

팔당댐 유역의 데이터 기반 수질 예측 모형 구성을 위한 사전 불확실성 분석

Preliminary Uncertainty Analysis to Build a Data-Driven Prediction Model for Water Quality in Paldang Dam

이은정¹ · 금호준^{2*}

¹청주대학교 환경공학과 박사, ²국립재난안전연구원 안전연구실 선임연구원

Eun Jeong Lee¹ and Ho Jun Keum^{2*}

¹Ph.D, Department of Environmental Engineering, Cheongju University, Cheongju 28503, Korea

²Senior Researcher, Safety Research Division, National Disaster Management Research Institute, Ulsan 44538, Korea

Received 22 January 2022, revised 8 February 2022, accepted 11 February 2022, published online 31 March 2022

ABSTRACT: For water quality management, it is necessary to continuously improve the forecasting by analyzing the past water quality, and a Data-driven model is emerging as an alternative. Because the Data-driven model is built based on a wide range of data, it is essential to apply the correlation analysis method for the combination of input variables to obtain more reliable results. In this study, the Gamma Test was applied as a preceding step to build a faster and more accurate data-driven water quality prediction model. First, a physical-based model (HSPF, EFDC) was operated to produce daily water quality reflecting the complexity of the watershed according to various hydrological conditions for Paldang Dam. The Gamma Test was performed on the water quality at the water quality prediction site (Paldangdam2) and major rivers flowing into the Paldang Dam, and the method of selecting the optimal input data combination was presented through the analysis results (Gamma, Gradient, Standar Error, V-Ratio). As a result of the study, the selection criteria for a more efficient combination of input data that can save time by omitting trial and error when building a data-driven model are presented.

KEYWORDS: Data uncertainty analysis, Gamma test, Physics-based model, Water quality

요 약: 수질분야에서 물재해 안정성 강화를 위해 과거와 현재의 수질을 분석하여 예측하는 기술을 지속적으로 고도화하는 것이 필요하며 데이터 기반의 예측 모형이 하나의 대안으로 대두되고 있다. 데이터 기반 모형은 복잡하고 광범위한 자료의 양을 기반으로 구축되기 때문에 보다 신뢰도 있는 결과를 얻을 수 있는 입력자료의 조합을 위한 상관관계 분석방법의 적용이 필수적이다. 본 연구에서는 보다 신속하고 정확한 데이터 기반의 수질 예측 모형을 구성하기 위한 선행단계로 Gamma Test를 적용하였다. 먼저 팔당댐의 다양한 수문조건에 따른 해당 유역의 복잡성과 정밀성이 재현된 과거와 현재의 일단위 수질을 최대한 확보하고자 물리적 기반 모형 (HSPF, EFDC)을 구동하였다. 팔당댐 수질예측지점과 팔당댐으로 유입되는 주요 하천의 수질을 대상으로 Gamma Test를 수행한 후 해석결과 (Gamma, Gradient, Standar Error, V-Ratio)를 통해 최적의 자료조합을 선정하는 방법을 제시하였다. 본 연구의 결과는 데이터 기반 모형 구축 시 반복적인 수행과정을 생략하여 시간을 단축하면서 보다 효율적으로 최적의 입력자료를 선정할 수 있는 정량적인 기준을 보여준다.

핵심어: Gamma test, 데이터 불확실성 분석, 수질, 물리적 기반 모델.

*Corresponding author: hojunkeum@korea.kr, ORCID 0000-0003-2433-3646

© Korean Society of Ecology and Infrastructure Engineering. 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.

1. 서론

최근 유역관리를 위하여 빅데이터 기반의 분석 기법이 적용되고 있다. 이 중 예측 분석(Predictive Analytics)은 과거의 패턴 식별을 통한 예측 기능으로 과거의 관측 자료 또는 물리적 기반 모형의 결과를 학습시켜 데이터 기반 모형(Data-driven model)을 개발하는 것으로 홍수, 수질오염 예측 등 물재해에 대한 사전관리를 위해 적용이 가능하다(Poul et al. 2019). 특히 수자원 분야에서 다양한 데이터 분석 및 기계학습의 방법론이 적용되고 있으며 데이터 기반 모형은 수문학의 기본인 물리학적 프로세스를 직접 고려하지 않고 과거의 데이터를 사용한다는 특징이 있다(Keum et al. 2019). 수질의 경우, 오염원의 불확실성과 비선형적 관계로 인해 데이터 기반 모형을 활용한 수질 예측이 새로운 모델링 툴로 평가받고 있으나 대량의 수질자료 간 상호연관성과 의도적으로 지나치게 단순화된 가정 등이 존재하기 때문에 다양한 유형의 불확실성과 모호성에 대해 지적된 바 있다(Kim and Seo 2015). 데이터 기반 모형인 ANFIS(Adaptive Neuro-Fuzzy Inference System)의 경우 각 입력자료의 규칙이 3개라고 하면 퍼지이론에 따라 총 규칙의 수는 $3n$ 개가 된다. 입력변수가 증가할수록 이와 비례하여 모형의 정확도가 증가하는 것은 아니며 퍼지 규칙수가 기하급수적으로 증가함으로써 모의시간이 길어지게 되는 단점이 있다(NDMI 2018). 따라서 예측결과와 정확도를 높이면서 모의시간 또한 적절하게 단축시킬 수 있는 입력자료 조합을 선정하기 위한 입력변수와 출력변수 간의 상관관계 분석은 필수적인 선행 단계라고 할 수 있으며, 분석방법으로는 주성분 분석, 군집분석, AIC, BIC 분석, K-fold Cross Validation, Gamma Test 등이 있다.

데이터 기반의 수질 예측 모형을 구축한 선행연구 사례를 보면 대부분 월단위 이상의 수질 관측자료를 입력 자료로 구성하였다(Faruk 2010, Najah et al. 2013, Sarkar and Pandey 2015, Khadr and Elshemy 2017). 국내 수질측정망 중 가장 많은 자료의 수로 공인된 수질을 제공하고 있는 것은 8일 간격의 총량측정망 자료라고 할 수 있는데 해당 자료만으로 데이터 기반 모형을 구축하기에는 자료의 개수가 부족하며 특정 유량구간에 수질 측정이 편중되어 있을 경우 해당 유량구간의 수질영향이 과다하게 반영되는 결과를 초래하는 한계점을 가지고 있다고 보고된 바 있다(Park and Oh 2012,

MOE 2015, Lee et al. 2018).

본 연구에서는 비선형성 특성을 가지는 수질 예측 모형의 입력자료를 최대한 확보하고자 지형, 오염원, 기상, 수문 등 유역환경조건을 반영하고 실제 관측자료로부터 보검증을 완료한 물리적 기반 모형으로부터 도출하였으며, 일단위로 확보된 입력자료에 대한 사전 불확실성 점검과 함께 신뢰도 향상을 위하여 Gamma Test를 수행하여 최적의 자료 조합을 선정하고자 한다. Gamma Test는 Stefánsson et al. (1997)에 의해 소개된 이론으로 $\text{NH}_3\text{-N}$, BOD 등의 수질 예측을 위한 모형 구축 시 핵심 입력인자 조합 선정을 위해 적용된 바 있다(Chang et al. 2015, Lee and Kim 2021). 연구대상유역은 우리나라 수도권 시민의 상수원인 팔당댐 유역으로 수질예측을 위한 목표지점은 EFDC (Environmental Fluid Dynamic Code), 팔당댐으로 유입되는 주요 하천별 수질은 HSPF (Hydrological Simulation Program - Fortran) 모형을 적용하였다. 즉, 목표지점(팔당댐)의 수질과 총 9개의 유입하천별 수질에 대하여 수질항목별 Gamma Test의 다양한 적용 결과를 고찰하였다.

2. 연구 방법

2.1 연구 대상 유역

본 연구의 대상 유역은 한강 대권역 내 충주댐하류, 섬강, 남한강하류, 의암댐, 홍천강, 팔당댐, 경안천, 한강잠실 등 9개 중권역을 포함하는 약 2,550만 명 수도권 시민의 상수원인 팔당댐의 영향권이다. 팔당댐은 국내 최대 규모의 취수원으로 수질측면에서 매우 큰 관심을 받고 있으며, 수질보전을 위한 엄격한 규제로 인해 목표수질 달성여부 또는 수질변화가 매우 중요한 유역이다(Kim and Lee 2019).

팔당댐은 23,800 km^2 의 넓은 유역면적을 가진 데 비하여 수면적이 36.2 km^2 이고, 유효 저수량이 2억 4,400만 톤으로 물그릇이 작아 체류시간이 평균 5-6일, 평균 수심 약 6.5 m로 하천형 인공호로서 하천 기준인 BOD 가 목표수질 중 하나로 평가되고 있다(Kim et al. 2009).

본 연구에서 수질예측 목표지점으로 설정한 팔당댐 2지점은 수도권 광역상수도 취수구 인근에 위치하고 있으며 우리나라 물환경분야의 최상위계획에서 관리하고 있는 목표기준은 1a등급이다. 해당 지점은 Fig. 1과 같이 남한강(한강D), 북한강(북한D)을 비롯한 경

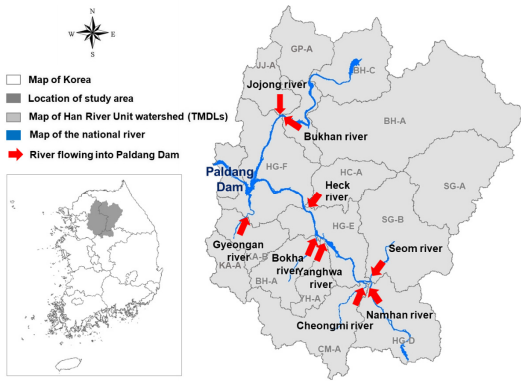


Fig. 1. Map showing the study area in Pandang Dam, Korea.

안천(경안B), 조종천(조종A), 양화천(양화A), 청미천(청미A), 섬강(섬강B), 흑천(흑천A), 복하천(복하A) 등 많은 유입하천의 영향을 받고 있다.

2.2 물리적 기반 모형의 적용

본 연구에서 다양한 유량조건에서의 일단위 수질을 생산하기 위해 물리적 기반 모형을 적용하였다. 먼저 수질예측의 목표지점으로 설정한 팔당댐2지점의 경우 호소모형인 EFDC 모형을 구축·구동하였으며, 팔당댐으로 유입되는 주요 하천별 수질을 모의하기 위해 유역모형인 HSPF 모형을 적용하였다. 또한 EFDC 모형 구축시 HSPF 모형의 모의자료를 초기 및 경계조건으로 입력하여 구동하였다.

HSPF 모형은 강우로 인한 비점오염물질 유출과정을 하천 내에서의 수리학적 현상, 퇴적물, 화학물질의 상호작용과 결합시켜 모의할 수 있는 유역모델과 수질 모델이 결합된 형태를 가진 유역유출모델이다 (U.S. EPA 2001). HSPF 모형의 사용을 위해 GIS 기반으로 방대한 자료를 관리하고 다양한 모델들을 지원해주는 통합관리시스템인 BASINS를 이용하여 기본정보를 입력하였다. 먼저 DEM으로부터 Flow Direction과 Flow Accumulation을 계산하여 하천망을 생성시킨 후 유역의 출구지점을 outlet으로 지정하여 분할하였으며, 토지피복도는 Urban or Built-up land, Wetland, Agricultural land, Forest Land, Pasture Land, Barren Land, Water 등 7개 항목으로 분류한 후, 모형 내 Landuse and Soil Definition Utility를 이용하여 분할된 소 유역에 대한 각각의 토지이용 정보를 추출하였다. 기상

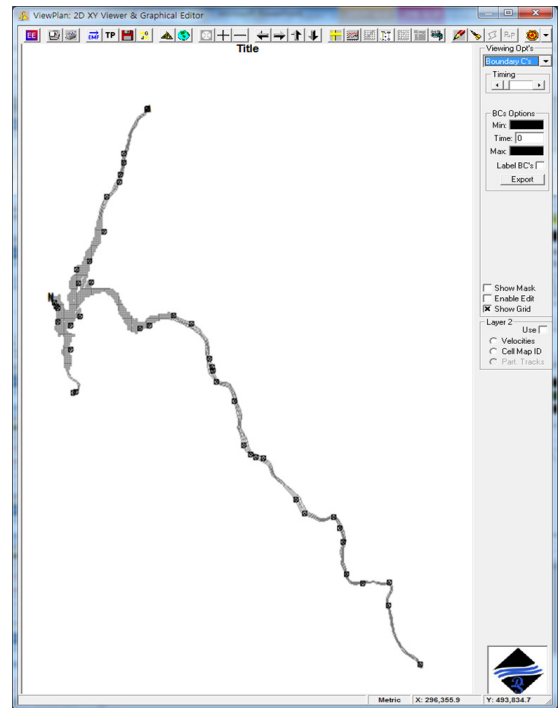


Fig. 2. Simulation interval of EFDC model and inflow boundary point of HSPF model.

자료는 강수량, 기온, 이슬점온도, 운량, 일사량, 풍속, 증발산량 등 7개 항목의 시간단위 관측값을 입력한 후 WDMUtil 내의 Disaggregate Functions 중 Evapotranspiration을 활용하여 증발산량을 생성하였다. 점오염부하량 자료는 2008 - 2018년의 오염원 조사자료로부터 오염부하량을 산정하여 입력하였으며, 일평균 방류량이 500 m³ 이상인 공공하수처리시설은 일별 방류량과 방류수질 (BOD, SS, T-N, T-P) 자료를 수집하여 입력하였다. 댐 방류량, 광역 취수자료는 일단위, 지방상수도는 월단위의 취수자료를 조사하여 입력하였다.

EFDC 모형은 연안, 하구, 호소, 습지, 저수지 등 유동 및 물질수송을 모의하는 3차원 수치모델로서 미국 VIMS (Virginia Institute of Marine Science)에서 개발되었다 (Tetra Tech, Inc. 2007). 팔당댐 유역의 EFDC 모형 구성은 북한강, 남한강, 경안천에서 팔당댐까지를 대상으로 하였다. 격자구성을 위해 필요한 지형자료는 한강수계 횡단면도를 AutoCAD에 적용하여 구성하였으며, 격자 크기가 작아질수록 대상 유역의 격자수가 많아져 모의시간이 길어지기 때문에 팔당댐은 200 m × 200 m 크기의 cell로 구성 (Cartesian Coordinate)하고 그 외의 하천은 Orthogonal Curvilinear Coordinate 방

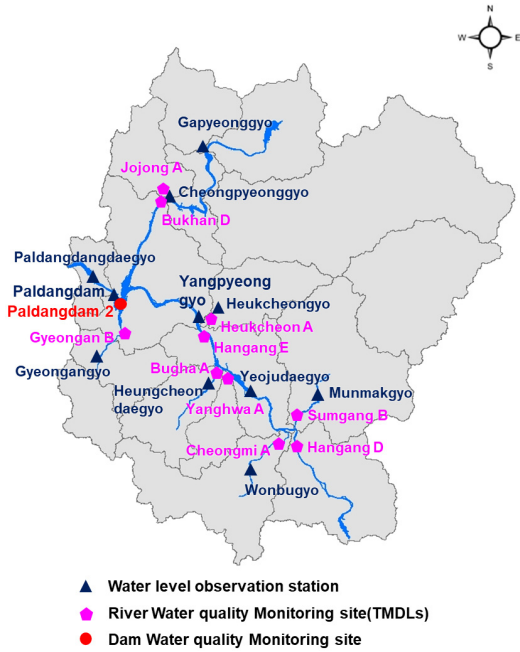


Fig. 3. Calibration and validation sites of physics-based model.

법을 사용하여 수평방향으로 2,700개의 격자를 구성하였다. HSPF 모형에서 나오는 유량 및 수질 결과파일을 EFDC 모형으로 입력하기 위한 경계지점은 40개로 지정하였다 (Fig. 2).

2.3 물리적 기반 모형의 보정 및 검증 방법

HSPF 모형의 보정 및 검증에 사용한 관측자료는 먼저 유량은 문막교, 원부교, 여주대교, 흥천대교, 흑천교, 양평교, 가평교, 청평교, 경안교, 팔당대교 등 10개의 수위관측소 자료를 이용하였다. 수질은 한강물환경연구소에서 각 총량관리 단위구역 말단에서 측정된 연간 약 40회의 수질자료를 활용하였다. EFDC 모형의 경우 팔당댐에서 수위보정을 수행하였으며 수질은 팔당댐 2지점에서 관측된 자료를 활용하였다 (Fig. 3). HSPF, EFDC 모형의 보정과 검증기간은 각각 2015 - 2018년, 2011 - 2014년도 자료를 이용하였다.

유량의 보정 및 검증 결과에 대한 적정성을 평가하기 위하여 결정계수 (R^2)와 % difference를 산정하였으며 Tables 1 and 2의 기준을 근거로 판단하였다. 수위의 경우, % difference와 관측값과 모의값의 평균값을 비교하였다. 수질은 수질항목별로 관측값과 모의값의 평균

Table 1. The General calibration/validation target or tolerance for HSPF application (Donigian 2000)

Criteria	Very good	Good	Fair	Poor
R^2	> 0.8	0.8 - 0.7	0.7 - 0.6	< 0.6

Table 2. The % difference value range for Model performance (Donigian 2000)

Constituent	Very good	Good	Fair
Hydrology / Flow	< 10	10 - 15	15 - 25
Water Quality	< 15	15 - 25	25 - 35

값을 제시하였으며 % difference, RMSE (Root Mean Square Error)를 산정하여 관측값에 대한 수질 보정 및 검증 결과의 적정성을 평가하였다.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (P_i - O_i)^2}{N}} \quad (Eq. 1)$$

P_i : predicted values, O_i : observed values

2.4 Gamma Test

Gamma Test는 연속적인 비선형 기법을 사용하여 데이터를 모델링 할 때 산정되는 최소 평균제곱오차를 추정하는 기법으로 최초로 Stefánsson et al. (1997) 의해 발표된 후, 많은 연구자들에 의해 기법이 발전되고 구체적으로 정립되었다 (Keum et al. 2019). Gamma Test를 위한 기본적인 개념은 비선형 해석을 이용한 기존의 자료 전처리 위한 분석방법과는 차별성이 있다. Eq. 2와 같이 분석 대상 자료가 있다는 가정 하에 분석을 실시한다.

$$(x_i, y_i), 1 \leq i \leq M. \quad (Eq. 2)$$

여기서, 입력 벡터 $x_i \in R_n$ 는 폐집합 $C \in R_n$ 에 한정된 벡터이며, 일반적인 모형의 결과 또는 종속변수에 해당되는 출력값 $y_i \in R_n$ 는 스칼라이다. 벡터 x 값들은 출력값 y 에 영향을 주며 예측을 위한 입력값으로 사용할 수 있는 독립변수이다. x 와 y 관계에 대한 기본적인 가정은 다음과 같다.

$$y = f(x_1 \dots x_m) + r \quad (Eq. 3)$$

여기서, f 는 지수 평활법 함수이고 y 은 노이즈를 나타

내는 확률 변수이다. Eq. 3에서 x 값은 입력자료, y 값은 예측되는 결과를 나타낼 수 있다. Gamma Test는 각 벡터 $x_i (1 \leq i \leq M)$ 에 대한 k 번째 ($1 \leq i \leq p$) 가장 가까운 변수 $N[i, k]$ 를 기반으로 비선형 분석을 실시한다. 각각의 입력자료 x 에 대해서 k 개의 최근접 이웃자료와의 평균 제곱근거리를 산정하게 된다. 이에 대한 수식은 Eq. 4와 같다. 독립변수 x 에 대응되는 결과(또는 목표)값이 되는 y 에 대해서도 동일한 방식으로 k 개의 최근접 이웃자료와의 평균 제곱근거리를 산정하게 된다 (Eq. 5).

$$\delta_M(k) = \frac{1}{M} \sum_{i=1}^M |x_{N(i,k)} - x_i|^2 \quad (1 \leq k \leq p) \quad (\text{Eq. 4})$$

$$\gamma_M(p) = \frac{1}{2M} \sum_{i=1}^M (y_{N[i,k]} - x_i)^2 \quad (1 \leq p \leq p_{\max}) \quad (\text{Eq. 5})$$

각 변수들에 대한 최근접 이웃자료와의 물리적 거리의 관계를 산정하였을 때에, 자료 양과 동일한 수만큼의 δ 와 γ 의 값들이 생성되게 되며, 두 값을 통하여 τ 를 산정하기 위한 회귀식을 생성하게 된다. 이를 통하여 산정된 감마 통계량 τ 는 지수 평활법 데이터 모형에 의해 설명될 수 없는 인공신경망 모형의 결과값에 대한 분산의 추정치이다 (Eq. 6).

$$\gamma = A\delta + \tau \quad (\text{Eq. 6})$$

$$V_{ratio} = \frac{\tau}{\sigma^2(y)} \quad (\text{Eq. 7})$$

산정된 τ 는 특정 입력 자료조합에 따른 불확실도 V_{ratio} 지표를 산정할 때에 사용된다. Eq. 7에서 $\sigma^2(y)$ 는 출력 y 의 분산이며, 이는 출력이 부드럽고 신뢰성 있게 모델링 될 수 있는지에 대하여 출력 범위와 독립적으로 예측력에 대한 판단이 가능하도록 만들어준다. 0에 가까운 V_{ratio} 는 주어진 출력 y 의 높은 예측 가능성을 나타낸다. 위의 과정을 통하여 Gamma Test는 예상 노이즈 분산을 근사하는 평균제곱오차로 모델을 구축하는 데 필요한 데이터의 양을 결정하는데 도움을 줄 수 있다 (Rauf et al. 2016).

3. 결과 및 고찰

3.1 물리적 기반 모형의 보정 및 검증결과

3.1.1 유량 및 수위

모형의 보정은 모형의 초기조건과 매개변수를 유역의 실제 여건에 맞게 환경을 구성하는 과정으로 모의된 값과 관측된 값이 가장 근접할 때의 매개변수 값을 추정하는 것이다 (Lee 2013). 보정기간 동안 시행착오법으로 HSPF, EFDC 모형의 모의 결과가 관측값과 최대한 근접하도록 하였으며 검증을 수행함으로써 보정기간에 맞춰진 모형 매개변수들의 타당성을 검토하였다.

먼저 HSPF 모형에서 유량의 보정 및 검증기간은 각각 2015 - 2018년, 2011 - 2014년으로 4개년씩 지정하여 R^2 , %difference 값을 산정하였으며, Table 3에서 보는 바와 같이 10개의 수위관측소에 모두 Very good으로 모형의 모의치가 실측치를 매우 잘 반영하는 것으로 나타났다. 강우 시의 첨두값이 대부분 관측값보다 약간 작게 모의되어 저·갈수기의 유량을 보다 잘 재현하는 것으로 나타났다. Table 4는 EFDC 모형에서 팔당댐 수

Table 3. The HSPF model statistical performance measures evaluated based on daily flows at the monitoring stations for calibration and validation periods

Site	Calibration (2015 - 2018)		Validation (2011 - 2014)	
	R ²	% Difference	R ²	% Difference
Munmakgyo	0.78	10.05	0.81	11.73
Wonbugyo	0.82	5.74	0.78	12.51
Yeojudaegyo	0.83	6.92	0.96	9.94
Heungcheondaegyo	0.81	7.03	0.94	11.39
Heukcheongyo	0.86	9.48	0.82	10.83
Yangpyeonggyo	0.89	9.29	0.94	8.94
Gapyeonggyo	0.91	4.02	0.84	0.74
Cheongpyeonggyo	0.92	6.57	0.90	4.59
Gyeongangyo	0.89	7.77	0.94	5.25
Paldangdaegyo	0.94	6.67	0.93	9.81

Table 4. The Water level efficiency evaluation of EFDC model for calibration and validation periods

Site	Calibration (2015 - 2018)			Validation (2011 - 2014)		
	mean OBS	mean SIM	% Diff	mean OBS	mean SIM	% Diff
Paldangdam	25.13	25.12	0.26	25.16	25.16	0.26

위에 대해 보정 및 검증한 결과로 관측값과 모의값의 평균, %difference 값을 산정한 것으로 Very good으로 평가되었으며 여름철 수위가 급격하게 증가하거나 감소하는 일부를 제외하고 모의값과 실측값이 적절히 일치하였다.

3.1.2 수질

HSPF 모형의 수질 보정을 위해 수온, DO를 보정하

후 BOD, TOC, T-N, T-P, Chl-a에 대하여 선행연구를 참고하여 보정을 수행하였으며 (Lee 2013), 본 연구에서는 Gamma Test 조합을 위해 고려한 BOD, TOC, T-P, Chl-a 항목만을 제시하였다. Table 5에서 제시한 바와 같이 보정과 검증 결과 모두 % difference 값이 대부분 15 이하로 산정되어 “Very Good”으로 평가되었다. 모의값과 실측값이 같으면 0에 가까운 값을 나타내는 RMSE의 경우 0.02 - 35.23의 범위로 산정되었다.

Table 5. The Water quality calibration and validation results of HSPF and EFDC models

Water-shed	WQ item	Calibration results (2015 - 2018)				Validation results (2011 - 2014)			
		meanOBS	meanSIM	%diff	RMSE	meanOBS	meanSIM	%diff	RMSE
Hanjung D (HSPF)	BOD	0.9	0.9	2.67	0.17	0.8	0.8	0.28	0.16
	TOC	2.2	2.1	1.34	0.30	1.9	1.9	0.25	0.41
	T-P	0.033	0.029	10.71	0.01	0.033	0.031	3.67	0.03
	Chl-a	5.452	5.467	0.27	1.02	5.253	4.930	6.55	1.66
Sumgang B (HSPF)	BOD	1.5	1.4	8.62	0.76	1.7	1.6	2.48	0.81
	TOC	3.1	3.1	0.65	0.75	2.8	2.7	4.48	2.19
	T-P	0.053	0.052	2.16	0.02	0.108	0.103	5.09	0.06
	Chl-a	13.714	14.280	4.12	4.01	11.800	11.452	2.95	7.44
Cheongmi A (HSPF)	BOD	3.5	2.9	16.37	3.15	2.5	2.6	9.00	2.82
	TOC	4.4	4.8	10.40	1.57	3.8	4.0	4.62	2.30
	T-P	0.134	0.113	15.18	0.08	0.121	0.114	5.26	0.05
	Chl-a	46.785	42.858	8.39	35.23	18.552	19.590	5.90	10.68
Yanghwa A (HSPF)	BOD	2.0	1.8	8.98	0.72	2.1	2.0	8.09	1.27
	TOC	3.9	3.6	7.50	0.94	4.0	3.3	17.77	2.33
	T-P	0.192	0.148	20.75	0.18	0.183	0.160	12.91	0.09
	Chl-a	12.228	11.988	1.97	2.48	10.312	8.881	19.69	6.31
Bugha A (HSPF)	BOD	4.4	4.5	1.16	1.47	3.0	3.2	5.50	1.26
	TOC	4.4	4.4	0.11	0.47	3.9	3.8	2.09	2.31
	T-P	0.168	0.150	10.80	0.08	0.166	0.148	10.75	0.06
	Chl-a	21.008	18.586	11.53	13.87	8.304	8.775	5.67	3.02
Heukcheon A (HSPF)	BOD	2.3	0.9	14.59	0.70	1.2	1.0	14.07	0.74
	TOC	3.5	1.9	7.97	0.55	1.9	1.9	2.30	0.38
	T-P	0.063	0.030	10.19	0.01	0.034	0.034	1.82	0.02
	Chl-a	29.636	5.569	0.18	5.26	6.098	4.915	9.39	3.95
Gyeongan A (HSPF)	BOD	2.3	2.0	11.39	1.16	2.6	2.5	3.61	0.71
	TOC	3.5	3.4	4.22	0.90	3.6	3.4	6.11	0.90
	T-P	0.063	0.045	18.25	0.04	0.0749	0.065	17.95	0.03
	Chl-a	29.636	24.010	18.98	24.90	22.363	19.731	11.77	14.90
Bukhan D (HSPF)	BOD	1.0	1.0	4.67	0.35	1.0	1.0	3.45	0.25
	TOC	2.1	2.0	5.59	0.45	1.9	1.8	5.03	0.56
	T-P	0.015	0.016	7.01	0.01	0.021	0.019	10.77	0.02
	Chl-a	7.574	6.940	8.37	2.48	7.666	7.150	6.73	4.14
Jojong A (HSPF)	BOD	0.9	0.8	19.25	0.44	1.0	0.9	7.10	0.51
	TOC	1.9	1.7	7.23	0.35	1.74	1.6	5.27	0.69
	T-P	0.026	0.018	29.42	0.02	0.028	0.022	10.41	0.02
	Chl-a	4.293	3.767	12.26	1.87	5.008	4.337	13.40	2.95
PaldangDam2 (EFDC)	BOD	1.4	1.3	3.96	0.48	1.3	1.1	15.16	0.57
	TOC	2.4	2.3	0.52	0.44	2.3	2.1	6.57	0.76
	T-P	0.026	0.026	1.20	0.01	0.035	0.030	12.70	0.03
	Chl-a	12.714	9.858	14.46	5.88	14.424	11.996	14.83	6.64

Table 6은 팔당댐으로 유입되는 주요 하천인 남한강에 대한 HSPF 보정 및 검증 결과를 그림과 함께 추가로 제시한 것이며 모의치가 관측치의 경향을 잘 반영하고 있는 것을 확인할 수 있다. EFDC 모형은 팔당댐 2지점을 대상으로 BOD, TOC, T-P, Chl-a의 수질항목을 보정 및 검증을 수행한 결과, % difference 값은 모두 15 이하로 산정되어 “Very Good”으로 평가되었으며, RMSE는 0.01 - 6.64의 범위로 산정되었다. 이에 대한 그림은 Table 7에 수질항목별로 도식화하여 제시하였다.

HSPF와 EFDC 모의 결과 모두 복잡한 수질 메커니즘과 유역의 공간적인 특성을 고려할 때 전반적으로 적

절히 모의된 것으로 판단된다.

3.2 데이터 기반 수질 예측 모형 구성을 위한 불확실성 분석

3.2.1 Gamma Test 적용

데이터 기반의 모형을 이용한 수질예측에 있어 중요한 사항은 유입하천별 다양한 오염원 및 유량조건에 따른 수질특성과 시간적 분포에 따른 입력자료의 조합을 어떻게 구성할 것인지 그리고 어느 정도의 자료를 사용할 것인지 결정하는 것이다. 즉, 예측하고자 하는 대상

Table 6. The calibration and Validation Results of HSPF model in Hangang-D (NamHangang)

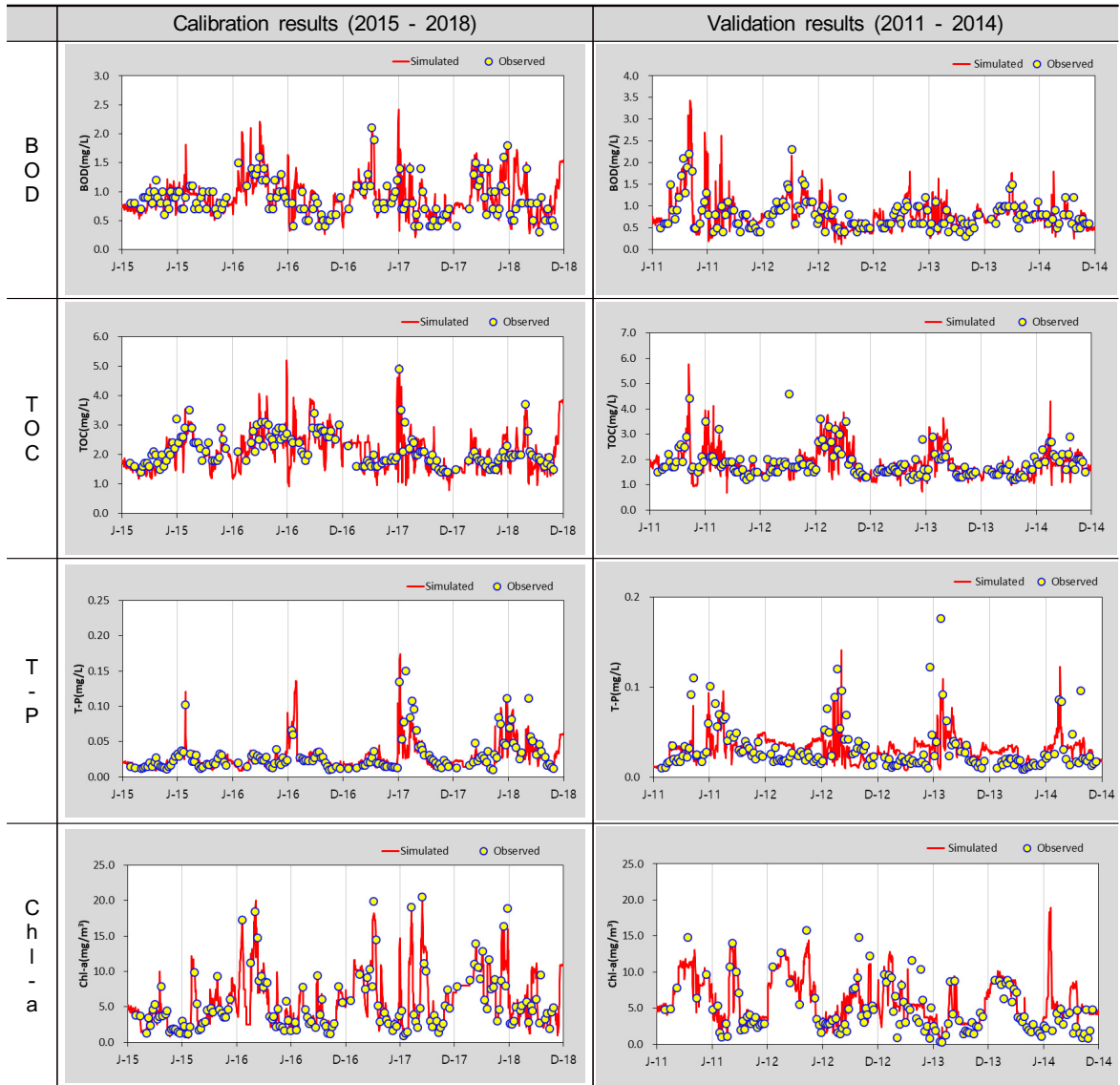
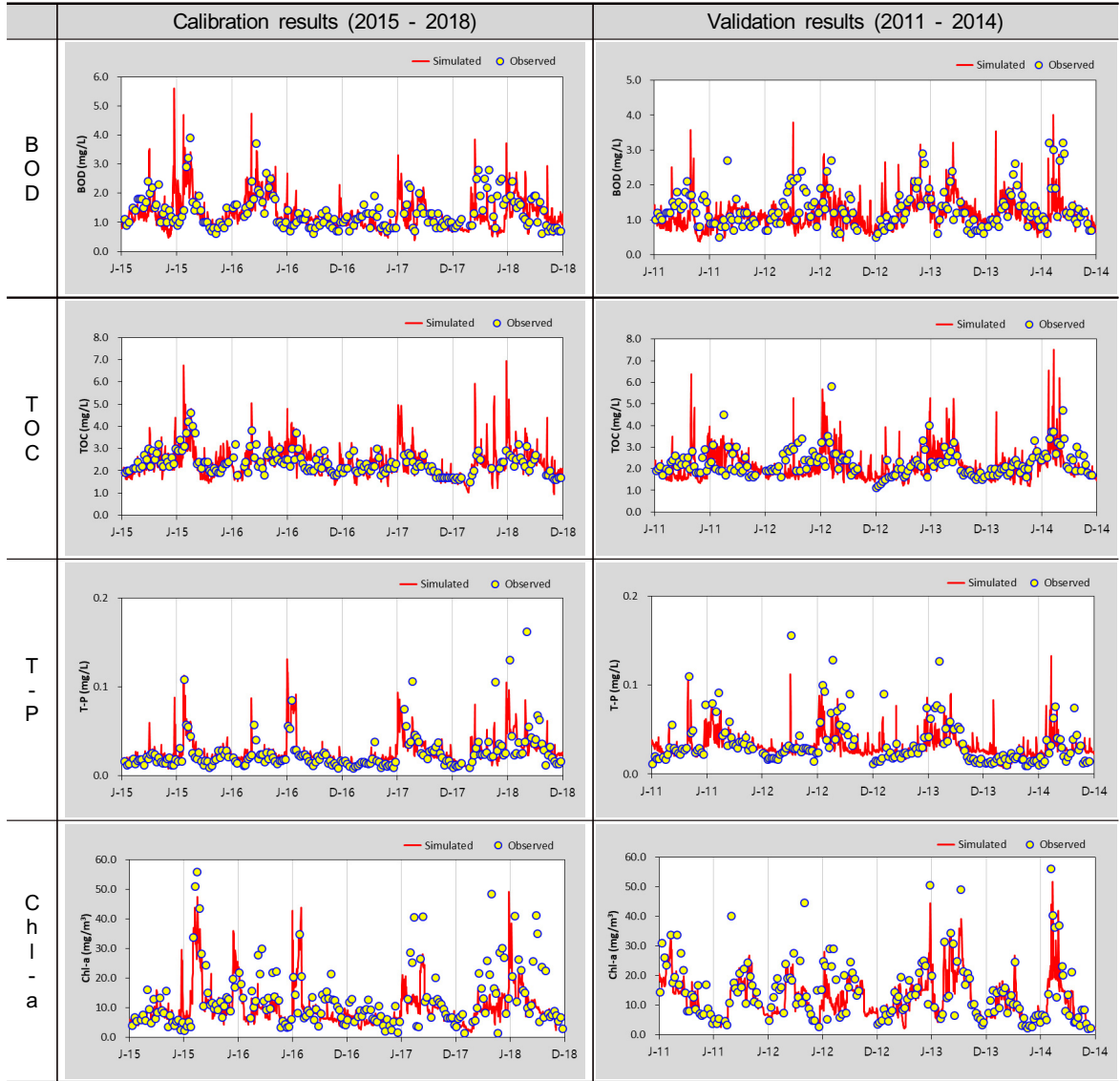


Table 7. The calibration and Validation Results of EFDC model in Paldang Dam – 2

지점의 수질과 가장 상관성이 있는 유입하천별 수질항목의 자료 조합을 어떻게 구성하는 것이 데이터 기반 모형의 구성에 있어 모의시간을 줄이고 보다 신뢰도 있는 결과를 제시할 수 있는지에 대하여 검토하였다. 먼저 팔당댐 수질을 예측하기 위한 미지의 데이터 기반 모형을 구축하고자 입력 및 출력자료로 앞 절에서 보정 및 검증과정을 거친 물리적 기반 모형으로부터 일단위 수질을 생산하였다. 즉, 하천별 HSPF 모의 수질과 같은 일자의 EFDC 모의 수질을 Gamma Test에 적용하여 최적의 입력자료를 선정하고자 계획하였다.

수질 예측을 목적으로 하는 경우 수질예보와 같이 특

정항목에 대한 예·경보발령 등 비상시를 모의하고, 단기적인 예측으로 각종 수질사고에 대비하기 위한 것으로 안정성, 신속성이 중요시된다(Seo and Choi 2007). 하지만 EFDC 모형의 경우 안정화 기간이 오래 소요되고 모형의 시작시간 역시 초기조건에 영향을 미쳐 모델링 기간에 대한 고찰이 필요하며, 본 연구대상지역을 1년간 모의하는데 약 3시간 소요되는 등 구동시간이 오래 걸리는 단점이 있다(Oh 2016). 이에 따라 해당 유역의 복잡성과 정밀성을 포함한 과거 수질자료를 기반으로 보다 안정되고 신속하게 표출할 수 데이터 기반 모형 구축을 위한 선행 연구단계의 결과를 제시하였다.

Fig. 4는 Gamma Test를 적용한 흐름도를 제시한 것으로 입력값에 대한 목표값의 정확도를 측정하기 위해서 Gamma Test의 Near Neighbours는 10개의 값을 사용하였다. Gamma Test를 수행 후에 계산되는 결과는 Gamma, Gradient, Standard error, V-ratio로 Gamma의 경우 값이 클수록 자료간의 상관성이 낮고 불확실도가 높음을 나타내고 Standard error는 자료의 표준편차를 나타낸다. Gradient의 경우 값이 급하게 나올수록 수질항목간 관계의 복잡함을 나타내며, V-ratio는 0값에 가까울수록 입력자료를 통한 수질의 예측력이 높음을 나타낸다 (Niknia et al. 2014).

3.2.2 Gamma Test의 결과 해석

본 연구의 수질예측 목표지점은 남한강을 비롯한 총 9개의 하천이 유입되어 영향을 미치는 지점으로 수질항목별 자료가 방대하여 예측하고자 하는 팔당댐2지점의 수질항목과 유입하천별 수질 (BOD, TOC, T-P, Chl-a 등)에 대한 상관성 분석을 수행하였다. 즉, 유입되는 모든 하천을 고려했을 경우와 유입 하천을 하나씩 소거하였을 때의 Gamma, V-ratio값을 비교함으로써 상관성은 더 높고, 자료의 양은 더 적게 입력자료 조합을 구성하는 것을 목표로 하였다. Table 8을 보면 BOD의 경우

모든 하천을 고려했을 경우의 Gamma값은 0.008163이며, 섬강, 흑천, 청미천, 북한강의 제외된 경우의 Gamma값을 보면 0.008163 보다 더 큰 값으로 산정되었는데 이는 모든 하천을 고려했을 때보다 4개 하천의 BOD를 제외하면 상관성이 더 낮아지고 불확실도가 높음을 의미한다. V-ratio값 또한 섬강, 흑천, 청미천, 북한강의 4개 하천을 제외했을 때 값을 보면 모든 하천을 고려했을 때의 0.032652 보다 더 크게 산정되었다. V-ratio값은 0에 가까울수록 입력자료를 통한 수질의 예측력이 높기 때문에 4개의 하천을 입력자료로 포함시켜야 함을 확인할 수 있다. 이에 따라 팔당댐 2지점의 BOD 수질에 가장 큰 영향을 끼치는 하천은 섬강, 흑천, 청미천, 북한강이라고 해석할 수 있다. 이와 같은 방법으로 타 수질항목의 상관성을 분석한 결과 팔당댐 2지점의 TOC는 양화천과 복하천을 제외한 7개의 하천 (남한강, 섬강, 청미천, 흑천, 경안천, 북한강, 조종천 등)에 대한 상관성이 큰 것으로 분석되었다(Table 9). T-P는 흑천, 경안천, 북한강의 3개의 하천이 Chl-a는 남한강, 청미천, 북한강, 조종천 등 4개의 하천에 대한 영향이 팔당댐2지점의 수질에 영향이 큰 것으로 나타났다 (Tables 10 and 11). 수질항목별 Gamma Test 결과를 보면 4개 수질항목 모두 북한강의 영향이 지배적임을 확인할 수 있다. 이는 팔당댐2

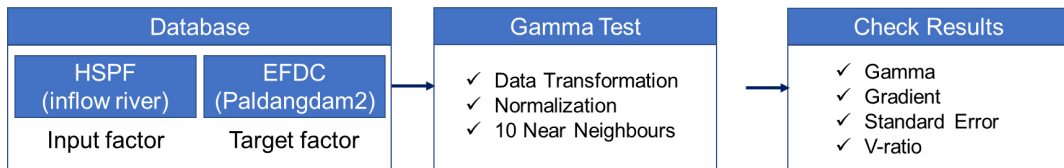


Fig. 4. Flowchart for the Application of Gamma Test.

Table 8. The gamma test results to derive the input variable combination for BOD

Except River	Gamma	Gradient	Standard Error	V-Ratio	selection of combination
Sum river	0.013059	0.018835	0.005882	0.052237	O
Cheongmi river	0.011939	0.014942	0.003154	0.047757	O
Heck river	0.012842	0.015249	0.004110	0.051368	O
Bukhan river	0.008517	0.019977	0.003838	0.034070	O
Namhan river	0.003927	0.020346	0.003253	0.015707	-
Yanghwa river	0.002057	0.019614	0.002236	0.008229	-
Bokha river	0.006438	0.017395	0.002915	0.025750	-
Gyeongang river	0.007967	0.016834	0.003337	0.031867	-
Jojong river	0.006400	0.017111	0.003402	0.025602	-
Considered all rivers	0.008163	0.014786	0.003551	0.032652	

Table 9. The gamma test results to derive the input variable combination for TOC

Except River	Gamma	Gradient	Standard Error	V-Ratio	selection of combination
Namhan river	0.017451	0.015394	0.002623	0.069803	O
Sum river	0.018379	0.017108	0.002903	0.073515	O
Cheongmi river	0.015360	0.014885	0.002732	0.061441	O
Heck river	0.021490	0.013224	0.004254	0.085960	O
Gyeongan river	0.016661	0.014477	0.003899	0.066645	O
Bukhan river	0.021403	0.015439	0.004600	0.085613	O
Jojong river	0.015753	0.014658	0.003868	0.063013	O
Yanghwa river	0.012567	0.016115	0.002921	0.050269	-
Bokha river	0.008611	0.016971	0.004239	0.034445	-
Considered all rivers	0.012819	0.013927	0.003933	0.051277	

Table 10. The gamma test results to derive the input variable combination for T-P

Except River	Gamma	Gradient	Standard Error	V-Ratio	selection of combination
Heck river	0.023798	0.006581	0.003716	0.095193	O
Gyeongan river	0.021062	0.007258	0.002986	0.084246	O
Bukhan river	0.021683	0.008428	0.002060	0.086730	O
Namhan river	0.010426	0.011722	0.002499	0.041705	-
Sum river	0.019251	0.010854	0.002264	0.077003	-
Cheongmi river	0.017406	0.008742	0.002710	0.069622	-
Yanghwa river	0.014449	0.009860	0.001491	0.057797	-
Bokha river	0.013963	0.010049	0.003338	0.055850	-
Jojong river	0.015351	0.009320	0.004021	0.061405	-
Considered all rivers	0.019741	0.006931	0.003163	0.078965	

Table 11. The gamma test results to derive the input variable combination for Chl-a

Except River	Gamma	Gradient	Standard Error	V-Ratio	selection of combination
Namhan river	-0.014298	0.022509	0.004178	-0.057193	O
Cheongmi river	-0.012170	0.019897	0.003807	-0.048680	O
Bukhan river	-0.013756	0.021848	0.003441	-0.055025	O
Jojong river	-0.013271	0.020326	0.004577	-0.053083	O
Sum river	-0.007803	0.021972	0.006141	-0.031212	-
Yanghwa river	-0.002884	0.017988	0.003419	-0.011537	-
Bokha river	-0.008345	0.018870	0.004165	-0.033380	-
Heck river	-0.006458	0.017875	0.004555	-0.025834	-
Gyeongan river	-0.003850	0.017476	0.004701	-0.015399	-
Considered all rivers	-0.009145	0.016943	0.004836	-0.036580	

지점과의 수직거리가 짧고 인근의 위치한 경안천 유입 유량과 비교했을 경우 유량 규모가 더 크기 때문에 이에 대한 영향으로 판단된다.

4. 결론

본 연구에서는 다양한 수문조건에 따른 해당 유역의

복잡성과 정밀성이 재현된 과거와 현재의 일단위 수질을 최대한 확보하고자 물리적 기반 모형을 활용하였으며 이를 기반으로 신속하고 안정된 데이터 기반 모형을 구축하기에 앞서 Gamma Test 기법을 적용하여 최적의 입력자료 조합을 선정하였다.

팔당댐 2지점의 효율적인 수질 예측을 위해 BOD는 섬강, 흑천, 청미천, 북한강의 영향이 가장 크게 나타났고, TOC는 남한강, 섬강, 청미천, 흑천, 경안천, 조종천, 북한강에서 유입되는 영향이 민감한 것으로 분석되었다. T-P의 경우 흑천, 경안천, 북한강, Chl-a는 남한강, 청미천, 조종천, 북한강의 영향이 분석되었다.

본 연구의 결과는 유입하천별 시계열 형태의 수질자료에 대한 불확실성 및 민감도를 사전에 검토하여 반복적인 모의 수행을 지양하고 신뢰도 있는 데이터 기반의 예측 모형을 구축하기 위한 일련의 과정이라고 할 수 있다. 본 연구에서 Gamma Test를 통해 도출한 수질항목별 입력자료 조합으로 데이터 기반 모형을 구축한다면 유입하천별 수질만 있으면 EFDC 모형을 구동하지 않고도 팔당댐 호소 수질의 예측이 가능할 것으로 판단된다.

현재 모델링 관련 연구의 추세는 서로 다른 유형의 모형을 결합하여 최적의 방식을 도출하는 것이다. 이러한 시점에서 본 연구에서 제시한 결과를 토대로 데이터 기반 모형을 구축할 것이며, 이는 물재해 안전성을 강화하는 사전관리 중심의 수질예측 기술을 한 단계 높은 수준으로 진화시킴으로써 보다 빠르고 신뢰도 높은 의사 결정 지원이 가능할 것으로 기대된다.

감사의 글

이 논문은 2019년도 정부(교육부)의 재원으로 한국연구재단 기초연구사업의 지원을 받아 수행된 연구입니다 (2019R1A6A3A01090378).

References

Chang, F.J., Tsai, Y.H., Chen, P.A., Coynel, A., and Vachaud, G. 2015. Modeling water quality in an urban river using hydrological factors-Data driven approaches. *Journal of environmental management* 151: 87-96.

Donigian, A.S. 2000. HSPF Training Workshop Handbook and CD, Lecture #19, Calibration and Verification Issues, Slide #L19-22, EPA Headquarters, Washington Information Center, Presented and prepared for U.S. EPA, Office of

Water, Office of Science and Technology.

Faruk, D.Ö. 2010. A hybrid neural network and ARIMA model for water quality time series prediction. *Engineering applications of artificial intelligence* 23(4): 586-594.

Keum, H.J., Kim, H.I., and Kim, B. 2019. Uncertainty analysis of rainfall scenarios for the prediction of flood disasters in urban areas. *Journal of the Korean Society of Hazard Mitigation* 19(2): 255-264. (in Korean)

Khadr, M. and Elshemy, M. 2017. Data-driven modeling for water quality prediction case study: The drains system associated with Manzala Lake, Egypt. *Ain Shams Engineering Journal* 8(4): 549-557.

Kim, J.K., Lee, S.H., Bang, H.H., and Hwang, S.O. 2009. Characteristics of algae occurrence in Lake Paldang. *Journal of Korean Society of Environmental Engineers* 31(5): 325-331. (in Korean)

Kim, S.E. and Seo, I.W. 2015. Artificial neural network ensemble modeling with exploratory factor analysis for streamflow forecasting. *Journal of Hydroinformatics* 17(4): 614-639.

Kim, Y.S. and Lee, E.J. 2019. Establishment of Target Water Quality for TOC of Total Water Load Management System. *Journal of Korean Society on Water Environment* 35(6): 520-538. (in Korean)

Lee, E. and Kim, T. 2021. Predicting BOD under Various Hydrological Conditions in the Dongjin River Basin Using Physics-Based and Data-Driven Models. *Water* 13(10): 1383.

Lee, E.J. 2013. Application of total water load management system using watershed model and load duration curves. Doctor's thesis, Cheongju University, Choengju, Korea. (in Korean)

Lee, E.J., Kim, T.G., and Keum, H.J. 2018. Application of FDC and LDC using HSPF model to support total water load management system. *Journal of Korean Society on Water Environment* 34(1): 33-45. (in Korean)

Ministry Of Environment (MOE). 2015. Nam River Water Management Big Data Analysis to Prepare Water Quality Improvement Plan. (in Korean)

Najah, A., El-Shafie, A., Karim, O.A., and El-Shafie, A.H. 2013. Application of artificial neural networks for water quality prediction. *Neural Computing and Applications* 22(1): 187-201.

National Disaster Management Research Institute (NDMI), 2018, Developing an Analytic Framework for Real-time Inundated Hazard Area. (in Korean)

Niknia, N., Moghaddam, H.K., Banaei, S.M., Podeh, H.T., Omidinasab, F., and Yazdi, A.A. 2014. Application of gamma test and neuro-fuzzy models in uncertainty analysis for prediction of pipeline scouring depth. *Journal of Water Resource and Protection* 6(05): 514.

Oh, C.S. 2016. Notes on applying the EFDC model for simulating the water quality of Saemangeum watershed.

- Korean National Committee on Irrigation and Drainage 55: 42-48. (in Korean)
- Park, J.D. and Oh, S.Y. 2012. Methodology for the identification of impaired waters using LDC for the management of total maximum daily loads. *Journal of Korean Society on Water Environment* 28(5): 693-703. (in Korean)
- Poul, S., Manguerra, H., and Slawewski, T. 2019. A Watershed Management Perspective, Digest of water industry and business of Korean society on water environment. *Big Data Anal* 20-23.
- Rauf, A., Ahmed, S., Ghumman, A.R., Ahmad, I., Khan, K.A., and Ahsan, M. 2016. Data Driven Modelling for Real-time Flood Forecasting. In *Proceedings of the 2nd International Multi-Disciplinary Conference, Gujrat, Pakistan* 19-20.
- Sarkar, A. and Pandey, P. 2015. River water quality modelling using artificial neural network technique. *Aquatic procedia*, 4: 1070-1077.
- Seo, I.W. and Choi, N.J. 2007. Water Quality Modeling and Water Quality Forecasting. *Environmental Engineering Research* 55(1): 57-84. (in Korean)
- Stefánsson, A., Končar, N., and Jones, A.J. 1997. A note on the gamma test. *Neural Computing & Applications* 5(3): 131-133.
- Tetra Tech, Inc. 2007. *The Environmental Fluid Dynamics Code User Manual US EPA Version 1.01*.
- U.S.EPA. 2001. *HSPF User's Manual*.