인공위성 원격 탐사 정보가 자료 기반 모형의 미계측 유역 하천유출 예측성능에 미치는 영향 분석

서지유* • 정하은* • 원정은* • 최시중** • 김상단*+

*부경대학교 지구환경시스템과학부 환경공학전공 **한국건설기술연구원 수자원하천연구본부

Analysis of the Impact of Satellite Remote Sensing Information on the Prediction Performance of Ungauged Basin Stream Flow Using Data-driven Models

Seo, Jiyu^{*} · Jung, Haeun^{*} · Won, Jeongeun^{*} · Choi, Sijung^{**} · Kim, Sangdan^{*†}

^{*}Division of Earth Environmental System Science (Major of Environmental Engineering), Pukyong National University, Busan, Korea ^{**}Department of Hydro Science and Engineering Research, Korea Institute of Civil Engineering and Building Technology (KICT), Goyang, KoreaKorea

(Received : 4 April 2024, Revised : 6 May 2024, Accepted : 6 May 2024)

요 약

부족한 하천유출 관측 데이터는 모델 보정 작업을 어렵게 만들어 모델의 성능 향상을 제한한다. 위성 기반 원격탐사 자료는 수문 관련 데이터의 확보에 적극적으로 활용될 수 있으므로 새로운 대안이 될 수 있다. 최근에는 여러 연구를 통하여 기존의 개념적/물리적 모델보다는 인공지능을 이용한 해법이 더 적절하다는 평가를 받고 있다. 본 연구에서는 다양한 순환 신경망들과 의사결정나무 기반 알고리즘들을 결합한 자료 기반 접근 방식을 제안하였다. 또한 인공지능 학습을 위하여 인공위성 원격탐사 정보의 활용성을 조사하였다. 본 연구에서 위성영상은 MODIS와 SMAP의 자료가 사용된다. 공적으로 공개된 25개 유역의 자료를 사용하여 제안된 접근 방식을 검증하였다. 전통적인 지역화 접근법에 서 착안하여 모든 유역의 자료를 통합하여 하나의 자료 기반 모델을 학습하는 전략을 채택하였으며, Leave-one-out cross-validation 지역화 설정을 이용하여 하나의 모델이 다양한 유역의 하천유출을 예측함으로써 제안된 접근 방식 의 잠재력을 평가하였다. GRU + Light GBM 모델이 대상 유역에 적합한 모델 조합으로 판명되었으며(25개 미계측 유역 일 하천유량 예측 모형효율계수 평균 0.7187) 하천유출이 매우 작은 시기를 제외하면 우수한 미계측 유역의 하 천유출 예측 성능을 보여주었다. 인공위성 원격탐사 정보의 영향력은 최대 10% 정도로 파악되었으며, 위성 정보의 추가 적용이 풍수기 또는 평수기보다는 저수기 또는 갈수기의 하천유출 예측에 더 큰 영향을 미쳤다. 핵심용어 : 인공지능 모형, 자료 기반 모형, 인공위성 원격탐사 정보, 미계측 유역 예측

*To whom correspondence should be addressed.

Division of Earth Environmental System Science (Major of Environmental Engineering), Pukyong National University, Busan, Korea E-mail : skim@pknu.ac.kr

[•] Ji Yu Seo Division of Earth Environmental System Science (Major of Environmental Engineering), Pukyong National University, Busan, Korea/Ph.D. Student(gu426@naver.com)

[•] Ha Eun Jung Division of Earth Environmental System Science (Major of Environmental Engineering), Pukyong National University, Busan, Korea/Ph.D. Student(q0881@naver.com)

[•] Jeong Eun Won Division of Earth Environmental System Science (Major of Environmental Engineering), Pukyong National University, Busan, Korea/Research Fellow(wjddms8960@naver.com)

[•] Si Jung Choi Department of Hydro Science and Engineering Research, Korea Institute of Civil Engineering and Building Technology (KICT), Goyang, Korea/Senior Researcher(sjchoi@kict.re.kr)

[•] Sang Dan Kim Division of Earth Environmental System Science (Major of Environmental Engineering), Pukyong National University, Busan, Korea/Professor(skim@pknu.ac.kr)

Abstract

Lack of streamflow observations makes model calibration difficult and limits model performance improvement. Satellite-based remote sensing products offer a new alternative as they can be actively utilized to obtain hydrological data. Recently, several studies have shown that artificial intelligence-based solutions are more appropriate than traditional conceptual and physical models. In this study, a data-driven approach combining various recurrent neural networks and decision tree-based algorithms is proposed, and the utilization of satellite remote sensing information for AI training is investigated. The satellite imagery used in this study is from MODIS and SMAP. The proposed approach is validated using publicly available data from 25 watersheds. Inspired by the traditional regionalization approach, a strategy is adopted to learn one data-driven model by integrating data from all basins, and the potential of the proposed approach is evaluated by using a leave-one-out cross-validation regionalization setting to predict streamflow from different basins with one model. The GRU + Light GBM model was found to be a suitable model combination for target basins and showed good streamflow in 25 ungauged basins is 0.7187) except for the period when streamflow is very small. The influence of satellite remote sensing information was found to be up to 10%, with the additional application of satellite information having a greater impact on streamflow prediction during low or dry seasons than during wet or normal seasons.

Key words : Artificial intelligence model, Data-driven model, Prediction of ungauged basin, Satellite remote sensing information

1. 서 론

하천유출 관측 데이터의 부족은 세계적으로 큰 주목을 받 고 있다. Global Runoff Data Center의 지도 추적 하천유 출 자료(Global Runoff Data Center, 2020)를 살펴보면 매우 적은 수의 관측소가 있는 광대한 지역을 발견할 수 있 다. 특히 중국, 인도, 중앙아메리카, 남아메리카, 오세아니 아, 아프리카와 같은 지역에서는 일 유출 관측치가 기록되 더라도 공개적으로 접근할 수 있는 경우는 거의 없다. 이와 같은 자료 부족으로 인해 수문 모델의 보정에 큰 어려움이 발생하고 있다. 현재 대부분의 하천유출 예측은 과거 유출 기록을 기반으로 광범위한 매개변수 조정이 필요한 수문 모델을 기반으로 하고 있으나(Choi et al., 2021; Feng et al., 2021), Worldbank (2018)에 따르면 심지어 산업화한 고소득 국가에서도 수문 관측소의 수가 감소하고 있다. 미 국에서는 30년 이상의 유출기록을 가진 2,500개 이상의 하 천 관측소가 중단되었다고 보고되었다(USGS, 2018), 부족 한 관측 데이터는 모델 보정 작업을 어렵게 만들어 모델의 성능 향상을 제한한다.

이러한 도전적인 상황에서 위성 기반 원격탐사 자료는 수문 관련 데이터를 확보하는 데 적극적으로 활용될 수 있음에 따라(Fekete et al., 2015), 이를 이용한 미계측 유역에서의 하 천유출 예측은 수문학 분야의 중요한 과제 중 하나가 되고 있 다(Hrachowitz et al., 2013). 위성으로 관측된 토양수분, 식생, 및/또는 증발산 자료의 활용은 PUB에 잠재적으로 유 용할 것으로 예상된다. 예를 들어 위성 관측된 토양수분은 자료 동화 접근법과 결합하여 하천유출 모의 개선을 위해 사용되었으며(Yan and Moradkhani, 2016), 최근 다른 연 구에서는 수문 모델의 보정에도 이용된 바 있다(Choi et al., 2021). 실제증발산은 일반적으로 계측 유역에서 하천유 량 기반의 보정 프로세스를 보조하거나(Wambura et al., 2018), 종종 단독으로 수문 모형의 보정에 사용되어 하천유량 의 예측 성능개선 여부가 실험되었다(Kunnath- Poovakka et al., 2021). 반면, 위성 관측으로부터 산정된 다양한 식생지 수는 생태 수문 모형과 결합하여 입력 매개변수 또는 보정 을 위한 관측된 자료로 사용되었다(Choi et al., 2023). 자료 기반 접근 방식은 특별한 가정 없이 자료에서 직접

기상 입력과 하천유출 출력 사이의 관계를 정량화한다. 최 근 Hochreiter and Schmidhuber(1997)가 개발한 LSTM (Long-Short Term Memory)을 이용하면, 미국 전역의 하 천유출 예측에서 기존의 개념적/물리적 모델을 능가하는 높은 성능을 달성할 수 있음이 보고되었다(Fang et al., 2020; Gauch et al., 2021; Lees et al., 2021, Choi et al., 2022b). 학습에 사용되는 유역의 다양성이 클수록 인공지 능이 기상 입력과 하천유출 사이의 관계를 더 일반적으로 학습할 수 있다는 관점에서 자료 기반 접근 방식은 미계측 유역의 하천유출 예측을 위한 좋은 대안이 될 수 있다 (Kratzert et al., 2022). 수문 분야에서 딥러닝을 이용한 미 계측 지역의 예측 연구는 Kratzert et al. (2018)의 연구를 통해 상당한 진전을 이루었으며(Arsenault et al., 2023), 기상 자료와 유역 특성만을 사용하여 미계측 유역의 하천 유출을 예측하는 가능성을 제시하고 있다. Choi et al. (2022a)의 연구에서도 LSTM 지역화 모델이 우리나라의 13개 댐 유역에서 뛰어난 결과를 얻었음을 보고하였다.

지역화를 위한 자료에는 유역의 물리적 속성이 포함될 수 있다. Kratzert et al. (2019)은 기상 자료뿐만 아니라 정적 인 유역 속성을 고려하는 Entity-Aware-LSTM이라는 모 델을 제안하였으나, 우리나라 유역에 관한 연구(Choi et al. 2022a)에서는 유역 속성의 포함이 모델 성능에 큰 역할을 하지 않을 수도 있음이 보고되었다. 다른 한편으로, Won et al. (2023)의 연구에서는 순환 신경망(Recurrent Neural Network, RNN)과 의사결정나무(Decision Tree, DT) 기 반의 기계학습기법을 결합한 모델을 사용하여 유역 속성 정보가 미계측 유역의 예측에 미치는 상대적인 중요도가 5% 내외인 것으로 제시되었다.

앞서 언급한 모든 자료 기반 연구는 모델 개발을 위해 지 상에서 관측한 자료를 사용하였다. 본 연구에서는 인공위성 원격탐사 자료를 사용하여 RNN과 DT를 이용한 인공지능 을 학습하면 얼마만큼의 성능개선을 기대할 수 있을지를 살펴보고자 하였다. 구체적으로 우리나라 25개 유역을 대 상으로 위성 기반 정보를 결합하여 RNN과 DT를 이용한 인공지능 모델을 개발하고, 그 성능을 지상에서 얻을 수 있 는 유역 속성 자료로 개발된 모델과 비교하였다. 이를 위한 인공지능 모델은 RNN과 DT 기반 알고리즘의 통합된 접 근 방식으로 구성된다. 더불어 우리가 비교적 쉽게 구성할 수 있는 RNN 모델들과 DT 기반 알고리즘들의 성능 차이 를 살펴보았다. 일반적으로 많이 사용되는 RNN 모델들 및 DT 기반 알고리즘들의 조합을 테스트 해 봄으로써 우리나 라에 가장 적절한 RNN-DT 결합 모델 조합을 탐색해 보 고자 한다.

2. 자료 및 방법

2.1 대상 유역 및 자료

본 연구에서는 우리나라 25개의 유역을 대상으로 진행되 었으며, 이에 관한 지리적 위치는 Fig. 1에서 확인할 수 있 다. 모델 훈련 및 테스트를 위하여 2016년부터 2020년까지 의 기상 자료와 하천유출 자료가 사용되었다.

기상 자료는 기상청의 기상자료개방포털(https://data.kma.go.kr /cmmn/main.do)을 통해 수집되었다. 각 유역에 영향을 미치 는 기상 관측소는 Fig. 1에서 검은 점으로 표시되어 있으며, 해당 관측소에서 관측된 일-단위 강수량 및 기상 요소(최저 및 최고 기온, 풍속, 이슬점 등)을 수집하였다. 이후에는 티센 망을 활용하여 유역 내 공간 평균 시계열을 생성하였다. 또한, 모델의 입력 변수 중 하나인 잠재증발산은 일일 기상 요소를 기반으로 Penman-Monteith 방법(Allen et al., 1998; Won et al., 2020)을 활용하여 계산하였다.

하천유출 자료는 환경부의 국가수자원관리종합정보시스 템(http://www.wamis.go.kr/)에서 입수하였으며, 연구 대 상 유역에 대한 Curve Number (CN), 포화투수계수 (Ks), 불투수율 (IMP) 등의 토양 및 토지 특성은 농촌진흥청과 환경부에서 제공하는 토양도 및 토지 피복도를 이용하여 획득하였다. 대상 유역에 대한 간략한 정보는 Table 1에 요 약되어 있다.

본 연구에서는 위성 기반 자료를 활용하여 유역의 속성을 정의하였다. 이를 위해 두 가지 자료, 즉 잎면적지수(Leap Area Index, LAI)와 실제증발산(ET)은 Land Processes Distributed Active Archive Centre(https://lpdaac.usgs.gov/) 에서 제공하는 Level-4 MODIS(Moderate Resolution Imaging Spectroradiometer) global LAI(MOD15A2H)와 ET(MOD16A2) 자료를 활용하였다. 이러한 MODIS 자료는 500m의 공간 해상도와 8일의 시간 해상도를 가지고 있다. 본 연구에서는 8일 동안의 자료가 일별로 변화가 없다는 가정을 적용하였 다. MODIS 자료의 경우 실제 관측자료를 이용한 품질관 리가 필요한 때도 있지만, 본 연구에서는 MODIS에서 얻 은 원자료를 그대로 사용하였다.

또한, 토양수분 정보는 National Snow and Ice Data Center Distributed Active Archive Center (https://nsidc.org/data/ data-programs/nsidc-daac)에서 제공하는 SMAP(Soil Moisture Active Passive) L4_SM을 사용하였다. 이 자료는 NASA 유역 지표면 모델로부터 SMAP L-band 밝기 온도 관측을 활용하여



Fig. 1. Study watersheds

ID	Watershed	Area (km²)	CN	K_{s} (mm/d)	IMP	P (mm/yr)	PET (mm/yr)
1	Seomjingam Dam	763	67	133.8	0.0750	1,354	1,006
2	Namgang Dam	2,282	63	156.4	0.0582	1,487	1,077
3	Andong Dam	1,591	59	177.5	0.0579	1,104	1,020
4	Goisan Dam	677	67	132.5	0.0464	1,315	1,043
5	Habcheon Dam	929	57	191.8	0.0629	1,259	1,054
6	Gwangdon Dam	121	69	127.7	0.0362	1,250	934
7	Unmun Dam	302	67	134.4	0.0516	1,155	1,136
8	Yeoju Wonbugyo	520	62	179.9	0.0807	1,188	1,048
9	Daejeon Wonchongyo	609	60	176.9	0.1312	1,295	1,023
10	Okcheon Sangyeri	491	63	168.1	0.0564	1,362	1,023
11	Hwasun Sinseonggyo	411	62	153.1	0.0646	1,403	1,032
12	Namyangju Jingwangyo	202	60	178.3	0.0878	1,219	1,045
13	Yesan Chunguidaegyo	221	70	129.1	0.0919	1,235	1,088
14	Boryeong Dam	162	56	187.1	0.0729	1,065	952
15	Hampyeong Hakyagyo	115	72	115.8	0.0879	1,034	895
16	Yeongweol Buksangri	1,616	60	171.4	0.0417	1,151	1,024
17	Mungyeong Gimyongri	612	62	162.7	0.0473	1,328	1,129
18	Buyeo Jicheongyo	209	62	161.7	0.0571	1,228	992
19	Cheongju Heungdeokgyo	168	66	158.9	0.1220	1,186	1,079
20	Jangheung Gamcheongyo	152	65	153.6	0.0707	1,399	1,026
21	Yongdam Dam	930	61	174.6	0.0784	1,449	1,010
22	Hoiseong Dam	208	52	216.7	0.0458	1,159	1,025
23	Buyeo Seokdonggyo	156	68	122.4	0.0768	1,247	1,000
24	Soyangang Dam	2,694	51	208.0	0.0586	1,231	1,038
25	Chungju Dam	6,661	62	156.5	0.0491	1,205	1,040

Table 1. Hydro-meteorological information for study watersheds.

생성된 토양수분 자료로, 지표면부터 지표면 아래 1,000mm 까지의 정규화된 토양수분 자료(표면 및 뿌리 영역 토양수분 에 대한 전체 추정치)를 제공한다. 이 자료의 시간 해상도는 3시간이며, 공간 해상도는 9km이다. 이와 같은 3시간 간격의 자료를 본 연구에서는 일별 자료로 평균화하여 활용하였다. 끝으로, 모든 위성 자료는 해당 유역의 공간 평균 자료로 재구성되었다.

2.2 방법

우리 연구의 주된 목적은 자료 기반 모델의 PUB(Predictions of Ungauged Basins) 성능을 테스트하는 것이다. 이때, 미계 측 유역의 경우에는 관측자료가 없으므로 그 유역에 대한 맞춤형 학습이 불가능하다. 프로세스 기반의 모델을 이용할 때는 인근 계측 유역에서 추정된 모델 매개변수를 미계측 지역으로 전이하는 지역화 접근법이 적용된다. 그러나 자료 기반 모델은 학습 과정에서 특정 계측 유역의 정보만을 이용할 경우, 데이터 부족에 따른 overfitting과 out-of-distribution prediction에 직면할 수 있다. 이러한 문제들은 학습할 자 료의 양이 작거나 예측자료와 학습자료의 패턴이 크게 다 를 때 발생한다. 특히, RNN 기반의 모델은 순수하게 학습 자료로부터 입력과 출력 시퀀스 사이의 관계를 학습한다. 학습자료와 예측자료가 유사한 수문학적 변동성을 가질 때 RNN은 의미 있는 예측을 제공할 수 있지만, 그렇지 않을 때는 경험한 적이 없는 패턴을 올바르게 추정할 수 없다 (Choi et al., 2022a). 관측된 데이터를 활용하여 수문 프로 세스를 학습하는 자료 기반 모델의 특성으로 인하여, 많은 양의 훈련 데이터는 모델이 더 일반적이고 추상적인 패턴 을 학습하는 데 도움이 된다(Schmidhuber, 2015; Lee et al., 2021). 모델이 강우-유출 프로세스를 더 명확하게 학 습하기 위해서는 가능한 많은 유역의 자료를 활용하는 것 이 바람직하므로, 본 연구에서는 사용할 수 있는 모든 유역 의 자료를 통합하여 하나의 자료 기반 모델을 학습하는 전 략을 채택하였다. 이는 적절한 수문학적 모델을 이용하여 미계측 유역의 하천유출을 예측하는 전통적인 지역화 접근 법의 자료 기반 버전이라고 정의할 수 있다.

실험적으로는 25개의 유역을 대상으로 Leave-one-out Cross-Validation (LOOCV) 방법을 적용하였다. 이 방법 은 특정한 1개 유역을 미계측 유역으로 가정하고, 나머지 24개 유역의 자료를 사용하여 모델을 학습한다. 그다음, 학 습된 모델을 이용하여 가정된 미계측 유역의 하천유출을 예측하고 예측성능을 평가한다. 이러한 프로세스는 모든 25개 유역이 테스트될 때까지 반복된다. 결과적으로, 25개 의 모델이 생성되며, 각 미계측 유역에 대한 예측성능을 평 가하는 데 사용된다. 이 방법을 통해 자료 기반 모델의 미 계측유역 예측성능을 신뢰성 있게 평가할 수 있다.

2.3 학습 전략

본 연구에서는 자료 기반 모델을 통한 미계측 유역에서 예측성능 향상과 원격 탐사 정보의 영향을 살펴보기 위해 두 가지 학습 전략을 적용하여 비교 및 분석하고자 하였다. 1) scheme MG; 2) scheme MGRS.

scheme MG에서는 기상 자료와 유역의 지형 정보를 활 용한다. 사용되는 입력 속성에는 유역 면적, curve number, 포화투수계수, 불투수면적 비율 등이 포함된다. 이 scheme 은 기상 데이터와 지형 정보만으로 모델을 학습하여 하천 유출을 예측한다. scheme MGRS에서는 scheme MG의 학 습 과정에 미계측 유역에서 유출 예측성능 향상에 도움이 될 것으로 예상되는 원격 탐사 정보를 추가로 사용한다.

모델은 성능에 대한 원격 탐사 정보의 효과를 명확하게 식별하기 위하여 RNN과 DT 기반 알고리즘을 결합한 2단 계로 구성된다. 첫 번째 단계인 RNN 모델은 시간적 패턴 을 학습하기 위해 사용된다. 유역 *i*에서 *t*-day의 하천유출 자료를 예측 대상(target)으로 설정하고, $I_{t-1}^{i}, I_{t-2}^{i}, \cdots$, I_{t-365}^{i} 을 입력자료로 하는 네트워크를 구성하였다. 즉, $Q_t = f(I_{t-1}, I_{t-2}, \cdots, I_{t-365})$. RNN 모델의 경우, scheme MG에서는 기상 자료만이 입력자료로 사용되며, scheme MGRS에서는 기상 자료와 원격 탐사 자료가 동시 에 입력자료로 사용된다. 두 번째 단계에서는 입력 속성들 사이의 속성 중요도를 표출할 수 있는 DT 기반 모형이 사 용되며, 기상 정보, 지형 정보, 1차 예측 유량 정보, 원격 탐사 정보의 상대적인 영향을 명시적으로 식별할 수 있게 된다. 하천유출의 시간적인 패턴만 훈련함으로써 발생하는 오류를 줄이기 위해, t-day의 기상 자료와 1단계 모형의 출력을 이용하여 t-day의 유출자료를 예측하는 DT 기반 모형이 구성된다. Scheme MG에서는 유역의 지형 정보가 추가 입력 속성으로 사용되며, Scheme MGRS에서는 지형 정보와 원격 탐사 정보가 함께 활용된다.

최근에 다양한 학습 알고리즘들이 제안되어 실제 적용되 고 있으므로 수많은 유망한 기계 학습 또는 딥러닝 기법 중 에서 어떤 것이 우리 자료에 더 적합할 것인지를 식별할 필 요가 있다. 본 연구에서는 1단계 RNN 모형화 단계에서는 LSTM과 GRU(Gated Recurrent Unit)를 각각 선택하였으 며, 2단계 DT 기반 모형화 단계에서는 Random Forest, XGBoost, LightGBM(Light Gradient Boosting Machine) 을 선택하여 각 모형 조합에 따른 성능을 테스트하였다. 즉, 2×3 = 6개의 모형 조합에 대한 성능을 살펴보았다. 본 연구에서 수행된 전체적인 모형화 작업 흐름도를 Fig. 2에 나타내었다. scheme M의 LSTM+RF 조합을 예시로 다시 설명하면, 아래와 같다.

- 1) 1단계: 365일 전부터 하루 전까지의 기상 자료를 이 용하여 당일의 하천유량을 예측하는 LSTM 모델
- 2) 2단계: 1단계에서 예측된 하천유량과 지형자료, 당일의 기상 자료를 이용하여 당일의 하천유량 최종 예측

본 연구에는 Python 3.9를 주요 프로그래밍 언어로 사용하 였다. 딥러닝 프레임워크로는 Tensorflow 2.9.1을, 기계 학습 을 위한 라이브러리로는 Scikit-learn 1.2.1, Xgboost 1.7.4, Lightgbm 3.3.5를 활용하였다. 1단계 RNN 모형에서는 총 2개의 은닉층과 각 층의 셀/은닉 상태 길이가 30인 네트워크 로 구성되었다. 과적합을 방지하기 위해 30%의 드롭아웃을 설정하였으며, validation_split = 0.25 상태에서 에포크가 10 번째 지속될 때까지 학습 개선 효과가 나타나지 않는 경우 학습을 조기 종료시켰다. 배치 당 표본 수는 512로 설정했고, 최적값을 찾아가는 optimizer로 adam을 사용하였다. 손실 함 수로는 평균 제곱 오차(Mean Squared Error, MSE)를 사용하 였다. 2단계 DT기반 모형에서는 5-fold cross-validation 기



Fig. 2. Modeling workflow diagram



(a) Box-plot for meteorological characteristics, where P(mm) is annual precipitation, PET(mm) is potential evapotranspiration, and AET(mm) is actual evapotranspiration



(b) Box-plot for geographical characteristics, where Area(mm) is watershed area, CN is curve number, Ks is saturation hydraulic conductivity, IMP is imperviousness ratio

Fig. 3. Meteorological and geographical characteristics for 25 watersheds.



Fig. 4. Remote sensing information box-plot for 25 watersheds, where SM is daily soil moisture and LAI is leap area index.

법을 적용하여 모형을 평가하였다. 이외에 모든 하이퍼 파라 미터들은 기본값이 적용되었다. 따라서 하이퍼 파라미터를 보정하여 각 모형에 적용한다면 더 좋은 예측 성능을 기대해 볼 수 있으므로, 본 연구의 결과는 우리가 얻을 수 있는 가장 보수적인 결과일 것이다.

3. 결과 및 고찰

3.1 학습 전략 비교 및 최적 모형 조합 선정

본 연구에서는 지역화 접근법에서 착안하여 미계측 유역 의 하천유량을 예측하기 위해 다수의 계측 유역의 정보를 통합하여 RNN과 DT 기반 알고리즘의 결합 모형을 학습 하였다. 본 절에서는 scheme MG과 scheme MGRS의 성 능 차이 및 대상 유역에 적합한 최적 모형을 살펴보고자 하 였다. 먼저, 적용된 25개 유역의 연평균 기상 자료 및 지형 자료의 box-plot을 Fig. 3에, 토양수분 및 LAI 자료의 box-plot을 Fig. 4에 도시하였다. 연평균 강수량, curve number, 포화투수계수, 불투수율의 분포는 적절하게 흩어 져 있으나, 연평균 잠재증발산량과 유역 면적의 공간적인 변동성은 상대적으로 작게 분포되어 있음을 발견할 수 있 다. 토양수분과 잎면적지수의 분포는 유역별로 차이가 있음 을 살펴볼 수 있다. 인공지능 모형은 제공된 자료로부터 수 문 현상을 학습하므로, Figure 3과 4에 제시된 분포에서 유 의미하게 벗어나는 특성을 가진 미계측 유역에 적용할 때 는 많은 주의가 필요하다고 말할 수 있다.

Fig. 5는 각각의 학습 전략 및 모형 조합에 의해 도출된 25개 미계측 유역의 하천유출의 예측성능을 보여주고 있다. 각 패널에서 한 개의 점은 특정 모형에서 도출된 하나의 미계측 유역에 대한 예측성능을 의미한다. 유역에 의존하여 편차는 존재하지만, 모든 모형에서 약간이나마 scheme MG의 성능보다는 scheme MGRS의 성능이 더 우수한 경향을 발견할 수 있다. 적용된 모델 조합에 따라 다양하지

만, 모델 평균적으로 NSE를 기준으로 정량적으로 살펴보 면, 25개 유역 중 16개 유역에서 위성 자료를 도입함에 따 른 성능개선 효과를 얻을 수 있었으며, 최대 16% 이상의 성능개선(Watershed 10)이 이루어진 것을 살펴볼 수 있었 다. 참고로 위성 자료를 도입함에 따라 성능이 하락한 나머 지 9개 유역의 경우에는 성능 하락 정도가 최대 7% 미만 이다(Watershed 20). 즉, 이러한 결과는 위성 정보를 예측 모형에 추가하는 것이 기상 자료 및 지형자료만을 이용하 는 것보다는 미계측 유역의 예측 성능을 높이는 데 약간이 나마 긍정적인 기여가 있음을 의미한다. 여기서 R2는 결정 계수이며, NSE는 Nash - Sutcliffe 모델 효율계수이다(Nash and Sutcliffe, 1970). 참고로 NSE가 0.5 이상이면 좋은 성능 이라고 볼 수 있다(Ritter and Munoz-Carpena, 2013; Choi and Kim, 2021).

다음으로 25개의 미계측 유역에서 scheme MGRS에 의해 학습된 모델들로 예측된 하천유출의 성능을 좀 더 상세히 살펴 보았다. 모든 미계측 유역들 및 모델들의 결과로부터 R²는 0.4328(GRU+XGBoost at Watershed 19)~0.9331 (GRU+ LightGBM at Watershed 1)의 분포를 보였으며, 25개 미계 측 유역들에 대한 평균 R2는 0.7148(LSTM+ XGBoost)에서 0.7518(GRU+LightGBM)의 범위를 나타냈다. NSE의 경우, 가장 낮은 성능(NSE = 0.3810)은 LSTM+XGBoost를 적용 한 Watershed 10에서 발견되었으며, 가장 높은 성능(NSE = 0.9192)은 LSTM+XGBoost를 적용한 Watershed 5에서 기 록되었다. 25개 미계측 유역들에 대한 평균 NSE는 0.6775 (LSTM+XGBoost)에서 0.7187(GRU+LightGBM)의 범위를 나타냈다. 모델들의 NSE 성능을 비교를 위하여 scheme MG 와 함께 Fig. 6에 나타내었다.

모델 조합 평균적으로 볼 때, 25개 미계측 유역들에 대한 NSE의 범위는 0.4583(Watershed 7)~0.8983(Watershed 5) 이었으며, 평균 NSE는 0.7005이었다. 25개 미계측 유역 중 17개 유역(68%)의 NSE가 0.65 이상이었으며, 25개 유역 중 92%에서 최소한의 예측 정확도인 0.5 이상의 NSE를 얻을



Fig. 5. Prediction performance of streamflow in 25 ungauged basins derived by learning strategy and model combination.





Fig. 6. Box-plot for prediction performance (NSE) of scheme MG and MGRS.



Fig. 7. Average prediction performance of streamflow by learning strategy and model combination.

수 있었다. 가장 나쁜 예측 성능을 보인 모델은 GRU+ XGBoost이었다. NSE는 0.2840~0.8972의 범위를 보였으며, 평균 NSE는 0.6672이었다. 적용 유역들의 56%에서 0.65이 상의 NSE을 얻었으며, NSE가 0.5이상인 유역들의 수는 22개 이었다. GRU+XGBoost의 경우, Watershed 19에서 특히 좋 은 않은 NSE를 기록하였다. 또한 Watershed 7은 모든 모델에 서 성능이 만족스럽지는 않았다. 흥미로운 사실은 LSTM+ XGBoost에서 가장 낮은 성능과 가장 높은 성능이 동시에 발견되었다는 것이다. 이러한 사실은 모델 성능의 우수성을 살펴보고자 할 때는 가능한 많은 유역에서 테스트해야 할 필 요성을 말해주는 것이다.

각각의 학습 전략 및 모형 조합에 대한 25개 미계측 유역 의 평균적인 예측 성능은 Figure 7과 같다. 2개의 학습 전 략 모두 GRU+LightGBM 모형 조합이 우수한 예측 성능 을 보여주었으며, 가장 예측 성능이 높은 것은 scheme MGRS(R² = 0.7518, NSE = 0.7187)로 나타났으며, 성능

한국습지학회지 제26권 제2호, 2024

지표에 기반한다면 GRU+LightGBM 모형 조합이 대상 유 역에 가장 적절한 최적 조합으로 볼 수 있다.

Fig. 8은 scheme MGRS의 학습 전략에 의해 훈련된 GRU + Light GBM 결합 모델의 결과를 보여주고 있다. Watershed 1, 7, 22를 각각 하나씩 미계측 유역으로 가정한 후, 해당 유역을 제외한 나머지 유역들의 기상 자료와 하천 유출 자료가 모델의 학습자료로 사용되었다. Watershed 1을 미계측 유역으로 간주한 결과는 scheme MGRS의 GRU + Light GBM 모델에서 가장 우수한 성능을 보여주는 경우이 며(R² = 0.9311, NSE = 0.8761), Watershed 7의 결과는 가 장 열등한 성능을 보여주는 경우이다(R² = 0.5545, NSE = 0.4669). Watershed 22의 결과는 scheme MGRS의 GRU + Light GBM 모델을 25개의 미계측 유역에 적용한 결과 중에 서 평균적인 성능을 보여주는 경우이다(R² = 0.7217, NSE = 0.7206). 하천유출이 매우 작은 시기를 제외하면 우수한 미계측 유역의 예측성능을 보여주고 있다.



(c) Watershed 22

Fig. 8. Prediction results from GRU + LightGBM with scheme MGRS. In the left panel, the red square is the observed streamflow, and the black dotted line is the streamflow predicted by the model.

3.2 위성 원격 탐사 정보의 영향 분석

위성 정보가 미계측 유역의 예측에 미치는 영향을 좀 더 살펴보기 위하여, DT 기반 알고리즘으로부터 특성 중요도 를 추출하였다. 특성 중요도는 관심이 있는 특성(즉, 입력 자료)에만 무작위로 노이즈를 주고 예측하였을 때 성능 평 가지표가 얼마나 감소하는지를 측정하여 정량화된다. 정확 도가 떨어질수록 예측을 위해 해당 특성이 더 중요하므로, mean decrease accuracy(MDA)라고 불린다. 본 연구에서 MDA는 Python의 Scikit-learn 패키지에서 permutation_ importance method를 이용하여 산출하였으며, 그 결과를 정규화하여 Fig. 9에 나타내었다.

모형에 따른 편차 또는 유역에 따른 편차가 거의 없음을 발견할 수 있으며, 이러한 결과는 미계측 유역의 예측에 미 치는 입력자료들 사이의 상대적인 영향력이 모형 및 유역 에 따라 큰 차이가 없음을 의미한다. 당일의 하천유출을 예 측하고자 할 때, RNN을 이용하여 선행 기상 자료로부터 1 차 예측된 하천유출이 가장 큰 영향을 미치고 있음을 살펴 볼 수 있다. 이는 당일 하천유출의 적어도 60% 이상은 선 행 기상 자료로부터 설명이 가능하다는 사실을 말해주는 것이다. 또한 당일 기상 자료의 상대적인 영향력은 약 20%

정도이며, 유역의 지형자료가 미치는 상대적인 중요도는 5% 내외인 것으로 나타났다. 위성 정보의 영향력은 최대 10% 정도로 파악되었다. 지형자료의 낮은 상대적 중요도 는 본 연구에서 적용한 대상 유역들 사이의 지형적 특성이 전반적으로 큰 이질성이 없기 때문으로 분석될 수 있다. 지 형적인 특성이 더 다양한 유역들을 포함한다면 Fig. 9와 유 사한 특성 중요도가 나타나지 않을 개연성이 매우 클 것이 다. 그러나, 뒤집어서 생각한다면, 지형적 특성이 너무 이질 적인 유역들을 한 바구니에 넣고 모형을 학습한다면 그 모 형들의 정확도가 본 연구에서와 같은 수준으로 유지될 수 있을지는 장담하기 어렵다. 따라서 Fig. 9와 같은 특성 중 요도의 결과는 Fig. 3에서 포괄하고 있는 범위의 기상학적 지형적 특성을 보이는 미계측 유역에 적용할 때 Fig. 8에서 보여주고 있는 수준의 예측성능을 기대할 수 있을 것이라 는 사실을 말해주는 것이라고 해석할 수 있을 것이다. 참고 로 지형 특성 중에서는 불투수율의 상대적인 중요도가 가 장 높음을 살펴볼 수 있다. 세 개의 위성 정보(실제증발산, 토양수분, 잎면적지수) 중에서는 토양수분의 상대적인 중요 도가 가장 높게 나타났다. 이는 미계측 지역의 유출 예측을 위해서는 지형자료와 더불어서 토양수분 정보의 정확한 획 득이 중요한 역할을 할 것이라는 추론을 가능하게 한다.



Fig. 9. Normalized MDA results by input data characteristics of model combination. P is the daily precipitation, PET is the daily potential evapotranspiration, AET is the daily actual evapotranspiration, SM is the daily soil moisture, LAI is the daily leap area index, Area is the watershed area, CN is the curve number, Ks is the saturated hydraulic conductivity, IMP is imperviousness, and Q* is the streamflow first predicted by RNN, M is the sum of the normalized MDA results for P and PET, which means the importance of meteorological characteristics, R is the sum of the normalized MDA results for the remote sensing information characteristics (AET, SM, and LAI) and G is the sum of the normalized MDA results for the geographical characteristics (Area, CN, Ks, and IMP).

scheme MG와 scheme MGRS의 예측성능을 더 상세하게 살펴보기 위하여 관측 유황곡선(flow duration curve, FDC) 과 예측된 FDC를 살펴보았다(Fig. 10). 하천유출의 초과확 률을 설명하는 FDC는 유역의 수문 프로세스를 나타내는 중 요한 지표 중 하나이다(Ruiz-Perez et al., 2016; Choi et al., 2020). 우리의 분석을 강화하기 위해 관측 및 예측된 FDC를 3개의 확률 구간으로 구분한 후(see Table 2), 구분 된 수문학적 조건에서 예측성능을 비교하였다(Yilmaz et al., 2008). 세 구간의 수문학적 조건 중에서 저유량 구간(Table 2에서 segment LFC)에 해당하는 하천유출의 percent Bias (pBias)를 계산하여 Fig. 11에 제시하였다.

scheme MG과 scheme MGRS 모두 segment HFC 및 segment MFC에 해당하는 수문학적 조건에서는 관측 FDC 를 상대적으로 잘 재현하고 있으나, segment LFC에 해당하 는 구간에서는 하천유출을 과대평가하고 있음을 발견할 수 있다(즉, pBias > 0). segment LFC에서 유역의 수문순환 프 로세스는 강수량보다는 증발산과 지표하 흐름에 의해 주도 된다. Fig. 9에서 살펴보았듯이, 잠재증발산과 포화투수계수 의 상대적 중요도가 강수량에 비하여 매우 작으므로, 증발산 과 지표하 흐름에 의해 주도되는 수문순환 프로세스를 모델

Table 2. Segment classification for each range of flow exceedance probabilities.

Segment	Hydrological condition class	Flow exceedance probability range		
HFC (%)	High flow condition	[0, 0.02]		
MFC (%)	Midium flow condition	[0.02 0.70]		
LFC (%)	Low flow condition	[0.70 1]		



Fig. 10. Flow duration curves at Watershed 25.

이 적절하게 표현하는 것은 어려움이 있다. 이러한 결과는 본 연구에서 구성된 모델이 건기에 미계측 유역의 하천유출 을 제대로 구현하기에 한계가 있음을 말해주고 있다. 이에 대한 원인은 다양할 수 있겠으나, 우선 관측자료의 한계도 중요한 원인 중 하나일 것이다. Fig. 8에서 볼 수 있듯이, 관 측된 하천유출 자료 중 작은 값들은 계단 형태의 일정한 값 들로 기록되어있다. 이는 건기의 하천유출 관측자료의 신뢰 성이 높지 않음을 의미한다. 일정한 하천유출 값을 갖는 거 동을 RNN이 학습하고 재현하기에는 많은 어려움이 있는 것으로 보고되고 있다(Kratzert et al., 2018). 그러나 Fig. 10 의 결과는 scheme MG보다는 scheme MGRS가 상대적으로 더 좋은 성능을 나타내고 있음을 말해주고 있다. 이는 위성 정보의 추가 적용이 wet period 또는 normal period보다는 dry period에 하천유출을 예측하는 데 더 크게 기여하고 있 음을 말해준다.



4. 결 론



본 연구에서는 인공위성 원격탐사 정보가 자료 기반 모형 의 미계측 유역 하천유량 예측 성능에 미치는 영향을 분석 하였으며, 이를 위해 순환 신경망과 의사결정나무 기반 알 고리즘을 결합한 모델을 적용하였다. 우리나라 25개 유역 의 기상 자료, 지형자료, 위성 정보 및 하천유출 자료를 이 용하여, 24개 유역의 자료를 통합하여 학습된 모델로부터 나머지 1개의 가정된 미계측 유역의 하천유출을 예측하는 과정을 총 25번의 반복 수행하였다.

선행 기상 자료와 위성 정보로부터 순환 신경망을 이용하 여 당일의 하천유출이 1차 예측된 후, 1차 예측된 하천 유 출자료, 당일의 기상 자료와 위성 정보, 유역의 지형 특성 들이 의사결정나무 기반 알고리즘과 결합하여 하천유출을 최종 예측하였다. 본 연구에서 순환 신경망은 LSTM과 GRU가, 의사결정나무 기반 알고리즘은 Random Forest, XGBoost, and LightGBM가 각각 적용되었다. 이러한 수 치 실험으로부터 위성 정보의 포함이 자료 기반 모델을 이 용한 미계측 유역의 예측성능에 어떤 영향을 미치는 지와 일반적으로 많이 적용되는 딥러닝 및 기계 학습 모델 사이 의 성능 차이가 조사되었다. 연구 결과, GRU + Light GBM 모델이 대상 유역에 적합한 모델 조합으로 판명되었 으며(25개 미계측 유역 일 하천유량 예측 모형효율계수 평 균 0.7187) 하천유출이 매우 작은 시기를 제외하면 우수한 미계측 유역의 하천유출 예측 성능을 보여주었다. 인공위성 원격탐사 정보의 영향력은 최대 10% 정도로 파악되었으 며, 위성 정보의 추가 적용이 풍수기 또는 평수기보다는 저 수기 또는 갈수기의 하천유출 예측에 더 큰 영향을 미쳤다.

다수의 계측 유역들의 정보를 통합하여 학습된 모델로부 터 달성한 미계측 유역의 예측성능은 본 연구에서 적용한 두 개의 학습 전략에 상관없이 적용할 수 있는 임계 수준 이상이었다. 이러한 결과는 자료 기반 모델을 이용하여 미

157

계측 유역의 하천유량을 양호하게 예측할 수 있음을 드러 낸다. 그러나 이러한 결과는 모델 학습 시에 유사한 기후 및 지형 조건을 가진 유역의 자료를 이용하였다는 전제가 있으며, 따라서 학습된 범위 안에 있는 미계측 유역에 적용 하는 것이 바람직할 것이다.

본 연구에서 위성 정보의 추가 활용이 미계측 유역의 예 측성능을 획기적으로 개선하지는 못하였으며, 추가된 위성 정보의 모델 예측 프로세스 안에서의 중요도도 높지 않음 을 인지하였다. 다만, 건기의 하천유출 예측성능 개선에 약 간의 도움이 됨을 확인할 수 있었다. 하지만, 본 연구에서 적용한 순환 신경망과 의사결정나무 기반 알고리즘의 결합 은 미계측 유역의 건기 하천유출을 예측하는데 여전히 한 계를 노출하였다. 건기와 우기에 유역의 수문학적 프로세스 가 달라진다는 사실에 착안한다면, 이후 연구에서는 건기와 우기를 구분하여 학습하는 전략을 마련하는 것도 미계측 유역에서 예측성능 개선에 도움이 될 수 있을 것으로 예상 된다. 또한, 현재 위성 자료를 포함하여 더 많은 대형 샘플 데이터 세트들이 학습을 위해 제공되고 있다. 이러한 기회 는 미계측 유역의 예측 연구에서 자료 기반 모델의 활용을 촉진할 수 있을 것이며 이를 위한 더 많은 연구가 필요하 다. 또한, 자료 기반 모델을 위한 적절한 하이퍼 파라미터 추정에 관한 연구도 추후 진행될 필요가 있을 것이다.

사 사

본 결과물은 환경부의 재원으로 한국환경산업기술원의 수생태계 건강성 확보 기술개발사업의 지원을 받아 연구되 었습니다.(2022003050007)

References

- Allen, R., Pereira, L., Raes, D. and Smith, M. (1998). Crop evapotranspiration–Guidelines for computing crop water requirements–FAO Irrigation and drainage paper 56. Fao, Rome, 300(9), D05109.
- Arsenault, R., Martel, J., Brunet, F., Brissette, F. and Mai, J. (2023) Continuous streamflow prediction in ungauged basins: long short-term memory neural networks clearly outperform traditional hydrological models, Hydrology and Earth System Sciences, 27, pp. 139–157, [DOI https://doi.org/10.5194/hess-27-139-2023]
- Choi, J. and Kim. S. (2021) Conceptual eco-hydrological model reflecting the interaction of climate-soil-vegetationgroundwater table in humid regions, Journal of Korea Water Resources Association, 54, pp. 681–692.
- Choi, J., Kim, U. and Kim, S. (2023) Ecohydrologic model with satellite–based data for predicting streamflow in ungauged basins, Science of The Total Environment, 903, 166617, [DOI https://doi.org/10.1016/j.scitotenv.2023.166617]

Choi, J., Lee, J. and Kim, S. (2022a) Utilization of the Long

Short-Term Memory network for predicting streamflow in ungauged basins in Korea, Ecological Engineering, 182, 106699, [DOI https://doi.org/10.1016/j.ecoleng.2022. 106699.]

- Choi, J., Seo, J., Won, J., Lee, O. and Kim, S. (2020) Effects of hydro-climate conditions on calibrating conceptual hydrologic partitioning model, Journal of Korean Society on Water Environment, 36, pp. 568–580.
- Choi, J., Won, J., Jang, S. and Kim, S. (2022b) Learning enhancement method of Long short-term memory network and its applicability in hydrological time series prediction, Water, 14, 2910, [DOI https://doi.org/10.3390/ w14182910.]
- Choi, J., Won, J., Lee, O. and Kim, S. (2021) Usefulness of global root zone soil moisture product for streamflow prediction of ungauged basins, Remote Sensing, 13(4), 756, [DOI https://doi.org/10.3390/rs13040756.]
- Fang, K., Kifer, D., Lawson, K. and Shen, C. (2020) Evaluating the potential and challenges of an uncertainty quantification method for long short-term memory models for soil moisture predictions, Water Resources Research, 56, e2020WR028095. [DOI https://doi.org/10.1029/ 2020WR028095.]
- Feng, D., Lawson, K. and Shen, C. (2021) Mitigating prediction error of deep learning streamflow models in large data-sparse regions with ensemble modeling and soft data, Geophysical Research Letters, 48, e2021GL092999. [DOI https://doi.org/10.1029/2021GL092999.]
- Fekete, B., Robarts, R., Kumagai, M., Nachtnebel, H., Odada, E. and Zhulidov, A. (2015) Time for in situ renaissance. Science, 349(6249), pp. 685–686.
- Gauch, M., Mai, J. and Lin, J. (2021) The proper care and feeding of CAMELS: How limited training data affects streamflow prediction. Environmental Modelling & Software, 135, 104926. https://doi.org/10.1016/j.envsoft.2020.104926.
- Global Runoff Data Center (GRDC) (2020). BfG–River discharge data–Station maps. Retrieved from https://www.bafg.de/GRDC/EN/02_srvcs/21_tmsrs/st ationMaps.html?nn=201566.
- Hochreiter, S. and Schmidhuber, J. (2017) Long Short–Term Memory, Neural Computation, 9, pp. 1735–1780.
- Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M., Ehret, U. et al. (2013) A decade of predictions in ungauged basins (pub)—a review. Hydrological Sciences Journal, 58, pp. 1198–1255.
- Kratzert, F., Gauch, M., Nearing, G. and Klotz, D. (2022) Neural Hydrology – A Python library for Deep Learning, Journal of Open Source Software, 7, 4050, [DOI https://doi.org/10.21105/joss.04050.]
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K. and Herrnegger,

M. (2018) Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrology and Earth System Sciences, 22, pp. 6005–6022, [DOI https://doi.org/10.5194/hess-22-6005-2018, 2018.]

- Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A., Hochreiter, S. and Nearing, G. (2019) Toward improved predictions in ungauged basins: Exploiting the power of machine learning, Water Resources Research, 55, pp. 11344–11354, [DOI https://doi.org/10.1029/2019WR 026065.]
- Kunnath–Poovakka, A., Ryu, D., Eldho, T. and George, B. (2021) Parameter uncertainty of a hydrologic model calibrated with remotely sensed evapotranspiration and soil Moisture. Journal of Hydrologic Engineering, 26(3), 04020070.
- Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G. and Dadson, S. (2021) Bench marking data-driven rainfall – runoff models in Great Britain: A comparison of long short-term memory (LSTM) – based models with four lumped conceptual models, Hydrology and Earth System Sciences, 25, pp. 5517–5534. [DOI https://doi.org/10.5194/ hess-25-5517-2021.]
- Lee, O., Won, J., Seo, J. and Kim, S. (2021) Evaporative demand drought index forecasting in Busan–Ulsan–Gyeongnam region using machine learning methods, Journal of Korea Water Resources Association, 54, pp. 617–628.
- Nash, J. and Sutcliffe, J. (1970). River flow forecasting through conceptual models part I—A discussion of principles. Journal of hydrology, 10, pp. 282–290.
- Ritter, A. and Munoz–Carpena, R. (2013). Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness–of–fit assessments. Journal of Hydrology, 480, pp. 33–45. [DOI https://doi.org/ 10.1016/j.jhydrol.2012.12.004.]
- Ruiz-Perez, G., Gonzalez-Sanchis, M., Del Campo, A. and Frances, F. (2016). Can a parsimonious model implemented

with satellite data be used for modelling the vegetation dynamics and water cycle in water-controlled environments? Ecological Modelling, 324, pp. 45–53.

- Schmidhuber, J. (2015) Deep learning in neural networks: An overview. Neural networks, 61, pp. 85–117.
- USGS (2018) Streamgage history. Available at https://water. usgs.gov/nsip/history.html.
- Wambura, F., Dietrich, O. and Lischeid, G. (2018) Improving a distributed hydrological model using evapotranspirationrelated boundary conditions as additional constraints in a data-scarce river basin. Hydrological processes, 32(6), 759–775.
- Won, J. and Kim, S. (2020) Future drought analysis using SPI and EDDI to consider climate change in South Korea, Water Supply, 20, pp. 3266–3280, ws2020209, [DOI https://doi.org/ 10.2166/ws.2020.209.]
- Won, J., Seo, J., Lee, J., Choi, J., Park, Y., Lee, O. and Kim, S. (2023) Streamflow predictions in ungauged basins using recurrent neural network and decision tree–based algorithm: application to the southern region of the Korean peninsula, Water, 15, 2485. [DOI https://doi.org/10.3390/ w15132485.]
- Worldbank (2018) Assessment of the state of hydrological services in developing countries. Available at https://www.gfdrr.org/sites/default/files/publication/state-of-hydro logical-services_web.pdf.
- Yan, H. and Moradkhani, H. (2016) Combined assimilation of streamflow and satellite soil moisture with the particle filter and geostatistical modeling. Advances in Water Resources, 94, 364–378.
- Yilmaz, K., Gupta, H. and Wagener, T. (2008) A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resources Research, 44, W09417. [DOI https://doi.org/10.1029/ 2007WR006716.]