

# 미계측 지역 하천 유출량 예측을 위한 앙상블 러닝의 수문학 및 수학적 프레임워크

김동환\* 

육군3사관학교 건설공학과 조교수

## Hydrological and Mathematical Framework of Ensemble Learning Regression for Estimating Discharge in Ungauged River Basins

Donghwan Kim\* 

Assistant Professor, Department of Civil Engineering, Korea Army Academy at Yeongcheon, Yeongcheon, Republic of Korea

**Abstract:** Accurate estimation of river discharge ( $Q$ ) globally is essential for water resources management, including flood control and drought management. However, the number of stream gauges globally used to calculate  $Q$  is decreasing. Additionally, estimating  $Q$  of transboundary rivers or rivers with unique hydrological characteristics, such as the Mekong River, is challenging using traditional hydrological methods. The most representative methods for estimating  $Q$  have been the empirical power function using at-a-station hydraulic geometry (AHG) proposed by Leopold and Maddock (1953) and the method using Manning's equation proposed by Manning (1889). Recently, Kim et al. (2019a; 2019b; 2021) improved the accuracy of  $Q$  estimation in the Congo and Mekong River Basins using ensemble learning regression for estimating  $Q$  (ELQ). However, despite ELQ's superior performance, its mathematical and hydrological framework has not been studied in detail. This review study analyzed relevant papers to understand the mathematical and hydrological significance of ELQ, which differentiates it from existing  $Q$  prediction techniques. We also analyzed cases cited in other international papers. Through this analysis, we expect to draw the contribution of the ELQ method for estimating  $Q$  to remote sensing hydrology domestically and internationally.

**Keywords:** Ensemble learning, ELQ, River discharge, Congo River, Mekong River, Satellite altimetry, Weak learner

**요약:** 전 세계적으로 하천 유출량( $Q$ )을 정확히 예측하는 것은 홍수조절, 가뭄관리 등 수자원 관리에 필수적인 요소이지만, 전 세계적으로  $Q$ 를 산출하고 추정하기 위한 현장 게이지의 숫자는 감소추세에 있다. 또한 공유하천이나 메콩강처럼 특수한 수문학적 특징을 보이는 하천의 유량은 기존 수문학적 방법으로 유출량을 산정하기에는 어려움이 따른다. 전통적으로 하천 유출량을 구하는 대표적인 방법들로 Leopold and Maddock (1953)이 제시한 at-a-station hydraulic geometry (AHG)를 이용한 경험 멱함수(empirical power function)와 Manning (1889)이 제안한 Manning's equation을 이용한 방법 등이 있다. 최근 Kim et al. (2019a; 2019b; 2021)은 앙상블 러닝 회귀법(ensemble learning regression for estimating  $Q$ , ELQ)이라는 머신러닝 기법을 통해 콩고강과 메콩강 일대의 하천 유출량 예측 정확도를 향상시켰다. 그러나 ELQ의 우수성에도 불

Received: January 16, 2025

Revised: February 4, 2025

Accepted: February 5, 2025

Published: February 28, 2025

**Corresponding author:**

Donghwan Kim

E-mail: raintear@kaay.ac.kr

구하고 ELQ의 수학적·수문학적 프레임워크(framework)에 대해 자세히 연구된 바는 없다. 본 리뷰 연구에서는 ELQ 관련 논문들을 분석하여, ELQ가 기존 하천 유출량 예측기법과 차별화되는 수학적·수문학적 의미를 살펴보고, 여러 국제논문에서 인용된 사례를 분석했다. 이로써 ELQ기법이 원격탐사 수문학에 기여하는 바를 도출하여, 향후 국내·외 원격탐사 수문학에 활용되어 보다 정확한 하천 유출량을 예측하는 방법이 되기를 기대한다.

**주요어:** 앙상블 러닝, 앙상블 러닝 회귀법, 하천 유출량, 콩고강, 메콩강, 위성 고도계, 약한 학습자

## 1. 서론

전 세계적으로 하천 유출량(river discharge,  $Q$ )을 정확히 측정하고 예측하는 것은 홍수조절, 가뭄관리 등 수자원 관리에 매우 필수적인 요소이지만, 전 세계적으로  $Q$ 를 산출하고 추정하기 위한 현장 게이지(stream gauge)의 숫자는 감소 추세에 있다(Collins et al., 2024). 따라서, 현재 운영되고 있는 현장 게이지에 의해 취득되는 데이터는 점점 희소해지고 있으며, 데이터를 획득하는데 많은 제한 사항이 따른다(Du et al., 2020). 예를 들면, Congo Basin은 강 유역 기준으로 Amazon Basin에 이어 두 번째로 넓고(~3.7 million km<sup>2</sup>), 유출량 기준으로 전 세계에서 두 번째이지만(~40,600 m<sup>3</sup>/s), 20세기 초까지 수 백 개의 현장 게이지가 있었던 것이 최근에는 약 10개 정도로 현저히 줄어든 상태이다(Alsdorf et al., 2016; Kim, 2021). 또한, 공유하천(두 나라 이상의 국경을 지나 흐르는 하천, transboundary river)의 경우에는 수자원에 대한 주변 국가들의 이익이 상충되는 경우가 많고, 이러한 경우 현장 게이지가 운영되더라도 데이터를 획득하거나 주변국까지 데이터를 공유하기가 매우 어렵다. 이러한 공유하천은 전 세계적으로 약 261개 이상, 전 세계 담수유량의 약 60%를 차지한다(Kim et al., 2020).

한편, 동남아시아에 위치해 있는 메콩강(Mekong River)의 경우 Tonle Sap Lake (TSL)와의 상호작용에 따라 특수한 수문학적 특징을 보인다. 우기(wet season)에는 메콩강의 수위가 상승하여 TSL의 수위보다 높아져 메콩강의 물이 호수로 역류하면서 호수 및 floodplain의 면적이 넓어지고, 건기(dry season)에는 반대로 메콩강의 수위가 낮아지면서 호수의 물이 다시 메콩강으로 흘러들어가 호수 및 floodplain의 면적은 줄어드는 경향을 보인다(Chang et al., 2020). 이러한 메콩강 유역의 연간 유량 변화 및 계절에 따른 유량 역전 현상은 기존의 수문학적 방법을 통해 유량을 산출하기에 많은 제한사항이 발생한다(Chang et al., 2019). 이렇듯 전 세계적으로 미계측 지역의 하천 유출량 산정을 위해 지금까지 다양한 방법이 연구되어 왔지만, 상대적으로 우리나라 원격탐사분야 학회지에서는 이러한 연구들이 많이 이루어지지 않았다. 2025년 1월 기준 대한원격탐사학회(<https://www.kjrs.org/main.html>)에서 검색어로 river discharge를 입력한 결과 검색되는 논문은 총 6편으로써 연안 해저 지하수 유출 특성에 관한 연구 1건(Kim and Jung, 2021), 홍수량 및 홍수면적에

대한 피해범위 산정에 관한 연구 4건(Son and Kim, 2019a; 2019b; Park and Lee, 2018a; Kim et al., 2018), 위성 영상을 활용한 가뭄 모니터링 1건(Park and Lee, 2018b)이며, 미계측 지역의 하천 유출량 산정에 대한 연구는 부족했다.

하천 유출량  $Q$ 는 식(1)과 같이 하천 흐름의 평균 유속(mean velocity of flow)과 단면적(cross-sectional wetted area)의 곱으로 정의된다(Chow, 1959).

$$Q = v \cdot A \tag{1}$$

여기서  $Q$ 는 하천 유출량이며,  $v$ 는 평균유속,  $A$ 는 단면적을 나타낸다. 한편, 이러한 하천 유출량을 구하는 방법은 전통적으로 두 가지 방법에 의해 주로 산출되었는데, 하나는 Leopold and Maddock (1953)이 제시한 at-a-station hydraulic geometry (AHG)를 이용한 방법으로, 그들은 하천 유출량에 대해 식(2-4)와 같은 경험 멱함수(empirical power functions)를 제안했다.

$$w = a \cdot Q^b \tag{2}$$

$$d = c \cdot Q^h \tag{3}$$

$$v = k \cdot Q^m \tag{4}$$

여기서  $Q$ 는 하천 유출량이며,  $w, d, v$ 는 각각 하천의 폭, 깊이, 유속을 말하며,  $a, b, c, h, k, m$ 은 경험상수로써,  $a \cdot c \cdot k$ 와  $b + h + m$ 은 각각 1이다(Dingman, 2007). 하천의 수리변수(hydraulic variables)들인 폭, 깊이, 유속 중 한 가지를 이용하면, 식(2-4)를 이용하여 하천 유출량을 계산할 수 있다.

다른 하나는 Manning (1889)이 제안한 경험식인 Manning's equation을 이용한 방법이다.

$$Q = \frac{1}{N} \cdot A \cdot R^{\frac{2}{3}} \cdot \sqrt{S} \tag{5}$$

여기서  $Q$ 는 하천 유출량이며,  $n$ 은 Manning's roughness coefficient이며,  $A$ 는 흐름의 단면적,  $R$ 은 동수경사(hydraulic radius),  $S$ 는 channel slope를 말한다. 일반적으로 Manning's equation을 적용할 때 하천의 흐름은 등류(uniform flow)이며, 하천 바닥의 경사와 하천 표면의 경사가 같은 것으로 가정한다.

지금까지 하천 유출량의 계산은 크게 위 두 가지 방법으로 산출되

**Table 1.** Constraints to estimate  $Q$  using hydraulic variables based on RS methods

Hydraulic variables	Constraints to estimate $Q$
Altimetry-derived $H$	<ul style="list-style-type: none"> <li>• Repeating cycles of 10 to 35 days</li> <li>• Sparse orbital spacings</li> </ul>
$W_e$ derived from satellite imagery	<ul style="list-style-type: none"> <li>• Less accurate results than those obtained from altimetry-derived <math>H</math></li> </ul>
$S$ derived from SRTM DEM	<ul style="list-style-type: none"> <li>• Rarely used to estimate <math>Q</math> due to unavailability (i.e., difficult to obtain)</li> </ul>

SRTM: Shuttle Radar Topography Mission, DEM: Digital Elevation Model.

었으나, 최근 원격탐사(remote sensing, RS) 기법의 발전에 따라 위 두 가지 식에 포함된 수리 변수들을 직접측량이 아닌 원격탐사 방법으로 측정할 수 있게 되었고, 이에 따라 지난 수십 년간 원격탐사에 의한 하천 유출량 기법이 발전하게 되었다. Kim et al. (2019a)은 원격탐사 등에 의한 하천 유출량 측정 기법을 아래와 같이 분류했다.

- 1) Velocity-area method (속도-면적 기법): 하천 유출량의 정의에 따라 유출량을 하천의 유속과 단면적의 곱으로 계산하는 방법이며, Tarpanelli et al. (2015)은 Moderate-Resolution Imaging Spectrometer 영상을 이용하여 유속과 단면적을 얻어 하천 유출량을 구했다.
- 2)  $H$ - $Q$  rating curve (수위-유출량 관계 곡선): AHG에 기반한 방법으로 많은 연구자들이 하천 수위와 유출량에 대한 관계곡선을 얻은 후 하천 유출량을 구했다(Coe and Birkett, 2004; Kouraev et al., 2004; Normandin et al., 2018; Bogning et al., 2018).
- 3) Effective river width ( $W_e$ )- $Q$  rating curve (유효 하천폭-유출량 관계 곡선): 하천 유출량은 유효 하천폭과 유출량의 관계식에서 얻어진다(Temimi et al., 2005; Smith and Pavelsky, 2008; Beighley and Gummati, 2011; Yamazaki et al., 2014; Scherelis et al., 2023; Huang et al., 2018).
- 4) Manning's equation: Manning's equation과 그것을 일부 수정하여 얻은 식을 이용하여 하천 유출량을 계산한다.  $S$  (surface water slope),  $W_e$ ,  $H$ ,  $v$  등 수리변수들의 곱으로 하천 유출량을 구한다(Bjerklie et al., 2005; Jung et al., 2010).
- 5) At-many-stations hydraulic geometry (AMHG): 하천 유출량은 in-situ  $Q$  데이터 없이 위성영상에서 얻어진 수리변수들을 이용한 scaling law를 이용해 얻어진다(Gleason and Smith, 2014; Gleason et al., 2014).
- 6) Hydrologic and/or hydrodynamic model (수문 모델 및 수문 동역학 모델): 하천 유출량은 원격탐사 관측값과 수문 모델 및 수문 동역학 모델 값에 의해 얻어진다(Brakenridge et al., 2012; Andreadis et al., 2007; Liu et al., 2015; Tarpanelli et al., 2013).

Table 1은 원격탐사로부터 얻어진 수리 변수들( $H$ ,  $W_e$ ,  $S$ )을 이용하여  $Q$ 를 예측할 때 발생하는 제약사항들을 정리한 것이다. 즉, 위성

고도계(radar altimetry)를 이용하여 하천 수위 정보를 얻을 수 있는데, 이러한 방법의 단점은 위성의 재방문 주기(repeating cycles)가 10~35일로 다소 길다는 점과 위성 고도계의 ground track 간격이 수십~수백 km에 이른다는 점이다. 예를 들면, Envisat과 Jason-1 위성 고도계의 경우, 적도에서 ground track 간격은 각각 80 km와 315 km이다. 한편, 유효 하천 폭( $W_e$ )을 이용한 하천 유출량 예측은 비교적 쉽게 위성 영상에서 그 정보를 추출할 수 있지만 그 결과물의 정확성이 상대적으로 낮다는 단점이 있다(Smith and Pavelsky, 2008; Gleason et al., 2014). 또한, 하천 표면 경사( $S$ )를 위성 영상에서 추출하여 수정된 Manning's equation을 사용하여  $Q$ 를 산출할 수 있지만,  $S$ 를 추출하는 방법은 상대적으로 어려워 이러한 방법을 사용하여 하천 유출량을 계산한 사례는 매우 드문 실정이다(Jung et al., 2010; LeFavour and Alsdorf, 2005). 그러나 미국의 National Aeronautics and Space Administration (NASA)와 프랑스 우주청 French National Space Agency (CNES)가 공동 추진하는 임무인 Surface Water and Ocean Topography (SWOT) 위성이 2022년 12월 16일 미국 캘리포니아주 반덴버그 우주군 기지에서 발사됨에 따라, 앞으로는 하천 표면 경사 등의 수리변수 추출이 비교적 쉬워질 것으로 예상된다.

이와 같이 하천 유출량을 예측하기 위하여 획득 가능한 다양한 수리 변수들과 모델이 존재하지만 각 선택지에 따른 단점이 존재하며, 어떠한 단일 변수나 단일 모델이 다른 모든 다른 변수나 모델들보다 성능이 월등하게 뛰어날 수는 없다. 이는 수학적으로 No free lunch theorem 이라고 한다(Wolpert, 1996). 한편, Kim et al. (2019a) (이하 K19)은 하천 유출량을 계산하기 위한 앙상블 러닝 회귀법(ensemble learning regression for estimating river discharge, ELQ)을 개발함으로써 기존 AHG 또는 Manning's equation에 기반한 다른 하천 유출량 계산법보다 정확한 유출량 알고리즘을 개발했다. 또한, 이 기법은 아프리카 콩고 강 유역에 적용된 데에 이어 ELQ를 메콩강 유역에 적용하여 기존 하천 유출량 계산보다 더욱 정확한 결과를 보였다(Kim et al., 2019b) (이하 K20). 더 나아가 ELQ를 in-situ  $Q$  데이터에 의존하지 않고 수문학 모델과 결합하여 더 정확한 하천 유출량을 산정하였다(Kim et al., 2021) (이하 K21).

최근 몇 년 간 인공지능, 머신러닝 및 딥러닝 기법은 수리 수문학 분야에도 많은 영향을 끼쳤다. 이에 따라 하천 유출량을 산정하고 수

리 수문학 분야에 이러한 기법들을 사용한 논문들이 다수 소개되고 있다(Wang et al., 2024; Lu et al., 2023; Meddage et al., 2022; Zounemat-Kermani et al., 2021). 예를 들면 Lu et al. (2023)은 두 개 저수지의 이전 유출량 값과 유역 강수량 데이터를 활용하여 하천 일일 유출량을 머신러닝 기법으로 예측하여 향상된 정확도를 보여 주었으며, Wang and Peng (2024) 또한 미국 본토를 대상으로 수문학적 시뮬레이션과 머신러닝 기법을 사용하여 하천 유출량의 정확도를 향상시켰다. 이와 같이 최근 머신러닝 기법의 사용은 수문 모델 및 수문 동역학 모델에서 다양한 변수를 결합하는 과정에서 향상된 결과를 보였다. 그러나 이러한 수문 모델을 유지 관리하는 것은 시간과 비용이 많이 소요되기 때문에, 미세측 지역 하천에 대한 유출량 예측에는 적합하지 않을 수 있다. 또한, 이러한 머신러닝 기법을 사용한 최근 논문들은 머신러닝의 다양한 기법을 혼합하여 더 나은 결과 값들을 다수 도출하였지만, 그 기법을 위한 입력 데이터 변수에 대한 고려사항이나 수학적·수문학적 고찰이 다소 부족한 실정이다.

이에 따라 본 논문에서는 원격탐사 데이터를 사용한 하천 유출량 산정 및 예측 분야에서 머신러닝을 사용한 기법인 ELQ의 수문학 및 수학적 프레임워크에 대하여 고찰하고자 한다. ELQ 기법을 이용한 하천 유출량 예측 관련 논문(K19-K21)은 지금까지(2024년 11월 기준) 많은 국제논문에서 약 94회 인용된 바 있지만, 아직 그 수학적·수문학적 프레임워크에 대해 자세히 소개된 바가 없다. 따라서 본 연구에서는 발표된 ELQ 논문(K19-K21)에 대해 간단히 소개하고, ELQ가 기존 하천 유출량 예측 기법과 차별화되는 수학적·수문학적 의의를 알아본다. 또한, 다른 국제논문에 인용된 사례를 분석하여 ELQ 기법이 원격탐사에 기반한 수문학에 기여한 바를 도출하고, 향후 국내·외 원격탐사 수문학(remote sensing hydrology) 분야에 활용되어 보다 정확한 하천 유출량을 예측할 수 있도록 ELQ 기법을 소개하고자 한다.

## 2. 연구 자료 및 방법

### 2.1. 앙상블 러닝

앙상블 러닝 회귀법(ELQ)은 K19에서 위성 고도계(satellite altimetry) 데이터를 이용하여 하천 유출량을 머신러닝 기법으로 산출한 것으로써, 기존 하천 유출량 계산 방법보다 더욱 정확한 결과를 보였다. 본 장에서는 ELQ에 적용된 앙상블 러닝(ensemble learning, EL)에 대해 간략히 소개한다.

EL은 다중 분류 시스템(multiple classifier system) 또는 위원회 결정 기반 학습(committee-based learning)이라고도 불리며, 어떠한 문제를 해결하기 위해 여러 모델을 훈련(train), 결합(combine)하여 보다 나은 모델을 만드는 과정이다(Zhou, 2021). EL은 머신러닝(machine learning, ML)의 한 방법이며, ML은 계산적인 방법에 의한 경험 학습을 통해 모델이나 시스템의 성능을 향상시키는 기법을 말한다(Zhou,

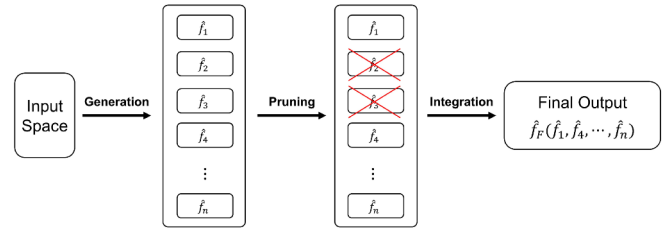


Fig. 1. Conceptual design of the ensemble learning.

2021). 통상적으로 EL은 분류(classification)에 관한 문제에 많이 사용되지만 회귀(regression) 문제에도 사용될 수 있다.

앙상블 러닝은 통상적으로 기본 학습자(base learners) 생성 단계인 ensemble generation과 생성된 기본 학습자들을 결합하는 단계인 ensemble integration의 두 단계로 이루어진다(Zhou, 2015). 일부 문헌에서는 그 두 단계 사이에 기본 학습자들 중 일부를 필터링하는 단계인 ensemble pruning을 넣기도 한다(Mendes-Moreira et al., 2012). 앙상블 러닝의 과정을 살펴보면 Fig. 1과 같다. Fig. 1에서 보는 바와 같이 1단계 앙상블 생성 단계는 다수의 기본 학습자(base learners or candidate functions,  $f_i$ )가 만들어지는 단계이다.

이 단계에서 학습자들은 동종 학습자(homogeneous functions) 또는 이종 학습자(heterogeneous functions)로 나누어지고, 이는 데이터를 훈련시키기 위해 각각 같은 모델 또는 다른 모델을 사용한다. 예를 들면, 유출량을 구하기 위해  $H-Q$  relationship 단일 모델을 사용할 수도 있지만,  $W_e-Q$  및  $S-Q$  relationships 등 여러 모델을 사용하여 유출량을 구할 수도 있다. 일반적으로 이종의 다양한 앙상블 후보 학습자(ensemble candidates)를 사용하는 것이 앙상블 러닝의 성능을 향상시키지만(Margineantu and Dietterich, 1997), 이종의 predictors를 얻는 것이 현실적으로 어려울 경우도 있다. 예를 들면, 2006년에 발사된 Advanced Land Observing Satellite (ALOS) Phased Array type L-band Synthetic Aperture Radar (PALSAR)의 경우, 재방문 주기(repeat cycle)가 46일로 비교적 긴 편이며, 이를 이용하여 만든  $W_e$ 의 시간 해상도는 상대적으로 낮을 수밖에 없다. 반면, 2008년에 발사된 Jason-2의 경우에는 궤도주기(orbiting cycle)가 10일로 상대적으로 짧기 때문에 시간 해상도가 비교적 높다고 할 수 있다.

한편, 최근 핀란드 국적의 지구관측 위성회사인 ICEYE의 경우, 2023년 6월까지 27기의 X-band Synthetic Aperture Radar (SAR) 위성을 쏘아 올렸으며, 같은 해 11월 4기의 SAR 위성을 발사함으로써 2018년부터 현재까지 31기의 SAR 위성을 운용중으로, 상대적으로 더 높은 1-22일의 재방문 주기를 가진다(<https://earth.esa.int/eogateway/missions/iceye>). ICEYE의 3세대 위성군인 4기의 위성은 50 cm급의 공간해상도를 얻을 수 있는 spot fine 모드와 지상변화를 거의 실시간으로 파악할 수 있는 dwell 모드르 갖춘 것으로 알려져 있다. 따라서 ICEYE의 소형 SAR 군집 위성군을 활용하면 하천 유출량 예측



을 위한 이종 데이터인  $W_i$  및  $S$  등을 좀 더 조밀한 시간 간격으로 얻을 수 있을 것으로 예상된다.

## 2.2. 위성 고도계 데이터

ELQ 관련 연구(K19-K21)에서는 하천 유출량을 예측하기 위한 기본 학습자료로써 AHG 경험 멱함수 중 식(3), 즉 하천 수위(깊이)와 유출량 관계 곡선을 사용했다( $H$ - $Q$  rating curve). 이 때 하천 수위를 얻기 위해 위성 고도계(radar altimetry) 데이터가 사용되었다. 위성 고도계 관련 연구 역시 한국의 원격탐사 연구에서 다소 생소한 분야로써, 2025년 1월 기준 대한원격탐사학회(<https://www.kjrs.org/main.html>)에서 검색어로 altimetry를 입력한 결과 총 3건의 연구논문과 1건의 리뷰 논문이 검색되었다. 관련 논문은 해수면 변화(Hwang et al., 2021), 해양 에디 경계 변화 관측(Kim et al., 2024), 북극 용빙호(melt ponds) 탐지(Kim et al., 2021), 원격탐사에서 지하수 수자원 적용 사례 관련 리뷰(Lee, 2017) 등 4건이었다.

위성 고도계는 보통의 위성이 취득하는 영상과 달리 1차원적인 고도 데이터를 취득한다. 실제로 취득하는 데이터는 위성에서 발사된 펄스가 지표면(통상적으로 수표면)을 맞고 되돌아오는 지연 시간(time delay)을 측정한다. 지연 시간이 측정되면 위성과 수표면과의 거리를 측정할 수 있고, 수위는 위성의 reference ellipsoid 고도에서 여러가지 보정과정을 거쳐 구해진다. 세계 최초의 위성 고도계는 1973년에 발사된 미국 NASA의 Skylab이며, 1993년 TOPEX/Poseidon 위성이 발사된 이후부터 고도계 자료를 활용한 해수면 높이 관측이 활발해지기 시작했다(Hwang et al., 2023; Lee et al., 2009). 그 이후 많은 위성 고도계가 발사되면서 성능이 향상되었다. 현재 운용 중인 위성 고도계는 Jason-3 등을 비롯하여 총 10개이다(<https://www.aviso.altimetry.fr/en/missions/current-missions.html>). ELQ 관련 연구에서는 콩고강을 지나는 Envisat radar altimetry (K19/K21, 2002년~2010년) 데이터와 메콩강을 지나는 Envisat radar altimetry (K20, 2002년~2010년) 및 Jason-2 radar altimetry (K20, 2008년~2016년) 데이터를 사용했다. Envisat은 2002년 3월 1일에 발사되었으며 ERS-1과 ERS-2를 후속하여 발사된 위성 고도계이며 35일의 repeat cycle을 가지고 Jason-2는 TOPEX/Poseidon과 Jason-1을 후속하여 발사된 위성 고도계로써 10일의 repeat cycle을 가진다(K20). Envisat의 경우 Geophysical Data Record (GDR) 18 Hz 자료를, Jason-2의 경우 GDR 20 Hz 자료를 각각 사용했다. 위성 고도계에서 반사된 펄스를 처리하기 위해 retracking 모델이 사용되는데, 이 때 내륙지방에 가장 적합하다고 알려진 ICE-1 retracking algorithm을 사용했다(Kim et al., 2017a; Yuan et al., 2017; Frappart et al., 2006).

## 2.3. 앙상블 러닝 회귀법의 개발

최근 수십 년간 원격 탐사 수문학이 발전하게 되면서 하천 유출량을

정확하게 예측하고자 하는 수요에 부응하여 많은 기법들이 소개되었다(Durand et al., 2016; Bogning et al., 2018; Jung et al., 2010; Sichangi et al., 2016; Gleason and Smith, 2014; Gleason et al., 2014). 그러나 이러한 방법들은 서론에서 소개한 바와 같이 원격탐사 데이터를 사용하지 않거나, 기존의 대표적인 방법인 AHG와 Manning's equation을 사용한 방법이었다(Tarpanelli et al., 2013; 2017; Huang et al., 2018; Paris et al., 2016).

한편, ELQ 기법은 원격탐사 데이터인 위성 고도계 데이터와 기존 방법인 AHG를 사용하면서 기존 방법과 대별되는 방법을 사용했다. 기존 방법과 다른 두 가지 특징은 첫째, 기존 방법은 단일식이나 변수들의 곱을 통해 유출량을 구한 반면, ELQ 기법은 ML의 한 방법인 앙상블 러닝을 이용하여 유출량을 구한다. 둘째, 기존 방법은 공간적으로 동일한 한 개의 지역의 수리변수들과 데이터를 이용하여 유출량을 구했다면, ELQ는 공간적으로 멀리 떨어진, 즉, 수백 km 떨어진 몇 개 지점들의 데이터를 결합하여 모델의 성능을 향상시킨다.

앙상블 러닝 회귀 모델의 기본 학습자는 식(3)을 변형하여 구해진다. 식(3)  $d = c \cdot Q^h$ 에서  $d$ 는 하천의 깊이,  $c$ 와  $h$ 는 경험상수,  $Q$ 는 하천 유출량을 나타낸다. 이 때, 하천의 깊이는 위성고도계로 측정이 직접적으로 불가능하며, 위성 고도계로 측정 가능한 것은 하천의 상대적인 높이( $H$ )이다. 따라서 K19에서는 위성 고도계를 통해 각 virtual station (VS)에서 얻어진 높이(altimetry-derived  $H$ )를 이용하여 하천의 깊이를 추정하였다. 세부적인 추정 방법은 K19를 참조 바란다. 위성 고도계 데이터를 통해 하천 깊이( $d$ )를 추정하면 식(3)  $d = c \cdot Q^h$ 을  $Q$ 에 대한 식으로 나타낼 수 있고, 이는 식(6)과 같다.

$$\hat{f}_{ij} = \left( \frac{d_i}{c_j} \right)^{1/h_j} \quad (6)$$

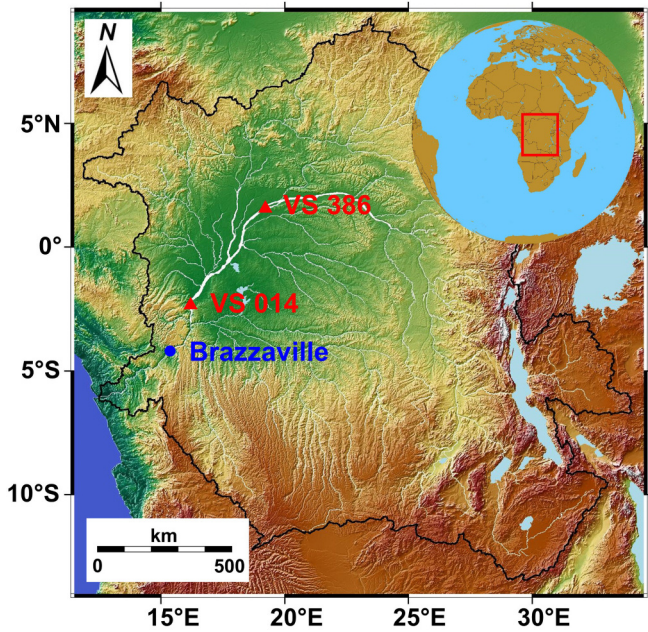
여기서  $\hat{f}_{ij}$ 는 ELQ의 기본 학습자료로써 하천 유출 예측량( $\hat{Q}_{ij}$ )을 나타내며,  $d_i$ 는 각 VS에서 얻어진 하천의 깊이,  $c_j$ 와  $h_j$ 는 식(3)에서 계수와 지수를 나타내고,  $i$ 는 VS의  $i$ 번째 관측값을, 그리고  $j$ 는 여러 기본 학습자 중  $j$ 번째 학습자 또는  $j$ 번째 VS를 나타낸다.

앙상블 러닝 회귀 모델은 이론적으로  $n$ 개의 기본 학습자를 가질 수 있으나 과적합 등을 고려하여 3개의 후보 학습자를 이용하였고, 식(7)과 같이 기본 학습자를 결합하였다(K19).

$$f_{ELQ} = \alpha + w_1 \cdot \hat{f}_{i1} + w_2 \cdot \hat{f}_{i2} + w_3 \cdot \hat{f}_{i3} + \varepsilon_i \quad (7)$$

여기서  $f_{ELQ}$ 는 최종 앙상블 러닝 회귀 모델이며,  $\hat{f}_{i1}, \hat{f}_{i2}, \hat{f}_{i3}$ 은 세 개의 기본 학습자들이고,  $w_1, w_2, w_3$ 는 각 기본 학습자의 가중치이다.  $\alpha$ 와  $\varepsilon_i$ 는 모델의  $y$ 절편과 오차항을 의미한다.  $i$ 는 각 virtual station (VS)의  $i$ 번째 관측값이다.

K19에서는 2006년부터 2010년까지 Envisat 위성고도계 데이터와 Congo Basin에서 유지되고 있는 in-situ station 중의 하나인 Brazzaville 유량 게이지의 in-situ  $Q$  데이터를 먼저 ELQ를 통해 훈련 시킨 후



**Fig. 2.** Map of the Congo River Basin. The back boundary represents the Congo River Basin. The red triangles indicate the farthest and closest VSs used in K19 and K21. The blue circle denotes the Brazzaville in-situ *Q* station.

(2002년 11월~2006년 11월), ELQ의 예측 값을 2006년 12월부터 2010년 9월까지의 기간 동안 검증하였고, 그 결과 root-mean-square error (RMSE)는 기존 AHG 모델 대비 약  $823 \text{ m}^3\text{s}^{-1}$ 만큼 감소하였으며, relative RMSE (RRMSE)는 약 2.08% 감소하였다. 이 때, ELQ의 기본 학습자는 콩고강의 본류에 위치한 9개의 VSs에서 추출된 높이 데이터를 사용했다. Fig. 2에서 보는 바와 같이 예측하고자 하는 Brazzaville station에서 최소 수 십 km에서 500~600 km 이상 떨어진 지역의 VS 데이터를 사용했다. 이것은 원격탐사 수문학에서 처음 시도되었으며, Irrgang et al. (2020)은 원격탐사 수문학에서 ELQ (K19)가 최초로 머신러닝과 위성고도계 데이터를 결합하여 기존 방법보다 성능 향상을 가져왔다고 평가했다. 이러한 사실은 ELQ가 단지 머신러닝 기법을 원격탐사 수문학에 적용한 것을 넘어서서, 기존의 하천 유출량 예측 패러다임을 결정적으로 변화시킨 것에 더 큰 의미가 있다.

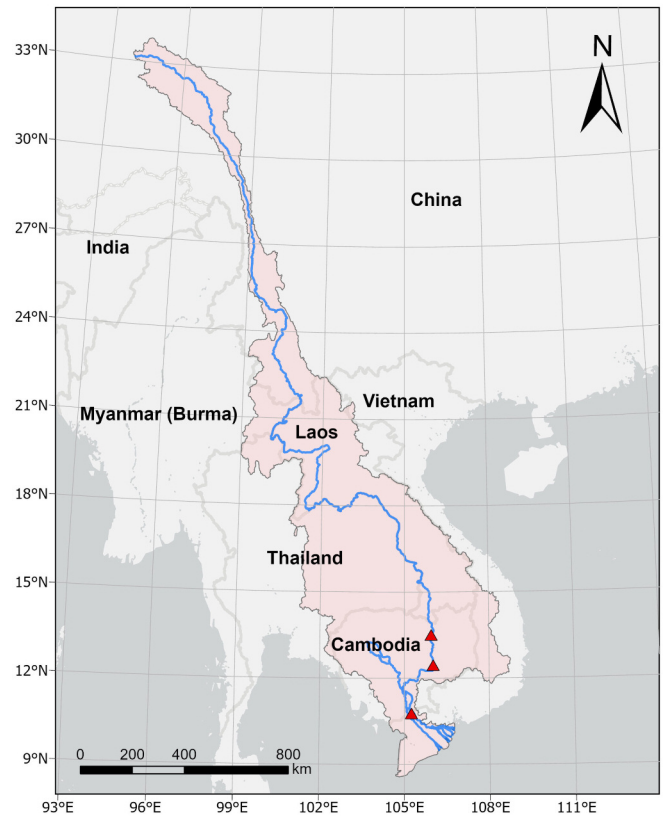
즉, 기존에는 AHG 또는 downstream hydraulic geometry (DHG)의 단일 관계에 의해 하천 유출량을 계산했다. DHG란 관측된 수리변수들과 하류 방향으로 떨어진 유출량의 관계가 식(2-4)와 같이 성립하는 수리 기하학적인 관계를 의미한다(Chow, 1959; Jung et al., 2013; Lee and Yoo, 2024). 한편, DHG의 관계에서도 통상적으로 수 십 km 내에 있는 수리 변수들을 획득하여 하천 유출량을 산정하였으나, K19에서는 수 백 km 이상 공간적으로 이격된 DHG를 기본 학습자료로 하여 세 개의 다른 DHG를 선형적으로 결합하여 유출량을 예측한 것은 매우 이례적인 사례로 평가되며, 이러한 새로운 기법으로 하천 유

출량 예측 정확도를 향상시켰다.

**2.4. 앙상블 러닝 회귀법의 응용**

K19에서 ELQ 기법을 소개한 후, K20과 K21에서 ELQ에 대한 응용이 이루어졌다. 먼저 K20에서는 ELQ를 메콩강 일대에 적용시켰다. 메콩강은 세계에서 12번째로 긴 강이며, 동남아시아에서 가장 큰 강 유역으로써, 특히 메콩 삼각주(Mekong Delta) 지역은 메콩강 유역에서 가장 인구가 많은 지역 중의 하나로 약 1,600만 명의 인구가 살고 있으며, 이는 메콩강 유역에 살고 있는 인구의 약 23%를 차지한다 (Chang et al., 2020).

한편, 앞서 서론에서 밝힌 바와 같이, 공유 하천의 경우에는 수자원에 대한 주변 국가들의 이익이 상충되는 경우가 많다. Fig. 3에서 보는 바와 같이, 메콩강 유역은 6개의 나라(중국, 미얀마, 라오스, 태국, 캄보디아, 베트남)의 국경을 지나고 있다. 따라서 현장 게이지가 운영되더라도 그 데이터를 획득하기 어려울 경우가 있다. 또한, 메콩강 하류 유역(Lower Mekong River Basin, LMRB)은 TSL의 수문학적 복잡성으로 인해 하천 유출량을 더욱 예측하기 힘들다. LMRB에 위치한 TSL은 1년 내내 호수인 지역과 seasonally flooded areas로 이루어지는 floodplain으로 둘러싸여 있다. 호수의 면적은 건기(11월~3월)



**Fig. 3.** The Mekong River Basin: the red triangles represent the three in-situ *Q* stations, Stung Treng, Kratie, and Tan Chau from top to bottom.



에는 길이 120 km, 폭 35 km로 약 2,500 km<sup>2</sup>의 면적을 보이는 반면, 우기(5월~10월)에는 길이 250 km, 폭 100 km로 약 17,500 km<sup>2</sup>의 면적을 보인다(Cambell et al., 2009). 메콩강의 TSL의 흐름은 일명 flow reversal로 불리는데, 우기가 끝나갈 무렵인 9월~10월에 메콩강의 수위가 낮아지면, 우기동안 floodplain에 저장된 물은 호수에서 메콩강 하류 삼각주 지역으로 흐르고, 건기가 끝날 무렵인 5-6월 부터 비가 내리기 시작하면서 메콩강의 수위가 상승하고, 이내 강의 흐름이 역류하여 물은 TSL로 흘러간다(Olson and Morton, 2018). 이러한 TSL의 flow reversal은 메콩강 유역의 주민들에게는 우기 동안에 홍수 피해를 감소시키고, 건기 동안에는 농업용으로 사용할 수 있는 관개용수 제공과 함께 메콩 삼각주 아래의 바닷물이 메콩 삼각주로 침투하는 정도를 완화시키는 등 지역 주민에게 많은 이점을 제공하지만, 동시에 메콩강의 유출량을 예측하는데 있어 복잡성을 증가시키고, 따라서 고해상도의 hydrodynamic model을 필요로 할 수 있다(Mekong River Commission, 2009).

K20에서는 K19에서 개발된 ELQ를 기본으로 메콩강에 적용시켜 보다 정확한 하천 유출량을 계산하였다. K20에서 사용된 in-situ  $Q$  데이터는 Stung Treng, Kratie, Tan Chau 세 곳에서 획득하였으며, 기본 학습자는 Envisat altimetry (2002-2010년 시기)와 Jason-2 altimetry (2008-2016년 시기) 데이터를 사용했다. 이러한 Multi-mission 위성 고도계 데이터를 사용할 때 장점은 비교적 장기간에 대해서 미세측 유역의 하천 유출량을 산정할 수 있다는 점이다. 그리고 Tan Chau의 경우, in-situ  $Q$  데이터는 2003-2006년, 2013-2016년에 대해 획득한 상태로 2007-2012년까지 6년의 기간 동안에는 미세측 구간으로 남아있다. 이렇게 현장 게이지의 운영이 일시적으로 중단된 경우, ELQ 기법을 사용하면, 하천 유출량의 예측 정확도를 보다 효과적으로 향상시킬 수 있다.

또한 TSL로 인한 메콩강의 수문학적 복잡성 때문에 세 개의 in-situ  $Q$  station 중 가장 하류에 위치하고 있는 Tan Chau에서의 하천 유출량 예측(2003-2006년)에서 단일 AHG에 의한 유출량은 Nash-Sutcliffe Efficiency (NSE) 기준 0.84/0.83 (training/validation)을 보였지만, ELQ 기법을 사용한 예측은 NSE 0.97/0.96 (training/validation)을 보였다.

한편, K20에서 TSL보다 고위도에 위치하고 있는 Stung Treng과 Kratie에서는 기존 모델과 ELQ 기법과의 성능차이는 미미했다. 상대적으로 성능이 좋지 않은 ELQ 결과는 바로 기본 학습자의 다양성(diversity) 부족에서 기인한다(Zhou, 2009; Brown et al., 2005a). 즉, 기본 학습자의 상관관계(correlation coefficient,  $r$ )가 증가하면, 기본 학습자의 다양성은 감소하므로, ELQ의 결과에 부정적인 영향을 미친다. 한편 Tan Chau에서는 일반적으로 ELQ 결과가 AHG 모델보다 우수한데, 이것 또한 기본 학습자의 다양성과 관련이 있다.

일반적으로 ELQ 기법은 in-situ  $Q$  데이터가 존재하는 하천 유역에서 기본 학습자를 학습하고, 그것들을 결합시켜 ELQ를 생성하였다.

K21에서는 in-situ  $Q$ 가 존재하지 않거나, 여러가지 사유로 인해 현장 게이지가 작동하지 않을 경우에 ELQ 기법을 사용하여 하천 유출량 정확도를 향상시켰다. K21 또한 콩고강 유역을 연구지역으로 하였고, 수리 수문 모델인 Hillslope River Routing (HRR)을 사용하여 HRR-derived  $Q$ 를 구한 후, Envisat 위성 고도계 데이터와 함께 훈련시켜 ELQ 모델을 생성하였다. 기존에 원격탐사 데이터와 수문모델 등을 사용하여 콩고강 유역을 연구한 몇몇 논문에서 유출량 정확도는 RRMSE 기준 약 12.2-37.2%였다(Nijssen et al., 2001; Beighley et al., 2011; Sichangi et al., 2016). 그 중 Sichangi et al. (2016)은 in-situ  $Q$  데이터를 사용하여 RRMSE 12.2%의 정확도를 보였고, in-situ  $Q$  데이터를 사용하지 않았을 경우에는 약 20-40% 안팎의 RRMSE를 보였다.

미계측 유역의 하천은 세 가지로 정의될 수 있는데, 첫째는 완전히 예측되지 않은 순수한 미계측 유역의 하천, 둘째, 현장 게이지가 과거에는 설치되어 있었으나 예산부족 등 여러가지 사유로 더 이상 운영되지 않거나(Shiklomanov et al., 2002), 셋째, 기계적 결함이나 오류로 인해 현장 게이지가 일시적으로 중단되어 예측을 하지 못한 경우(Loukas and Vasiliades, 2014) 등으로 나뉜다. K21에서는 세 가지 경우 중 둘째와 셋째의 경우  $Q$ 를 예측하는 것을 목표로 하였다. K21에서는 Kinshasa에서의 현장 게이지가 1990년 이후 운영을 중단했다고 가정하고, 이후 하천 유출량을 예측한다. 단, Kinshasa의 현장 게이지 데이터는 1990년 이후에도 획득이 가능하나 K21에서는 사용되지 않았다. 따라서 ELQ의 기본 학습자는 HRR-derived  $Q$ 를 사용하여 얻었다. 한편, K19와 K21에서 앙상블 결합은 조금 다른 방식으로 이루어졌다. K19에서는 기본 학습자들의 결합 과정 간에 in-situ  $Q$ 와 estimated  $Q$ 의 차이를 최소화하는 학습 과정을 거치게 되며, 이 과정을 통해 각 기본 학습자의 가중치가 각각 결정되었다. 그러나 K21에서는 각각의 기본 학습자의 가중치를 동일하게 하는 방식인 Basic Ensemble Method (BEM)를 식(8)과 같이 적용했다.

$$f_{BEM} = \alpha + \frac{1}{N} \sum_{j=1}^N \hat{f}_{ij} + \varepsilon_i \quad (8)$$

여기서  $f_{BEM}$ 은 basic ensemble function이며,  $\hat{f}_{ij}$ 는 기본 학습자,  $i$ 와  $j$ 는 각각 관측값의 수와 기본 학습자의 수를 나타내고,  $\alpha$ 와  $\varepsilon$ 은 각각 intercept와 오차항을 나타낸다.  $f_{BEM}$ 은 앙상블 러닝에서 기본 학습자들의 예측값의 평균을 나타낸다(Mendes-Moreira et al., 2012). 일반적으로  $f_{BEM}$ 은 향상된 예측치를 제공하는데 이것은  $f_{BEM}$ 이 앙상블 러닝에서 모든 기본 학습자들을 과적합(overfitting) 없이 훈련시키기 때문이다(Perrone and Cooper, 1992). K21에서  $f_{BEM}$ 을 사용하는 또 다른 이유는 기본 학습자의 결정계수(coefficient of determination,  $R^2$ )가 0.6 미만일 때는  $f_{ELQ}$ 가 제대로 작동하지 않기 때문이다(K19).

K21에서 현장 게이지 데이터와 비교한 HRR에서 얻어진 하천 유출량의 정확도는 RRMSE 기준, 약 15.72/18.00% (training/validation)

의 정확도를 보인 반면, ELQ에서 얻은 가장 성능이 좋은 유출량의 정확도는 약 9.89/6.34% (training/validation)였다. 따라서, 기존 in-situ Q 데이터를 사용하지 않은 하천 유출량 예측에 비해서 상당히 향상된 성능을 보여 주었다. 이는 근본적으로 ELQ가 단일 학습자보다 여러 학습자를 결합함으로써 부가적인 정보와 데이터를 결합함으로써 성능을 향상시키고, 실제로 원격탐사에서 위성에서 얻어진 데이터에서 결측치가 발생하는 경우도 있기 때문에, 시공간상 다른 데이터를 결합하여 보다 정확한 예측치를 얻을 수 있기 때문이다.

### 3. 연구 결과 및 토의

#### 3.1. ELQ의 수학적 프레임워크

앙상블 러닝은 앞서 말한 바와 같이 어떠한 문제를 해결하기 위해 여러 모델을 학습시켜 생성하고 그것들을 결합하는 일련의 방법을 말한다(Zhou, 2021). 여기서 어떠한 문제란 통상적으로 분류(classification) 또는 회귀(regression) 문제를 말한다. 또한, Zhou (2012)는 앙상블 기법이란 여러 학습자를 훈련시키고, 대표적으로 Boosting과 Bagging 등으로 대표되는 결합 방식 등을 사용하는 머신러닝의 한 방법이라고 설명했다. 따라서 앙상블 러닝의 생성과 결합에서 대두되는 두 가지 이슈 중 첫 번째는, ‘어떠한 문제를 해결한 것인가?’이고, 두 번째는 ‘어떠한 방식으로 결합할 것인가?’이다.

##### 3.1.1. 약한 학습자와 앙상블

앙상블 러닝은 Fig. 1과 같이 다수의 기본 학습자를 생성한 다음, 특정 결합 방식을 이용하여 그들을 결합한다. 앙상블 러닝은 단일 학습자에 의한 결과보다 나은 결과를 가져다 준다. 왜일까? 직관적으로 생각하면, 여러 다른 특성을 가진 것을 섞으면 그 결과는 최악의 것보다는 낫지만, 최고의 것보다는 더 안 좋은 결과를 가져온다(Zhou, 2021). Fig. 4는 앙상블 기법의 여러가지 예를 나타낸 것이다. 예시는 Zhou (2021)의 예를 이해하기 쉽게 국어, 수학, 영어의 예로 재구성하였다. 우리는 Alex, Brian 그리고 Cyril 3명의 학습자 또는 분류기(classifier)가 있다. 그 분류기는 각각 이진(binary) 분류를 실시한다. 즉 국어, 영어, 수학 과목에 대해 통과(O) 또는 낙제(X)를 할 수 있다. 여기서 앙상블 분류(classification of EL)는 투표(voting)를 통해 결정된다. 즉 3명 중 다수의 사람이 통과하면 그 과목에서 통과를 하는 것이고, 다수의 사람이 낙제를 하면 그 과목에서 낙제를 하는 것이다. Fig. 4 (Case #1)에서 보면 각각의 분류기는 약 66.6%의 성공률을 갖는 반면 앙상블은 100%의 성공률을 보인다. Fig. 4 (Case #2)에서는 각각의 분류기는 Fig. 4 (Case #1)와 동일한 성공율을 보이지만 각각의 분류기가 정확히 일치하므로, 앙상블 또한 66.6%의 성공률을 보이기 때문에 앙상블은 개별 분류기에 비해 전혀 도움이 되지 않는다. Fig. 4 (Case #3)에서는 각각의 분류기는 33.3%의 성공률을 보이지만, 앙

	Korean	Math	English
Alex	O	O	X
Brian	X	O	O
Cyril	O	X	O
Ensemble	O	O	O
▪ Case #1: Ensemble helps.			
	Korean	Math	English
Alex	O	O	X
Brian	O	O	X
Cyril	O	O	X
Ensemble	O	O	X
▪ Case #2: Ensemble doesn't help.			
	Korean	Math	English
Alex	O	X	X
Brian	X	O	X
Cyril	X	X	O
Ensemble	X	X	X
▪ Case #3: Ensemble hurts.			

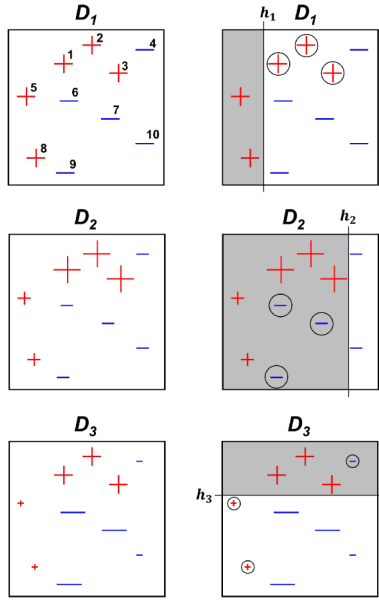
Fig. 4. Example of the ensemble (modified from Fig. 8.2 in Zhou, 2021).

상블은 이보다 더 나빠져 0%의 성공률을 보인다(Zhou, 2021). 이러한 예는 개별 학습자들(individual learners)이 ‘정확하고 다양해야 (accurate and diverse)’ 개별 학습자들의 앙상블이 좋은 결과를 가져다 준다는 것을 알 수 있다(Zhou, 2021).

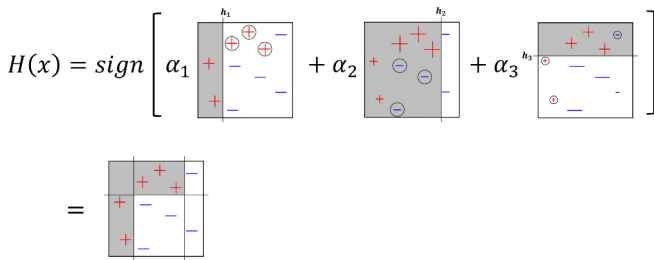
이처럼 여러 학습자를 결합시키는 경우, 앙상블의 능력은 개별 분류기의 일반화 능력보다 훨씬 강한 경우가 많으며, 이는 특히 weak learners (약한 학습자)가 결합할 때 훨씬 더 능력이 강해지는 경우가 많다(Zhou, 2021). 예를 들면, Freund and Schapire (1995)는 Adaptive Boosting (AdaBoost)이라는 meta-learning 알고리즘을 개발했는데, 이는 다수의 weak learners를 결합하여 강한 학습자(strong learner)를 만드는 앙상블 기법이다. Weak learner (boosting)라는 용어는 머신러닝에서 다수 분류기를 이용하여 오차를 줄이는 EL 기법의 하나로써 Kearns and Valiant (1988; 1989)가 ‘weak learnability’라는 개념을 처음 발표하면서 알려졌다. 그 이후 다수의 연구자들에 의해 boosting이 개발되었고, 그 대표적인 알고리즘이 AdaBoost이다(Freund and Schapire, 1995).

AdaBoost 알고리즘의 작동 개념에 대한 예가 Fig. 5와 Fig. 6에 설명되어 있다. 이 경우, 우리는 10개의 샘플을 -1 또는 +1로 분류하려고 한다. 편의상 -1과 +1은 숫자를 생략하고 부호로만 나타내기로 한다. 먼저 하나의 weak learner는 샘플들을 수직선이나 수평선을 이용하여 그 선의 왼쪽이나 오른쪽 또는 위나 아래를 (+) 또는 (-)로 분류한다. Fig. 5에서 각 행(row)의 그림은 시간(또는 라운드)  $t=1, 2, 3$ 일 때의 분류를 나타낸다. 회색 영역은 가설( $h_t$ )에 의해 예측된 양(+의 영역)을 나타내고, 각 학습 라운드( $D_t$ )에서 잘못 분류된 샘플을 원으로 표기했다. 각 샘플의 크기는  $D_t$ 에서 각각의 가중치를 나타낸다. Fig. 6는 각각의 라운드에서 얻어진 결과를 바탕으로 AdaBoost 알고





**Fig. 5.** Conceptual design of AdaBoost (modified from Fig. 1.1 in Schapire and Freund, 2012). Each row indicates one round. The darker shading represents the region of the domain predicted to be positive by the weak hypothesis  $h_i$ .



**Fig. 6.** The combined classifier for the AdaBoost shown in Fig. 5 (modified from Fig. 1.2 in Schapire and Freund, 2012).

리즘을 개념적으로 나타낸 그림이다. 그림에서 보는 바와 같이, 어떠한 단일 수직선 또는 수평선도 샘플을 완벽하게 분류하지 못하지만, AdaBoost는 3번의 반복(iteration)을 통해 양(positive)과 음(negative)으로 완벽히 분류할 수 있다.

이 때 약한 학습자인  $h_i$ 는 weak function으로 나타내면 다음의  $h_i: X \rightarrow \{-1, +1\}$ 과 같다. 그리고 최초 훈련 데이터는 10개의 샘플로  $(x_1, y_1), (x_2, y_2), \dots, (x_{10}, y_{10})$ 이며,  $x_i, y_i$ 는  $X$ 평면(input space)에 포함되어 있고, 이것들은 label set (output space)인  $Y=\{-1, +1\}$ 로 분류된다. 첫 번째 라운드( $D_1$ )의 경우에는 샘플의 크기가 같고, 이는 모든 샘플의 가중치가 같다는 것을 나타낸다. 이 때 첫 번째 약한 학습자인  $h_1$ 이 수직선으로 그어지면서 왼쪽을 양(+), 오른쪽을 음(-)으로 분류한다. 결과적으로 잘못 분류된 샘플들이 발견된다. 두 번째 라운드( $D_2$ )에서는 잘못 분류된 샘플들이 커지고(가중치가 높아짐), 제대로 분류된 샘플들은 작아진다(가중치가 작아짐). 그리고 두 번째 약한 학습자

인  $h_2$  수직선이 그어지고 세 개의 잘못된 샘플들을 찾는다. 마지막 라운드( $D_3$ )에서는 세 번째 약한 학습자인  $h_3$  수평선이 그어지고, 전 단계에서 잘못 분류되었던 세 개의 샘플들을 올바르게 분류한다. 종합하면, 이렇게 각 라운드에서 AdaBoost는 약한 학습자인  $h_i$ 를 잘못 분류된 샘플을 고려한 가중치  $\alpha_i$ 를 이용하여 훈련시킨다. 그러면 최종적으로 다수의 약한 학습자  $h_i$ 와 가중치  $\alpha_i$ 가 강한 분류기인  $H(x)$ 를 만들어 내며 이는 식(9)로 나타낼 수 있다.

$$H(x) = \text{sign} \left( \sum_{i=1}^T \alpha_i h_i(x) \right) \tag{9}$$

이 때 AdaBoost 알고리즘은 Fig. 5에서 보는 바와 같이 적응적(adaptive)이다. 즉 약한 학습자는 회차를 반복하면서 가중치에 의해 업데이트 된다. 개념적으로 ELQ의 결합 알고리즘을 Adaboost 알고리즘의 형태로 나타내면, ELQ는 약한 학습자(weak learners)인 다수의 rating curves로 구성된 기본 학습자들이 여러 회차를 반복해서 가중치를 구하는 대신, 다수의 rating curves가 단일회차로 최소제곱법(least squares)을 통해 가중치를 구하여 약한 학습자를 결합하는 방식으로 설명할 수 있다(Kim et al., 2017b). 다만, 이러한 설명의 한계로써, AdaBoost는 통상적으로 분류(classification)의 문제이고, ELQ는 회귀(regression)의 문제이므로 weak learners 또는 base learners에 대한 생성 및 결합방식에 대한 접근법과 설명은 정확히 일치하지 않으며, 따라서 분류와 회귀에 대한 앙상블 기법은 다소 독립적으로 발전되어왔다(Mendes-Moreira et al., 2012).

### 3.1.2. 다중선형회귀와 앙상블 러닝

ELQ는 식(7)과 같이 기본 학습자의 결합 방식으로 가중평균방식을 사용한다. 이는 다중선형회귀(multiple linear regression or multivariate linear regression, MLR)와 비슷하지만 다른 특성을 갖는다. MLR은 종속변수(dependent variable)와 독립변수(independent variable) 사이에 다음과 같은 관계가 성립한다.

$$y_i = \beta_0 + \beta_1 \cdot x_{i1} + \dots + \beta_j \cdot x_{ij} + \varepsilon_i \tag{10}$$

여기서  $y_i$ 는 종속변수,  $x_i$ 는 독립변수,  $\beta$ 는 MLR의 계수이며,  $\varepsilon$ 은 오차항을 말한다.  $i$ 와  $j$ 는 각각 관측수와 설명변수(explanatory variables)의 개수를 말한다. 식(10)을 벡터형태로 나타내면  $Y = X\beta + \varepsilon$ 로 나타낼 수 있고, 이를 다시 행렬 형태로 나타내면 식(11)과 같이 나타낼 수 있다.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1j} \\ 1 & x_{21} & \dots & x_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nj} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{11}$$

계수벡터인  $\beta$ 를 구하기 위해 최소제곱법을 이용하면 식(12)와 같이 mean squared error (MSE)를 최소화하는 값을 구할 수 있다.

$$MSE(\hat{y}) = E[(\hat{y} - y)^2] \tag{12}$$

따라서  $\beta$ 는 식(13)에 의해 구할 수 있다.

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{13}$$

따라서  $f: X \rightarrow \hat{Y}$ 을 얻을 수 있다.

$$\hat{f}(X) = X\hat{\beta} \tag{14}$$

MLR과 EL은 모두 예측 모델링에 사용되지만 MLR은 하나의 종속 변수를 다수의 독립변수의 선형 조합으로 독립변수(설명변수)의 계수를 구하는 방법이다. 반면, EL은 여러 개의 개별 모델을 별도로 훈련시킨 다음 다수의 모델들을 결합시켜 더 나은 모델을 만드는 머신러닝 기법 중 하나이다. 두 방식의 차이점 중 하나는 MLR의 계수의 크기는 각 독립변수의 설명력과 전혀 관계가 없다는 것이고, MLR의 계수의 합 또한 임의적(arbitrary)이라는 것이다. 예를 들면 식(7)에서 ELQ의 가중치의 합은 1이 되어, 각각의 가중치는 각 모델들의 설명력을 표현해 주지만 MLR의 경우 계수는 전체 모델에서 각 변수가 어떤 기여를 하는지 설명하지 못한다. 즉 MLR은 단일 모델이라고 볼 수 있고, EL은 다수 모델의 결합이라고 볼 수 있다.

### 3.1.3. ELQ의 정확성과 다양성

Zhou (2021)가 말한 것처럼 EL의 성능은 정확성과 다양성에 기반을 둔다. 그렇다면 'ELQ의 정확성과 다양성은 어떻게 측정될 수 있는가?'가 궁금할 것이다. ELQ를 최초 개발한 K19에서는 성능을 보장하는 지표로써 보상정도(degree of compensation,  $I_{DoC}$ )라는 것을 제시했다. 보상정도는  $I_{DoC} = (1 - r_{mm})$ 으로 정의된다. 여기서  $r$ 은 상관계수(correlation coefficient)이며,  $m$ 과  $n$ 은 결합된 VSs의 조합번호를 말한다. 다시 말해, 보상정도는 1에서 두 기본 학습자의 상관관계를 뺀 값이다. 즉, 두 기본 학습자의 상관관계가 높을수록  $I_{DoC}$ 는 작아져서 성능향상 또한 작아지고, 반대로 두 기본 학습자의 상관관계가 작을수록  $I_{DoC}$ 는 커져서 성능향상이 커진다.

제시된 보상정도 지표 개발의 이론적 토대는 바로 편향-분산 분해(bias-variance decomposition)이다(Krogh and Vedelsby, 1995; Brown et al., 2005b; Kim, 2019). 이론적으로 앙상블의 일반화 오차(generalization error of ensemble)는 두 가지 항, 분산(variance)과 편향의 제곱(squared bias) 항으로 나눌 수 있으며, 식(15)와 같이 나타낼 수 있다.

$$\sum_{j=1}^N w_j (\hat{f}_j - Q)^2 = \sum_{j=1}^N w_j (\hat{f}_j - f_{ELQ})^2 + (f_{ELQ} - Q)^2 \tag{15}$$

여기서  $(f_{ELQ} - Q)$ 는 ELQ 예측값과 관측된 유출량의 차이이며,  $j$ 는 후보 앙상블 함수의 수이다. 식(15) 등식의 왼쪽 항(편향의 제곱)은 각각의 후보 앙상블 함수의 오차의 제곱을 가중 평균한 합(sum of weighted squared errors of the individual candidate ensemble function)이고, 등식의 오른쪽 첫째 항(분산)은 개별 후보 앙상블 함수와 최종 앙상블 함수 사이의 오차의 제곱을 가중 평균한 합(sum of weighted

squared errors between the individual candidate ensemble function and final ensemble function)이다. 여기에서 후자를 ensemble ambiguity라고 한다(Krogh and Vedelsby, 1995). 식(15)를 보면, 편향과 분산의 tradeoff를 발견할 수 있다. 예를 들면,  $f_{ELQ}$ 의 편향이 크면, 식(15)에서 분산항(ensemble ambiguity)이 작아진다. Ensemble ambiguity가 작아진다는 것은 비슷한 함수들이 앙상블에 추가된다고 해도, 추가되는 함수들이 더 이상 앙상블의 성능 향상에 영향을 미치지 못하는 것을 의미한다. 즉, 분산항이 0에 가까워지면, ELQ의 오차는 개별 후보 함수들의 오차의 제곱의 가중평균과 같아지므로 ELQ의 성능향상을 기대하기 어렵다(Krogh and Vedelsby, 1995). 따라서 ELQ의 앙상블 후보함수들 중에서 유사성을 판단하는 기준이 바로 보상정도가 되는 것이다.

후보함수들의 다양성과 더불어 고려해야 할 것은 개별 함수들의 정확성이다. EL이 weak learners에 기반한다고 해도 개별 학습자들의 정확성이 크다면, 그 앙상블도 더욱 정확할 가능성이 높다. 따라서 K21에서는 기본 학습자의 힘(power of base learner,  $P_{BL}$ )이라는 지표를 도입했다.  $P_{BL}$ 은  $f_{ELQ}$ 에서 결합된 개별 학습자들의 NSE 평균으로 정의된다. K21에서는  $P_{BL}$ 이 클수록 하천 유출량 예측성능이 좋았다. 한편, K21에서는  $f_{BEM}$ 을 사용했기 때문에 각 개별학습자들의 가중치는 모두 동일했으나, 다수의  $f_{BEM}$ 을 평균함으로써  $f_{BEM}$ 이  $f_{ELQ}$ 와 같이 다른 가중치를 갖는 함수로 전환됨을 확인하였다. 즉, 다수의  $f_{BEM}$ 을 평균한 함수는  $f_{BEM}$ 의 앙상블 평균(ensemble mean of  $f_{BEM}$ )이라고 K21에서 정의하였으며, 그것은 다시  $f_{ELQ}$ 의 형식으로 전환된다.

$$\overline{f_{BEM}} = \frac{1}{M} \sum_{k=1}^M (f_{BEM})_k = \alpha + w_1 \cdot \hat{f}_{i1} + w_2 \cdot \hat{f}_{i2} + w_3 \cdot \hat{f}_{i3} + \epsilon_i = f_{ELQ} \tag{16}$$

여기서  $\overline{f_{BEM}}$ 는 몇몇  $f_{BEM}$ 의 앙상블 평균이고,  $(f_{BEM})_k$ 는  $k$ 번째  $f_{BEM}$ 이며,  $M$ 은 결합된  $f_{BEM}$ 의 수이다. 식(16)은 다수의  $f_{BEM}$ 을  $P_{BL}$ 과  $I_{DoC}$  기준으로 앙상블 평균을 취하게 되면  $f_{BEM}$ 의 개별 학습자의 가중치가 최초에는 동일하더라도 최종 함수인  $\overline{f_{BEM}}$ 에서는 결합된 개별 학습자의 가중치가 차등을 이루어 좀 더 나은 후보 학습자를 가려 낼 수 있다는 장점이 있다. 따라서 이러한 방법들을 ELQ 기법을 통해 구현해 내면, 미계측 유역(in-situ  $Q$ 가 없는 유역)의 하천 유출량을 예측하는데 큰 도움이 될 수 있을 것이다.

### 3.2. ELQ의 수문학적 프레임워크

앙상블은 개별 학습자보다 일반적으로 뛰어나고 그 성능은 정확도와 다양성에 의해 결정된다. 미계측 유역의 하천 유출량을 산정하는 알고리즘으로 ELQ가 K19에서 소개되었으며, K20과 K21에서 ELQ가 응용되어 적용되었다. 이러한 ELQ 논문 세 편을 인용한 국제논문은 24년 11월까지 76편의 논문에 94회 인용되었다. 본 장에서는 ELQ 논문을 인용한 논문에 대해서 간략히 언급하고, 그에 따라 ELQ가 원격탐사 수문학에 어떠한 기여를 했으며, ELQ가 갖는 수문학적 의의

**Table 2.** List of included journals and the corresponding count of citations

Journal Title	Subject Area	Count
Remote Sensing	Remote Sensing	11
Remote Sensing of Environment	Remote Sensing	1
IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing	Remote Sensing	1
International Journal of Applied Earth Observation and Geoinformation	Remote Sensing	2
Remote Sensing Application: Society and Environment	Remote Sensing	1
International Journal of Digital Earth	Remote Sensing	1
Water	Water	1
Water Resources Management	Water	1
Reviews of Geophysics	Geophysics	1
Geophysical Research Letters	Geophysics	2
Science of The Total Environment	Environmental Sciences	1
Water Resources Research	Environmental Sciences	1
Journal of Hydrology	Environmental Sciences	5
Hydrologic Sciences Journal	Environmental Sciences	3
Environmental Modelling & Software	Environmental Sciences	2
Advances in Space Research	Earth and Planetary Sciences	2
Artificial Intelligence Review	Computer Sciences	1
Not Included	N/A	57
Total		94

에 대해 고찰한다.

### 3.2.1. 원격탐사 수문학에서 ELQ 논문의 인용사례

ELQ (K19)는 원격탐사 관련 저널 중 최고의 권위를 갖고 있는 저널 중 하나인 Remote Sensing of Environment에 2019년 처음 게재되었다. 그 이후 Envisat과 Jason-2의 다중 위성 고도계 미션을 이용하여 메콩강에 적용한 논문인 K20이 Remote Sensing에 2019년 11월에 게재되었다. 또한, 수문학 모델인 HRR과 위성 고도계를 이용하여 콩고강에 적용하여 in-situ Q의 도움없이 유출량을 산정한 논문인 K21은 Advances in Space Research 논문에 게재되었다. 이와 같이 ELQ 관련 세 편의 논문은 2019년 이후 총 76편의 국제 논문에서 94회 인용되었으며, 우리는 Google Scholar와 ResearchGate를 통해 ELQ관련 논문 인용을 확인하고 분석했다. 총 94회의 인용 횟수 중 K19가 60회, K20은 18회, K21은 16회 인용되었다. 각 저널에 인용된 횟수는 Table 2와 같다. ELQ 관련 논문은 각각 원격탐사, 지구물리학, 환경과학 및 컴퓨터 과학 등 폭 넓은 분야의 국제논문으로부터 인용되었다.

ELQ 논문들을 인용한 76편의 논문들을 인용 주제별로 묶으면 크게 다섯 개의 주제로 나눌 수 있다. 그 다섯 개의 주제는 ① 일반주제, ② 머신러닝 및 앙상블 러닝, ③ 원격탐사 방법이나 머신러닝 방법을 이용한 수리변수의 도출, ④ 머신러닝 방법을 이용한 하천 유출량의 예측, ⑤ ELQ이다. 이러한 주제들은 Table 3에 나타나 있다. 일반사항

**Table 3.** Research areas and the corresponding count of citations

Research Area	Count
General / Water strategy	7
ML / EL	10
Inferring hydrological variables using RS and/or ML/EL	31
ML technique to estimate discharge	19
ELQ	9
Total	76

및 물 전략 분야에 7편의 논문, ML/EL에 관한 분야에서 10편의 논문, 수리변수 예측 및 하천 유출량 예측 분야에 각각 31편, 19편 그리고 논문에 ELQ를 직접적으로 언급한 수리·수문학 분야 논문 등이 9편이었다.

ELQ 논문들은 수리·수문학 분야에서 많이 인용되었는데, 앞서 언급한 바와 같이 Irrgang et al. (2020)은 논문에 하천 유출량 예측 기법으로 ELQ를 직접 언급하면서, ELQ (K19)가 최초로 머신러닝과 위성 고도계 데이터를 결합하여 기존 방법보다 우수한 결과를 내놓았다고 평가했다. Akpoti et al. (2024)은 아프리카 대륙의 하천 유출량에 대한 도전과 기회에 대한 논문에서 역시 ELQ를 언급하며, 기존 하천 유출량 예측 기법보다 향상된 방법을 제시했으며, 이러한 ELQ와 같은 기법은 SWOT 미션에서 수집된 데이터를 포함하여 향후 원격탐



사수문학에 많은 도움이 될 것으로 평가했다. Kechnit et al. (2024) 또한 ELQ를 언급하며, ELQ가 하도의 구조를 예측하고, 다양한 수리변수들을 결합하여 하천 유출량을 예측한다고 평가하였으며, 다른 몇몇 논문들도 ELQ를 언급하여 ELQ가 기존 방법과 대비되는 특징들을 기술했다(Garkoti and Kundapura, 2021; Dhote et al., 2021).

또한, 수리·수문학 분야에서 원격탐사 방법을 사용하여 수리변수들을 예측하는 연구논문(31편)에서 ELQ 논문을 주로 인용했다. 그 중 하천의 바닥 등 하도 및 수중의 깊이와 모양을 추정하는 bathymetry 분야(Kwon et al., 2022; Gehring et al., 2022; Póssa et al., 2020)에서 ELQ 논문이 인용되었으며, Bhuyian et al. (2020)은 bathymetry 분야의 연구에서 K19가 여러 위성 고도계를 결합함으로써 데이터의 시간 해상도에 대한 정밀도를 향상시켰다고 평가했다. 또한, ELQ 기법은 수리변수 중 하나인 Total Water Storage (TWS)의 변화(Gyawali et al., 2022; Wang et al., 2021; Yu et al., 2021; Sun et al., 2020), 수표면 면적 및 높이가 변화(Chen et al., 2020; Markert et al., 2019; Carr et al., 2019; Li et al., 2023), 수문학 모델에 대한 검토(Shastri et al., 2020), 기타 수리변수들을 예측하는 리뷰 논문(Uereyen and Kuenzer, 2019; Chawla et al., 2020) 등에 인용되었다.

하천 유출량을 예측하는데 있어 ELQ 기법을 머신러닝을 이용한 새로운 유출량 예측기법으로 평가하였는데, 그 중 Zounemat-Kermani et al. (2021)은 수문학 분야의 앙상블 러닝 패러다임이라는 제목의 리뷰 논문에서 ELQ (K19)를 EML 기법을 사용하여 기존의 방법보다 더 우수한 성능을 가진 알고리즘으로 소개하였다. 그리고 수문학에 적용된 EML 방법들을 언급하면서 앙상블 결합 방식에 따라 분류하였고, ELQ를 stacking 방법에 의한 앙상블 결합방식으로 분류하였다. 또한 원격탐사 데이터와 수문학 모델, 강우 데이터 등을 이용한 유출량 향상 등에 대한 논문들(Fok et al., 2021; 2023; Papacharalampous et al., 2023)에서 ELQ 논문을 인용했다.

여러가지 다양한 주제의 논문들에서 ML/EL 기법을 적용한 연구로 ELQ 논문을 인용하였는데, 원격탐사를 통한 토양 염분 분류(Chen et al., 2024), 머신러닝 기법 중 boosting에 대한 연구(Kouadio et al., 2023) 등 다양한 ML/EL 기법 연구(Naitam et al., 2023; Shahhosseini et al., 2022; Shi et al., 2020; Wang et al., 2020)에서도 ELQ 관련 논문을 인용했다. 이러한 ELQ 기법은 수리·수문학 뿐만 아니라 컴퓨터 과학 분야 논문에서도 인용되었는데 Artificial Intelligence Review에 게재된 Chen et al. (2024)의 논문에서는 단일 rating curve보다 다수의 rating curve를 결합하여 만든 ELQ의 성능에 대해 평가하며 imbalanced learning의 최신 연구동향 등을 제시하였다. 이 밖에 Lawford et al. (2023a; 2023b)은 물 전략 분야에 있어 공유하천의 경우에 다양한 국가를 지나가기 때문에 다중 고도계 사용이 유용하다고 판단했다.

### 3.2.2. 원격탐사 수문학에서 ELQ의 학문적 의의

원격탐사 수문학은 주로 위성에 의한 원격탐사 방법을 이용하여 수리·수문학 연구에 적용하는 학문으로써, 주로 앞서 언급한 수리학적 변수들을 결합할 때 그 변수들을 원격탐사로부터 획득한 데이터를 활용하여 하천 유출량 등을 산정하였다. 그 후 Gleason and Smith (2014)가 AMHG 기법을 사용하여 in-situ  $Q$  데이터 없이도 하천 유출량을 산정하는 방법을 제안했다. 그러나 AMHG 기법은 in-situ  $Q$  데이터를 사용하지 않은 만큼, 측정된 유출량과의 정확도는 약 20-30%의 차이가 있었으며, 주로 광학 영상에 의한 하천의 유효 폭을 관측함으로써 유량에 따라 하천 폭의 변화가 작은 하천에 대해서는 부정확할 수 있다. 한편, K20에서는 메콩강에서 Stung Treng, Kratie, Tan Chau 세 곳에서 AMHG 기법에서 사용한 두 가지 변수의 관계( $d = cQ^h$ : rating curve에서  $\log c$ 와  $h$ 의 관계)를 그래프로 표현하였는데, 그 두 가지 변수의 결정계수는 0.94-0.98의 범위에 있었다. 이러한 높은 결정계수는 AMHG의 장점에도 불구하고, 하천 유출량의 정확도 향상에는 AMHG가 기여하는 바가 작다는 것을 의미한다.

ELQ (K19)가 원격탐사 수문학에 가장 큰 영향을 미친 것은 지리적으로 수 백 km 떨어진 지역의 위성 고도계 정보를 결합하여 한 지점의 하천 유출량을 더욱 정확하게 예측한다는 점이다. 이것은 Tobler의 지리학 제 1법칙(The first law of geography)에 정면으로 위배되는 논리로서, Tobler (1970)는 그의 논문에서 다음과 같이 말했다. "Everything is related to everything else, but near things are more related than distant things." 즉, 모든 것은 연관되어 있지만 가까운 것이 먼 것보다 더욱 연관되어 있다는 지리학 1법칙에 의해, 원격탐사 수문학에서도 하천 유출량은 주로 관측지점에서 매우 가까운 곳이나 주로 수 km에서 수십 km 내외의 데이터만을 사용하여 하천 유출량을 예측하였다. 이것은 Leopold and Maddock (1953)이 사용한 방법인 관측지점에서 매우 가까운 곳(AHG)이나 하류의 관측지점과 연계되는 곳(DHG)에서 rating curve를 이용하거나, Manning's equation을 이용하여 각종 수리변수들을 곱하는 방식으로 한 지점에서 하천 유출량을 산정하였다. 그 이후, 원격탐사 수문학에서도 그러한 방식들을 원격탐사 데이터로 치환하는 정도의 변형이 이루어졌다. 그러나 K19에서는 관측지점에서 수리변수들을 획득하는 것이 아니라, 심지어 수 백 km 떨어진 강의 하천 높이들을 한 식에 선형적으로 결합함으로써 더 나은 성능을 보여주었고, 이것은 원격탐사 수문학에서 기존의 패러다임을 바꾸는 결과였다. 현재까지도 이러한 방식으로 수 백 km 떨어진 수리변수들을 결합하여 더 나은 하천 유출량을 예측하는 것은 매우 드물다. 이는 ML/EL 기법을 단순히 적용하는 것이 아니라, ML/EL 기법을 적용하는 기본 학습자의 시공간적인 도메인이 기존의 도메인과는 전혀 다른 특성을 가진 것이 ELQ의 특징이다.

## 4. 결론

ELQ (K19)는 세계 최초로 ML 기법을 위성 고도계 데이터를 이용하여 원격탐사 수문학에 적용한 사례이다(Irrgang et al., 2020). K19가 ML 기법을 적용한 최초의 사례라는 사실보다 더 중요한 것은 기존 원격탐사 수문학이 머물러 있던 원칙, 즉 가까운 것의 데이터를 이용하여 단일 rating curve나 Manning's equation을 이용하여 하천 유출량을 산정하는 틀에서 벗어난 것에 대한 그 의미가 더욱 크다. K19 이후 K20에서는 한 걸음 더 나아가 다중 위성 고도계를 사용함으로써, 그 예측 가능 범위를 넓혔고, 또한 아프리카 콩고강 이후 동남아시아의 메콩강, TSL 때문에 복잡한 수문 환경을 가진 하천에서 보다 정확한 하천 유출량을 산정하였다. 또한, 메콩강과 같은 공유하천에서는 하천 유출량 데이터가 다른 나라와 공유되지 않는 경우도 빈번히 발생하기 때문에 각국의 물 전략에 있어서도 ELQ가 중요한 역할을 할 수 있다(Lawford et al., 2023a; 2023b).

K21에서는 in-situ  $Q$  가 존재하지 않는 상황에서도 HRR 수문모델과 결합하여 훨씬 정확한 ELQ 기법을 개발했으며, 그 과정에서 ELQ 성능에 영향을 미치는 성능지표를 개발했다. 앙상블 러닝에서 성능에 중요한 영향을 미치는 것은 개별 학습자의 정확성과 다양성이고, 이것은 ELQ에서  $P_{BL}$ 과  $I_{Doc}$ 로 확인할 수 있다. 또한, in-situ  $Q$  데이터를 획득할 수 없을 때에는  $f_{BEM}$  몇몇  $f_{BEM}$ 의 앙상블 평균인  $\overline{f_{BEM}}$ 을 통하여 정확한 ELQ 결과를 얻을 수 있었다.

향후 ELQ 연구에서는 현재 아프리카 콩고강과 동남아시아 메콩강 외에 다른 하천을 대상으로 ELQ를 적용하는 것이 필요하며, 기존 수문학 모델을 검보정하는 방식으로 다중 위성 고도계 데이터를 사용하는 것 또한 연구 과제 중 하나가 될 수 있다(Dhote et al., 2021; Du et al., 2020). 또한 지금까지 ELQ는 동종 학습자( $H-Q$  relationship)를 사용하여 하천 유출량을 추정하였으나, 최근 SWOT 미션 및 소형 SAR 군집 위성 등이 다양하게 개발됨에 따라 다양한 수리변수들을 활용한 이종 학습자( $H-Q / W_c-Q / S-Q$  relationships)를 기본 학습자로 한 ELQ 모델도 추후 개발되어야 할 것이다. 그리고 무엇보다도 in-situ  $Q$  데이터의 도움없이 하천 유출량을 예측하는 것 또한 도전적인 과제으로써, 이는 ELQ와 수문학 모델, AMHG 등 다양한 알고리즘과 모델, 데이터의 결합을 필요로 할 것이다. ELQ는 지금까지 원격탐사 수문학에 여러방면으로 기여했지만, ELQ 기법만으로는 하천 유출량의 정확한 예측은 쉽지 않다. 그러나 ELQ는 계산적으로 간단하면서 비교적 정확한 하천 유출량을 산정할 수 있는 기법 중 하나이다. 따라서 ELQ의 장점을 잘 보존하면서 그 알고리즘을 개선해 나갈 수 있는 연구가 추가적으로 필요하다.

## 사사

본 논문은 육군3사관학교 부설 충성대연구소 2024년도 국고연구비 지원(24-11) 및 2025년도 논문게재비 지원을 받았다.

## Conflict of Interest

No potential conflict of interest relevant to this article was reported.

## Supplementary Materials

The supplementary materials for this review article can be found online at [https://github.com/dkim-32/Review\\_ELQ](https://github.com/dkim-32/Review_ELQ).

## References

- Akpoti, K., Mekonnen, K., Leh, M., Owusu, A., Dembélé, M., and Tinonetsana, P., 2024. State of continental discharge estimation and modelling: Challenges and opportunities for Africa. *Hydrological Sciences Journal*, 69(15), 2124–2152. <https://doi.org/10.1080/02626667.2024.2402938>
- Alsdorf, D., Beighley, E., Laraque, A., Lee, H., Tshimanga, R. M., O'Loughlin, E., et al., 2016. Opportunities for hydrologic research in the Congo Basin. *Reviews of Geophysics*, 54(2), 378–409. <https://doi.org/10.1002/2016RG000517>
- Andreadis, K. M., Clark, E. A., Lettenmaier, D. P., and Alsdorf, D. E., 2007. Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model. *Geophysical Research Letters*, 34(10), L10403. <https://doi.org/10.1029/2007GL029721>
- Beighley, R. E., and Gummadi, V., 2011. Developing channel and floodplain dimensions with limited data: A case study in the Amazon Basin. *Earth Surface Processes and Landforms*, 36(8), 1059–1071. <https://doi.org/10.1002/esp.2132>
- Beighley, R. E., Ray, R. L., He, Y., Lee, H., Schaller, L., Andreadis, K. M., 2011. Comparing satellite derived precipitation datasets using the Hillslope River Routing (HRR) model in the Congo River Basin. *Hydrological Processes*, 25(20), 3216–3229. <https://doi.org/10.1002/hyp.8045>
- Bhuyian, M. N., and Kalyanapu, A., 2020. Predicting channel conveyance and characterizing planform using river bathymetry via satellite image compilation (RiBaSIC) algorithm for DEM-based hydrodynamic modeling. *Remote Sensing*, 12(17), 2799. <https://doi.org/10.3390/rs12172799>

- Bjerklie, D. M., Moller, D., Smith, L. C., and Dingman, S. L., 2005. Estimating discharge in rivers using remotely sensed hydraulic information. *Journal of Hydrology*, 309(1), 191–209. <https://doi.org/10.1016/j.jhydrol.2004.11.022>
- Bogning, S., Frappart, F., Blarel, F., Niño, F., Mahé, G., and Bricquet, J. P., 2018. Monitoring water levels and discharges using radar altimetry in an ungauged river basin: The case of the Ogooué. *Remote Sensing*, 10(2), 350. <https://doi.org/10.3390/rs10020350>
- Brakenridge, G. R., Cohen, S., Kettner, A. J., De Groeve, T., Nghiem, S. V., and Syvitski, J. P., 2012. Calibration of satellite measurements of river discharge using a global hydrology model. *Journal of Hydrology*, 475, 123–136. <https://doi.org/10.1016/j.jhydrol.2012.09.035>
- Brown, G., Wyatt, J. L., Harris, R., and Yao, X., 2005a. Diversity creation methods: A survey and categorisation. *Information Fusion*, 6(1), 5–20. <https://doi.org/10.1016/j.inffus.2004.04.004>
- Brown, G., Wyatt, J. L., and Tiño, P., 2005b. Managing diversity in regression ensembles. *Journal of Machine Learning Research*, 6, 1621–1650.
- Campbell, I. C., Say, S., and Beardall, J., 2009. Tonle Sap Lake, the heart of the lower Mekong. In: Campbell, I. C. (ed.), *Aquatic Ecology, The Mekong*, Academic Press, pp. 251–272. <https://doi.org/10.1016/B978-0-12-374026-7.00010-3>
- Carr, A. B., Trigg, M. A., Tshimanga, R. M., Borman, D. J., and Smith, M. W., 2019. Greater water surface variability revealed by new Congo River field data: Implications for satellite altimetry measurements of large rivers. *Geophysical Research Letters*, 46(14), 8093–8101. <https://doi.org/10.1029/2019GL083720>
- Chang, C. H., Lee, H., Hossain, F., Basnayake, S., Jayasinghe, S., Chishtie, F., et al., 2019. A model-aided satellite-altimetry-based flood forecasting system for the Mekong River. *Environmental Modelling & Software*, 112, 112–127. <https://doi.org/10.1016/j.envsoft.2018.11.017>
- Chang, C. H., Lee, H., Kim, D., Hwang, E., Hossain, F., Chishtie, F., et al., 2020. Hindcast and forecast of daily inundation extents using satellite SAR and altimetry data with rotated empirical orthogonal function analysis: Case study in Tonle Sap Lake Floodplain. *Remote Sensing of Environment*, 241, 111732. <https://doi.org/10.1016/j.rse.2020.111732>
- Chawla, I., Karthikeyan, L., and Mishra, A. K., 2020. A review of remote sensing applications for water security: Quantity, quality, and extremes. *Journal of Hydrology*, 585, 124826. <https://doi.org/10.1016/j.jhydrol.2020.124826>
- Chen, J., Liao, J., and Wang, C., 2020. Improved lake level estimation from radar altimeter using an automatic multiscale-based peak detection retracker. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 1246–1259. <https://doi.org/10.1109/JSTARS.2020.3035686>
- Chen, H., Wu, J., and Xu, C., 2024. Monitoring soil salinity classes through remote sensing-based ensemble learning concept: Considering scale effects. *Remote Sensing*, 16(4), 642. <https://doi.org/10.3390/rs16040642>
- Chen, W., Yang, K., Yu, Z., Shi, Y., and Chen, C. L., 2024. A survey on imbalanced learning: latest research, applications and future directions. *Artificial Intelligence Review*, 57, 137. <https://doi.org/10.1007/s10462-024-10759-6>
- Chow, V. T., 1959. Open-channel flow and its classifications. In: Chow, V. T. (ed.), *Open Channel Hydraulics*, McGraw-Hill, pp. 3–18.
- Coe, M. T., and Birkett, C. M., 2004. Calculation of river discharge and prediction of lake height from satellite radar altimetry: Example for the Lake Chad Basin. *Water Resources Research*, 40(10), W10205. <https://doi.org/10.1029/2003WR002543>
- Collins, E. L., David, C. H., Riggs, R., Allen, G. H., Pavelsky, T. M., Lin, P., et al., 2024. Global patterns in river water storage dependent on residence time. *Nature Geoscience*, 17, 433–439. <https://doi.org/10.1038/s41561-024-01421-5>
- Dhote, P. R., Thakur, P. K., Domeneghetti, A., Chouksey, A., Garg, V., Aggarwal, S. P., et al., 2021. The use of SARAL/AltiKa altimeter measurements for multi-site hydrodynamic model validation and rating curves estimation: An application to Brahmaputra River. *Advances in Space Research*, 68(2), 691–702. <https://doi.org/10.1016/j.asr.2020.05.012>
- Dingman, S. L., 2007. Analytical derivation of at-a-station hydraulic-geometry relations. *Journal of Hydrology*, 334(1), 17–27. <https://doi.org/10.1016/j.jhydrol.2006.09.021>
- Du, T. L., Lee, H., Bui, D. D., Arheimer, B., Li, H. Y., Olsson, J., et al., 2020. Streamflow prediction in “geopolitically ungauged” basins using satellite observations and regionalization at subcontinental scale. *Journal of Hydrology*, 588, 125016. <https://doi.org/10.1016/j.jhydrol.2020.125016>
- Durand, M., Gleason, C. J., Garambois, P. A., Bjerklie, D., Smith, L. C., Roux, H., et al., 2016. An intercomparison of remote sensing river discharge estimation algorithms from measurements of river height, width, and slope. *Water Resources Research*, 52(6), 4527–4549. <https://doi.org/10.1002/2015WR018434>
- Fok, H. S., Chen, Y., Ma, Z., Ferreira, V. G., and Tenzer, R., 2023. Geographically-weighted water balance approach for satellite-hydrologic runoff estimation in Mekong Basin under ENSO. *International Journal of Applied Earth Observation and Geoinformation*, 118, 103234. <https://doi.org/10.1016/j.jag.2023.103234>



- Fok, H. S., Chen, Y., Wang, L., Tenzer, R., and He, Q., 2021. Improved Mekong Basin runoff estimate and its error characteristics using pure remotely sensed data products. *Remote Sensing*, 13(5), 996. <https://doi.org/10.3390/rs13050996>
- Frappart, F., Calmant, S., Cauhopé, M., Seyler, F., and Cazenave, A., 2006. Preliminary results of ENVISAT RA-2-derived water levels validation over the Amazon Basin. *Remote sensing of Environment*, 100(2), 252–264. <https://doi.org/10.1016/j.rse.2005.10.027>
- Freund, Y., and Schapire, R. E., 1995. A decision-theoretic generalization of on-line learning and an application to boosting. In: Vitányi, P. (ed.), *Computational Learning Theory*, Springer, pp. 23–37. [https://doi.org/10.1007/3-540-59119-2\\_166](https://doi.org/10.1007/3-540-59119-2_166)
- Garkoti, A., and Kundapura, S., 2021. Deriving water level and discharge estimation using satellite altimetry for Krishna River, Karnataka. *Remote Sensing Applications: Society and Environment*, 22, 100487. <https://doi.org/10.1016/j.rsase.2021.100487>
- Gehring, J., Duvvuri, B., and Beighley, E., 2022. Deriving River discharge using remotely sensed water surface characteristics and satellite altimetry in the Mississippi River Basin. *Remote Sensing*, 14(15), 3541. <https://doi.org/10.3390/rs14153541>
- Gleason, C. J., and Smith, L. C., 2014. Toward global mapping of river discharge using satellite images and at-many-stations hydraulic geometry. *Proceedings of the National Academy of Sciences*, 111(13), 4788–4791. <https://doi.org/10.1073/pnas.1317606111>
- Gleason, C. J., Smith, L. C., and Lee, J., 2014. Retrieval of river discharge solely from satellite imagery and at-many-stations hydraulic geometry: Sensitivity to river form and optimization parameters. *Water Resources Research*, 50(12), 9604–9619. <https://doi.org/10.1002/2014WR016109>
- Gyawali, B., Murgulet, D., and Ahmed, M., 2022. Quantifying changes in groundwater storage and response to hydroclimatic extremes in a coastal aquifer using remote sensing and ground-based measurements: The Texas Gulf Coast aquifer. *Remote Sensing*, 14(3), 612. <https://doi.org/10.3390/rs14030612>
- Han, D., Kim, Y. J., Jung, S., Sim, S., Kim, W., Jang, E., et al., 2021. Detection of Arctic Summer melt ponds using ICESat-2 altimetry data. *Korean Journal of Remote Sensing*, 37(5–1), 1177–1186. <https://doi.org/10.7780/kjrs.2021.37.5.1.27>
- Huang, Q., Long, D., Du, M., Zeng, C., Qiao, G., Li, X., et al., 2018. Discharge estimation in high-mountain regions with improved methods using multisource remote sensing: A case study of the Upper Brahmaputra River. *Remote Sensing of Environment*, 219, 115–134. <https://doi.org/10.1016/j.rse.2018.10.008>
- Hwang, D. H., Bak, S., Jeong, M. J., Kim, N. K., Park, M. S., Kim, B. R., et al., 2021. Study on dimensionality reduction for sea-level variations by using altimetry data around the East Asia coasts. *Korean Journal of Remote Sensing*, 37(1), 85–95. <https://doi.org/10.7780/kjrs.2021.37.1.7>
- Hwang, D. H., Jung, H. C., and Lee, H., 2023. Estimation of coastal area and lake water level changes using high-resolution altimetry. *GEO DATA*, 5(3), 155–160. <http://doi.org/10.22761/GD.2023.0030>
- Irrgang, C., Saynisch-Wagner, J., Dill, R., Boergens, E., and Thomas, M., 2020. Self-validating deep learning for recovering terrestrial water storage from gravity and altimetry measurements. *Geophysical Research Letters*, 47(17), e2020GL089258. <https://doi.org/10.1029/2020GL089258>
- Jung, D., Paik, K., and Kim, J. H., 2013. Relationship between downstream hydraulic geometry and suspended sediment concentration characteristics. *Journal of Hydro-Environment Research*, 7(4), 243–252. <https://doi.org/10.1016/j.jher.2013.03.002>
- Jung, H. C., Hamski, J., Durand, M., Alsdorf, D., Hossain, F., Lee, H., et al., 2010. Characterization of complex fluvial systems using remote sensing of spatial and temporal water level variations in the Amazon, Congo, and Brahmaputra Rivers. *Earth Surface Processes and Landforms*, 35(3), 294–304. <https://doi.org/10.1002/esp.1914>
- Kearns, M., and Valiant, L. G., 1989. Cryptographic limitations on learning Boolean formulae and finite automata. In *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing*, New York, NY, USA, May 14–17, pp. 433–444. <https://doi.org/10.1145/73007.73049>
- Kearns, M. J., and Valiant, L. G., 1988. *Learning Boolean formulae or finite automata is as hard as factoring* (Technical Report TR-14-88). Harvard University Aiken Computation Laboratory.
- Kechnit, D., Tshimanga, R. M., Ammari, A., Trigg, M. A., Carr, A. B., Bahmanpouri, F., et al., 2024. Bathymetry and discharge estimation in large and data-scarce rivers using an entropy-based approach. *Hydrological Sciences Journal*, 69(15), 2109–2123. <https://doi.org/10.1080/02626667.2024.2402933>
- Kim, D., 2019. *Ensemble learning regression for estimating river discharge using remotely sensed data and hydrological model*. Doctoral dissertation, University of Houston, Houston, TX, USA.
- Kim, D., 2021. Opportunities for gravity and GNSS surveying to monitor changes in the terrestrial water storage in the Congo River Basin. *Remote Sensing in Earth Systems Sciences*, 4(3), 121–124. <https://doi.org/10.1007/s41976-021-00049-y>
- Kim, D., and Jung, H. C., 2021. The characteristics of submarine groundwater discharge in the coastal area of Nakdong River Basin. *Korean Journal of Remote Sensing*, 37(6–1), 1589–1597.

- <https://doi.org/10.7780/kjrs.2021.37.6.1.8>
- Kim, D., Lee, H., Beighley, E., and Tshimanga, R. M., 2021. Estimating discharges for poorly gauged river basin using ensemble learning regression with satellite altimetry data and a hydrologic model. *Advances in Space Research*, 68(2), 607–618. <https://doi.org/10.1016/j.asr.2019.08.018>
- Kim, D., Lee, H., Chang, C. H., Bui, D. D., Jayasinghe, S., Basnayake, S., et al., 2019b. Daily river discharge estimation using multi-mission radar altimetry data and ensemble learning regression in the lower Mekong River Basin. *Remote Sensing*, 11(22), 2684. <https://doi.org/10.3390/rs11222684>
- Kim, D., Lee, H., Jung, H. C., Hwang, E., Hossain, F., Bonnema, M., et al., 2020. Monitoring river basin development and variation in water resources in transboundary Imjin River in North and South Korea using remote sensing. *Remote Sensing*, 12(1), 195. <https://doi.org/10.3390/rs12010195>
- Kim, D., Lee, H., Laraque, A., Tshimanga, R. M., Yuan, T., Jung, H. C., et al., 2017a. Mapping spatio-temporal water level variations over the central Congo River using PALSAR ScanSAR and Envisat altimetry data. *International Journal of Remote Sensing*, 38(23), 7021–7040. <https://doi.org/10.1080/01431161.2017.1371867>
- Kim, D., Lee, H., Yu, H., Beighley, E., Durand, M. T., Alsdorf, D. E., and Hwang, E., 2017b. Weak learner method for estimating river discharges using remotely sensed data: Central Congo River as a testbed. In *Proceedings of the 2017 AGU Fall Meeting Abstracts*, New Orleans, LA, USA, Dec. 11–15, Abstract Number H33F-1973.
- Kim, D., Yu, H., Lee, H., Beighley, E., Durand, M., Alsdorf, D. E., et al., 2019a. Ensemble learning regression for estimating river discharges using satellite altimetry data: Central Congo River as a Test-bed. *Remote Sensing of Environment*, 221, 741–755. <https://doi.org/10.1016/j.rse.2018.12.010>
- Kim, D. Y., Kim, D., Jeong, Y., and Jo, Y. H., 2024. Validation of spatial boundary of the Ulleung warm eddy using altimetry. *Korean Journal of Remote Sensing*, 40(6–1), 1019–1026. <https://doi.org/10.7780/kjrs.2024.40.6.1.12>
- Kim, H. G., Kim, J. I., Yoon, S. J., and Kim, T., 2018. Development of a method for calculating the allowable storage capacity of rivers by using drone images. *Korean Journal of Remote Sensing*, 34(2–1), 203–211. <https://doi.org/10.7780/kjrs.2018.34.2.1.4>
- Kouadio, K. L., Liu, J., Kouamelan, S. K., and Liu, R., 2023. Ensemble learning paradigms for flow rate prediction boosting. *Water Resources Management*, 37(11), 4413–4431. <https://doi.org/10.1007/s11269-023-03562-5>
- Kouraev, A. V., Zakharova, E. A., Samain, O., Mognard, N. M., and Cazenave, A., 2004. Ob'river discharge from TOPEX/Poseidon satellite altimetry (1992–2002). *Remote Sensing of Environment*, 93(1), 238–245. <https://doi.org/10.1016/j.rse.2004.07.007>
- Krogh, A., and Vedelsby, J., 1995. Neural network ensembles, cross validation, and active learning. In: Tesauro, G., Touretzky, D., and Leen, T. (eds.), *Advances in Neural Information Processing Systems*, MIT Press, pp. 231–238.
- Kwon, S., Seo, I. W., Noh, H., and Kim, B., 2022. Hyperspectral retrievals of suspended sediment using cluster-based machine learning regression in shallow waters. *Science of the Total Environment*, 833, 155168. <https://doi.org/10.1016/j.scitotenv.2022.155168>
- Lawford, R., Unninayar, S., Huffman, G. J., Grabs, W., Gutiérrez, A., Ishida-Watanabe, C., et al., 2023a. Implementing the GEOSS water strategy: from observations to decisions. *International Journal of Digital Earth*, 16(1), 1439–1468. <https://doi.org/10.1080/17538947.2023.2202420>
- Lawford, R., Unninayar, S., Huffman, G. J., Grabs, W., Gutiérrez, A., and Koike, T., 2023b. A data-oriented strategy to support water resource managers and researchers. *Journal of the American Water Resources Association*, 59(5), 877–884. <https://doi.org/10.1111/1752-1688.13126>
- Lee, H., Shum, C. K., Yi, Y., Ibaraki, M., Kim, J. W., Braun, A., et al., 2009. Louisiana wetland water level monitoring using retracked TOPEX/POSEIDON altimetry. *Marine Geodesy*, 32(3), 284–302. <https://doi.org/10.1080/01490410903094767>
- Lee, J., 2017. Review of remote sensing studies on groundwater resources. *Korean Journal of Remote Sensing*, 33(5–3), 855–866. <https://doi.org/10.7780/kjrs.2017.33.5.3.8>
- Lee, J., and Yoo, C., 2024. Comparing approaches for obtaining downstream hydraulic geometry in two Korean basins: Focusing on peak flow velocity. *Journal of Hydrology*, 635, 131156. <https://doi.org/10.1016/j.jhydrol.2024.131156>
- LeFavour, G., and Alsdorf, D., 2005. Water slope and discharge in the Amazon River estimated using the shuttle radar topography mission digital elevation model. *Geophysical Research Letters*, 32(17), L17404. <https://doi.org/10.1029/2005GL023836>, 2005
- Leopold, L. B., and Maddock, T., 1953. *The hydraulic geometry of stream channels and some physiographic implications*. US Government Printing Office.
- Li, S., and Yang, J., 2023. Improved river water-stage forecasts by ensemble learning. *Engineering with Computers*, 39(5), 3293–3311. <https://doi.org/10.1007/s00366-022-01751-1>
- Liu, G., Schwartz, F. W., Tseng, K. H., and Shum, C. K., 2015. Discharge and water-depth estimates for ungauged rivers: Combining hydrologic, hydraulic, and inverse modeling with stage and water-

- area measurements from satellites. *Water Resources Research*, 51(8), 6017–6035. <https://doi.org/10.1002/2015WR016971>
- Loukas, A., and Vasilades, L., 2014. Streamflow simulation methods for ungauged and poorly gauged watersheds. *Natural Hazards and Earth System Sciences*, 14(7), 1641–1661. <https://doi.org/10.5194/nhess-14-1641-2014>
- Lu, M., Hou, Q., Qin, S., Zhou, L., Hua, D., Wang, X., et al., 2023. A stacking ensemble model of various machine learning models for daily runoff forecasting. *Water*, 15(7), 1265. <https://doi.org/10.3390/w15071265>
- Manning, R., 1889. On the flow of water in open channels and pipes. *Transactions of the Institution of Civil Engineers of Ireland*, 20, 161–207.
- Margineantu, D. D., and Dietterich, T. G., 1997. Pruning adaptive boosting. In *Proceedings of the Fourteenth International Conference on Machine Learning*, San Francisco, CA, USA, July 8–12, pp. 211–218. <https://dl.acm.org/doi/10.5555/645526.757762>
- Markert, K. N., Pulla, S. T., Lee, H., Markert, A. M., Anderson, E. R., Okeowo, M. A., et al., 2019. AltEx: An open source web application and toolkit for accessing and exploring altimetry datasets. *Environmental Modelling & Software*, 117, 164–175. <https://doi.org/10.1016/j.envsoft.2019.03.021>
- Meddage, D. P. P., Ekanayake, I. U., Herath, S., Gobirahavan, R., Muttill, N., and Rathnayake, U., 2022. Predicting bulk average velocity with rigid vegetation in open channels using tree-based machine learning: A novel approach using explainable artificial intelligence. *Sensors*, 22(12), 4398. <https://doi.org/10.3390/s22124398>
- Mendes-Moreira, J., Soares, C., Jorge, A. M., and Sousa, J. F. D., 2012. Ensemble approaches for regression: A survey. *ACM Computing Surveys (CSUR)*, 45(1), 1–40. <https://doi.org/10.1145/2379776.2379786>
- Mekong River Commission, 2009. *MRC management information booklet series No.2: The flow of the Mekong*. MRC Secretariat. <https://www.mrcmekong.org/publications/the-flow-of-the-mekong/>
- Naitam, A., Meghana, N., and Srivastav, R., 2023. Multimodel classification and regression technique for the statistical downscaling of temperature. *Stochastic Environmental Research and Risk Assessment*, 37(10), 3707–3729. <https://doi.org/10.1007/s00477-023-02472-7>
- Nijssen, B., O'Donnell, G. M., Lettenmaier, D. P., Lohmann, D., and Wood, E. F., 2001. Predicting the discharge of global rivers. *Journal of Climate*, 14(15), 3307–3323. [https://doi.org/10.1175/1520-0442\(2001\)014%3C3307:PTDOGR%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014%3C3307:PTDOGR%3E2.0.CO;2)
- Normandin, C., Frappart, F., Diepkilé, A. T., Marieu, V., Mougin, E., and Blarel, F., 2018. Evolution of the performances of radar altimetry missions from ERS-2 to Sentinel-3A over the inner Niger Delta. *Remote Sensing*, 10(6), 833. <https://doi.org/10.3390/rs10060833>
- Olson, K. R., and Morton, L. W., 2018. Tonle Sap Lake and river and confluence with the Mekong River in Cambodia. *Journal of Soil and Water Conservation*, 73(3), 60A–66A. <https://doi.org/10.2489/jswc.73.3.60A>
- Papacharalampous, G., Tyralis, H., Doulamis, A., and Doulamis, N., 2023. Comparison of tree-based ensemble algorithms for merging satellite and earth-observed precipitation data at the daily time scale. *Hydrology*, 10(2), 50. <https://doi.org/10.3390/hydrology10020050>
- Paris, A., Dias de Paiva, R., Santos da Silva, J., Medeiros Moreira, D., Calmant, S., Garambois, P. A., et al., 2016. Stage-discharge rating curves based on satellite altimetry and modeled discharge in the Amazon Basin. *Water Resources Research*, 52(5), 3787–3814. <https://doi.org/10.1002/2014WR016618>
- Park, S. J., and Lee, C. W., 2018. Monitoring of the drought in the upstream area of Soyang River, Inje-Gun, Kangwon-do using KOMPSAT-2/3 satellite. *Korean Journal of Remote Sensing*, 34(6–3), 1319–1327. <https://doi.org/10.7780/kjrs.2018.34.6.3.2>
- Park, S. J., and Lee, C. W., 2018. Simulation of the flood damage area of the Imjin River Basin in the case of North Korea's Hwanggang Dam discharge. *Korean Journal of Remote Sensing*, 34(6–1), 1033–1039. <https://doi.org/10.7780/kjrs.2018.34.6.1.15>
- Perrone, M. P., and Cooper, L. N., 1992. When networks disagree: Ensemble methods for hybrid neural networks. In *How we learn; How we remember: Toward an understanding of brain and neural systems: Selected papers of Leon N Cooper*, pp. 342–358. [https://doi.org/10.1142/9789812795885\\_0025](https://doi.org/10.1142/9789812795885_0025)
- Pössa, É. M., Maillard, P., and de Oliveira, L. M., 2020. Discharge estimation for medium-sized river using multi-temporal remote sensing data: a case study in Brazil. *Hydrological Sciences Journal*, 65(14), 2402–2418. <https://doi.org/10.1080/02626667.2020.1808220>
- Schapiro, R. E., and Freund, Y., 2012. Introduction and overview. In: Schapiro, R. E., and Freund, Y. (eds.), *Boosting: Foundations and Algorithms*, MIT Press, pp. 1–20. <https://doi.org/10.7551/mitpress/8291.003.0004>
- Scherelis, V., Doering, M., and Laube, P., 2023. HydroWidth: A small-scale approach to calculate river width and its variability. *Transactions in GIS*, 27(5), 1503–1525. <https://doi.org/10.1111/tgis.13083>
- Shahhosseini, M., Hu, G., and Pham, H., 2022. Optimizing ensemble weights and hyperparameters of machine learning models for regression problems. *Machine Learning with Applications*, 7,



100251. <https://doi.org/10.1016/j.mlwa.2022.100251>
- Shastry, A., Durand, M., Neal, J., Fernández, A., Phang, S. C., Mohr, B., et al., 2020. Small-scale anthropogenic changes impact floodplain hydraulics: Simulating the effects of fish canals on the Logone floodplain. *Journal of Hydrology*, 588, 125035. <https://doi.org/10.1016/j.jhydrol.2020.125035>
- Shi, Z., Chen, Y., Liu, Q., and Huang, C., 2020. Discharge estimation using harmonized Landsat and Sentinel-2 product: Case studies in the Murray Darling Basin. *Remote Sensing*, 12(17), 2810. <https://doi.org/10.3390/rs12172810>
- Shiklomanov, A. I., Lammers, R. B., and Vörösmarty, C. J., 2002. Widespread decline in hydrological monitoring threatens pan-Arctic research. *Eos, Transactions American Geophysical Union*, 83(2), 13–17. <https://doi.org/10.1029/2002EO000007>
- Sichangi, A. W., Wang, L., Yang, K., Chen, D., Wang, Z., Li, X., et al., 2016. Estimating continental river basin discharges using multiple remote sensing data sets. *Remote Sensing of Environment*, 179, 36–53. <https://doi.org/10.1016/j.rse.2016.03.019>
- Smith, L. C., and Pavelsky, T. M., 2008. Estimation of river discharge, propagation speed, and hydraulic geometry from space: Lena River, Siberia. *Water Resources Research*, 44(3), W03427. <https://doi.org/10.1029/2007WR006133>
- Son, A., and Kim, J., 2019a. The analysis of flood in an ungauged watershed using remotely sensed and geospatial datasets (I) - Focus on estimation of flood discharge. *Korean Journal of Remote Sensing*, 35(5–2), 781–796. <https://doi.org/10.7780/kjrs.2019.35.5.2.3>
- Son, A., and Kim, J., 2019b. The Analysis of Flood in an Ungauged Watershed using Remotely Sensed and Geospatial Datasets (II) - Focus on Estimation of Flood Inundation. *Korean Journal of Remote Sensing*, 35(5–2), 797–808. <https://doi.org/10.7780/kjrs.2019.35.5.2.4>
- Sun, Z., Long, D., Yang, W., Li, X., and Pan, Y., 2020. Reconstruction of GRACE data on changes in total water storage over the global land surface and 60 basins. *Water Resources Research*, 56(4), e2019WR026250. <https://doi.org/10.1029/2019WR026250>
- Tarpanelli, A., Amarnath, G., Brocca, L., Massari, C., and Moramarco, T., 2017. Discharge estimation and forecasting by MODIS and altimetry data in Niger-Benue River. *Remote Sensing of Environment*, 195, 96–106. <https://doi.org/10.1016/j.rse.2017.04.015>
- Tarpanelli, A., Barbetta, S., Brocca, L., and Moramarco, T., 2013. River discharge estimation by using altimetry data and simplified flood routing modeling. *Remote Sensing*, 5(9), 4145–4162. <https://doi.org/10.3390/rs5094145>
- Tarpanelli, A., Brocca, L., Barbetta, S., Faruolo, M., Lacava, T., and Moramarco, T., 2015. Coupling MODIS and radar altimetry data for discharge estimation in poorly gauged river basins. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(1), 141–148. <https://doi.org/10.1109/JSTARS.2014.2320582>
- Temimi, M., Leconte, R., Brissette, F., and Chaouch, N., 2005. Flood monitoring over the Mackenzie River Basin using passive microwave data. *Remote Sensing of Environment*, 98(2), 344–355. <https://doi.org/10.1016/j.rse.2005.06.010>
- Tobler, W. R., 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(sup1), 234–240. <https://doi.org/10.2307/143141>
- Uereyen, S., and Kuenzer, C., 2019. A review of earth observation-based analyses for major river basins. *Remote Sensing*, 11(24), 2951. <https://doi.org/10.3390/rs11242951>
- Wang, F., Chen, Y., Li, Z., Fang, G., Li, Y., Wang, X., et al., 2021. Developing a long short-term memory (LSTM)-based model for reconstructing terrestrial water storage variations from 1982 to 2016 in the Tarim River Basin, Northwest China. *Remote Sensing*, 13(5), 889. <https://doi.org/10.3390/rs13050889>
- Wang, S., and Peng, H., 2024. Multiple spatio-temporal scale runoff forecasting and driving mechanism exploration by K-means optimized XGBoost and SHAP. *Journal of Hydrology*, 630, 130650. <https://doi.org/10.1016/j.jhydrol.2024.130650>
- Wang, Y., Liu, D., Ma, S., Wu, F., and Gao, W., 2020. Ensemble learning-based rate-distortion optimization for end-to-end image compression. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(3), 1193–1207. <https://doi.org/10.1109/TCSVT.2020.3000331>
- Wolpert, D. H., 1996. The lack of a priori distinctions between learning algorithms. *Neural Computation*, 8(7), 1341–1390. <https://doi.org/10.1162/neco.1996.8.7.1341>
- Yamazaki, D., O'Loughlin, F., Trigg, M. A., Miller, Z. F., Pavelsky, T. M., and Bates, P. D., 2014. Development of the global width database for large rivers. *Water Resources Research*, 50(4), 3467–3480. <https://doi.org/10.1002/2013WR014664>
- Yu, Q., Wang, S., He, H., Yang, K., Ma, L., and Li, J., 2021. Reconstructing GRACE-like TWS anomalies for the Canadian landmass using deep learning and land surface model. *International Journal of Applied Earth Observation and Geoinformation*, 102, 102404. <https://doi.org/10.1016/j.jag.2021.102404>
- Yuan, T., Lee, H., Jung, H. C., Aierken, A., Beighley, E., Alsdorf, D. E., et al., 2017. Absolute water storages in the Congo River floodplains from integration of InSAR and satellite radar altimetry. *Remote Sensing of Environment*, 201, 57–72. <https://doi.org/10.1016/j.rse.2017.09.003>

- Zhou, Z. H., 2009. Ensemble. In: Liu, L., Özsu, M.T. (eds.), *Encyclopedia of Database Systems*, Springer, pp. 988–991. [https://doi.org/10.1007/978-0-387-39940-9\\_768](https://doi.org/10.1007/978-0-387-39940-9_768)
- Zhou, Z. H., 2012. *Ensemble methods: Foundations and algorithms*. CRC Press.
- Zhou, Z. H., 2015. Ensemble learning. In: Li, S. Z., Jain, A. K. (eds.), *Encyclopedia of Biometrics*, Springer, pp. 411–416. [https://doi.org/10.1007/978-1-4899-7488-4\\_293](https://doi.org/10.1007/978-1-4899-7488-4_293)
- Zhou, Z. H., 2021. *Machine learning*, Springer.
- Zounemat-Kermani, M., Batelaan, O., Fadaee, M., and Hinkelmann, R., 2021. Ensemble machine learning paradigms in hydrology: A review. *Journal of Hydrology*, 598, 126266. <https://doi.org/10.1016/j.jhydrol.2021.126266>