

하천유량 예측에서 자료 동화를 위한 기계학습의 활용

이채림* · 정하은* · 장수형** · 김상단*†

*국립부경대학교

**한국수자원공사

Harnessing machine learning for data assimilation in river flow prediction

Chaelim Lee* · Haeun Jung* · Suhyung Jang** · Sangdan Kim**

*Division of Earth Environmental System Science, Pukyong National University, Busan, Republic of Korea

**Water Resources and Environmental Research Center, Kwater Research Institute, Daejeon, Republic of Korea

(Received : 15 July 2025, Revised : 24 July 2025, Accepted : 24 July 2025)

요약

자료 동화를 사용하면 모델의 상태 변수를 업데이트하여 유역의 초기 상태를 Open-Loop 시뮬레이션보다 더 정확하게 나타낼 수 있다. 수문 분야에서 자료 동화는 종종 예보를 위한 전제 조건이다. 기계학습은 입력과 출력 사이의 모든 비선형 관계를 학습할 수 있다. 본 연구에서 기계학습은 수문 모델에서 시뮬레이션된 하천유량과 해당 상태 변수 사이의 관계를 학습할 수 있다는 가설을 세웠다. 학습이 완료되면 이 관계를 시뮬레이션된 하천유량이 아닌 관측된 하천유량에 적용하여 갱신된 상태 변수를 얻는 것에 사용할 수 있다. 이를 바탕으로 기계학습 기반의 자료 동화 접근 방식이 제안된다. 제안된 자료 동화 방법을 두 가지 상태 변수가 있는 장기 유출 모델에 적용하였다. R^2 , NSE, KGE, pBias와 같은 잘 알려진 성능지표를 사용하여 자료 동화된 하천유량 시계열을 Open Loop 하천유량, 앙상블 칼만 필터를 이용하여 자료 동화된 하천유량 및 관측된 하천유량과 비교했다. 기계학습은 자료 동화에 성공적으로 사용될 수 있으며, Open Loop 시뮬레이션 및 앙상블 칼만 필터에 비해 눈에 띄게 성능지표가 개선되었음을 보여주었다.

핵심용어 : 자료동화, 앙상블 칼만 필터, 장기유출모델, 기계학습, Random Forest

Abstract

Data assimilation can be used to update the state variables of a model, thereby representing the initial state of a watershed more accurately than in open-loop simulations. In hydrology, data assimilation is often a prerequisite for forecasting. Machine learning can learn all nonlinear relationships between inputs and outputs. In this study, we hypothesized that machine learning can learn the relationship between simulated river discharge and corresponding state variables in a hydrological model. Once learning is complete, this relationship can be applied to observed river flows rather than simulated river flows to obtain corrected state variables. Based on this, a machine learning-based data assimilation approach is proposed. The proposed data assimilation method was applied to a long-term continuous rainfall-runoff model with two state variables. Well-known performance metrics such as R^2 , NSE, KGE, and pBias were used to compare the assimilated river flow time series with open-loop river flow, river flow assimilated using an ensemble Kalman filter, and observed river flow. The results demonstrate that machine learning can be successfully applied to data assimilation, with significantly improved performance metrics compared to open-loop simulations and ensemble Kalman filter.

Key words : Data assimilation, Ensemble Kalman filter, Long-term continuous rainfall-runoff model, Machine learning, Random Forest

†To whom correspondence should be addressed.

Pukyong National University, Busan, Republic of Korea

E-mail : skim@pknu.ac.kr

• Chaelim Lee Pukyong National University, Busan, Republic of Korea/Ph.D. Candidate(coflarj1@pukyong.ac.kr)

• Haeun Jung Pukyong National University, Busan, Republic of Korea/Ph.D. Candidate(q0881@pukyong.ac.kr)

• Suhyung Jang Kwater Research Institute, Daejeon, Republic of Korea/ Head Researcher (kwaterjang@kwater.or.kr)

• Sangdan Kim Pukyong National University, Busan, Republic of Korea/Professor(skim@pknu.ac.kr)



This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

자료 동화는 모델의 상태 변수를 업데이트하여 초기(또는 Open-Loop) 시뮬레이션보다 더 정확하게 현실을 표현할 수 있도록 하는 모든 프로세스를 말한다. 수문 분야에서 자료 동화는 예보가 가능한 최상의 초기 모델 상태에서 시작되기 위해 예보를 위한 전제 조건인 경우가 많다(Choi et al., 2020). 직접 삽입이라고 하는 가장 기본적인 형태의 자료 동화는 하나 이상의 시뮬레이션된 상태 변수가 관측 가능해지면 관측값으로 대체하는 것이다. 예를 들어, 많은 수문 모델이 적설량을 상태 변수로 계산하지만, 이 상태 변수도 측정할 수 있다. 직접 삽입은 관측값에 전적인 신뢰성을 부여하고 Open-Loop 시뮬레이션에는 아무런 신뢰성을 부여하지 않는다(Boucher et al., 2020). 이때, Open loop 시뮬레이션은 시뮬레이션 모델에서 출력 결과를 다시 입력에 반영하지 않고, 사전에 정해진 입력값만으로 시스템을 실행하는 방식의 시뮬레이션을 말한다. 즉, 피드백이 없는 방식이다.

그러나 관측값도 불확실하므로 Open-Loop 시뮬레이션과 관측값 사이의 절충점을 찾는 체계적인 자료 동화 방법이 존재한다. 수문 분야에서 가장 잘 알려진 자료 동화 방법으로는 앙상블 칼만 필터(EnKF, Evensen, 1994)와 입자 필터(PF, Doucet et al., 2000)가 있다. 국내에서도 수문 모델에 자료 동화 기법을 적용한 최근 연구 사례를 찾아볼 수 있다(Choi et al., 2020; Choi and Kim, 2021; Lee et al., 2022; Lee et al., 2024).

기계학습은 입력과 출력 사이의 다양한 비선형 관계를 학습할 수 있다(Choi and Kim, 2025). 본 연구에서는 기계학습이 수문 모델에서 시뮬레이션된 하천유량과 해당 상태 변수 사이의 관계를 학습할 수 있다는 가설을 세웠다. 학습이 완료되면 이 관계를 시뮬레이션된 하천유량이 아닌 관측된 하천유량에 적용할 수 있으므로 갱신된 상태 변수를 얻는 데 사용할 수 있다. 이를 바탕으로 기계학습 기반의 자료 동화 접근법이 제안된다. 이 접근법은 모든 유형의 기계학습 방법을 사용하여 적용될 수 있지만, 본 연구에서는 수문 분야에서 많이 사용되는 기계학습모델에 중점을 두었다. 우리는 Random Forest(RF)라는 기계학습모델을 사용하였다. RF는 수문학에서 가장 널리 사용되는 기계학습모델 중 하나이며, 비교적 빠르고 간단하게 구현할 수 있다는 장점이 있다(Choi et al., 2022). RF는 개별 결정 트리의 앙상블이기 때문에, 모델 예측의 해석 가능성이 뛰어나며, 과적합을 줄이는 앙상블 구조 덕분에 안정성이 좋아서 환경 센서 이상 탐지 및 토양수분 예측 등 실제 환경 문제에서 여전히 많이 사용된다.

기계학습은 오래전부터 수문 관련 문제를 해결하는 데 사용됐으며, 가장 일반적인 응용 분야는 하천유량 예보이다(Lima et al., 2016). 하지만 극소수의 예외를 제외하고는 예보 운영 기관은 유량 예보에 기계학습을 제1번 옵션으로 사용하지 않는다. 기계학습 기반의 강우 유출 모델의 블랙박스 특성이 주된 이유일 것이다. 실제로 운영 예보 및 수자원 관리자는 자신의 결정에 책임이 있으므로, 이들은 수문 프로세스에

기반한 건전한 추론으로 이러한 결정을 뒷받침할 수 있어야 한다. 그런데도 자료 기반 모델의 사용을 장려하는 주된 이유는 많은 자료 기반 모델이 관련된 기본 물리적 프로세스를 완벽히 이해하지 않고도 비선형 프로세스를 수치상으로 포착할 수 있기 때문이다.

본 연구에서는 기계학습이 개념적 또는 물리 기반 모델을 대체하지 않으며, 상태 변수에 대한 추정치를 갱신하는 목적으로 사용된다. 즉, 기계학습모델이 수문 모델을 대체하는 것이 아니라 수문 모델을 지원하는 역할은 한다. 따라서 운영 기관은 프로세스 이해도를 저하하지 않으면서도 기계학습의 강력한 기능을 활용하여 예보를 개선할 수 있다.

엄격한 의미에서 살펴보면, Wahle et al. (2015)이 파랑 모델링에서 자료 동화를 위한 기계학습의 기능을 최초로 입증했다. 그들의 연구에서는 북해 독일 바이트 지역의 파도 높이, 방향, 주기 및 기타 특성을 시뮬레이션하기 위해 COSYNA WAM 사이클 4.5.4 모델을 사용했다. 이 모델은 상태 변수를 사용하여 시뮬레이션의 각 시간 단계에서 시스템의 상태를 설명한다. 그들은 먼저 기계학습 기법을 사용하여 파동 모델의 상태 변수와 시뮬레이션된 파동 매개변수 사이의 관계를 모델링했다. 이 관계를 학습한 후 기계학습을 반전시켜 파동 매개변수로부터 상태 변수를 계산할 수 있도록 했다. 이 접근 방식에서 우리가 착안한 것은 Open-Loop 상태 변수에서 파생된 시뮬레이션된 하천유량을 사용하여 기계학습모델을 학습시킬 수 있다는 사실이다. 이후, 자료 동화를 목적으로 기계학습의 적용성을 탐색한 연구들이 최근 다양하게 제안되고 있다(Boucher et al., 2020; Jeung et al., 2023; Matin, 2023; Jeong et al., 2024; Zhang et al., 2024).

본 연구의 목적은 기계학습 모델 중 하나인 RF를 사용하여 간단한 장기 강우 유출 모델에서 자료 동화(사용할 수 있는 유량 관측에 따른 상태 변수의 업데이트)를 수행할 수 있는지를 탐색하는 것이다. 이를 위해 다양한 성능지표를 이용하여 Open-Loop 시뮬레이션(자료 동화가 없는 시뮬레이션)과 RF를 이용하여 자료 동화된 시뮬레이션을 비교한다. 더불어서, 전통적인 자료 동화 기법으로 많이 사용되는 앙상블 칼만 필터를 이용하여 자료 동화된 시뮬레이션도 함께 비교한다.

2. 자료 및 방법

2.1 연구 구역

본 연구에서는 낙동강 유역에 있는 유역 면적 2,285km²의 남강댐 유역이 적용 지역으로 선정되었다(Fig. 1). 남강댐 유역의 일 유입량 자료는 수자원관리종합정보시스템(www.wamis.go.kr)에서 수집하여 사용하였으며, 유량 시뮬레이션 기간은 2009년부터 2018년까지이다. 남강댐 유역 인근 기상청에서 운영 중인 거창, 남원, 산청, 장수, 진주, 함천 지점의 일 기상 자료를 바탕으로 티센 가중치 방법을 이용하여 면적 평균 일 강우량 P 및 일 잠재증발산량 E 를 계산하였다. 일 잠재증발산량은 Penman-Monteith 방법(Allen et al., 1998)으로 계산하였다.

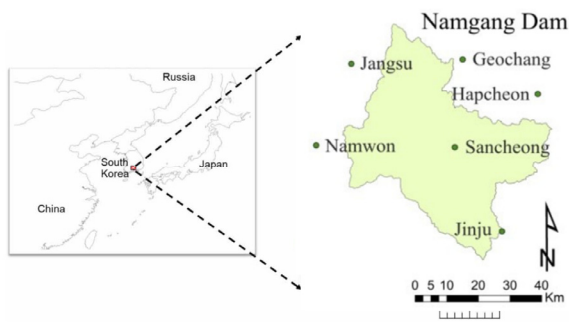
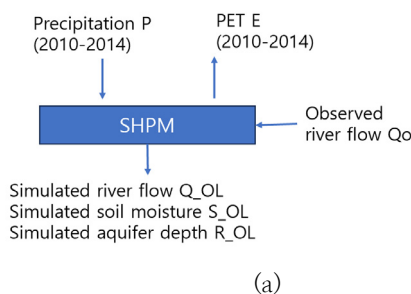


Fig. 1. Namgang Dam watershed

2.2 수치 실험 계획

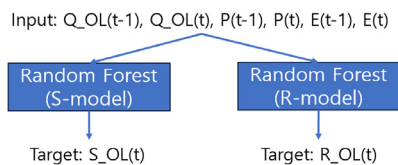
본 연구에 사용된 수치 실험 설정을 Fig. 2에 나타내었다. 본 연구에는 세 가지 주요 단계가 있다. 첫 번째는 사용된 단순 수문분할모형(Simple Hydrologic Partitioning Model, SHPM)의 매개변수를 추정하는 것이다. SHPM의 자세한 내용은 Choi et al. (2020)을 참고할 수 있다. 적용된 모형의 매개변수는 Metropolis-Hastings(MH) 알고리즘(Choi et al., 2023)을 이용하여 2010~2014년 자료를 바탕으로 추정되었다(Fig. 2(a)).

Calibrating



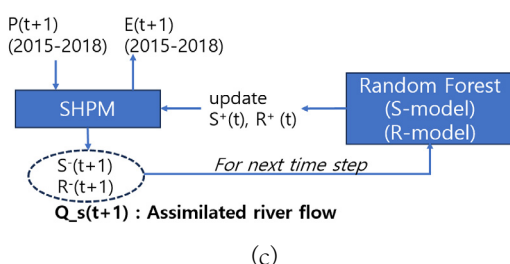
(a)

Learning



(b)

Assimilating



(c)

Fig. 2. Numerical experimental setup diagram
(a) Calibrating (b) Learning (c) Assimilating

두 번째는 RF가 수문 기상 변수(예: 여러 시차 이전(본 연구에서는 하루 전)의 하천유량 및 기상 변수(본 연구에서는 강수량과 잠재증발산량))와 수문 모델의 상태 변수 사이의 관계를 정확하게 모델링할 때까지 학습하는 것이다. 이러한 학습 프로세스가 Fig. 2(b)에 설명되어 있다. 이 단계의 주요 가정은 모델링된 하천유량과 상태 변수(본 연구에서는 토양수분과 대수층의 높이) 사이의 관계가 관측된 하천유량과 추정하고자 하는 그러나 알 수 없는 상태 변수 사이의 관계와 같다는 것이다. 이때, 강수가 발생한 날에는 토양수분을 목표 데이터로 하는 RF 모델이, 강수가 발생하지 않는 날에는 대수층의 높이를 목표 데이터로 하는 RF 모델이 구성되었다.

세 번째는 자료 동화이다. 이 단계에서는 하천유량과 상태 변수 사이의 관계를 이미 알고 있다고 가정한다. 그런 다음 시뮬레이션이 아닌 관측된 하천유량을 RF에 입력으로 제공하기만 하면 해당 상태 변수의 추정치를 얻을 수 있다. 이 과정은 Fig. 2(c)에 설명되어 있다. 이렇게 갱신된 새로운 상태 변수를 적절한 입력과 함께 수문 모델에 다시 전달하여 일련의 동화된 하천유량을 생성할 수 있다. SHPM의 경우 상태 변수의 실제 값을 알 수 없으므로, 관측된 하천유량 시계열과 자료 동화 후 해당 시뮬레이션 하천유량을 비교함으로써만 제한된 자료 동화 방법론의 성능을 평가할 수 있다. 이때, 강수가 발생한 날에는 토양수분을, 강수가 발생하지 않는 날에는 대수층의 높이를 갱신하는 자료 동화 전략이 사용되었다.

3. 결과 및 토론

3.1 MH 알고리즘을 이용한 SHPM의 매개변수 추정

SHPM의 보정은 2010년부터 2014년까지의 남강댐에 관측된 일 유입량 자료와 시뮬레이션 자료를 이용하여 수행되었으며, 검정은 2015년부터 2018년까지의 자료로 수행되었다. 일 유량을 기준으로 성능지표를 살펴보면, 보정 시에는 결정계수 R^2 는 0.77, Nash-Sutcliffe 효율계수 NSE는 0.76, Kling-Gupta 효율계수 KGE는 0.83, pBias는 -7.0%이었으며, 검증 시에는 $R^2 = 0.67$, NSE = 0.65, KGE = 0.79, pBias = -7.7%로 비교적 관측자료를 잘 재현하도록 매개변수가 추정되었다.

3.2 SHPM 시뮬레이션 결과를 이용한 Random Forest 모델 학습

SHPM으로부터 모의된 2010년부터 2014년 토양수분 S , 대수층 높이 R , 일 유량 Q 을 이용하여 RF 모델을 학습하였다. RF 모델은 2개가 구성되었다(S-model과 R-model). RF 모델의 t -일의 입력자료는 Q_{t-1} , Q_t , P_{t-1} , P_t , E_{t-1} , E_t 로 1일 선행 및 당일의 유량, 강수량, 잠재증발산량이 포함되며, RF 모델의 t -일의 출력자료는 강수가 발생한 날에는 S_t , 강수가 발생하지 않는 날에는 R_t 이다. Table 1은 다양한 선행 일수의 입력자료를 이용하여 구성된 RF 모델의 테스트 성능(2015~2018년 자료 적용)을 보여준다. RF 모델에 포함

Table 1. Performance of Random Forest models based on the number of preceding days of input data

lag (day)	S-model			R-model		
	R2	NSE	KGE	R2	NSE	KGE
0	0.91	0.90	0.87	0.86	0.63	0.44
1	0.97	0.97	0.94	0.98	0.98	0.93
2	0.98	0.98	0.94	0.99	0.99	0.96
3	0.98	0.98	0.94	0.99	0.99	0.95

되는 선행 일수를 테스트한 결과는 선행 1일이면 충분한 모델 성능을 확보할 수 있다는 것을 보여주었다. RF 모델의 하이퍼파라미터는 모두 기본값을 적용하였으며, 사용된 프로그래밍 언어는 Matlab R2022a 버전이다.

3.3 Random Forest를 이용한 SHPM의 상태 변수 갱신 및 하천유량 예측

Table 2는 4개 사례를 통하여 RF 모델을 이용한 자료동화(Data Assimilation using RF models, DARF) 절차가 어떤 방식으로 수행되는지를 보여준다.

먼저 2018년 9월 4일의 결과를 보면, 강수가 발생한 날이므로 Open-Loop에서 시뮬레이션된 토양수분 S^- 가 DARF의 S-model를 이용하여 S^+ 로 갱신되었다. 이때, Open-Loop로 시뮬레이션된 유량 Q_s 가 관측된 유량 Q_o 보다 작으므로, 토양수분은 양의 방향으로 갱신되어 이후 시뮬레이션에서 유량이 커지는 방향으로 상태 변수가 조정되었다. 반대로, 2018년 10월 5일의 결과를 보면, 강수가 발생한 날이므로 토양수분이 갱신되었다. 이때, $Q_s > Q_o$ 이므로, 토양수분이 작아지는 방향으로 상태 변수가 조정되었다. 2018년 10월 7일은 강수가 발생하지 않았으므로, R-model을 적용하여 대수층의 높이 R 이 갱신되었다. 이때, $Q_s < Q_o$ 이므로, 대수층 높이가 높아지는 방향으로 상태 변수가 조정되어 이후 시뮬레이션에서는 유량이 증가하는 방향으로 초기 상태가 설정되었다. 마지막으로 2018년 10월 8일은 강수가 발생하지 않았으므로, 대수층 높이가 갱신 대상이 되며, $Q_s > Q_o$ 이므로, 대수층 높이가 낮아지게 조정되었다. Table 2의 사례로부터 DARF는 현재 시뮬레이션된 유량과 관측 유량을 비교하여

Table 2. Implementation of data assimilation using RF models

date	2018.9.4	2018.10.5	2018.10.7	2018.10.8
P (mm)	59.5	70.0	0	0
Q_s (mm)	17.9	13.2	10.3	11.9
Q_o (mm)	47.4	3.5	18.0	7.6
S^- (-)	0.59	0.44	0.57	0.54
S^+ (-)	0.63	0.40	0.57	0.54
R^- (mm)	3.42	0.41	5.62	6.48
R^+ (mm)	3.42	0.41	7.23	4.96

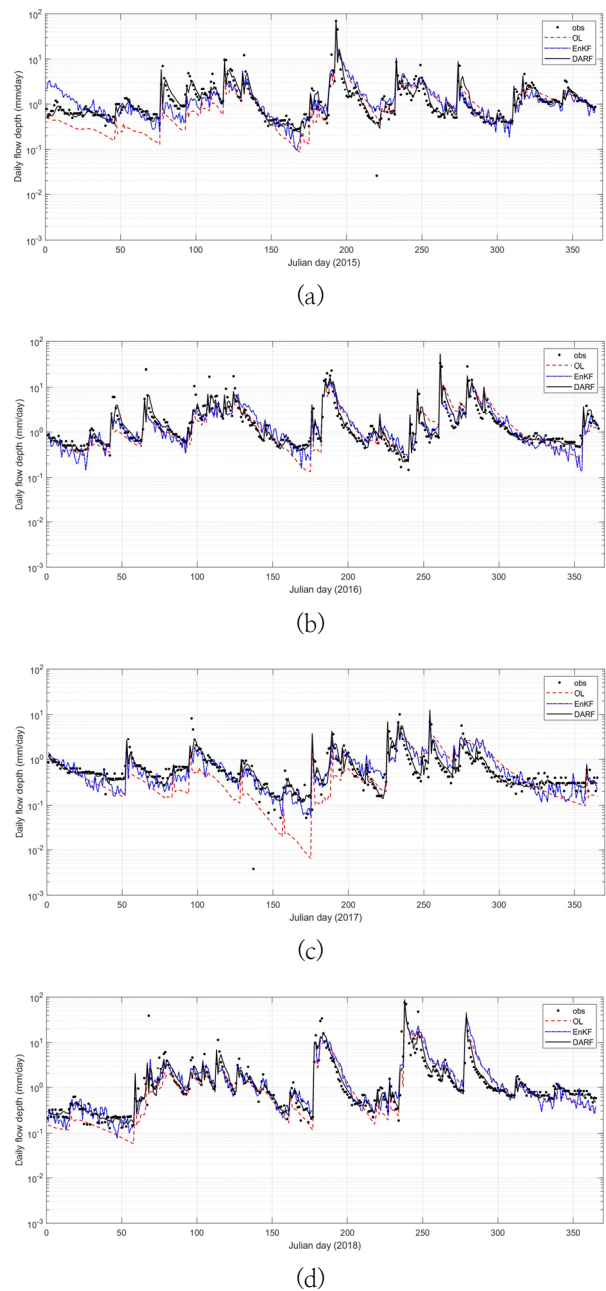


Fig. 3. Time series of various simulation results
(a) Year 2015 (b) Year 2016 (c) Year 2017 (d) Year 2018

상태 변수를 갱신함으로써, 이후 시뮬레이션에서 관측 유량을 더 잘 맞출 수 있도록 초기 상태를 조정하고 있음을 발견할 수 있다.

Fig. 3은 Open-Loop 시뮬레이션(OL) 시계열과 더불어서 EnKF에 의한 자료 동화로 도출된 시계열과 본 연구에서 제안하고 있는 DARF에 의한 자료 동화로 도출된 시계열을 관측 자료와 함께 보여주고 있다. OL보다는 자료 동화가 수행된 EnKF와 DARF가 관측된 유량을 더 잘 맞추고 있음을 시각적으로 확인할 수 있다. 특히 EnKF보다는 DARF가 감수부의 수문곡선을 특히 더 잘 따라가고 있다.

Table 4. Performance metrics of various simulation results

Performance Metrics		Methods	OL	EnKF	DARF
R^2			0.671	0.680	0.734
NSE			0.646	0.669	0.733
KGE			0.795	0.811	0.815
pBias (%)	all		-7.68	1.07	-0.92
	EH		-18.20	-14.59	-15.04
	H		2.05	8.03	7.60
	M		-4.52	6.48	1.75
	L		-9.20	4.12	1.18
	EL		-18.90	9.37	11.69

Table 4는 OL 시계열과 더불어서 EnKF에 의한 시계열과 DARF에 의한 시계열의 성능 지표를 보여주고 있다. 이때, 양의 pBias는 모의 결과가 관측자료를 과대 추정한 것을 의미한다. pBias(all)은 전체 시계열의 pBias이며, pBias(EH)는 관측 자료의 상위 10%에 해당하는 extremely high flow 구간의 자료만으로 계산한 pBias이다. pBias(H)는 상위 10~30%의 high flow 구간, pBias(M)은 30~70%의 medium 구간, pBias(L)은 하위 10~30%의 low flow 구간, pBias(EL)은 하위 10%에 해당하는 extremely low flow 구간의 pBias이다.

R^2 , NSE, KGE, pBias(all) 모두 OL, EnKF, DARF 순으로 성능지표가 개선되고 있다. 특히, 자료 동화의 구조 상 EnKF는 OL의 정확도에 더 의존적이므로, Table 4의 결과는 DARF의 우수성이 잘 나타내 준다. 유량 규모별 pBias의 비교에서는 high flow, medium flow, low flow 구간에서는 DARF가 OL뿐만 아니라 EnKF보다도 더 우수한 성능을 보이고 있다. Extremely high flow 및 Extremely low flow 구간에서는 EnKF가 DARF보다 더 우수한 성능을 보이고 있지만, 성능 지표의 수치적 차이가 크지 않음을 살펴볼 수 있다. 이 구간에서도 DARF는 OL보다 더 우수한 성능지표를 보여준다.

마지막으로 살펴본 것은 토양수분의 자료 동화와 대수층 높이의 자료 동화를 각각 분리하여 수행한 결과이다. 강수가 발생한 날에 토양수분만 자료 동화를 시행하였을 때는 $R^2 = 0.72$, NSE = 0.72, KGE = 0.81, pBias = -0.88%의 성능 지표를 얻을 수 있었으며, 강수가 발생하지 않은 날에 대수층의 높이만 자료 동화를 시행하였을 때는 $R^2 = 0.70$, NSE = 0.69, KGE = 0.81, pBias = -5.76%를 얻을 수 있었다. 강수 발생 여부에 따라서 토양수분과 대수층 높이를 자료 동화하였을 때보다는 성능지표가 낮아졌으며, 토양수분의 자료 동화가 모델의 성능을 높이는 것에 상대적인 공헌도가 더 높음을 발견할 수 있다.

4. 결 론

본 연구에서는 수문학적 예측을 위한 자료 동화 기술로 기계학습을 이용하는 방법을 제안하였다. 제안된 방법은 일상적으로 운영되는 현업 실무에서 기계학습의 매우 유망한

적용 방향이 될 수 있다. 실제로 본 연구에서 기계학습은 개념적 또는 물리 기반 모델을 대체하지 않는다. 이러한 특성은 프로세스의 이해와 모델 운영 실무에서 예측을 기반으로 한 결정의 정당화를 위해 매우 중요하다. 오히려 기계학습은 자료 동화를 통해 개념적 및/또는 물리 기반 모델을 지원할 수 있다.

본 연구에서 Random Forest 모델은 앙상블 칼만 필터 및 입자 필터와 같은 다른 자료 동화 방법과 정확히 동일한 역할(수문 모델을 Open Loop 모드에서 실행하여 얻은 추정값을 개선하기 위해 상태 변수를 갱신)을 수행한다.

본 연구를 통해 우리는 기계학습이 수문 모델에서 자료 동화 작업을 정확히 수행할 수 있음을 보여주었다. 이는 간단한 강우-유출 모델을 사용하여 입증되었고, 제안된 자료 동화 방법은 정확하고 신뢰할 수 있음을 보여주었으나, 본 연구는 단일 개념 모델 및 단일 유역(남강댐 유역)에만 초점을 맞췄다. 제안된 방법을 다른 개념 모델에 적용하는 것은 어렵지 않을 것이며, 더 복잡한 물리 기반 모델이나 분포형 모델로 확장하는 것도 가능할 것이다.

본 연구는 방법론 개발과 성능을 일차적으로 검증하는 것을 목표로 진행되었다. 추후 대규모 적용 연구를 진행하여 기계학습 기반의 자료 동화 방법이 앙상블 칼만 필터 및 입자 필터와 같은 전통적인 기술보다 우수한 성능을 발휘할 수 있는지 더 다양한 측면에서 살펴보는 것이 중요할 것이다.

Acknowledgement

이 논문은 국립부경대학교 자율창의학술연구비(2023년)에 의하여 연구되었음.

References

- Boucher, M. A., Quilty, J., and Adamowski, J. (2020). Data assimilation for streamflow forecasting using extreme learning machines and multilayer perceptrons. *Water Resources Research*, 56, e2019WR026226, <<https://doi.org/10.1029/2019WR026226>>.
- Choi, J. and Kim, S. (2021). Estimating time-varying parameters for monthly water balance model using particle filter: assimilation of stream flow data. *Journal of Korea Water Resources Association*, 54(6), pp. 365–379.
- Choi, J. and Kim, S. (2025). Data-driven model as a post-process for daily streamflow prediction in ungauged basins. *Heliyon*, 11(4), e42512, <<https://doi.org/10.1016/j.heliyon.2025.e42512>>.
- 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*, 166617, <<https://doi.org/10.1016/j.scitotenv.2023.166617>>.

- Choi, J., Lee, O., Won, J., and Kim, S. (2020). Stochastic Simple Hydrologic Partitioning Model Associated with Markov Chain Monte Carlo and Ensemble Kalman Filter. *Journal of Korean Society on Water Environment*, 36(5), pp. 353–363.
- Choi, J., Won, J., Jang, S., and Kim, S. (2022). Learning enhancement method of Long short-term memory network and its applicability in hydrological time series prediction. *Water*, 14(18), 2910, <<https://doi.org/10.3390/w14182910>>.
- Doucet, A., Godsill, S., and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3), 197–208.
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5), 10143–10162.
- Jeong, M., Kwon, M., Cha, J., and Kim, D. (2024). High flow prediction model integrating physically and deep learning based approaches with quasi real-time watershed data assimilation. *Journal of Hydrology*, 636, 131304.
- Jeung, M., Jang, J., Yoon, K., and Baek, S. (2023). Data assimilation for urban stormwater and water quality simulations using deep reinforcement learning. *Journal of Hydrology*, 624, 129973.
- Lee, G., Lee, S., Kim, B., Woo, D., and Noh, S. (2022). Comparative assessment and uncertainty analysis of ensemble-based hydrologic data assimilation using ai rGRdatassim. *Journal of Korea Water Resources Association*, 55(10), pp. 761–774.
- Lee, G., Woo, D., and Noh, S. (2024). Comparative assessment of sequential data assimilation-based streamflow predictions using semi-distributed and lumped GR4J hydrologic models: a case study of Namgang Dam basin. *Journal of Korea Water Resources Association*, 57(9), pp. 585–598.
- Lima, A. R., Cannon, A. J., and Hsieh, W. W. (2016). Forecasting daily streamflow using online sequential extreme learning machines. *Journal of Hydrology*, 537, 431–443.
- Martin, N. (2023). Dynamic Assimilation of Deep Learning Predictions to a Process-Based Water Budget. *Hydrology*, 10(6), 129.
- Zhang, J., Cao, C., Nan, T., Ju, L., Zhou, H., and Zeng, L. (2024). A novel deep learning approach for data assimilation of complex hydrological systems. *Water Resources Research*, 60(2), e2023WR035389.