 수질측정망으로 지정하여 정기적으로 수질측정을 실시하고 있다. 또한 농업용수의 수질오염 방지와 개선을 위해서 1997년부터 농업용수 수질개선시험연구사업을 착수하여 농업용수 수질개선공법 개발 및 현장 적용성을 평가한 바 있다(Jang, 2010). 최근 한국농어촌공사는 '농업용 담수호 수질개선 사업 추진계획'을 마련하고 500만 톤 이상의 대규모 담수호 16곳에 대해 오는 2025년까지 수질 개선방안·실증연구를 추진하고 있다. 이에 따라 RRI (2010)는 호소내 수질특성을 분석하여 저수지 회귀모델을 개발하고 간단한 농업용저수지 수질모델을 개발한 바 있다.

Lim *et al*. (2015)는 1998년부터 2008년까지 낙동강 물금지점의 수질, 수리 및 기상 인자를 통해 조류(algae)발생의 주요 영향인자를 평가하고, 다중회귀분석을 이용하여 낙동강 하류의 Chlorophyll-a 농도와 복합 영향인자들의 상관관계 분석하였으며 Chlorophyll-a 예측 모델을 도출하였다. Lee *et al*. (2022)는 서해권역에 위치한 국내 15개 대규모 담수호를 대상으로 장기간 축적된 다양한 수질 인자를 활용하여 기본통계분석, 요인분석, 주성분분석 등 다변량 통계분석을 실시하고 부영양화로 인한 식물성 플랑크톤의 증식과 관련된 Chlorophyll-a의 예측을 위한 다중선형 회귀분석을 진행하여 장래의 Chlorophyll-a를 예측하였으며 효율적인 담수호 특성별 수질 개선 대책 수립을 위한 보조 자료로 활용하고자 하였다. Kong (2019)은 2013~2017년간의 환경부 물환경측정망 자료를 활용하여 국내 81개 호소의 수질 특성으로 수질항목간의 관계를 통계학적 방법으로 분석한 바 있다.

해양수산부에서 연안 오염상황을 상시 측정하기 위하여 해양수질자동측정망을 운영하고 자료를 제공하고 있지만 해양환경을 직관적으로 평가하기 위한 수질평가지수(water quality index, WQI)는 제공하지 않는 것에 대하여, Kim *et al*. (2023)은 해양환경측정망 자료를 학습한 인공지능(artificial intelligent, AI) 모델을 이용하여 분석하고 다양한 알고리즘을 적용하였다.

환경부에서 운영하고 있는 물환경정보시스템(Water Environment Information System, WEIS; http://water.nier.go.

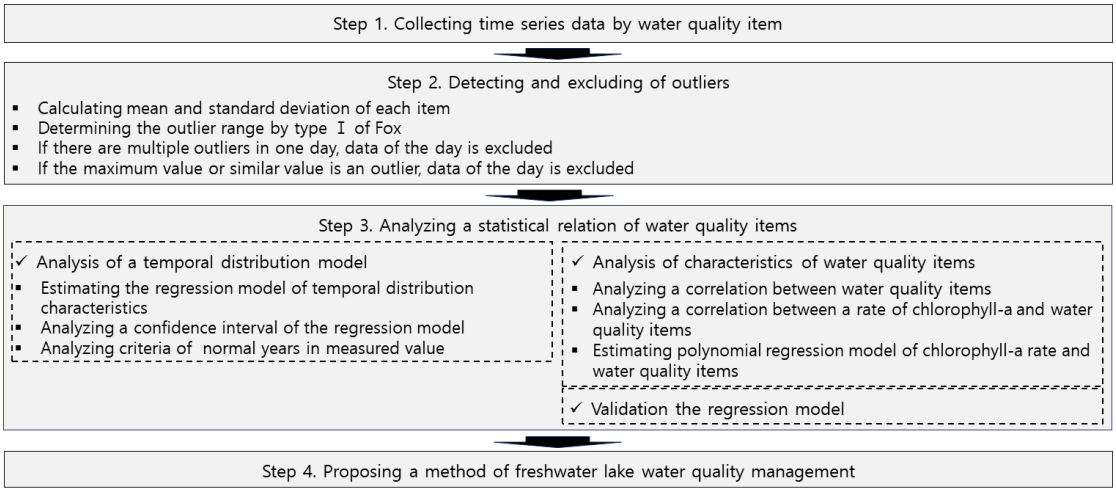


Fig. 1. Procedure of this study

kr)에서는 지리정보 기반으로 수질, 생물, 유량, 퇴적물 등의 종합정보를 제공하고 있으며, 간단한 자료조회를 통해 담수호에서 수행된 과거부터의 모든 측정된 수질자료를 수집할 수 있다. 본 연구에서는 담수호의 수질관리를 위하여 수질항목의 시계열적 변화 특성 및 수질항목간의 관계를 통계학적으로 분석하고, 담수호 수질 관리기준 및 관리방안 설정에 이용이 가능한 자료를 생성하고자 하였다. 또 담수호의 부영양화의 발생여부 및 진행정도를 파악하는 방법으로서 영양순환의 1차생산자인 식물플랑크톤의 농도를 이용할 수 있는데, Chlorophyll-a는 모든 조류에 보편적으로 존재하고 다른 방법에 비해 분석시간 적다는(Shim and Shin, 1989) 장점이 있다. 기존 연구에서는 수질인자와 Chlorophyll-a의 상관관계 분석에만 집중하였지만, 금회 연구에서는 수질인자의 시간적 분포 특성 및 시간 규모를 고려한 Chlorophyll-a의 예측을 시도하는 특징이 있으며, 따라서 측정자료를 선별하고 통계학적 분석 및 회귀모형을 통해 적용이 용이한 방법론을 제시하였고, 연구과정을 Fig. 1에 나타내었다.

2. 연구대상 및 연구방법

2.1 연구대상지점

대형 방조제에 의하여 축조된 대규모 담수호는 대체로 한국농어촌공사 관할에 해당되고, 일반현황은 RAWRIS을 통해 조회가 가능하며, 수질측정 자료는 환경부의 WEIS를 통해서 수집할 수 있다. 본 연구에서는 대규모 담수호 중 경기도 및 충남도에 위치하여 위도가 36.5°~37°로 유사한 3개소를 대상으로 하였다. 선정된 남양(NY), 대호(DH), 간월(GW) 담수호는 유역면적이 비슷하고 서해에 위치하여 지형 및 기상조건도 유사할 것으로 판단하였다. 이 담수호들은 WEIS에서 각 담수호별 3개 지점의 수질측정자료를 제공하고 있으므로 총 9개 지점을 연구대상지점으로 선정하였으며 Fig. 2 및 Table 1에서는 담수호의 위치와 특성을 나타내었다.

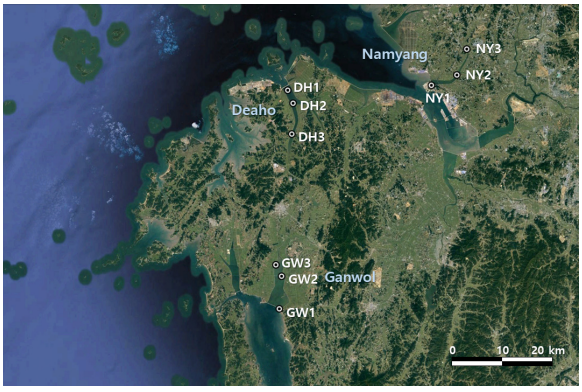


Fig. 2. Location of target study site

2.2 수질 항목

WEIS에서 호소의 수질측정결과는 월단위로 제공되나 매년 월별 일자는 일정하지 않으며, 동절기엔 결측이 많고, 하절기에도 불규칙하게 결측이 발생하고 있다. 한편 Chlorophyll-a의 변화를 예측하기 위해서, 본 연구에서는 Table 2에서 나타낸 바와 같이 제시된 수질항목 중 조류 발생의 지표로 활용되기 용이한 수질항목으로 연구대상을 한정하였다.

2.3 통계적 분석 방법

2.3.1 이상치

이상치는 전체적인 경향을 벗어나는 자료로 받아들여지는데, Fox (1972)는 관측오류로 인해 한 개의 관측치에 영향을 주는 것을 Type I, 관측기기의 세팅이나 기기 자체의 오류 가능성이 크며 상태변화가 극적인 것으로서 관측치 전반에 영향을 주는 것을 Type II로 정의하였다. 본 연구에서는 단순 오류에 의한 이상치로서 type I 이상치를 제외하였다.

2.3.2 상관 관계 분석

모집단에서 크기가 n 인 샘플을 채취한 경우, n 개의 자료점$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$이 얻어지는데, 이 두 변수 사이의 Pearson상관계수는 두 변수의 선형 상관관계를 계량화한 수치로서 양의 상관관계가 클수록 1에 근접하고, 음의 상관관계가 클수록 -1에 근접하며, 0에 근접하면 상관관계가 없다. 다만 p-value가 유의수준 이하인 경우 상관계수는 '0'이라는

Table 1. Location and size of freshwater lake

Name	Basin Area (㎢)	Water Surface Area (㎢)	Effective Storage Volume (10^3㎥)	Total Storage Volume (10^3㎥)	Average Depth (m)	Duration (year)	Amount (EA)
Namyang	208.35	8.0	20.407	31.489	4.26	2002~2022	725
Daeho	312.15	40.0	464.000	1,220.0	6.82	2002~2022	1751
Ganwol	488.00	28.8	49.200	1,159.0	4.64	2002~2022	762

Annotation: Water surface area was estimated at google earth. Average depth was calculated with total storage volume and water surface area

Table 2. Water quality items provided by the WEIS

Study Target Items	Non-target items
Temp (℃), COD (㎎/L), Chl-a (㎎/㎥), TN (㎎/L), TP (㎎/L), DTN (㎎/L), NH3-N (㎎/L), NO3-N (㎎/L), DTP (㎎/L), PO4-P (㎎/L)	DO (㎎/L), BOD (㎎/L), TOC (㎎/L), hydrogen ion, Phenolic (㎎/L), Electrical conductivity (μS/㎝), Coliform group (/100ml), Cadmium (㎎/L), CN (㎎/L), Lead (㎎/L), Chromium (㎎/L), Arsenic (㎎/L), Mercury (㎎/L), SS (㎎/L), Fecal Coliforms, Anionic Surfactant (㎎/L), Antimony (㎎/L)

Abbreviations: Temp:Temperature, DO:Dissolved Oxygen, BOD:Biochemical Oxygen Demand, COD:Chemical oxygen demand), Chl-a:Chlorophyll-a, TOC:Total Organic Carbon, TN:Total Nitrogen, TP:Total Phosphors, DTN:Dissolved TN, NH3-N:Ammonia Nitrogen, NO3-N:Nitrate Nitrogen DTP:Dissolved TP, PO4-P:phosphate

귀무가설을 기각하므로 유의하다. 샘플 x_i 의 평균이 $\bar{x}$, y_i 의 평균이 $\bar{y}$ 인 경우 Pearson상관계수 r_{xy} 를 구하는 공식은 Eq. (1)과 같다. 한편, 상관계수의 검증은 $T-test$ 을 수행하는 바, t_o 가 $t(n-2, \alpha/2)$ 보다 크면 상관계수는 '0'이라는 귀무가설을 기각하며 상관계수는 유의하다.

$$r_{xy} = \frac{\sum_i^n (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_i^n (x_i-\bar{x})^2}\sqrt{\sum_i^n (y_i-\bar{y})^2}} \quad (1)$$

2.3.3 다중 및 다항 회귀 분석

회귀식은 변수간의 함수관계를 가짐을 나타내는 것이고, 다중회귀분석에서 1개의 종속변수에 대하여 2개 이상의 독립변수 사이에 함수관계가 성립하면, n 개의 자료 점, i 개의 독립변수를 가정 시($x_{1,1}$, ⋯, $x_{i,1}$, y_1), ($x_{1,2}$, ⋯, $x_{i,2}$ y_2), ⋯, ($x_{1,n}$, ⋯, $x_{i,n}$ y_n)에 대하여 오차를 포함하는 회귀식이 Eq. (2)와 같이 성립된다. 반면 종속변수와 독립변수 간에 선형관계가 아닌 곡선관계가 성립하면 x 의 제곱이나 세제곱이 들어가는 곡선회귀가 적절하고 이런 경우 2차 모형 또는 3차 모형 등으로 표현할 수 있으며 독립변수가 k 개인 경우 2차 모형은 Eq. (3)과 같다.

$$\hat{y_j} = b_0 + b_1 x_{1,j} + b_2 x_{2,j} + \cdots + b_i x_{i,j}, \quad j = 1, 2, \cdots, n \quad (2)$$

$$\hat{y_j} = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{i=\le j}^{k} b_{i,j} x_i x_j \quad (3)$$

여기서 $\hat{y_j}$ 은 종속변수, b_i 는 추정된 회귀계수, $x_{i,j}$ 는 독립변수이다. 다중회귀모형 및 다항회위모형의 검증은 $F-test$ 을 수행하는 바, F_o 가 $F(k, n-k-1, \alpha)$ 보다 작으면 회귀계수는 '0'이라는 귀무가설이 채택되어 유의하지 않다.

3. 연구결과

3.1 이상치 검정

이상치의 확정을 위해서는 표본자료를 그래프로 표출하는 과정이 필수적이며 정규분포를 따르지 않는 경우는 자료를 변형하는 것이 도움이 된다(Robinson *et al*., 2005). 단변량 자료의 이상치를 검정하는 방법으로는 정규분포 가정에 근거한 z-score Test가 유용하다고 볼 수 있는데 z-score Test는 자료 내에 존재하는 잠재적 이상치의 분포를 개략적으로 평가하기 위해 적용되는 통계적 분석방법이다(Robinson *et al*., 2005; Park *et al*., 2015). Park *et al*. (2015)은 z-score Test에서 잠재적 이상치 분류기준을 Eq. (4)로 제시한 바, Eq. (1)에서 표현한 바와 같이 자료에 대해 평균값을 기준으로 3σ를 이상치를 분류하나 정규분포를 하지 않는 경우는 그 적용이 곤란하다.

$$z_i = \frac{x_i-\mu}{\sigma} \begin{cases} \text{if } z_i > 3 \text{ or } z_i < -3 & : potential\ \ outlier \\ \text{if } -3 \le z \le 3 & : normal\ \ data \end{cases} \quad (4)$$

본 연구에서는 첫 번째, 상기 연구대상 수질항목 10개에 대하여 각 항목별로 평균, 표준편차를 산정하고 3σ를 기준으로 이상치 후보로 선정하였다. 총 10개 항목에서 선정된 이상치 후보는 측정일자별로 동시에 검토하여 2개 이상의 이상치 후보가 있는 경우 해당 일자의 모든 관측치를 제외하였다. 두 번째, 각 수질항목별 최대치가 이상치 이면 해당 일자에 모든 관측치를 제외하였다. 이러한 과정은 수질자료의 시계열분석을 위해서 동시자료를 사용하고자 하는 목적으로 수행되었는바, 과다한 자료의 제외를 방지하고자 하는 것이다. 마지막으로 모든 자료를 수질항목별로 시계열 그래프로 도시한 후 해당 기간에서 극단적인 값으로 보이는 지점을 추가로 제외하였다.

3.2 수질항목의 시계열 특성 및 수질항목간의 상관성 분석

3.2.1 주요 수질항목의 시계열 특성

담수호는 수계의 말단부로서 상류하천을 포함하여 유역에서 발생하는 유출물질의 집합지로 볼 수 있다. 따라서 담수호

수질관리를 위해서는 연중 수질상태상황을 파악할 필요가 있다. Fig. 3에서는 수질항목별 시계열 변화를 나타내고 있는데, 수온시계열자료는 1년 단위로 주기성을 비교적 명확히 나타내고 있고, Chl-a, TN, TP 등의 수질항목들도 대체로 1년 단위의 주기성을 보여 주고 있으나 항목별 자료의 연간 진폭의 크기는 변화하고 연도별 평균도 다르게 변화한다. Chl-a (mg/㎥)는 평균이 18.608, 표준편차가 19.608이고, TN (mg/L)는 각각 2.910 1.235 이며, TP (mg/L)는 각각 0.060 및 0.046로서 연중 변화폭이 작지 않음을 알 수 있다.

자료의 시계열적 특성을 통계학적 방법으로 분석하고자 수온분포에 대해 회귀식을 검토하였다. 매년 일정한 패턴으로 반복되는 형태를 주기함수로 나타날 수 있는 회귀식은 최대 5차 다항식과 삼각함수형태로 가정한 바, Eq. (5) 및 Eq. (6)과 같다.

$$\hat{y_i} = b_0 + b\cos\left(2\pi\frac{x_i}{S} - \theta\right) \tag{5}$$

$$\hat{y_i} = b_0 + \sum_{k=1}^{j} b_k x_i^k,\ j = \text{from 2 to 5} \tag{6}$$

여기서 삼각함수에서 S는 자료의 주기, b_0, b, θ는 회귀계수이다. Table 3에서 가정된 회귀식의 F_o가 574 이상의 높은 값으로서 유의수준 0.05의 기각치를 초과하므로 모두 유의한 것으로 판정되며, R^2는 0.84~095을 보이나 차수가 4차 이상에서 R^2는 0.95를 나타낸다. 한편, Fig. 4에서 관측치 분포 중 회귀식의 주기적 변화를 보여주고 있는데 시점과 종점이 일치하려면 삼각함수형태의 주기함수가 적절하다고 판단하였다.

조류성장에 필요한 영양물질은 주로 C, N, P 등 이지만 N/P의 무게비로 P 또는 N 제한을 판단할 때 Guildford and Hecky

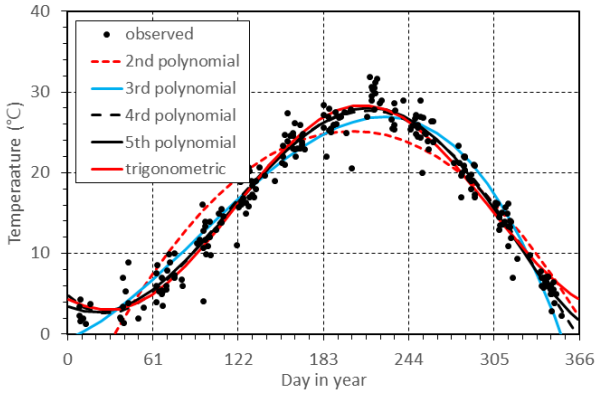
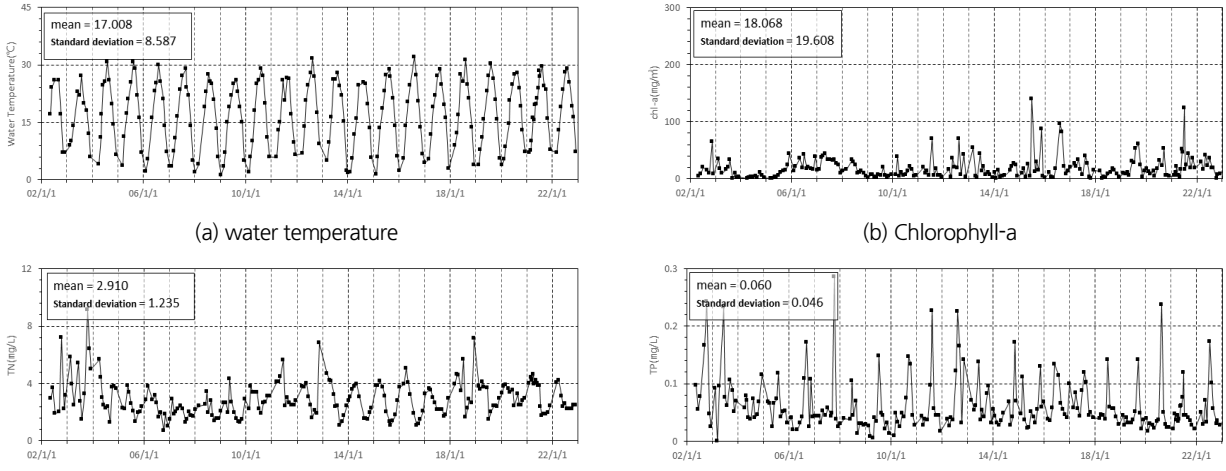


Fig. 4. Applicability of periodic regression equation for temperature distribution



(a) water temperature (b) Chlorophyll-a

(c) TN (d) TP

Fig. 3. Time series data of water quality items in NY1

Table 3. Applicability of periodic regression equation for temperature distribution

Classification	2nd polynomial	3rd polynomial	4th polynomial	5th polynomial	trigonometric
R^2	0.84	0.93	0.95	0.95	0.95
F_o	574.26	966.51	1,101.96	903.68	2,013.29
$F(\alpha, \phi_1, \phi_2)$	2.65	2.41	2.26	2.14	2.65

(2000)는 23과 9, Dzialowski *et al*. (2005)은 29와 8로 제시한 바 있다(Kong, 2019). 현재 남양호를 포함한 우리나라의 많은 담수호는 영양물질이 충분히 공급된 상태이고, Table 4에서 Janus and Vollenweider (1981)의 기준으로 담수호의 영향상태를 판단하면 TN는 부영양 상태, TP는 중영양 또는 중부영양 상태이다. 일반적으로 담수호 유역에서는 농업용지 면적이 크고 비료나 가축사료 등의 이용에 의한 비점오염원이 큰 상태이므로 조류생장이 유리한 상황이고 N이 P에 대하여 약 10~20배 정도로 N이 풍부한 상태로서 지점별 시기별 수온, 일사량 등의 요소에 따라 Chl-a 및 TP의 변화가 큰 것으로 추정된다. 따라서 우리나라 담수호의 영양상태를 보면 대체로 P가 제한인자로 판단할 수 있다.

Carlson (1977)은 투명도 (Secchi depth, SD)를 기반으로 TSI (trophic state index)를 고안한 후 투명도에 대한 관계로부터 Chl-a와 TP의 TSI를 도출하였다. 한편 우리나라에서는 환경부고시 제2013-134호(2013.11.5.)의 물환경 목표기준 평가규정 별표에 TSIKO가 포함되어 부영영화 정도를 평가하기 위한 공식적인 지수로 지정되었으며, TSIKO에 적용되는 수질은 호수 전 층의 연평균 COD, Chl-a 및 TP 농도이다(Kim and Kong, 2019). 따라서 부영양화의 주요 지표로서 COD, Chl-a 및 TP에 TN을 추가하여 연중변화를 통계적으로 분석하였다.

COD, Chl-a 및 TP는 하절기 증가, 동절기 감소의 유사한 경향을 나타내나 TN은 각 지점마다 상이한 경향을 나타내며, Fig. 5에서 NY1의 TN은 하절기인 9월 작은 값을 나타낸다. N와 P은 주로 농업용 비료를 통해 공급되는데, 비료살포 시기는 식물이 생장하기 전인 봄 또는 가을로 예상된다. 한편, 암노

Table 4. Classification of trophic state in the OECD eutrophication programme

Variables	Oligotorophic	Mesotrophic	Eutrophic	Hyper-Eutrophic
TN (mg/l)	0.31 - 1.63	0.36 - 1.39	0.39 - 6.1	-
TP (μg/l)	3.0 - 17.7	10.9 - 95.6	16.2 - 386	750-1200
Chl-a (μg/l)	0.3 - 4.5	3 - 11	2.7 - 78	100-150

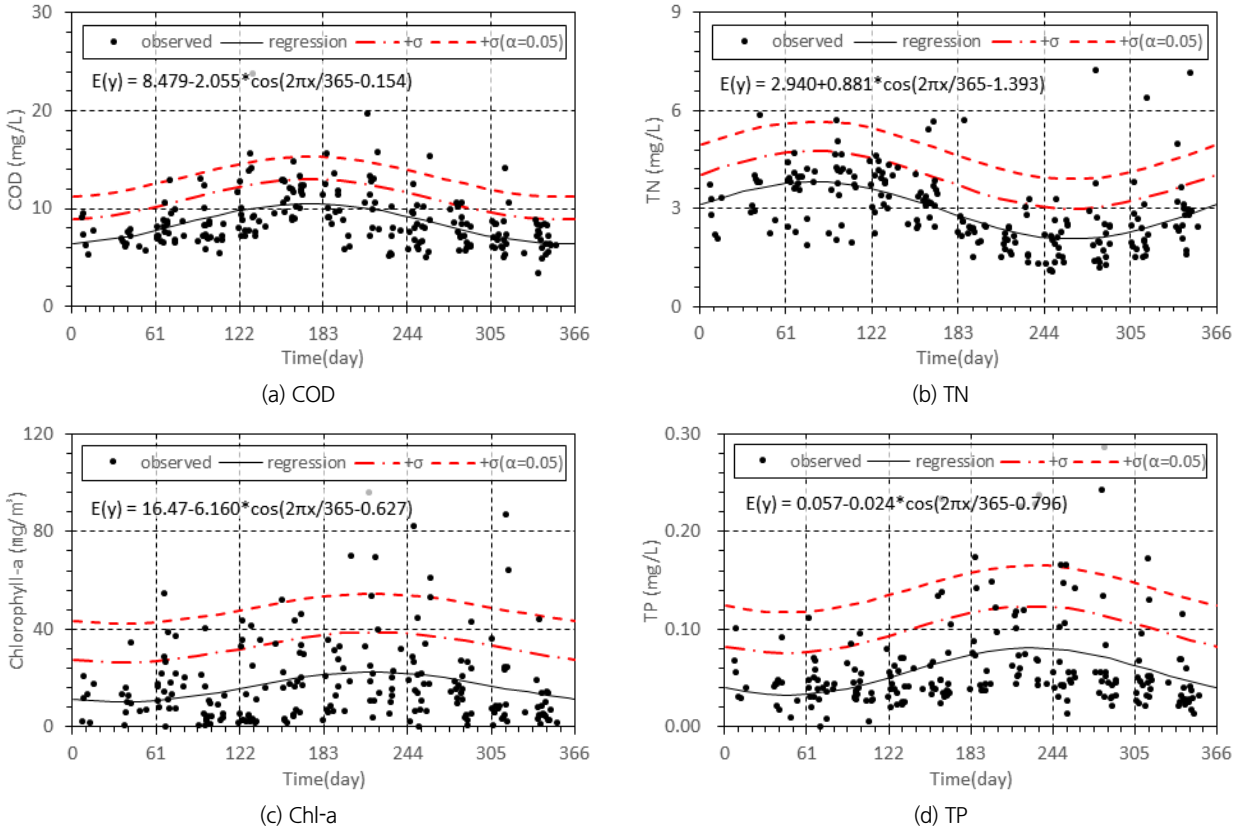


Fig. 5. Annual change trends of water quality items at major points in NY1

니아성 질소비료는 관개수 및 강우에 용해되어 유출이 용이한 반면 P은 물에 잘 녹지는 않으나 토립자에 강하게 흡착되어 인의 배출량은 작은 편이다(Kim and Chung, 2002). 따라서 P은 주로 축산농가나 하수종말처리장을 통해 배출될 것으로 추정되는 바, N와 P의 변화곡선 차이는 기상 및 물순환 과정에 따른 결과로 추정된다.

Table 5에서 회귀모형의 R^2 값은 대체로 NY, GW, DH 순으로 높았고 COD, TN이 Chl-a, TP보다 상대적으로 높게 나타났다. R^2 이 0~0.49로 대체로 낮았으나 $F-test$ 에서 유의수준(α) 0.05에서 기각치와 비교한 바, NY3, GW3의 TP에서만 유의하지 않고 다른 지점 및 항목에서 대체로 유의하게 나타났다. 즉 수질항목의 값들은 대체로 주기적인 경향을 나타내는 것으로 볼 수 있다.

Fig. 5에서 회귀식의 기대치를 기준으로 표준편차 선과 최상부에 위치한 유의수준 0.05에서의 신뢰구간을 제시한 바, 수질항목의 관측 값이 최상위의 선을 초과하는 것은 신뢰도 95%에서 표준편차를 벗어나는 값을 의미한다. 따라서 수질항목변화의 주기함수 유도는 관리자가 주의 및 조치가 필요한 시점을 정할 때 수질의 연중변화를 고려할 수 있는 방법으로 사용할 수 있다.

3.2.2 Chlorohpyll-a와 수질항목간의 상관관계

조류생장에 따른 식물성 조류의 양을 계측하는 방법은 광합성 색소 즉, Chl-a의 양을 측정하는 것이며, Chl-a의 양은 부영양 상태를 판정하는 중요한 기준이다. 녹조생장과 관련된 환경요소는 수온과 일조량 등이 있으며 유기물 생성에 필요한 영양물질 중 물속에서 흡수되는 것은 N과 P가 대표적이고 C는 물속에 포화된 이산화탄소이다. 수질항목 중 TOC는 조류생장의 결과로 볼 수 있고, 유기체의 증가는 BOD 또는 COD의 증가로 이어지는 것이므로 TOC, BOD, COD는 녹조생장의 활성화 예측이 아닌 결과 항목으로 판단하였다. Table 6에서는 C1~C9에서 Chl-a과 수질항목간의 상관관계를 나타내고 있는데, Chl-a와 수온은 양의관계, N계와는 음의 관계를, TP는 양의 관계를 보이지만, TN 또는 P계는 명확하지 않은 것으로 보인다. 한편 N1~N3에서 TN은 DTN, NO3-N과 양의 상관관계를 나타내고, P1, P2에서 TP도 높은 상관관계를 나타내고 있다. 유의수준 0.05에서 t 기각치 1.96~1.97과 비교한 바, Chl-a와 TN, DTP 및 PO4-P의 상관계수는 일부가 유의하지 않은 것으로 나타났으나 TN와 DTN, NO3-N, NH3-N의 상관관계 및 TP와 DTP 및 PO4-P의 상관관계는 모두 유의하였다. 따라서 TN과 다른 N계 및 TP와 다른 P계 수질항목은

Table 5. R^2 of Periodic Regression Model for Water Quality Items

Point	Point #1				Point #2				Point #3			
	COD	Chl-a	TN	TP	COD	Chl-a	TN	TP	COD	Chl-a	TN	TP
MY	0.25	0.06	0.32	0.13	0.19	0.03	0.49	0.04	0.27	0.09	0.26	0.00
DH	0.07	0.02	0.08	0.14	0.06	0.05	0.09	0.20	0.04	0.03	0.07	0.18
GW	0.08	0.06	0.05	0.12	0.11	0.14	0.49	0.13	0.21	0.09	0.18	0.01

Table 6. Comparision of Correlation Coefficient of between Chlorophyll-a and Items

Point	Magintude of Correlation Coefficients													
	C1	C2	C3	C4	C5	C6	C7	C8	C9	N1	N2	N3	P1	P2
NY1	0.26	0.44	-0.23	0.24	-0.23	-0.21	-0.19	0.16	0.05*	0.90	0.60	0.72	0.79	0.48
NY2	0.20	0.56	-0.15	0.13*	-0.12*	-0.06*	-0.17	-0.04*	-0.08*	0.92	0.75	0.68	0.79	0.60
NY3	0.33	0.45	-0.10*	0.13	-0.12*	-0.04*	-0.23	-0.14	-0.18	0.98	0.90	0.39	0.82	0.67
DH1	0.17	0.53	-0.09	0.21	-0.26	-0.25	-0.23	-0.06	-0.13	0.78	0.37	0.71	0.76	0.72
DH2	0.21	0.61	-0.12	0.31	-0.33	-0.30	-0.32	0.00*	-0.12	0.85	0.29	0.67	0.62	0.54
DH3	0.16	0.66	0.00*	0.44	-0.28	-0.30	-0.30	0.10	-0.06*	0.84	0.26	0.66	0.66	0.54
GW1	0.19	0.66	-0.01*	0.38	-0.29	-0.25	-0.22	-0.11*	-0.16	0.70	0.33	0.64	0.30	0.16
GW2	0.23	0.69	0.06*	0.40	-0.13	-0.14	-0.22	0.00*	-0.10	0.88	0.49	0.67	0.54	0.28
GW3	0.30	0.54	-0.26	0.04*	-0.40	-0.31	-0.41	-0.28	-0.29	0.92	0.71	0.63	0.83	0.81

Abbreviations: C1: Chl-a-Temp., C2: Chl-a-COD, C3: Chl-a-TN, C4: Chl-a-TP, C5: Chl-a-DTN, C6: Chl-a-NH3-N, C7: Chl-a-NO3-N, C8: Chl-a-DTP, C9: Chl-a-PO4-P, N1: TN-DTN, N2: TN-NH3-N, N3: TN-NO3-N, P1: TP-DTP, P2: TP-PO4-P

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

중복할 필요가 없는 것으로 판단되며, 각 수질항목의 특징을 고려하려면 N계 및 P계의 수질항목을 직접 사용하는 것이 적절한 것으로 판단된다.

3.2.3 Chlorohpyll-a의 변화율과 수질항목간 상관관계

부영양화 지표인 Chl-a의 증가를 예측하기 위해서는 시간에 따른 변화량을 추정해야 하므로 Chl-a의 변화율과 다른 수질항목 및 시간변화율의 관계를 분석하였다. 우선, WEIS 전국수질측정망에서 제공하는 자료는 대체로 월간자료에 해당되지만 측정일자는 불규칙하고, 결측도 있으므로 관측일자의 시간간격을 고려해야 한다. 자료의 간격을 45일 이내로 제한하고 가용한 자료를 선정한 후, 관측치의 차이를 시간간격으로 나누어 항목별 시간변화율을 산정하였다.

Table 7에서 Chl-a의 변화율과 수온, DTN, NH3-N, NO3-N, DTP 및 PO4-P은 대체로 양의 상관관계를 나타내고, Chl-a의 변화율과 Chl-a, DTN, NH3-N, NO3-N, DTP 및 PO4-P의 변화율과는 대체로 음의 상관관계를 나타낸다. Chl-a 변화율과 수질항목간의 상관관계가 Chl-a과 다른 수질항목간의 상관관계보다 낮게 나타났는데, 자료변환 과정에서의 자료 수 감소가 원인의 하나로 추정되며, 대체로 수온변화율, Chl-a, TP, NO3-N과 PO4-P의 변화율에서 상관관계가 비교적 높게 나타났다. 유의성 검정에서는 유의수준 0.05를 기준으로 Chl-a 변화율과 수온의 관계는 유의하지 않고 수온변화율 및 Chl-a는 대체로 유의한 것으로 나타났다. 그 이외의 경우, Chl-a 변화율은 수질항목 보다 수질항목 변화율과의 상관관계가 더 높은 유의성을 보여주지만 그 차이가 크지 않았다.

3.3 Chlorohpyll-a 예측

3.3.1 회귀모형분석

회귀모형의 변수 선택은 각 인자의 포함여부에 따른 결정계수의 변화를 보면서 시행착오과정을 반복하는 것이 이상적이며 추가 또는 제외 방식에 따라서 전진선택법(Forward selection) 또는 후진선택법(Backward elimination) 등이 있다. 다만 선택과정에서 많은 시간과 계산부하량이 소요되므로 본 연구에서는 선행 분석된 3.2.2 및 3.2.3의 상관관계 분석결과를 이용하여 높은 상관성을 보이는 수온변화율, Chl-a와 함께 모형의 단순화를 위해 TN, TP를 제외한 세부항목인 DTN, NH3-N, NO3-N, DTP, PO4-P를 고려하여 총 7개를 대상으로 회귀모형을 구성하였다. 여기서 Type 1 다중회귀모형, Type 2는 각 수질항목의 곱으로 구성하였지만 로그변환을 통해 다중회귀모형으로 변형하였다. Type 3는 다항회귀모형으로 2차 모형으로 결정하였지만 다항회귀모형을 구성할 때 모든 항을 포함시키는 것이 바람직한 것은 아니다. 이중에서 필요한 항만 선택하여 사용하는 것이 더 좋은 경우도 있다(Park, 2007). 본 연구에서 Type 3는 Type 1과 함께 상관성이 높은 수온변화율 및 Chl-a을 각 항에 곱한 항만 추가하여 다항회귀모형을 구성하였으며 Type 1, 2, 3 회귀모형은 Eqs. (7)~(9)와 같다.

$$\hat{y_i} = b_0 + \sum_{j=1}^{j=7} b_j x_{j,i} \tag{7}$$

$$\log \hat{y_i} = b_0 + \sum_{j=1}^{j=7} b_j \log x_{j,i} \tag{8}$$

Table 7. Comparision of Correlation Coefficient of between Chlorophyll-a change rate and Items

POINT	Magintude of Correlation Coefficients																		
	A1	R_A1	A2	R_A2	A3	A4	R_A4	A5	R_A5	A6	R_A6	A7	R_A7	A8	R_A8	A9	R_A9	A10	R_A10
NY1	0.04*	0.08*	-0.16	0.45	-0.59	0.12*	-0.15	-0.06*	0.13*	0.10*	-0.17	0.04*	-0.08*	0.06*	-0.08*	-0.06*	0.10*	-0.03*	-0.05*
NY2	-0.02*	0.07*	-0.20	0.49	-0.51	0.10*	-0.14	0.07*	-0.08*	0.11*	-0.18	0.08*	-0.15	0.09*	-0.03*	0.16	-0.23	0.15	-0.17
NY3	-0.06*	0.25	-0.17	0.48	-0.61	0.05*	-0.06*	0.05*	0.04*	0.06*	-0.10*	0.08*	-0.12*	-0.01*	0.07*	0.18	-0.32	0.17	-0.34
DH1	-0.03*	0.57	-0.06*	0.48	-0.31	0.09*	-0.07*	0.01*	0.31	0.11*	0.07*	0.05*	0.06*	0.10*	-0.46	0.08*	0.04*	0.03*	0.16
DH2	0.00*	0.42	0.02*	0.18	-0.20	0.14	-0.24	0.07*	-0.03*	0.13	-0.04*	0.05*	-0.11*	0.08*	-0.40	0.12*	-0.11*	0.05*	0.19
DH3	-0.01*	0.45	0.01*	-0.13*	-0.22	0.09*	-0.01*	0.05*	0.15	0.14	-0.35	0.15	-0.16	0.11*	-0.39	0.07*	-0.03*	0.06*	0.00*
GW1	0.02*	0.23	-0.40	0.71	-0.61	-0.04*	0.15	-0.20	0.29	0.13*	-0.16	0.12*	-0.17	0.09*	-0.10*	0.10*	-0.12*	0.11*	-0.19
GW2	0.06*	0.15	-0.24	0.51	-0.52	-0.09*	0.25	-0.22	0.40	0.01*	0.05*	0.05*	-0.04*	0.02*	-0.02*	0.03*	-0.05*	0.09*	-0.12*
GW3	-0.12*	0.17	-0.23	0.36	-0.63	0.09	0.00*	-0.04*	0.13*	0.16*	-0.19	0.13	-0.12*	0.17*	-0.23	0.13*	-0.21	0.12*	-0.22

Abbreviations: A1: Temp., A2: COD, A3: Chl-a, A4: TN, A5: TP, A6: DTN, A7: NH3-N, A8: NO3-N, A9: DTP, A10: PO4-P, R_: Temporal variation rate of water quality items, (value_before - value_after) / duration

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

Table 8. Statistics of regression analysis results

Point	R^2			F_o		
	Type 1	Type 2	Type 3	Type 1	Type 2	Type 3
NY1	0.37	0.18	0.42	25.00	25.00	9.11
NY2	0.29	0.18	0.33	26.86	26.86	9.83
NY3	0.44	0.28	0.55	29.29	29.29	10.78
DH1	0.42	0.50	0.55	28.14	28.14	10.33
DH2	0.24	0.24	0.38	25.71	25.71	9.39
DH3	0.26	0.44	0.43	25.71	25.71	9.39
GW1	0.39	0.19	0.43	27.71	27.71	10.17
GW2	0.30	0.24	0.38	27.71	27.71	10.17
GW3	0.46	0.28	0.49	27.43	27.43	10.06

$$\hat{y_i} = b_0 + \sum_{j=1}^{j=7} b_j x_{j,i} + \sum_{j=2}^{j=7} b_{6+j} x_{j,i} x_{1,i} + \sum_{j=3}^{j=7} b_{11+j} x_{j,i} x_{2,i} \quad (9)$$

여기서 x_1는 온도변화율, x_2는 Chl-a, x_3는 DTN, x_4는 NH3-N, x_5는 NO3-N, x_6는 DTP, x_7는 PO4-P이다.

Chl-a의 변화율을 추정하는 회귀모형에 3개 담수호에 2002년~2022년간 관측자료를 적용한 결과, Table 8에서 나타낸 바와 같이 대체로 Type 별 R^2는 Type 3, Type 1, Type 2 순이고 상·하류지점보다 중류지점이 낮으며, 담수호별로는 대호(DH)에서 가장 높았다. 다만 회귀모형의 결정계수가 높지 않은 것은 종속변수의 변화를 나타내기 위한 독립변수의 부족과 자료 자체의 부족으로 추정할 수 있으며 자료의 한계를 극복하기 위해서는 자료축적의 기간이 필수적이라고 판단된다. 다만 유의수준 0.05에서 $F-test$ 의 기각치 1.65~2.06에 대해 회귀모형이 모두 유의한 것으로 나타났다.

3.3.2 검증

Chl-a의 변화율을 추정하는 회귀모형의 적합도 분석에서 NY 및 GW 지점에서는 Type 1, Type 3에 비해 Type 2의 R^2가 비교적 큰 차이를 나타내므로, Type 1, Type 3 회귀모형에 대해서만 Validation을 수행하였다. 본 연구에서 사용한 Validation 방법은 교차검증(Cross-validation) 방법인데, 이 방법은 분류와 회귀분석에서 알고리즘의 일반화능력을 평가하는데 가장 폭 넓게 사용하는 방법 중 하나다(Bergmeir *et al*., 2015; Hastie *et al*., 2009; Moreno-Torres *et al*., 2012).

교차검증방법으로는 leave-p-out cross-validation, leave-one-out cross-validation, k-fold cross-validation 등 다양한 방법이 있으나, 지점별 수심평균 수질항목자료는 200개 내외로서 자료의 양이 충분하지 않으므로 전체 자료 중 비교적 최신 연도 5년의 자료를 추출하여 leave-one-out cross-validation을 수행하였다. 이 경우 전체 자료의 평가치는 각 연도별 평가치의 평균을 사용한다. 다만 본 연구에서는 사용하는 각 연도별 자료의 개수가 크지 않으므로 5개년 자료를 합하여 전체 평가치를 산정하였다.

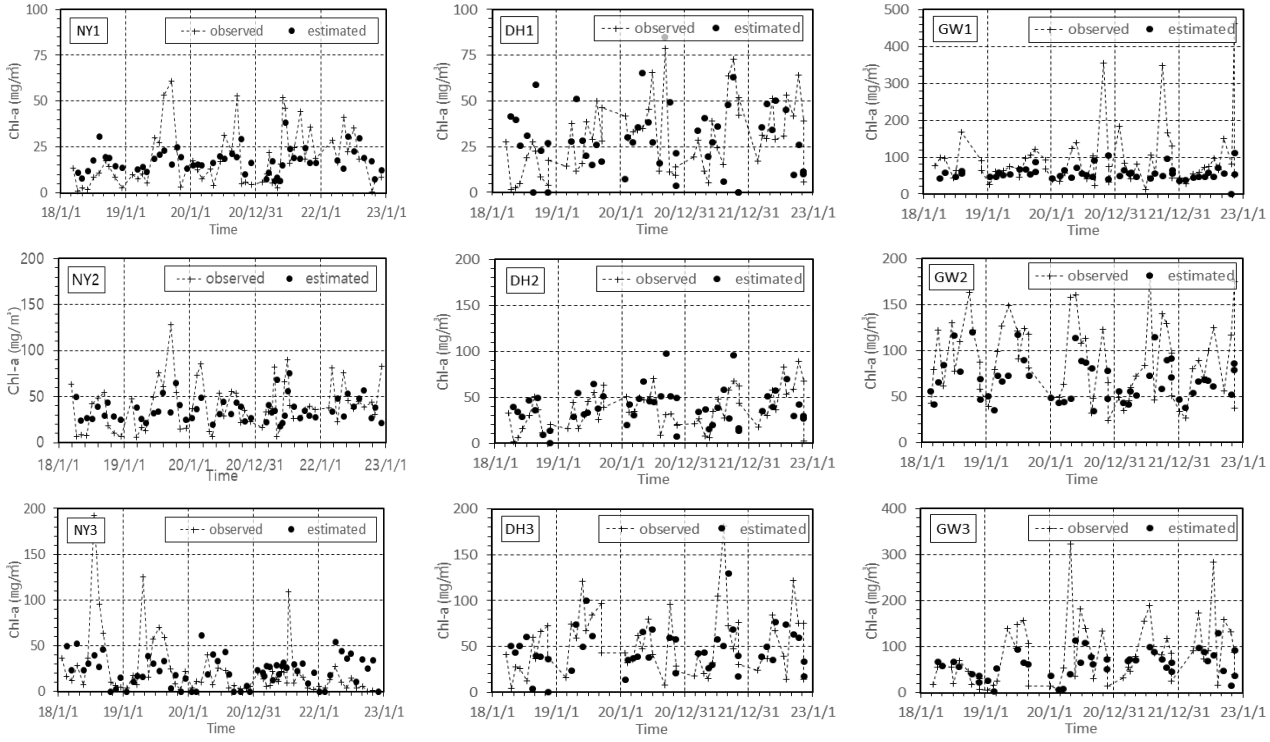
Table 9. R^2 of regression model in types and points

Classification	R^2								
	NY1	NY2	NY3	DH1	DH2	DH3	GW1	GW2	GW3
Type 1	0.19	0.26	0.25	0.91	0.96	0.53	0.27	0.49	0.25
Type 3	0.24	0.20	0.36	-	-	-	0.26	0.63	-

지점별 연도별 유도된 Type 1 회귀모형에 관측 일자의 수질 관측항목 값을 적용하여 산정한 Chl-a의 변화율로부터 다음 관측일자와의 기간을 적용하여 예측치를 선정하고 관측치와 비교하여 Fig. 5에 나타내었으며, 계산식은 Eq. (10)과 같고 $rate\ \ of\ \ Chl-a$는 Eq. (7)이다.

$$Chl-a_{i,est.} = Chl-a_{i-1,obs} + rate\ \ of\ \ Chl-a \times dt \quad (10)$$

Table 9에서 R^2은 Type 1에서 DH가 NY 및 GW는 보다는 높았지만 Type 3에서는 반대로 결과를 보인다. 회귀모형에서 더 많은 변수를 포함하는 Type 3가 Type 1에 비하여 유효성이 떨어지는 이유로는 사용되는 변수 개수 증가에도 사용 자료양이 같음으로 인한 상대적인 자료부족으로 추정된다. 한편, DH는 NY 및 GW보다 담수호 규모가 커서 변화율이 완만한 것과, 자료량도 2배 이상으로 유리한 점이 적합도 증가에 영향을 주는 것으로 판단된다. 유효성이 높은 회귀모형의 변수와 형태를 결정하기 위해서 다양한 변수와 형태 가정을 반복 수행하는 과정이 필요할 것으로 보이나 금회 연구에서는 방법론의 적용성을 검토하기 위하여 제한적인 범위로 적용하였다. Fig. 6에서는 3개 담수호 지점에서 Chl-a의 예측과 관측을 비교한 것인데, 대체로 예측치가 급격히 변화는 변화하는 관측치를 따라가지 못하지만 전체적으로 변화 경향을 따라가는 것으로 나타났고, 예측치 변화폭이 관측치 변화폭 보다 작은 것으로 나타났다. 따라서 담수호 또는 지점에 모두 적용이 가능한 단일의 회귀모형의 구성이 어렵다고 판단되는 바, 이러한 특징은 수질변화가 지역의 기상, 물리 및 생물학적 환경 특성의 영향이 크게 작용하기 때문으로 판단된다. 다만, 이 방법은 특정기간 후의 Chl-a를 근사치로 예측하여 관리에 필요한 시간을 확보하는데 도움이 될 것으로 판단된다.



(a) NY (b) DH (c) GW

Fig. 6. Comparison of observed and estimated values of Type 1

4. 결 론

본 연구에서는 수질통계학적 회귀분석방법을 사용하였으며 주요 결과는 다음과 같다.

1) 20년 이상의 분석대상 자료를 수집하고 Fox의 Type 1의 이상치를 설정하되, 해당 일에 이상치 개수가 복수인 경우, 자료의 최대치가 이상치인 경우에 한하여 해당 일자의 자료를 모두 제외하는 방식으로 자료손실을 최소화하였다.
2) 부영양화의 제한인자인 인과 주요 영양인 질소를 포함하여 7개의 수질항목을 선정한 후, 수질항목 자료의 시계열 변화특성을 통계학적으로 분석하였고, 주기함수에 의한 회귀분석을 통해 연중변화특성을 고려하여 수질관리 기준설정 및 조치시점을 결정할 수 있는 방법을 제시하였다.
3) 수질항목 중 Chl-a과 다른 수질항목간의 상관관계와 Chl-a의 변화율과 다른 수질항목 및 수질항목의 변화율간의 상관관계를 분석한 바, Chl-a는 TN 및 TP의 관계보다는 Chl-a와 DTN, NH3-N, NO3-N, DTP, PO4-P 관계를 회귀모형구성에 이용하는 것이 더 적절한 것으로 판단하였다.
4) Type 3의 수질항목의 변화율을 적용하기 위해서는 변수 증가에 따라 자료 양도 증가해야 하나, 자료측적기간이 부족으로 회귀모형의 적용성이 낮았다. 따라서 향후 충분한 자료가 축적된 경우 또는 자료가 많이 축적된 지점에 한하여 적용성의 재검토가 필요하다.
5) Chl-a의 변화율을 예측하는 제안된 회귀모형은 관측치가 급격히 변할 경우에는 차이가 있으나 Chl-a의 변화추세를 따라가는 것으로 보이며 근사적인 예측치를 제공함으로 예상되는 수질변화 양상을 담수호 관리에 이용할 수 있는 방법으로 판단된다. 다만 담수호의 규모 및 자료 양을 고려하여 회귀모형의 적용대상이 제안적으로 선정되어야 한다.

기존 연구의 Chl-a 예측은 대부분 다른 수질항목과의 관계분석을 통해 제시된 바, 시간적 개념을 포함하지 않으나, 본 연구에서는 Chl-a의 시간적 변화율에 기반을 둔 통계적인 분석을 수행하였고, 관측된 수질항목으로부터 특정한 기간 후의 Chl-a를 근사적으로 추정하는 방법을 연구하였다. 제안된 방법은 근사적 예측치를 제공함으로서 복잡한 물리적 모형에 더불어 간편하고 보조적인 수단이 될 것으로 기대된다. 다만, 수질 자료 및 시간의 제약으로 인해 충분한 검토가 수행되지 못하였으므로 관측 자료의 축적 및 모형변수에 대한 지속적인

보완을 통한 개선이 적용성을 높일 것으로 기대된다.

감사의 글

본 연구는 농림축산식품부의 재원으로 농림식품기술기획평가원의 농업기반 및 재해대응기술개발사업의 지원을 받아 연구되었습니다(320049-5).

Conflicts of Interest

The authors declare no conflict of interest.

References

Bergmeir, C., Hyndman, R.J., and Koo, B.S. (2015). *A note on the validity of cross-validation for evaluating time series prediction*. Working Paper, 10/15, Department of Econometrics and Business Statistics, Monash University, Victoria, Australia, pp. 1-16.

Carlson, R.E. (1977). "A trophic state index for lakes." *Limnology and Oceanography*, Vol. 22, No. 2, pp. 361-369.

Dzialowski, A.R., Wang, S.H., Lim, N.C., Spotts, W.W., and Huggins, D.G. (2005). "Nutrient limitation of phytoplankton growth in central plains reservoirs, USA." *Journal of Plankton Research*, Vol. 27, No. 6, pp. 587-595.

Fox, A.J. (1972). "Outlier in time series." *Journal of the Royal Statistical Society Series B*, Vol. 32, pp. 337-645.

Guildford, S.J., and Hecky, R.E. (2000). "Total nitrogen, total phosphorus and nutrient limitation in lakes and oceans: Is there a common relationship?" *Limnology and Oceanography*, Vol. 45, No. 6, pp. 1213-1223.

Hastie, T, Tibshirani, R., and Friedman, J. (2009). *Elements of statistical learning*. Springer, New York, NY, U.S.

Jang, J.Y. (2010). "Improvement measures and implementation status of agricultural water quality." *Rural Resources*, Vol. 52, No. 2, pp. 23-24.

Janus, L.L. and Vollenweider, R.A. (1981). *The OECD cooperative programmi on eutrophication*. Scientific Series No. 131-S, National Water Research Institute, Canada Centre for Inland Waters, Burlington, Ontario, Canada. pp. A1-A34.

Kim, B.C., and Kong, D.S. (2019). "Examination of the applicability of TOC to Korean Trophic State Index (TSIKO)." *Journal of Korean Society on Water Environment*, Vol. 35, No. 3, pp. 271-277.

Kim, H.S., and Chung, S.O. (2002). "Prediction of nutrient loading from Paddy Fields (I)." *Journal of the Korea Society of Agriculture Engineers*, Vol. 44, No. 4, pp. 51-61.

Kim, S.B., Kim, K.T., and Lee, J.S. (2023). "Real-time WQI prediction using AI-based models." *Journal of the Korean Society for Marine Environment and Energy*, Vol. 26, No. 1, pp. 66-80.

Kong, D.S. (2019). "Statistical analysis on water quality characteristics of large lakes in Korea." *Journal of Korean Society on Water Environment*, Vol. 35, No. 2, pp. 165-180.

Lee, D.H., Kang, E.T., Joo, J.C., Go, H.W., Ahn, C.M., Bae, Y.H., and Song, K.D. (2022). "Water quality analysis and Chl-a prediction of 15 large-scale freshwater lakes in Korea by multivariate statistical analysis." *Journal of Korean Society Enviornment Engineering*, Vol. 44, No. 12, pp. 589-602.

Lim, J.S., Kim, Y.W., Lee, J.H., Park, T.J., and Byun, I.G. (2015). "Evaluation of correlation between Chlorophyll-a and multiple parameters by multiple linear regression analysis." *Journal of Korean Society Enviornment Engineering*, Vol. 37, No. 5, pp. 253-261.

Moreno-Torres, J.G., Saez, J.A., and Herrera, F. (2012). "Study on the impact of partition-induced dataset shift on k-fold cross-validation." *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 23, No. 8, pp. 1304-1312.

Park, N.S., Choi, D.Y., Lee, Y.J., and Yoon, S.M. (2015). "Evaluation of pressure reducing valves performance using statistical approach in water distribution system: Case study." *Journal of Korean Society of Water and Wastewater*, Vol. 29, No. 4, pp. 519-531.

Park, S.H. (2007). *Regression analysis (3rd edition)*, Minyoungsa, pp. 290-291.

Robinson, R.B., Cox, C.D., and Odom, K. (2005). "Identifying outliers in correlated water quality data." *Journal of Environmental Engineering*, Vol. 131, No. 4, pp. 651-657.

Rural Research Institute (RRI) (2010). *A study on the application measure of water quality model in each type of agricultural reservoir*. 11-D150769-000014-01, Rural Research Institute.

Shim, J.Y., and Shin, Y.K. (1989). "Biomass of primary producer in the Chomsu Bay - Relationships between phytoplankton carbon, cell number and Chlorophyll -." *Journal of the Oceanological Society of Korea*, Vol. 24, No. 4, pp. 197-205.