D2 - Model Development

Improving hydrological model performance in a tidal river using deep learning: A case study of the Chao Phraya River in Thailand

Kritnipit Phetanan[a], Heewon Jeong[a], Ather Abbas[a], Daeun Yun[a], Jiye Lee[a], Kyung Hwa Cho[a, b*]

[a] Department of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology, UNIST-gil 50, Ulsan 44919, Republic of Korea

[b] Graduate School of Carbon Neutrality, Ulsan National Institute of Science and Technology, Republic of Korea

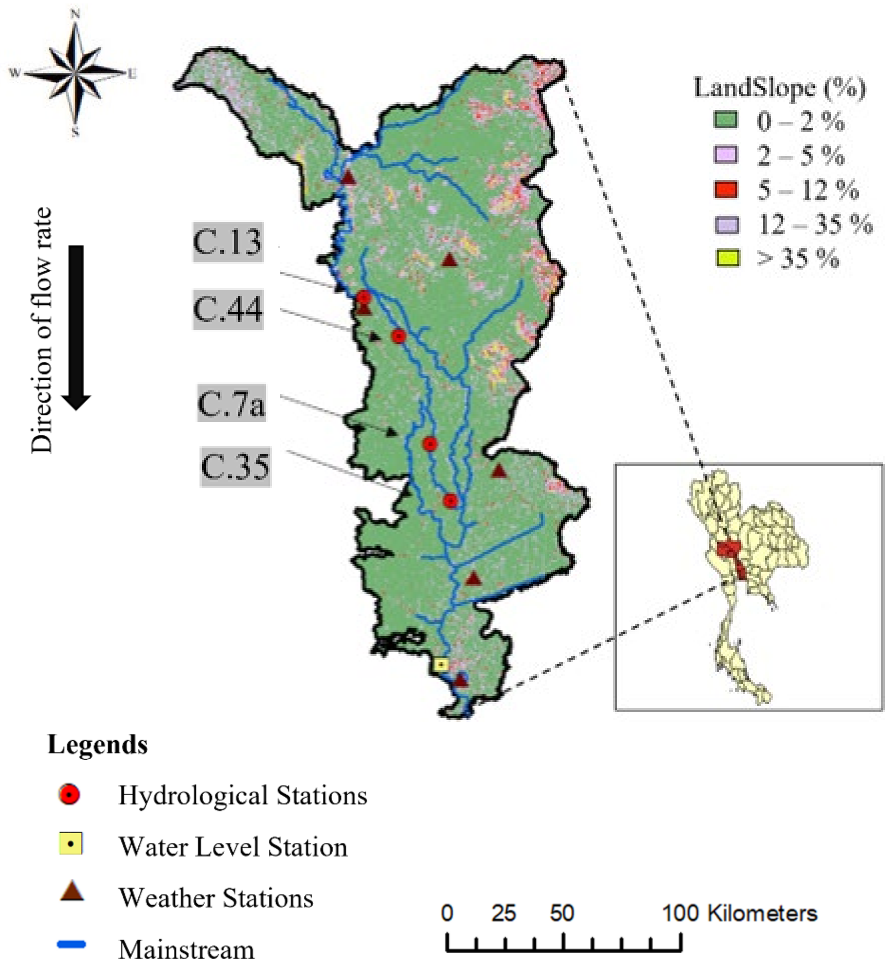
Kritnipit Phetanan | Ulsan National Institute of Science and Technology (UNIST)
MS candidate | Water-Environmental Informatics Laboratory (WEIL)

Email: Krit.fr41@unist.ac.kr

Contents

1. Introduction

Background



Study area: Chao Phraya River watershed

- Main water resource in the central region of Thailand, supplying irrigation and maintaining agriculture, also contributing to industrialization (Molle, 2007).
- The river is 372 km long, watershed covers an area of 21,725 km^2
- The **tidal effect** is caused by the mix of enclosed sea and tides around the upper Gulf of Thailand (Saramul & Ezer, 2014).

Tides in the Chao Phraya River

- The tides caused **changes** in the river flow rate making it **difficult** to simulate and predict the exact behavior of the river.
- Predicting and managing the water flow rate through modeling to protect water resources is needed.

→ **Need the improvement of flow rate simulation**

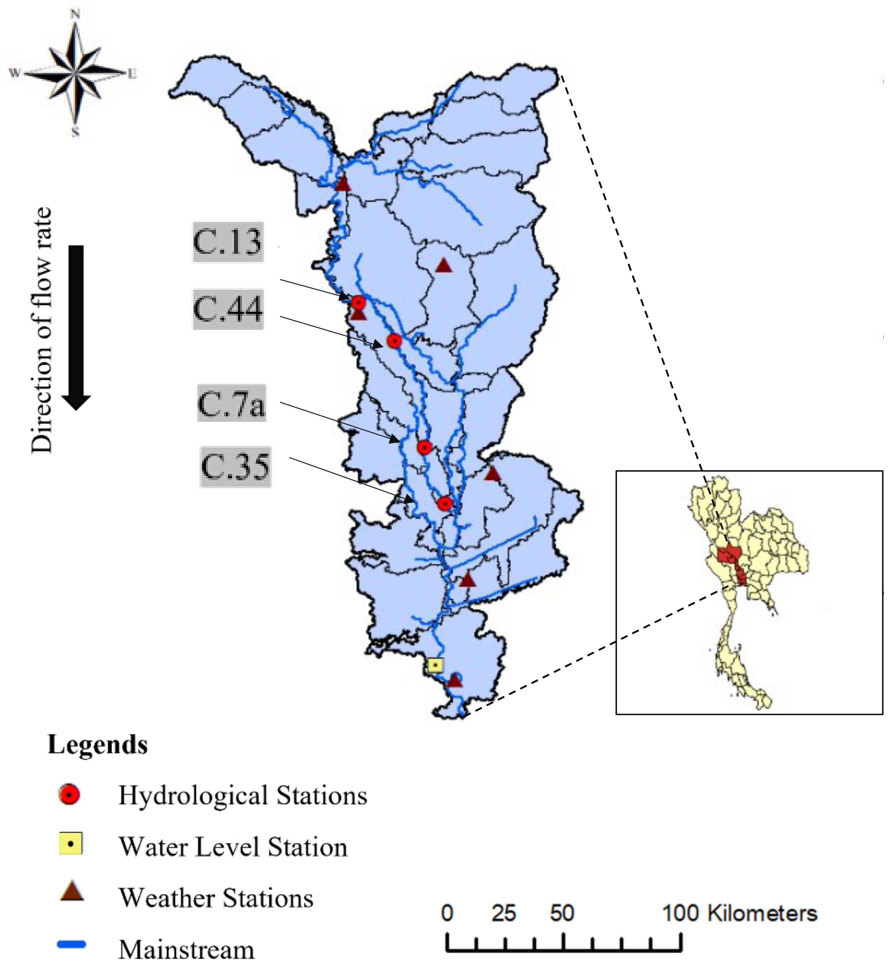
Research Objectives

1. To develop the model that coupled SWAT and deep learning model for improving the simulation of flow rate in the tidal river due to the lack of tidal consideration in the SWAT model.

2. To investigate the influence of tides on the deep learning model by the model interpretation.

2. Materials and Methods

Study area



Study area: Chao Phraya River watershed

- **4 hydrological stations** used as the outlets; **C.13** (15°16 E, 100°19 N), **C.44** (15°01 E, 100°33 N), **C.7a** (14°59 E, 100°45 N), **C.35** (14°36 E, 100°52 N)
- 6 meteorological stations covered in the watershed.
- Water level was obtained from the **Samsean station** (13°78 E, 100°50 N) and served as the nearest shoreline sampling location along the water flow path before it reaches the Gulf of Thailand.
- The data were obtained from
  - Royal Thai Survey Department (RTSD) – Digital Elevation Map
  - Land Development Department Thailand (LDD) – Land use, Soil data
  - Thai Meteorological Department Thailand (TMD) – Meteorological data
  - Royal Irrigation Department Thailand (RID) – Hydrological data

SWAT model

Modeling process

Step 1. Data preparation to build the SWAT model

Meteorological data

- **Daily data (2009 – 2021)**
  - Max/Min Temperature (˚C)
  - Wind speed (m/s)
  - Relative humidity (%)
  - Precipitation (mm/day)

Geospatial data

- **Watershed**
  - Digital elevation map
- **Hydrological response unit (HRU)**
  - Land use
  - Soil data

Monitoring data

- **Daily Inlet discharge (2012 – 2021)**
  - Inlet station (15°69 E, 100°13 N)

Step 2. Simulation of the hydrological outputs (flow rate) in the SWAT model

Module in SWAT

- **Warm up 2009 – 2011**
- **Calibration 2012 – 2018**
- **Validation 2019 – 2021**

→

Sensitivity analysis

- **Global sensitivity analysis (GSA)**

→

Multiple-site calibration

- **SWAT-CUP**
  - The sequential uncertainty fitting algorithm version 2 (SUFI-2)

SWAT model

Calibrated parameters

Hydrological Process	Parameter	Description	Range Min	Range Max
Mgt	CN2	Initial SCS runoff curve number for moisture condition II	35	98
Gw	ALPHA_BF	Baseflow alpha factor	0	10
	GW_DELAY	Groundwater delay time	0	500
	GWQMN	Threshold water level in shallow aquifer for baseflow	0	5000
	GW_REVAP	Groundwater re-evaporation coefficient	0.02	0.2
	REVAPMN	Threshold depth of water in the shallow aquifer for re-evaporation	0	1000
	RCHRG_DP	Deep aquifer percolation fraction	0	1
Rte	CH_K2	Effective hydraulic conductivity in main channel alluvium	5	130
	CH_N2	Manning's "n" value for the main channel	0.01	0.3
	ESCO	Soil evaporation compensation factor	0	1
	EPCO	Plant uptake compensation factor	0	1
	CANMX	Maximum canopy storage	0	100

Hydrological Process	Parameter	Description	Range Min	Range Max
Hru	LAT_TTIME	Lateral flow travel time	0	180
Sub	LAT_SED	Sediment concentration in lateral and groundwater flow	0	5000
	OV_N	Initial residue Cover	0.01	30
	CH_K1	Effective hydraulic conductivity in tributary channel alluvium	0	300
	CH_N1	Manning's "n" value for the main channel	0.01	30
Bsn	ESCO	Soil evaporation compensation factor	0	1
	EPCO	Plant uptake compensation factor	0	1
	SURLAG	Surface runoff lag coefficient	1	24

SWAT model

Global sensitivity analysis (GSA)

- Calculating the changes of the objective function from each parameter while varying all parameters simultaneously
- Understanding which factors have the most significant impact on the model's behavior or output.

Conventional calibration & Multi-site calibration

Differences / Calibration	Conventional calibration	Multi-site calibration
Calibration process	The calibration is site-specific.	Considering the spatial variability and interactions between different sites.
Parameter estimation	Each site has its own set of optimized parameters.	Expected to capture the patterns and processes of the entire region.
Complexity	Suitable when the goal is to assess site-specific performance	Suitable when capturing the spatial variability and regional performance of the model.
Data requirement	Site-specific data for each calibration location.	Data from multiple monitoring sites to capture the regional patterns and interactions

SWAT-LSTM model

Modeling process

Step 1. Data pre-processing for deep learning model

Meteorological data

- **Daily data (2012 – 2021)**
  - Max/Min Temperature (˚C)
  - Wind speed (m/s)
  - Relative humidity (%)
  - Precipitation (mm/day)

SWAT characteristics

- **Daily data (2012 – 2021)**
  - Curve number
  - Evapotranspiration
- **Static data**
  - Watershed area
  - Watershed elevation

Tide-related data

- **Daily data (2012 – 2021)**
  - Tidal data
  - The nearest water level to shoreline
- **Static data**
  - Distance from shoreline

Step 2. Hyper-parameter optimization (HPO)

Bayesian optimization method

- Seeking the optimal value for the following selections

- Units
- Epochs
- Dropout
- Batch size
- Lookback
- Loss function
- Optimizer
- Activation function

Step 3. Simulation of the flow rate using a deep-learning model and model interpretation

Input

Meteorological data

Daily data

10.3	5.6	0.0	0.0	0.0
0.0	17.1	4.5	19.2	0.0
20.5	0.0	12.3	52.1	25.8

SWAT characteristics

Daily data

11.5	14.5	12.3	17.4	10.4

Static data

10	10	10	10	10

Tide-related data

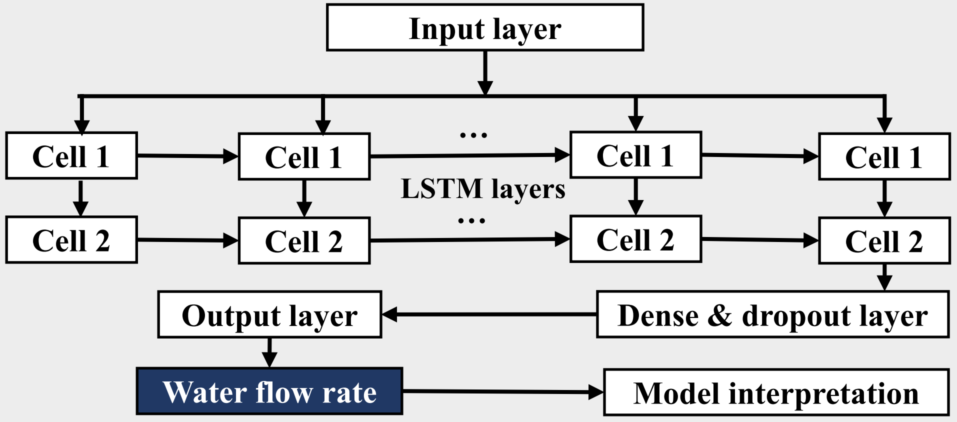
Daily data

2.2	3.1	4.5	1.4	2.6

Static data

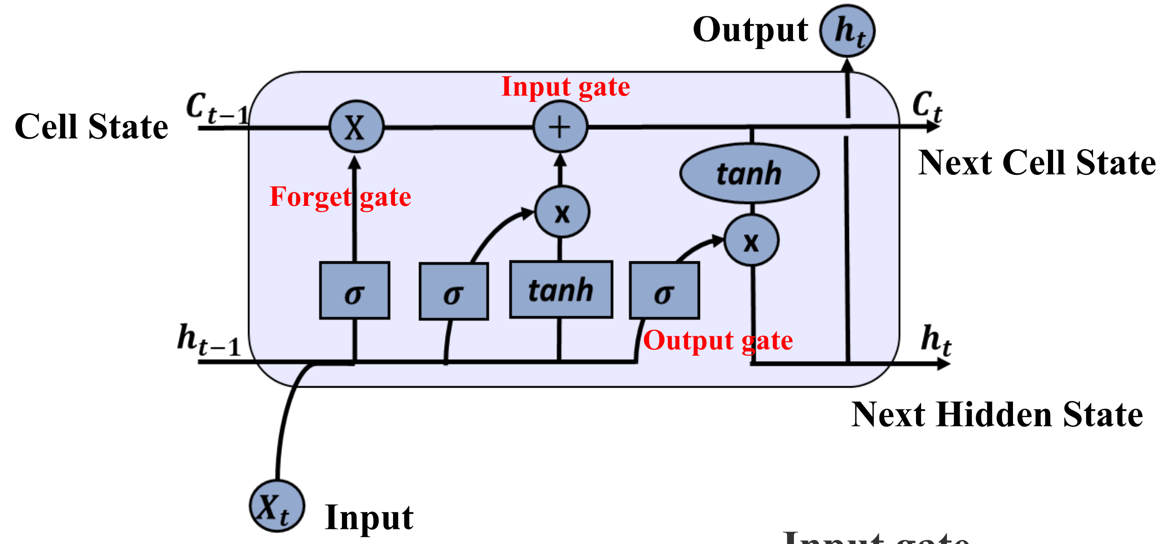
80	80	80	80	80

Long short-term memory (LSTM)

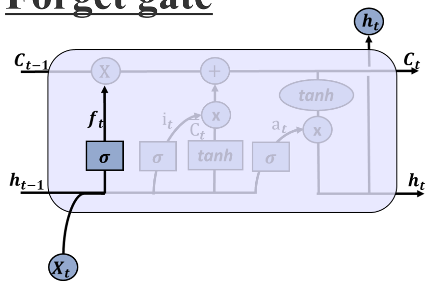


LSTM model

Model structure



Forget gate



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right)$$

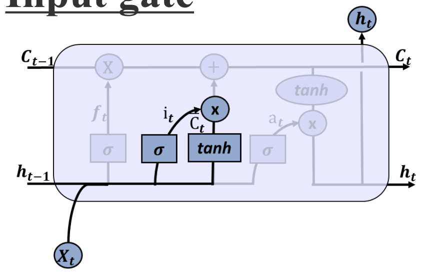
Vector with 0~1 values

Generated matrix with same dimension of hidden layer

Bias

Concatenated input vector

Input gate



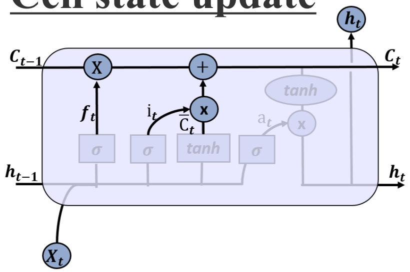
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right)$$

Decided which values to update

$$\tilde{C}_t = \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right)$$

Created a vector of new values that could be added to the state

Cell state update

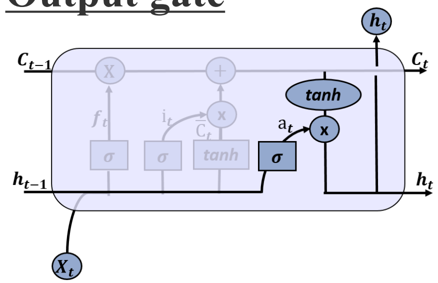


Added the decided value to update each state value

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Decided whichever information should be forgotten

Output gate



$$o_t = \sigma\left(W_o\ [h_{t-1}, x_t] + b_o\right)$$

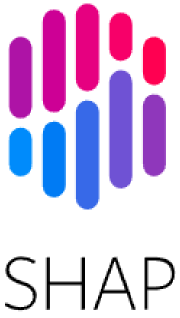
: output coefficient

$$h_t = o_t * \tanh\left(C_t\right)$$

Put tanh (-1~1) and determined output

Model interpretation

SHapley Additive exPlanations (SHAP) analysis



Source: Lundberg and Lee (2017)

- The goal of SHAP is to explain the prediction of an instance x by computing the contribution of each feature to the prediction.
- By quantifying the impact of each feature on the output, SHAP allows users to **identify the most influential features** and focus their attention on understanding and analyzing those features.
- In this study, we used **SHAP to investigate the influence of tides** on the deep learning model.

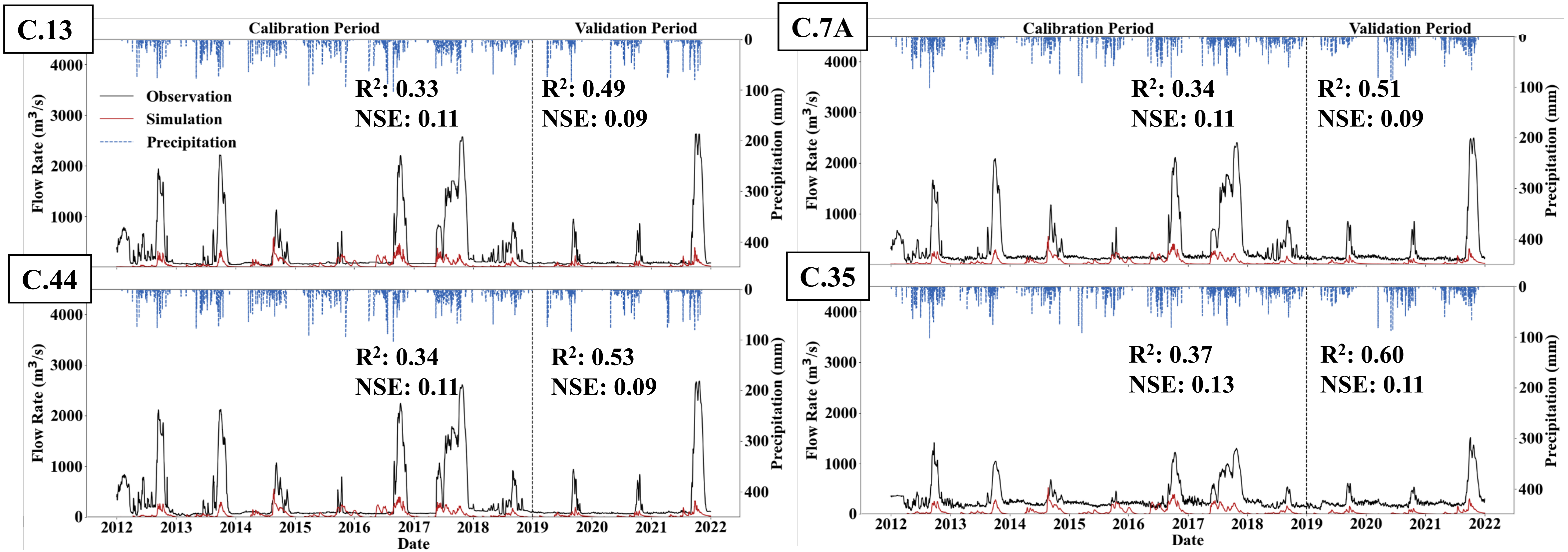
3. Results and Discussion

Sensitivity analysis results

Rank	Parameter	Fitted value	P-value	Rank	Parameter	Fitted value	P-value
1	GW_DELAY.gw	72.000000	0.0000	11	EPCO.bsn	0.300000	0.3961
1	CANMX.hru	50.000000	0.0000	12	GW_REVAP.gw	0.038000	0.4065
1	RCHRG_DP.gw	0.700000	0.0000	13	REVAPMN.gw	900.000000	0.4275
4	EPCO.hru	0.100000	0.0001	14	ESCO.hru	0.100000	0.4555
5	ALPHA_BF.gw	0.900000	0.0066	15	LAT_TTIME.hru	90.000000	0.5325
6	CH_K2.rte	17.500000	0.0114	16	SURLAG.bsn	0.300000	0.8404
7	CH_N1.sub	9.007000	0.0305	17	OV_N.hru	21.003000	0.8494
8	CH_N2.rte	0.271000	0.1891	18	CH_K1.sub	210.000000	0.8806
9	GWQMN.gw	3500.000000	0.2776	19	LAT_SED.hru	4500.000000	0.8985
10	ESCO.bsn	0.300000	0.3694	20	CN2.mgt	0.160000	0.9118

- Among the parameters considered, the groundwater delay time **(GW_DELAY)**, maximum canopy storage **(CANMX)**, and deep aquifer percolation fraction **(RCHRG_DP)** were identified as the three most sensitive factors influencing the simulation of water flow rates.
- The parameter GW_DELAY directly pertains to groundwater dynamics, RCHRG_DP represents a fraction indicating the percolation of water through the soil profile, and CANMX serves as a direct indicator of the hydrological response unit's hydrological response.

SWAT model: multi-site calibration results



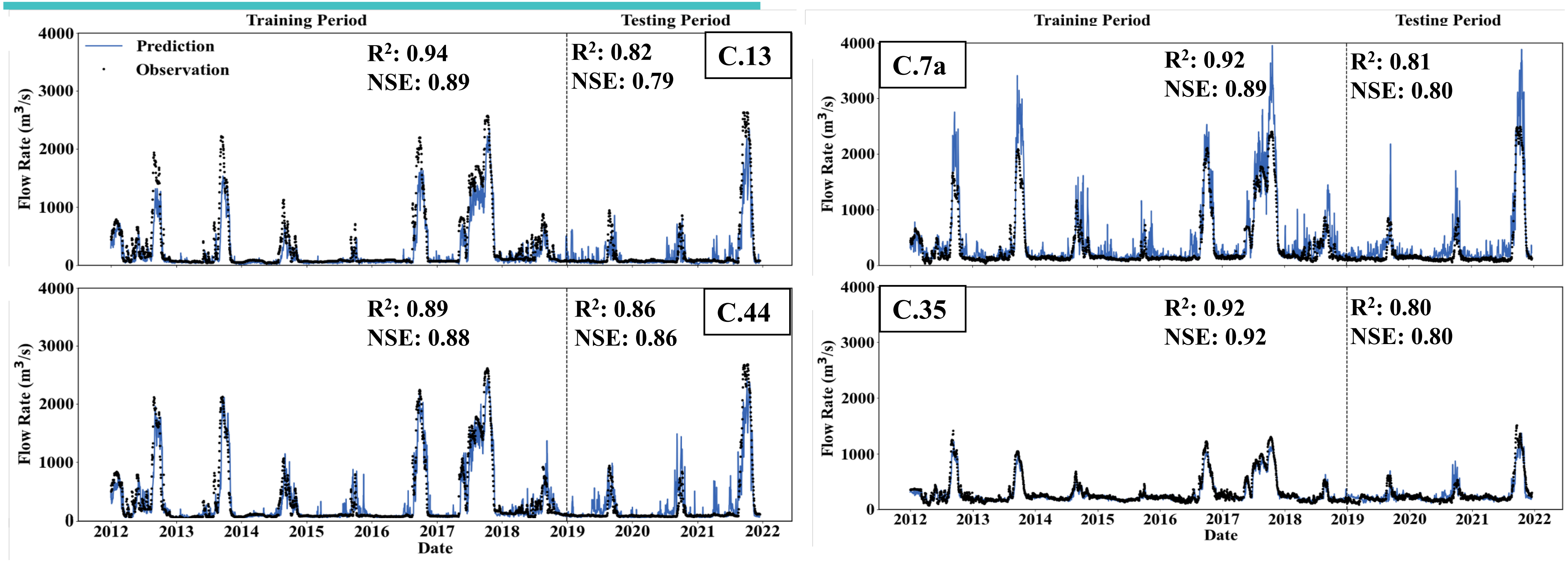
- The temporal comparison spanning 10 years between the observed data and simulated outcomes revealed that the **SWAT model** exhibited an **underestimation of the flow rate**, primarily attributable to the model's lack of consideration for tidal river dynamics.
- The simulation of the SWAT model in the context of a tidal river necessitates the application of **alternative approaches** (Upadhyay et al., 2022).

SWAT-LSTM: HPO results

Stations	Batch size	Dropout 1	Dropout 2	Dropout 3	Unit 1	Unit 2	Unit 3	Lookback	Optimizer
C.13	64	0.475	0.425	0.425	128	16	128	6	Adam
C.44	64	0.25	0.175	0.375	32	32	128	5	Adam
C.7a	64	0.45	0.35	0.325	64	32	16	5	Adam
C.35	128	0.4	0.175		16	16		12	Adam

- The loss function employed in the Bayesian optimization method to optimize the model parameters was determined based on a 7-year dataset, utilizing 30 epochs, and evaluating the mean squared error (MSE) metric.
- **Station C.35** stood out from the other stations in terms of the lookback of the LSTM model.
- This distinction highlights the significant influence of **14-day intervals of tides**, specifically spring tide and neap tide, on the water dynamics observed at this particular station.

SWAT-LSTM: flow rate simulation results



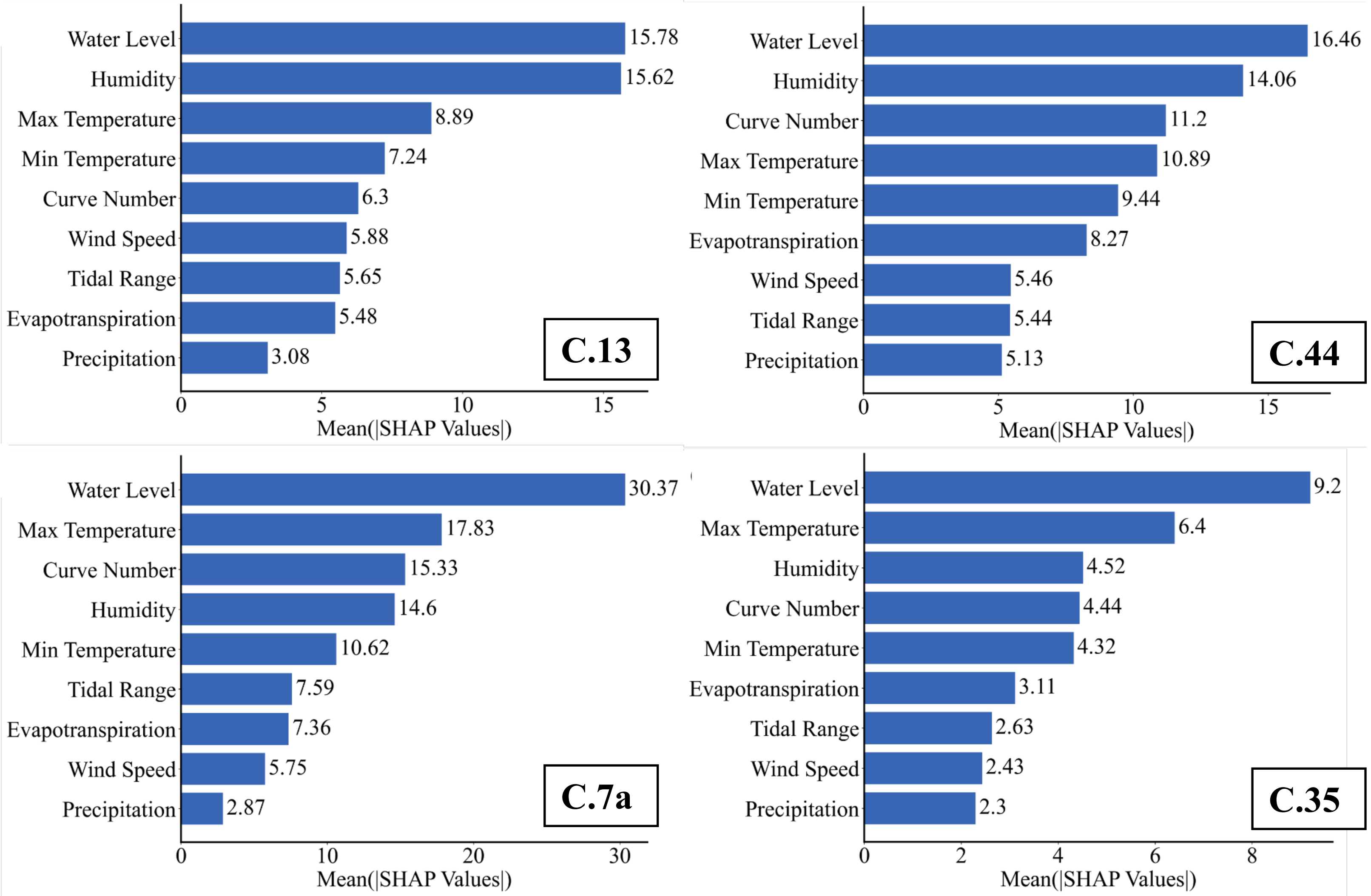
- The **SWAT-LSTM model** resulted in improved performance across all models, leading to **more accurate and reliable flow rate predictions**.
- The overall **R^2** and **NSE** values from all stations were higher than **0.80**, and **0.79** respectively.

Comparison of the model performance

Stations	Type of model	Calibrating/ Training performance		Validating/ Testing performance	
		R^2	NSE	R^2	NSE
C.13	SWAT	0.33	0.11	0.49	0.09
	SWAT-LSTM	0.94	0.89	0.82	0.79
C.44	SWAT	0.34	0.11	0.53	0.09
	SWAT-LSTM	0.89	0.88	0.86	0.86
C.7a	SWAT	0.34	0.11	0.51	0.09
	SWAT-LSTM	0.92	0.89	0.81	0.80
C.35	SWAT	0.37	0.13	0.60	0.11
	SWAT-LSTM	0.92	0.92	0.80	0.80

- **The inclusion of SWAT characteristic data and tide-related data** in the LSTM model led to notable changes in the Nash-Sutcliffe Efficiency (NSE) value.
- The utilization of SWAT characteristic data in the LSTM model incorporated to enhance the model's learning capacity by capturing relevant patterns and dynamics.

The importance of input features



- The analysis identified the **water level at the shoreline** as the most crucial input feature.
- Changes or variations in **water level** will have a relatively larger impact on the model's output compared to other features.
- In contrast, the analysis revealed that **precipitation** had the least impact as an input feature, its influence on the models in the tidal river context was relatively minimal compared to other factors.

4. Conclusion

Conclusion

Summary

1. A coupled approach combining the SWAT model with LSTM was developed. This integration aimed to mitigate the difficulty caused by tidal effects and improve the accuracy of flow rate simulations.
2. The implementation of the SWAT-LSTM model resulted in a significant enhancement of the overall model performance.
3. By employing the SHAP, the relative importance was revealed that the water level at the shoreline emerged as the most influential input feature, indicating its substantial impact on the model's predictions and overall performance.

Future Plan

1. To further enhance the modeling capabilities and accuracy, other deep learning approaches can be introduced in conjunction with SWAT.
2. Considering more with dams, minor outlets, and discharges can indeed contribute to the development of a more accurate model.

Thank You

Email: Krit.fr41@unist.ac.kr