2022 年吉林大学中日韩青年论坛
China-Japan-ROK Youth Forum 2022

The 5th China-Japan-South Korea Water Science Research Forum

K water AI Lab

"Building Secure and Smart Water Systems in the Uncertain Future"

Hybrid machine learning and remote-sensing data applications in Water Resources Management

November 19th ,2022

Sung Hoon KIM(金成勳), Yeonsu KIM, Young Don CHOI, Young Teck HUR and Joonwoo Noh

Table of content

Background

(source) https://hedgersabroad.com/seomjin_river_cycling/

Heavy rain in Korea turns tourist favo into underwater island

2020.08.07 10:36:45

[Photo by Yonhap]

Jara Island on Bukhan River in Gapyeong, Gyeonggi Province, known for its annual jazz festival and popular camping sites, was drenched with water on Thursday after water level of Bukhan River surged overnight due to heavy downpour that led Soyang River Dam to discharge its floodgate. Ahead of water release, Gapyeong County evacuated mobile facilities such as caravans to higher land area. Soyang River Dam has been discharging up to 3,000 tons of water per second since 3 p.m. on Wednesday to adjust water level after torrential rain.

It is still difficult to control natural phenomena such as floods.

There was a big flood in Korea last August in 2020. It's because of the longest monsoon ever.

Under these conditions, continuous development of flood and river forecasting technologies is still necessary.

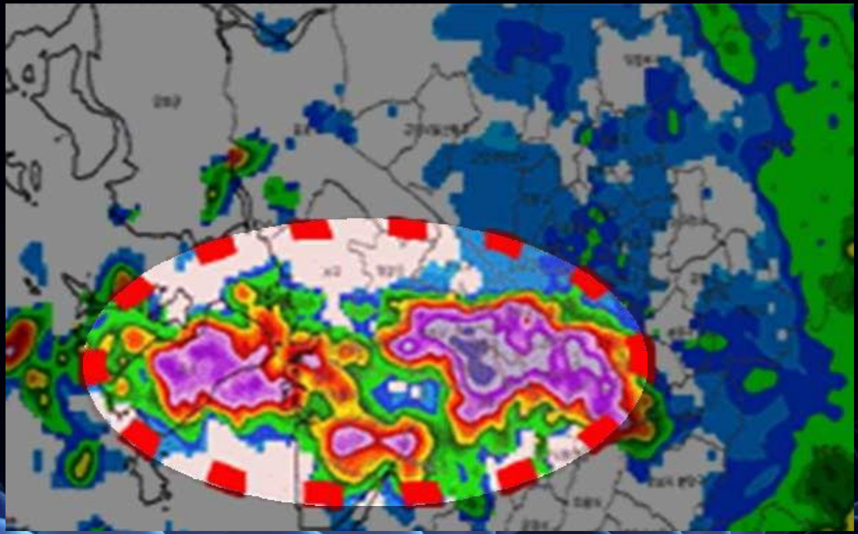
Background

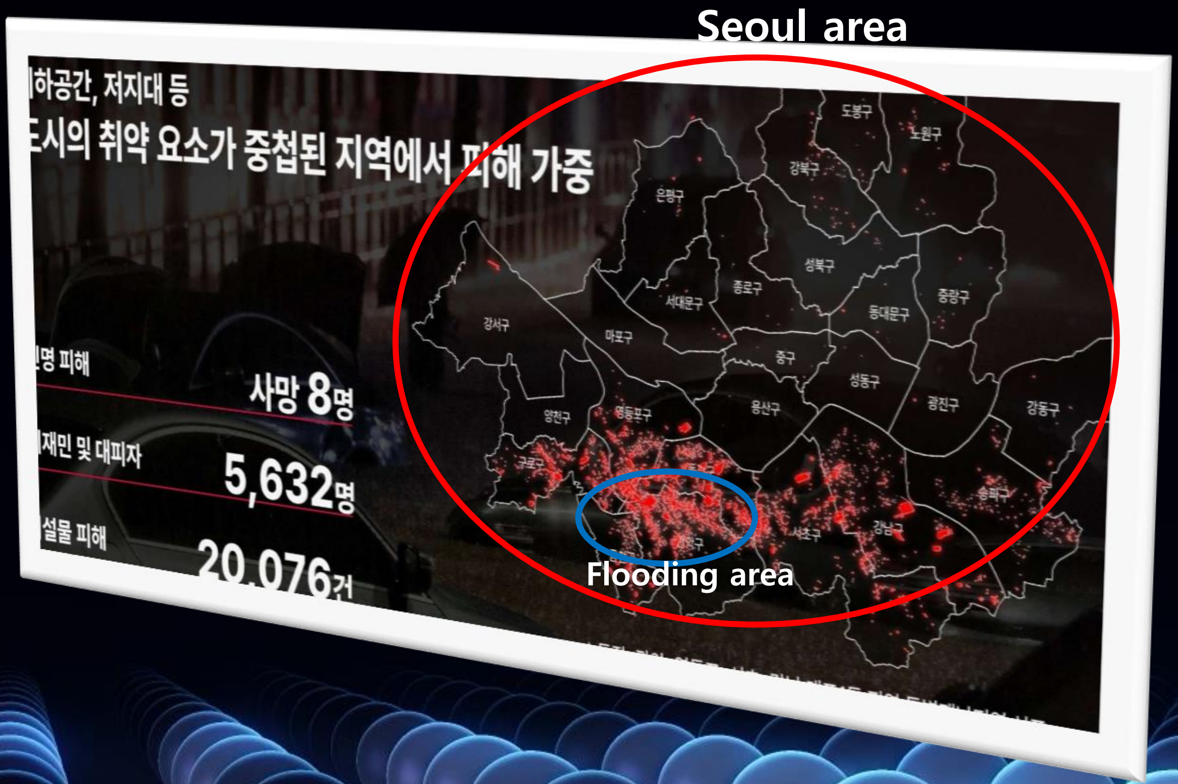
Recent Urban Flood Events (2022)

Background

Recent Urban Flood Events (2022)

Rainfall (mm)	동작구 (500 yr.)	강남구 (100 yr.)	서초구	구로구	서대문
1 hr.	141.5	116	110.5	101	33
1 day	381.5	326.5	354.5	289	129
2 day	515	439.5	480.5	410	229
accumulated	577	500	535.5	449.5	264

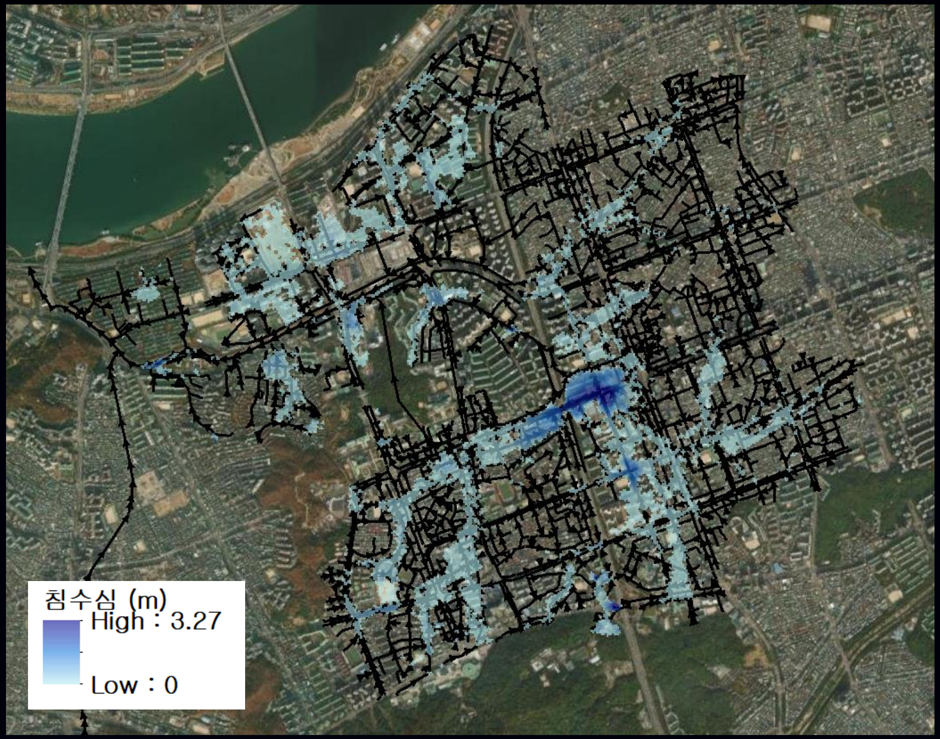




Background

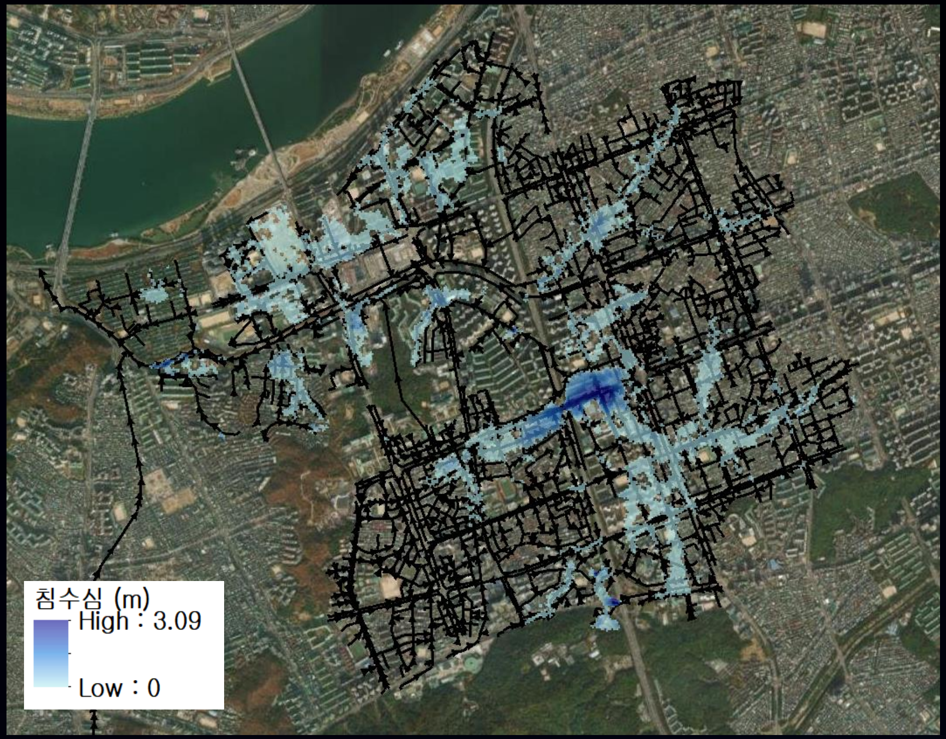
Before installing (110.5mm 60분 지속시간)

Flooding area : app. 235.04 ha



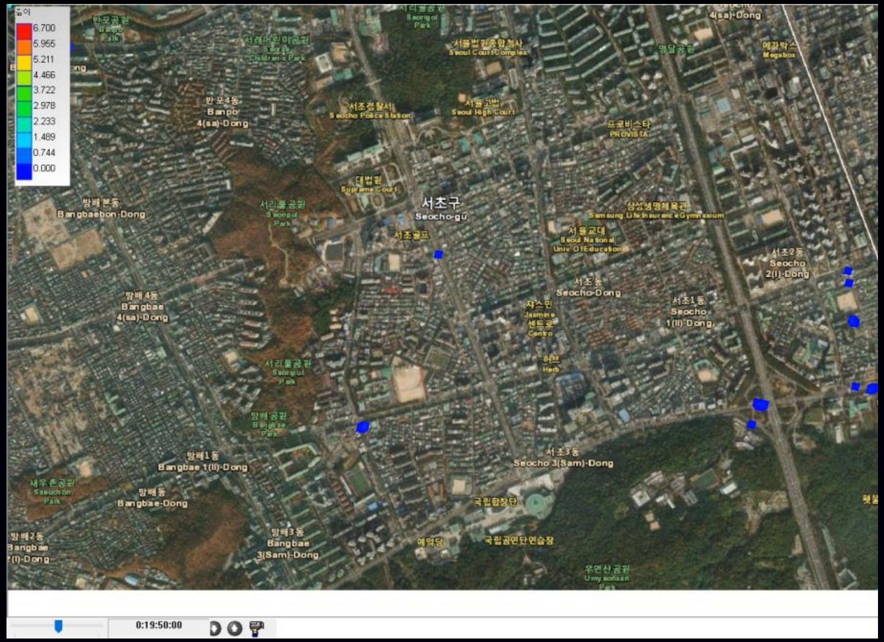
After installing (110.5mm 60분 지속시간)

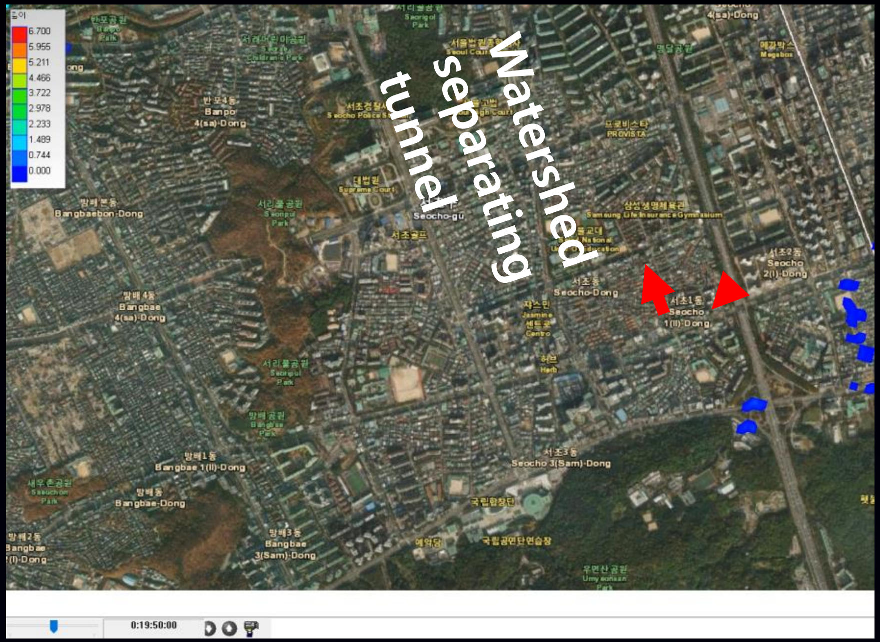
Flooding area : app. 204.31 ha



(source) 윤선권, 2022년 8월 서울시 홍수피해 현황 및 대책, Korea Water Resource Association

Background

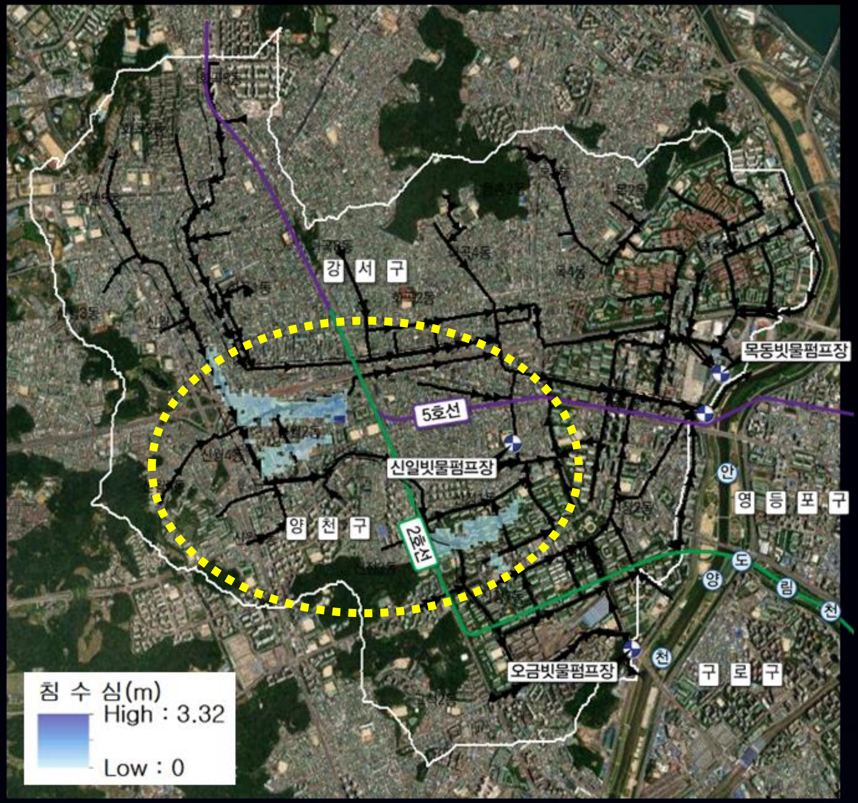




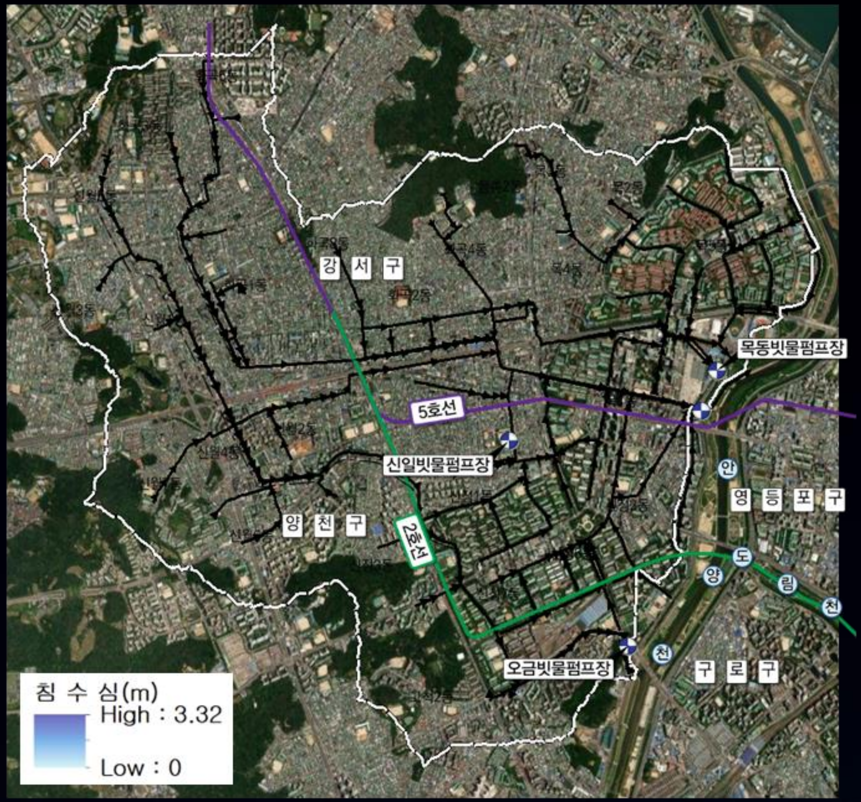
(source) 윤선권, 2022년 8월 서울시 홍수피해 현황 및 대책, Korea Water Resource Association

Background

Before (59.5mm 60min)



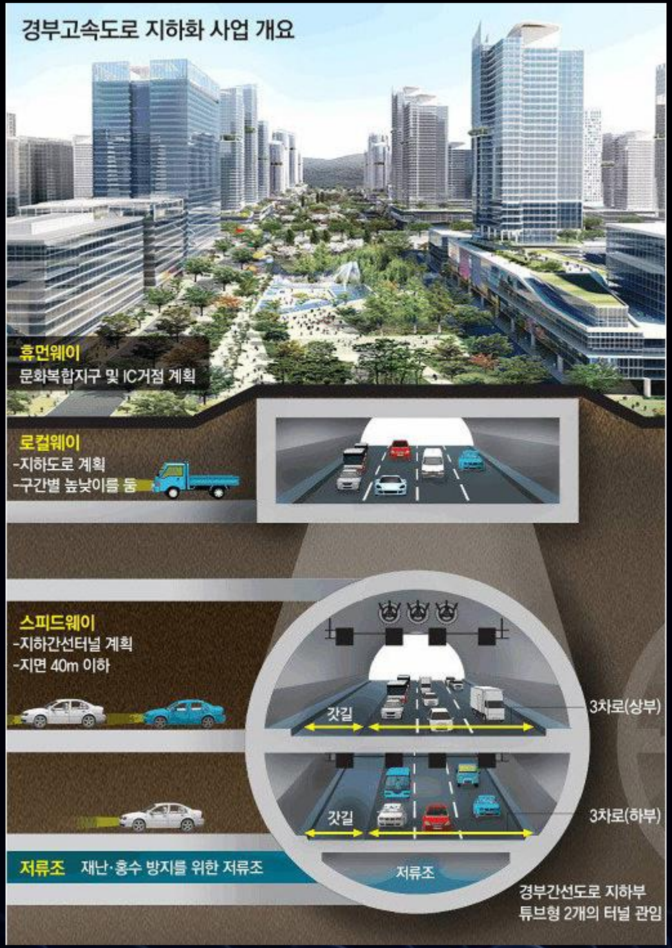
설치 후 (59.5mm 60min)



(source) 윤선권, 2022년 8월 서울시 홍수피해 현황 및 대책, Korea Water Resource Association

Background

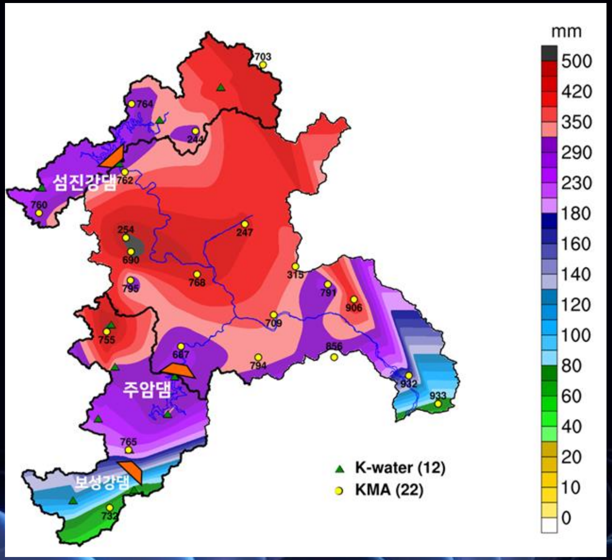
Seoul's ambition

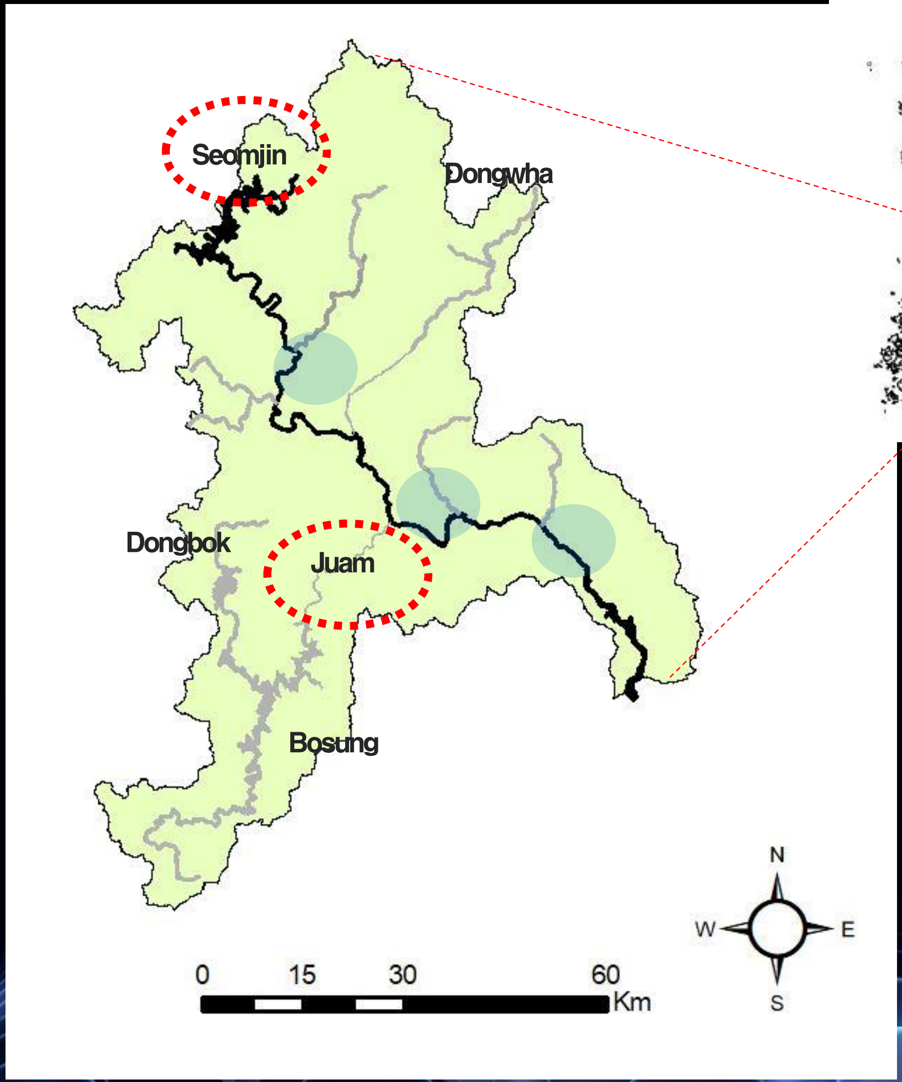


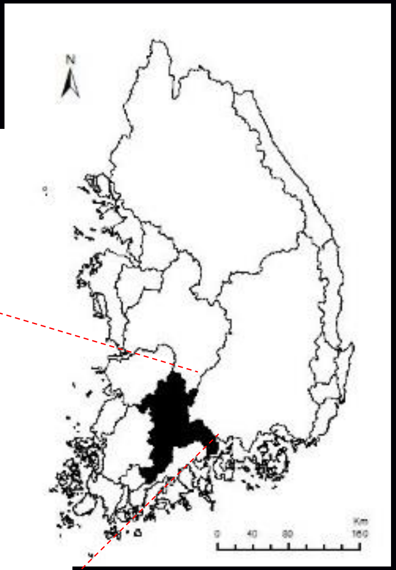
(source) 윤선권, 2022년 8월 서울시 홍수피해 현황 및 대책, Korea Water Resource Association

Flood in Seomjin River Basin (study area)

- Location : South west of Korean peninsula
- Basin Area : 4,911.9 km2
- River Length : 223.9 Km
- Main stream : Seomjin Dam in the upstream
- Tributaries : Juam, Bosung, Dongbok, and Dongwha dam
- Extreme flood from Aug. 7 to 8, 2020
- Average rainfall for 2 days : 347.8 mm



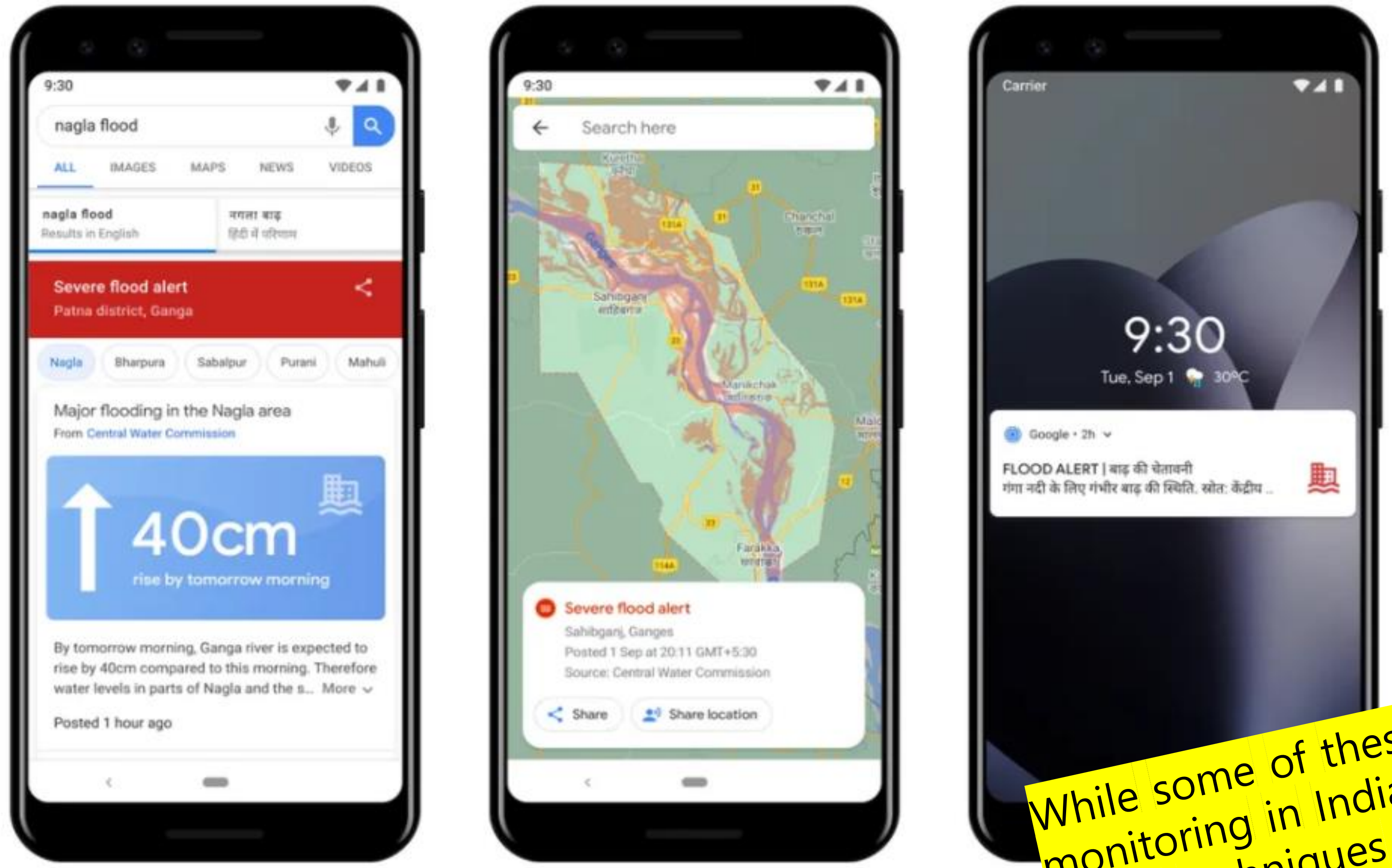




Extreme Flood over 100yr Period

Levee Breach in Namwon city

Machine Learning in Flood Management



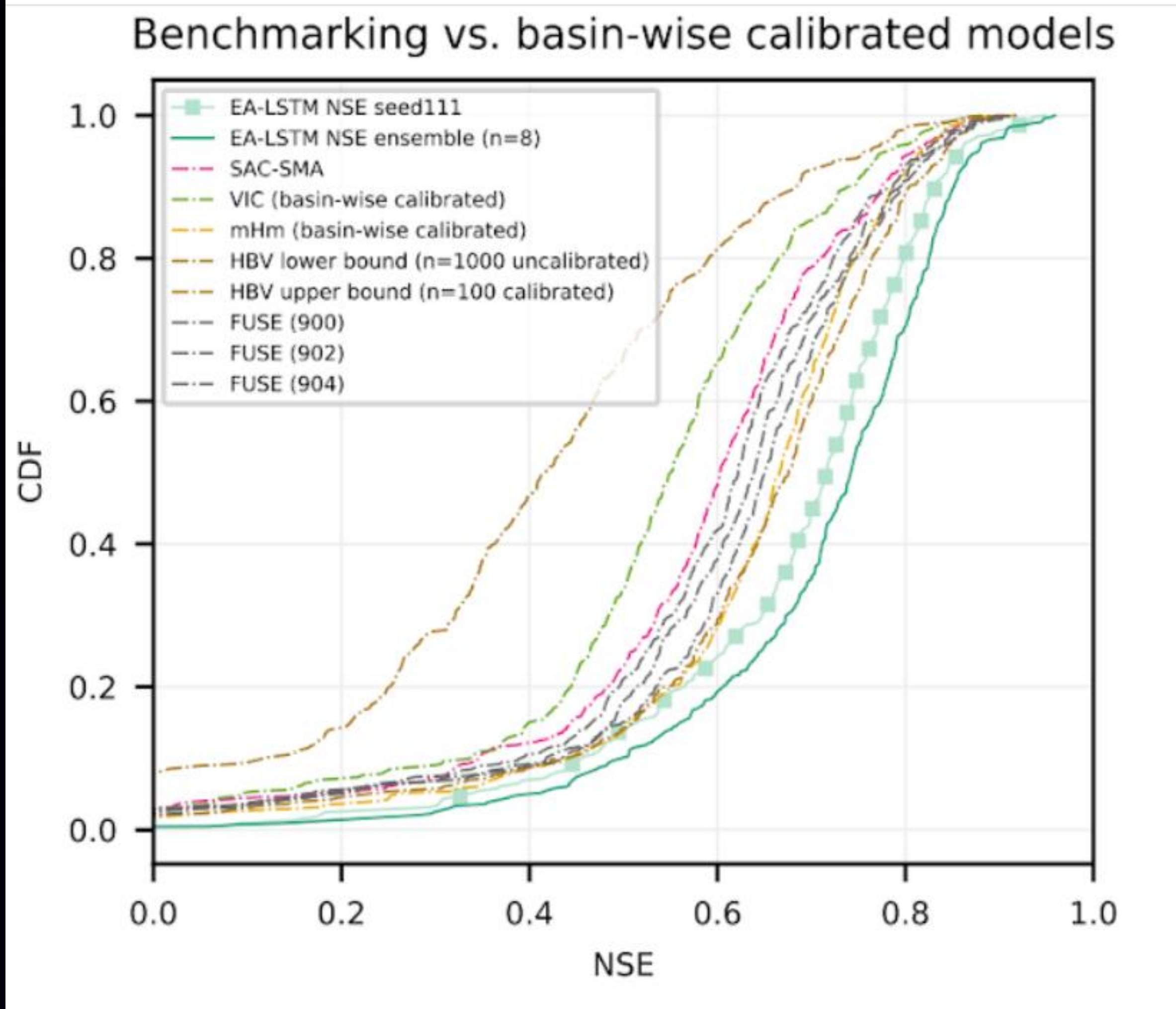
Google's new flood alerts offer information in some areas about the depth of the waters. | Ima

While some of these problems are specific to flood monitoring in India and other developing nations, some of the techniques Google has pioneered in India could change flood forecasting worldwide.

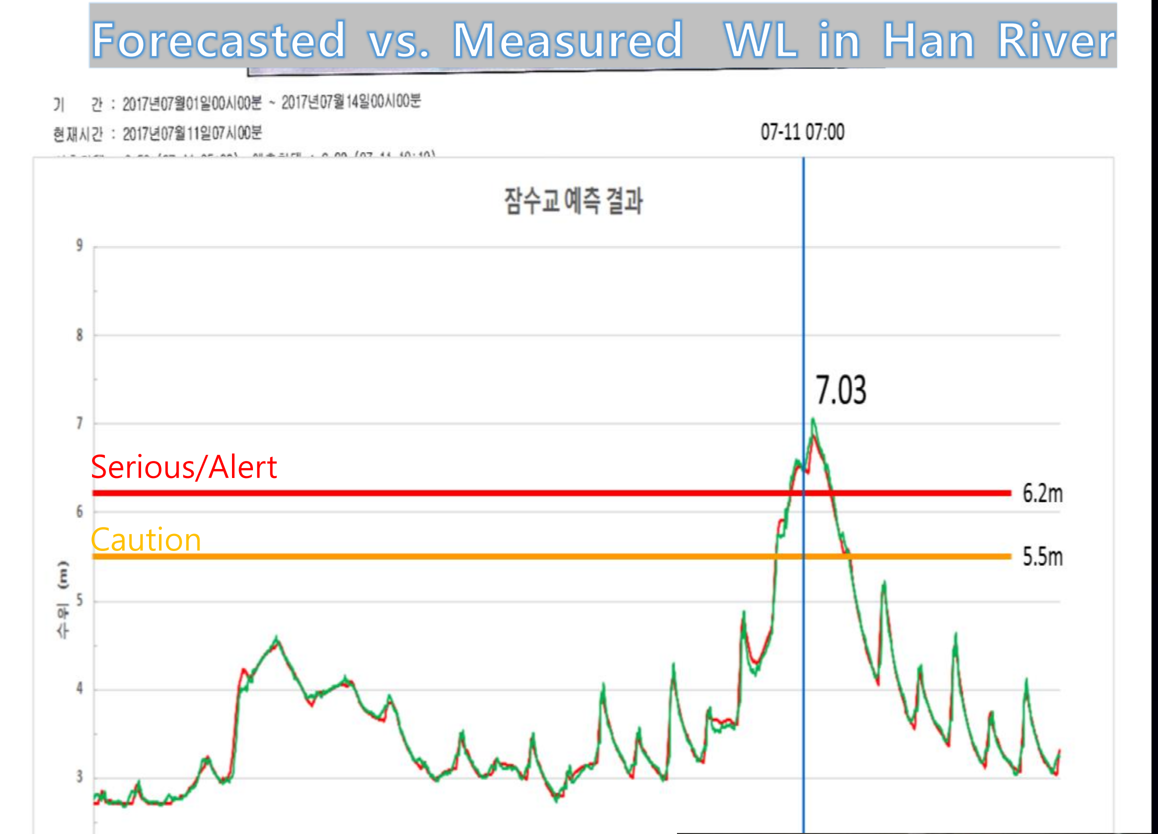
Sella Nevo, a senior software engineer at Google

(source) https://www.theverge.com/2020/9/1/21410252/google-ai-flood-warnings-india-bangladesh-coverage-prediction

Background



The distribution of NSE scores on basins across the United States for various models, showing the proposed EA-LSTM consistently outperforming a wide range of commonly used models.



(source) https://www.theverge.com/2020/9/1/21410252/google-ai-flood-warnings-india-bangladesh-coverage-prediction

(source) G.H.Lee, et. al. (2020), Kyungpook National University, Water Disaster Research lab.

Challenges

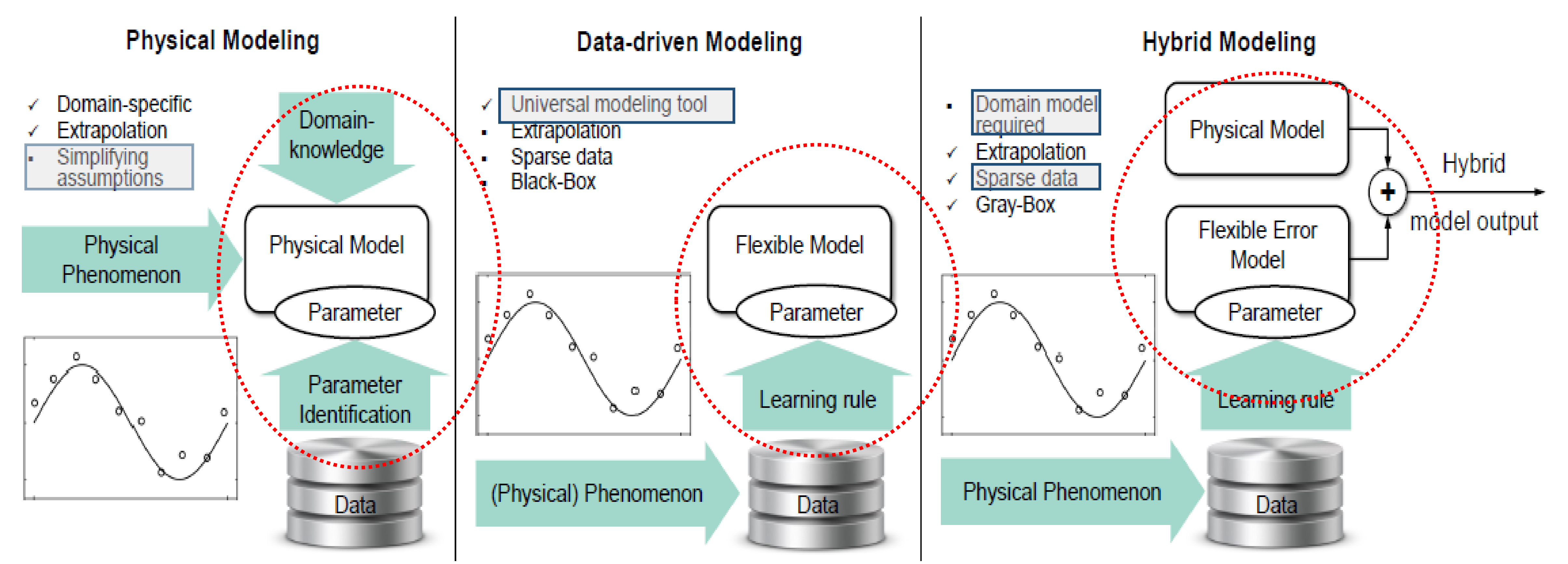
(Part-I)

- to evaluate practical applicability of various M/L methodologies for water resource management and operations with case studies
- to introduce a method that can partially surrogate river routing model
- to applying the reservoir operation optimization with scenarios based on hybrid machine learning model

(Part-II)

- to obtain more accurate and precise rainfall data by using remote sensing based data

Types of Modelling (Physical Model, DDM, and Hybrid)



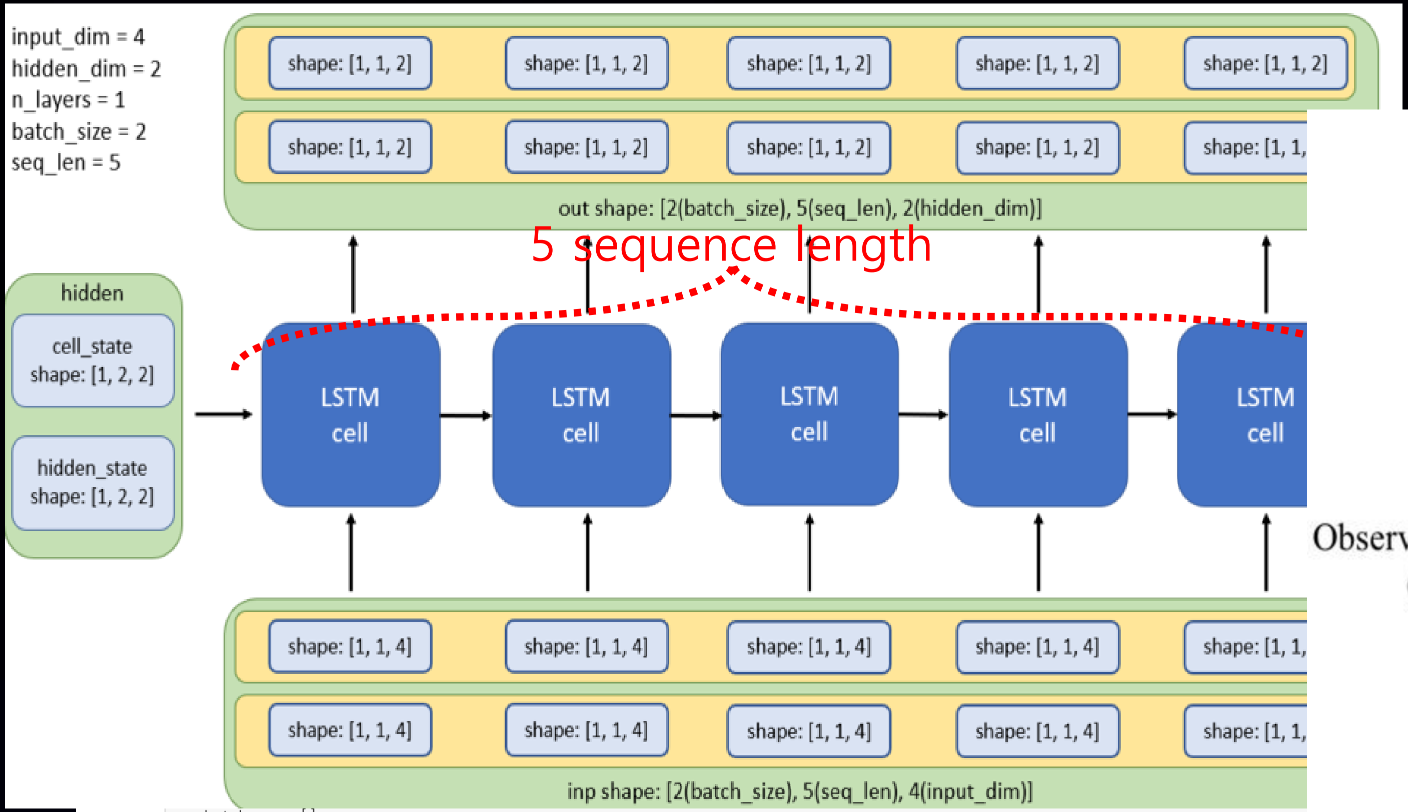
Deductive

y=function(x)

Inductive

y=function(x)

(Part I-a) : ML application for river routing

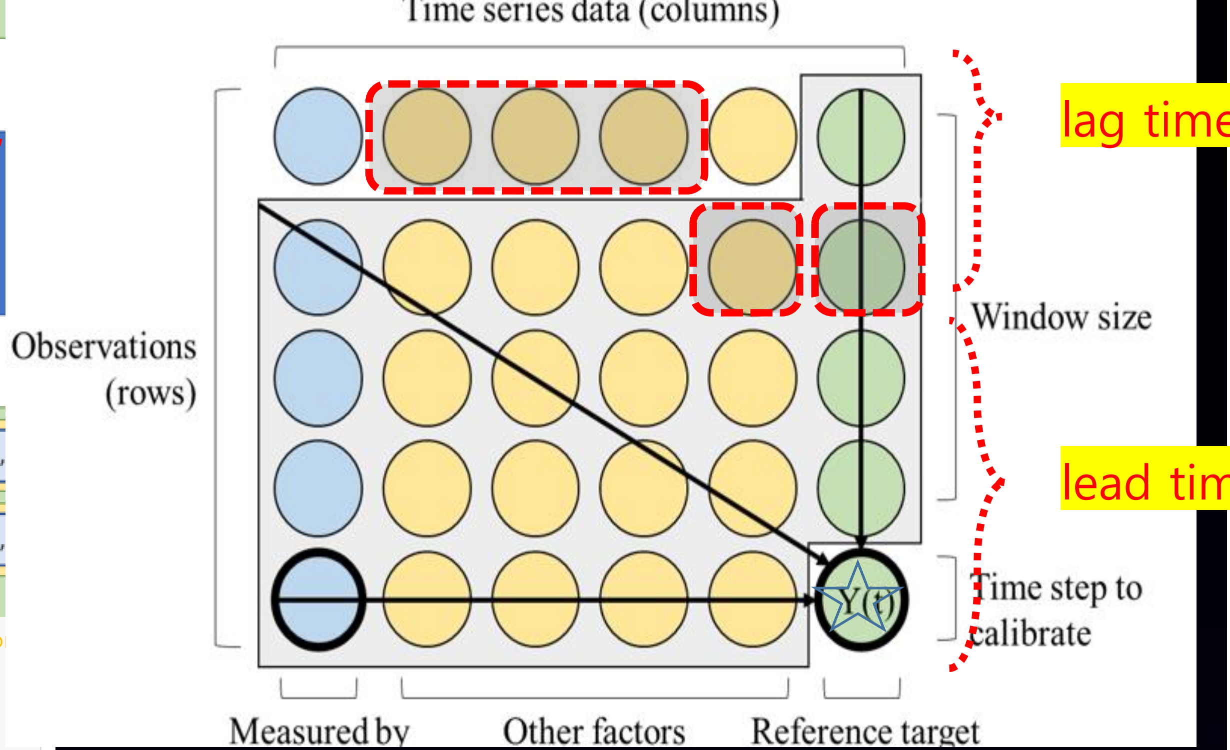



```
batch_xs = []
batch_ys = []

# make batches
for a in range(0, batch_size):
    arbit_x = data_x[a:a + sequence_length]
    arbit_y = data_y[a + sequence_length + 2]
    if a <= 100 :
        print(arbit_x, "->", arbit_y)
    batch_xs.append(arbit_x)
    batch_ys.append(arbit_y)

# make train, test datasets
train_size = 49464
vali_size = 8760
train_xs,train_ys = np.array(batch_xs[:train_size], np.float32), np.array(batch_ys[:train_size])
vali_xs, vali_ys = np.array(batch_xs[train_size:], np.float32), np.array(batch_ys[train_size:])

print(train_xs.shape, vali_xs.shape, train_ys.shape, vali_ys.shape)
```

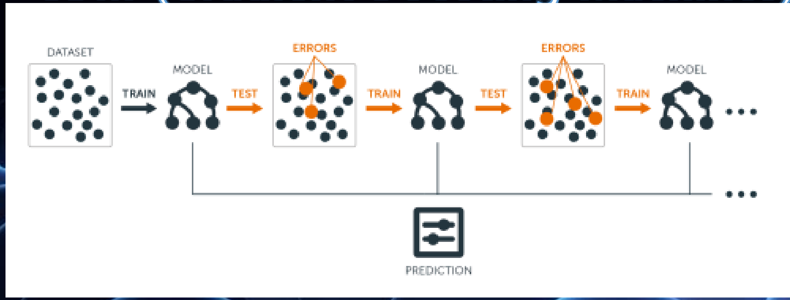
(Source) https://blog.naver.com/so

Schematic Diagram

(Source)
https://blog.naver.com/songblue61

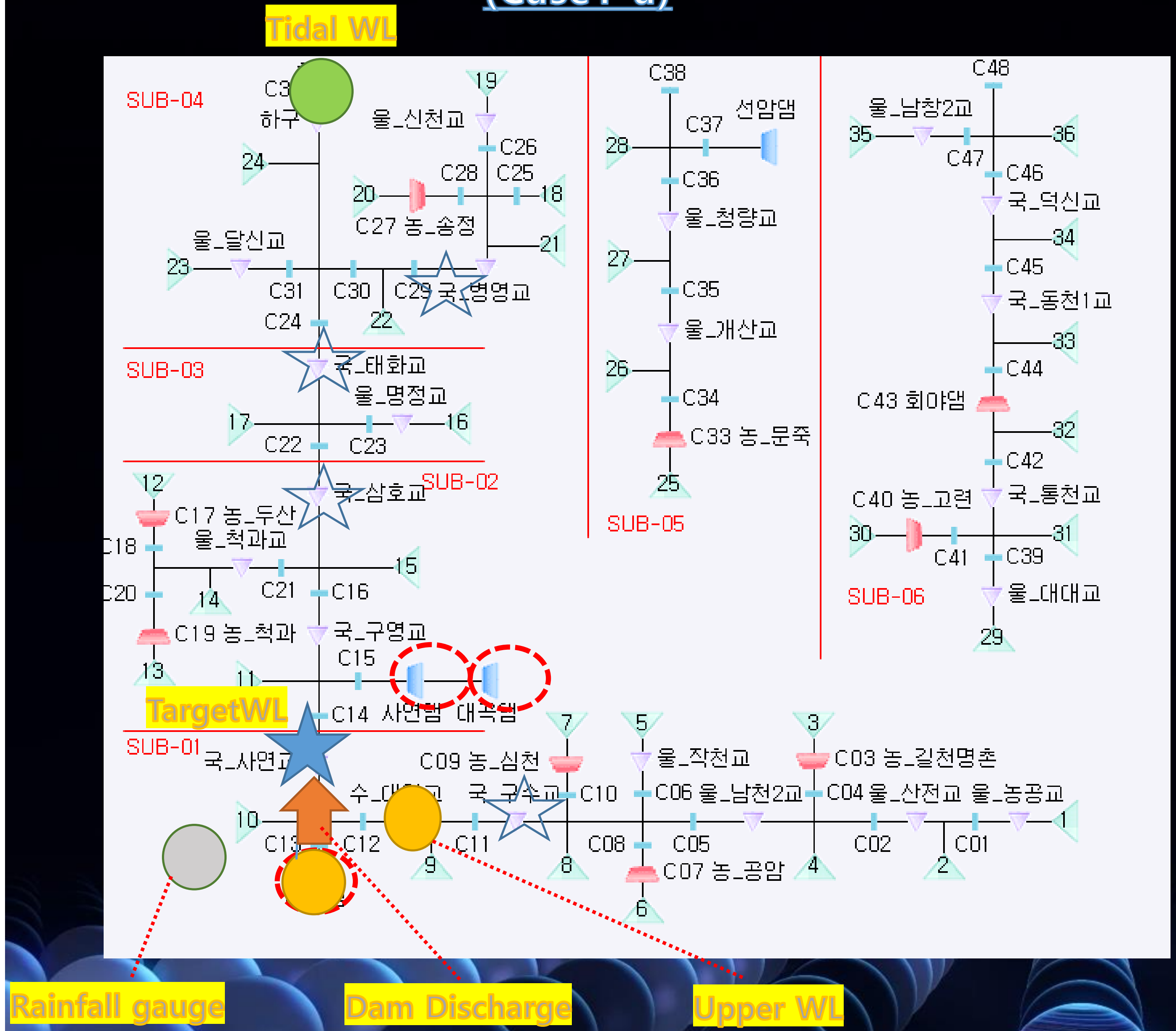


GBM (Gradient Boosting Machines)



(Part I-a) : ML application for river routing

(Case1-a)



Case(1-1) LSTM

Lead time : 1,2,3 hrs

Sequence_length = 2, 3, 6 (hrs)

Hidden_Layer = 3, 6, 10

Epochs = 5000

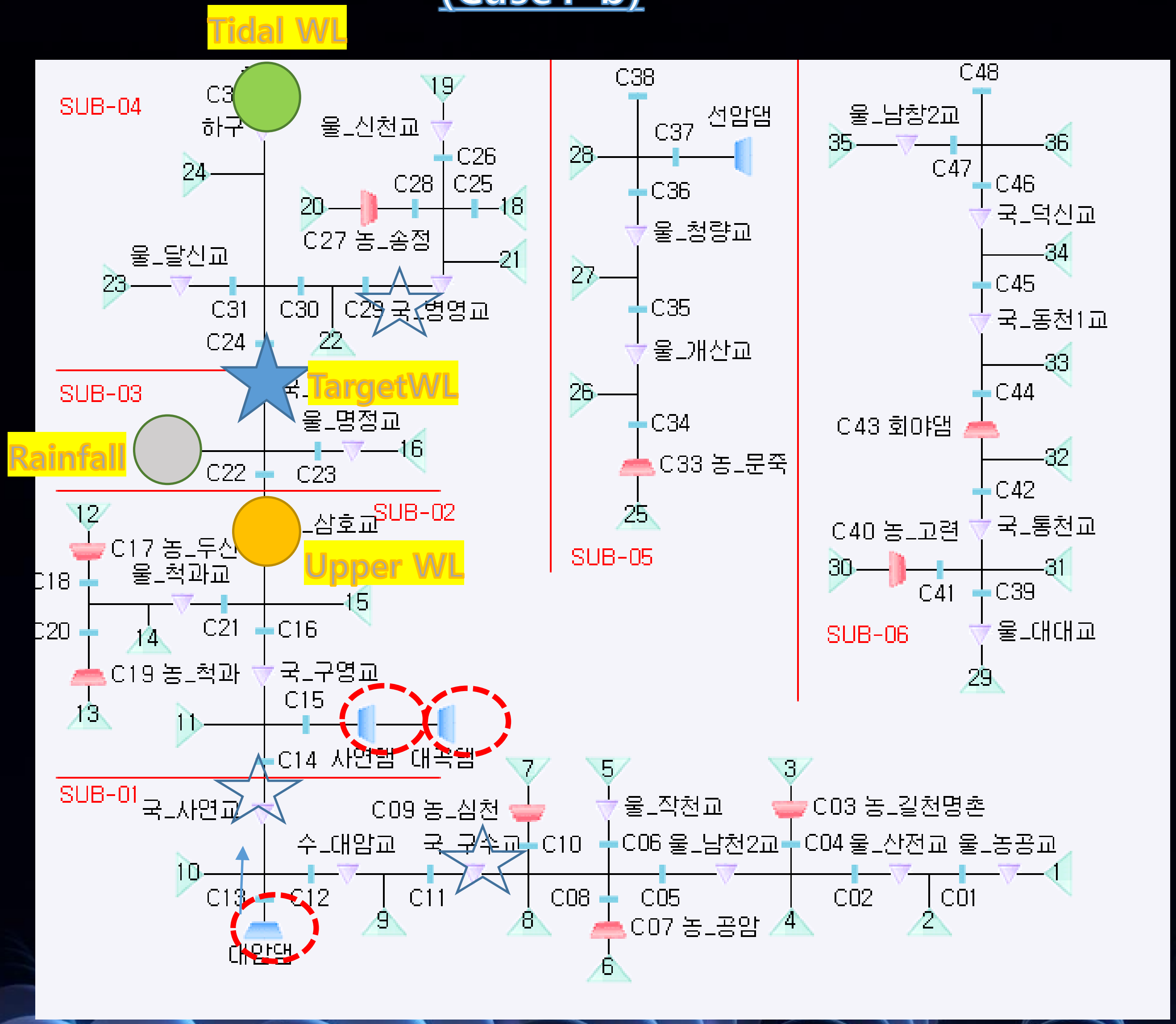
Learning_Rate = 0.001

Transfer_Function = reLU

Optimizer = atom

(Part I-a) : ML application for river routing

(Case1-b)



LSTM

Lead time : 1,3,5,9 hrs

Sequence_length = 1,2,4,8 (hrs)

Random Forest

Lead time : 3 hrs

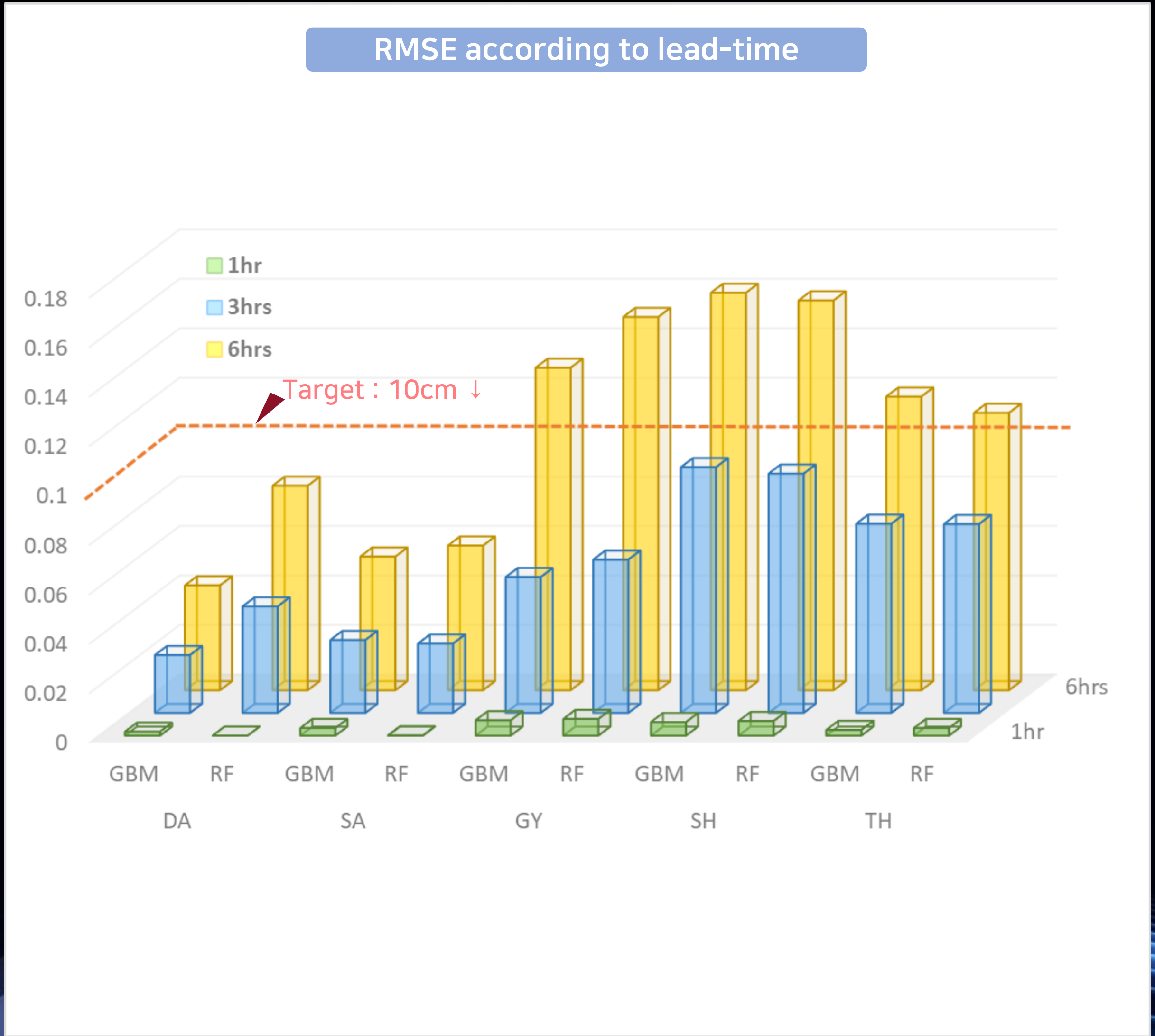
Regressor (n_estimators=30, random_state=15)

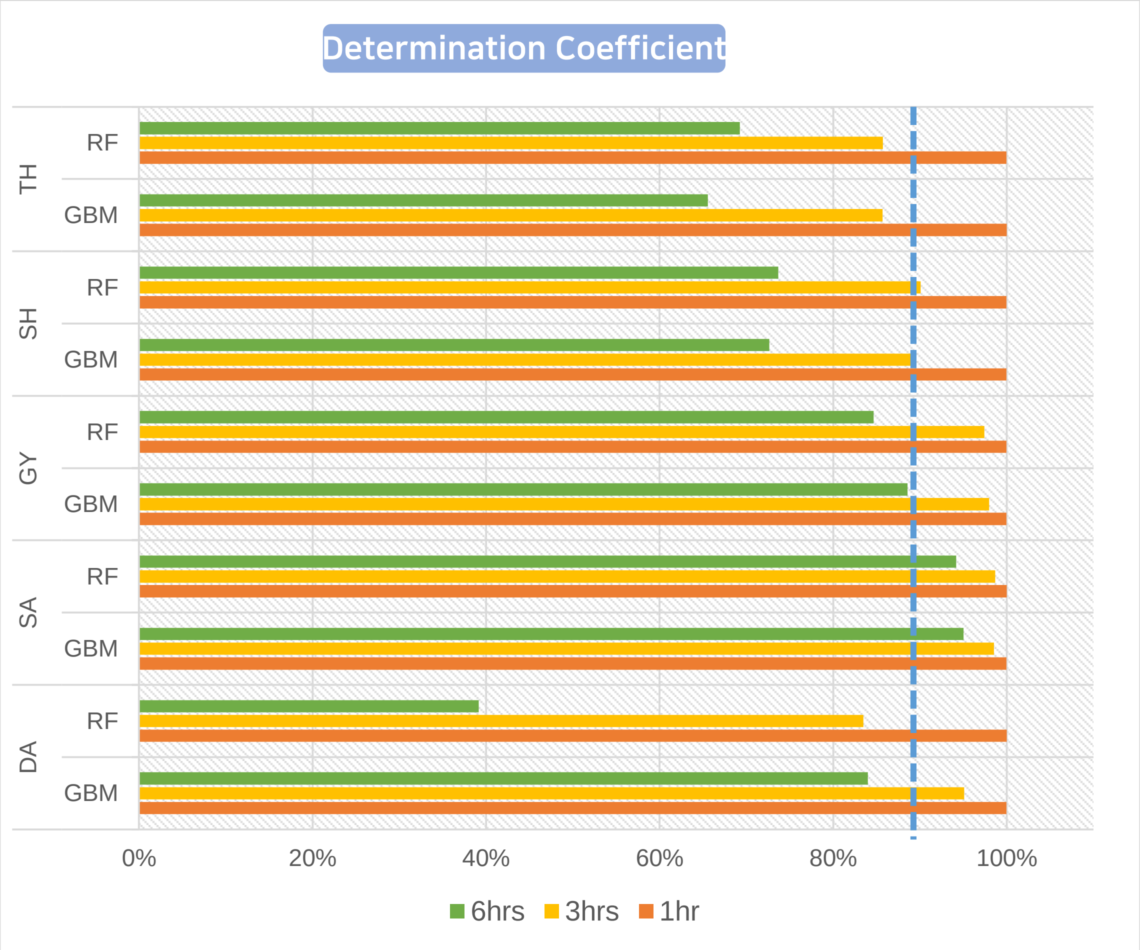
GBM

Lead time : 3 hrs

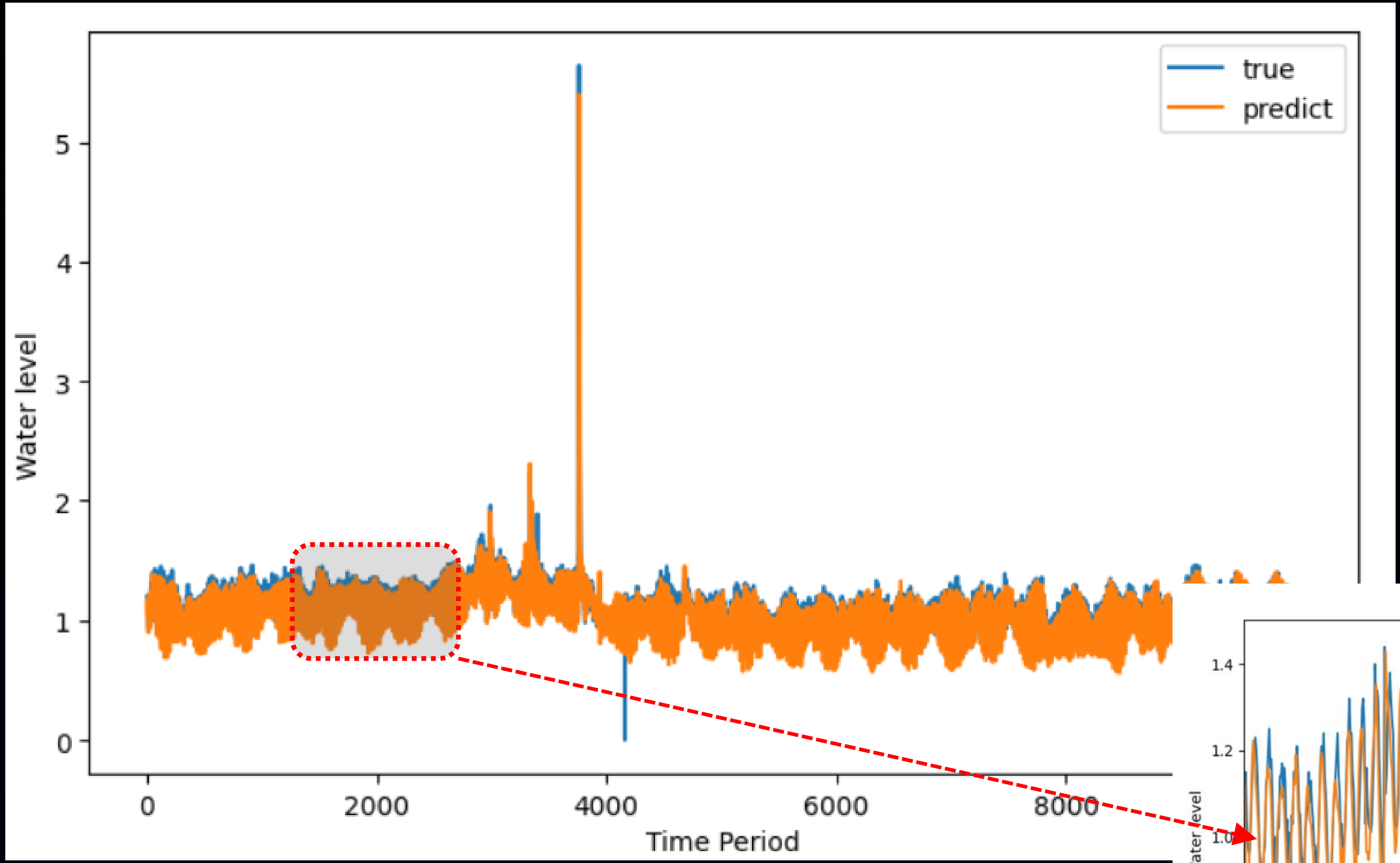
Regressor (n_estimators=100, max_depth=3)

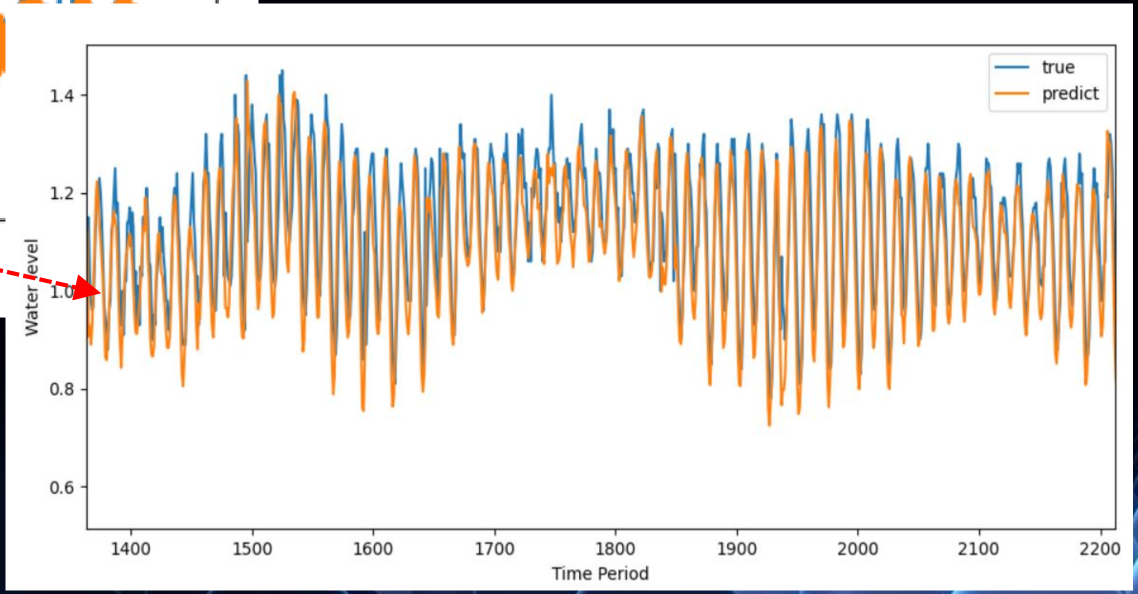
(Part I-a) : ML application for river routing





(Part I-a) : ML application for river routing



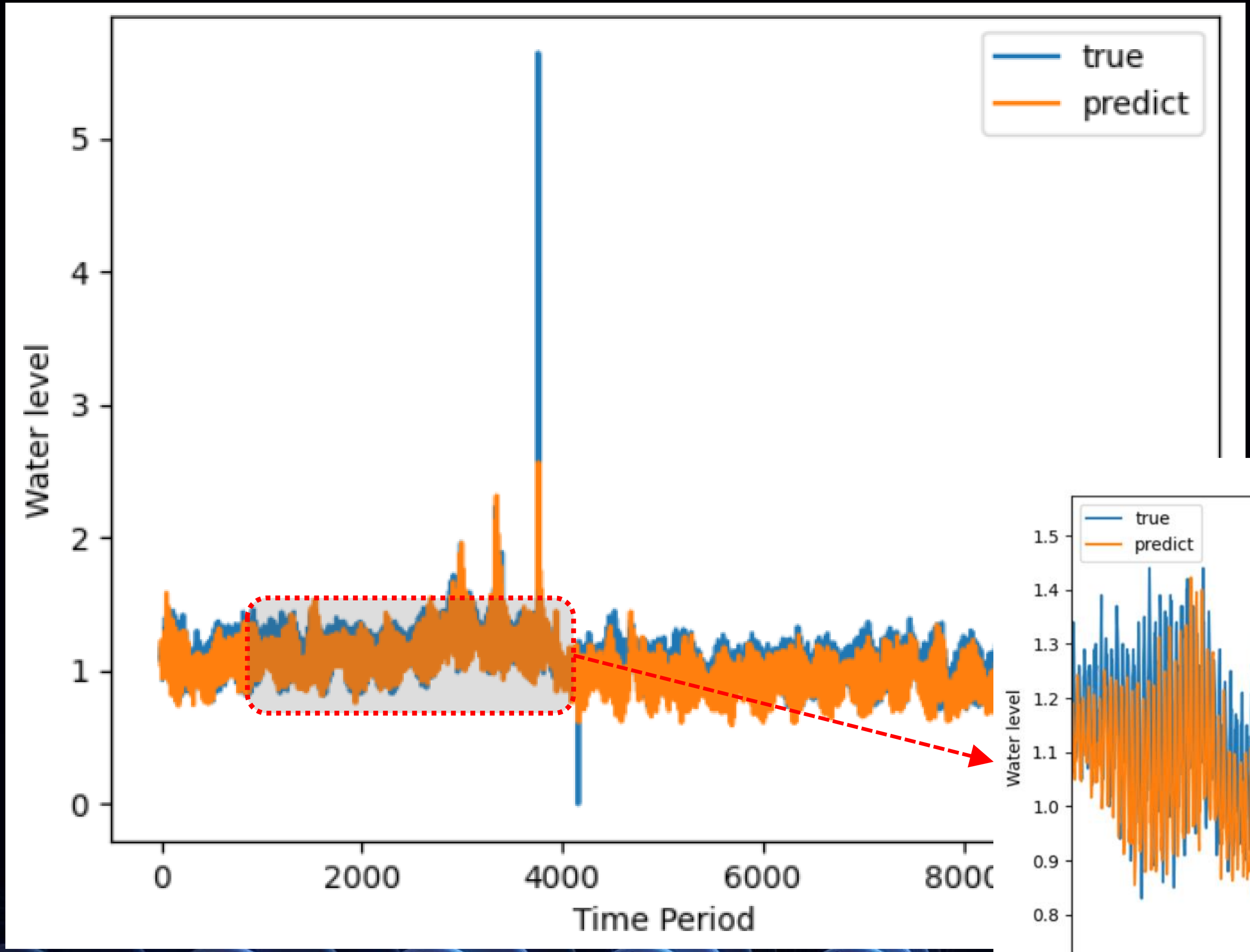


Lead time : 1 hr

(seq_len = 8, hidden = 7)

RMSE = 0.0158

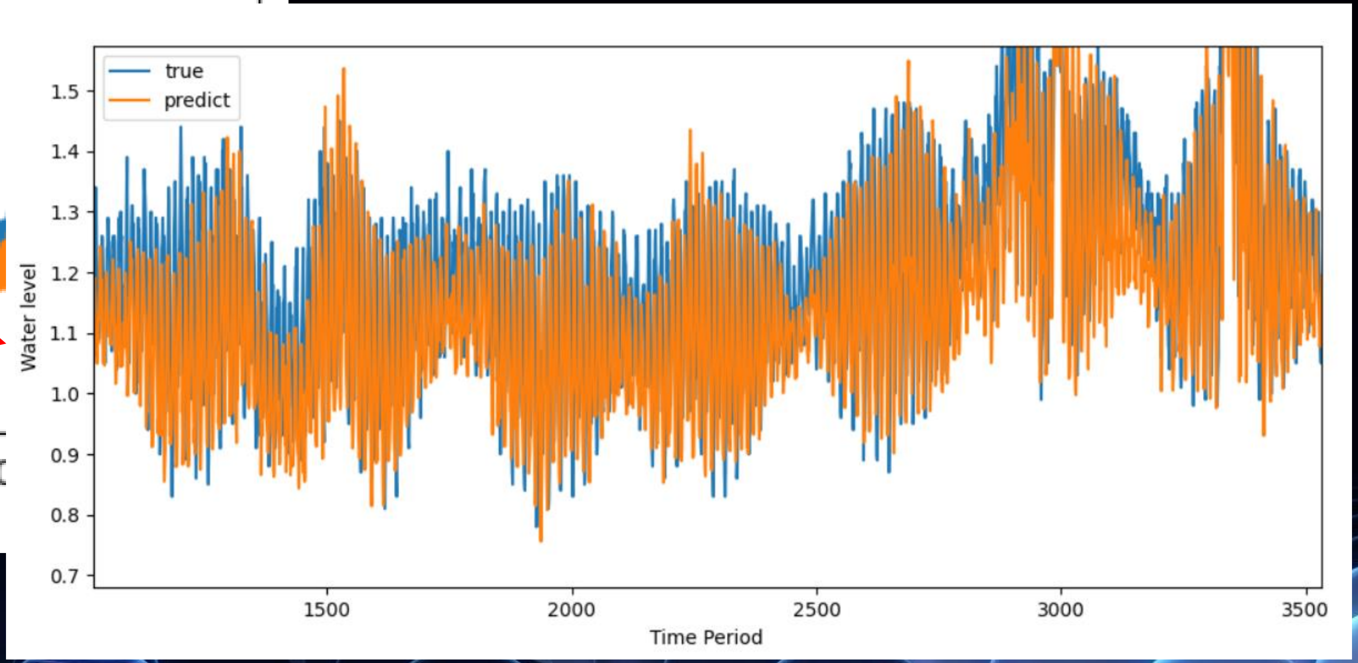
(Part I-a) : ML application for river routing



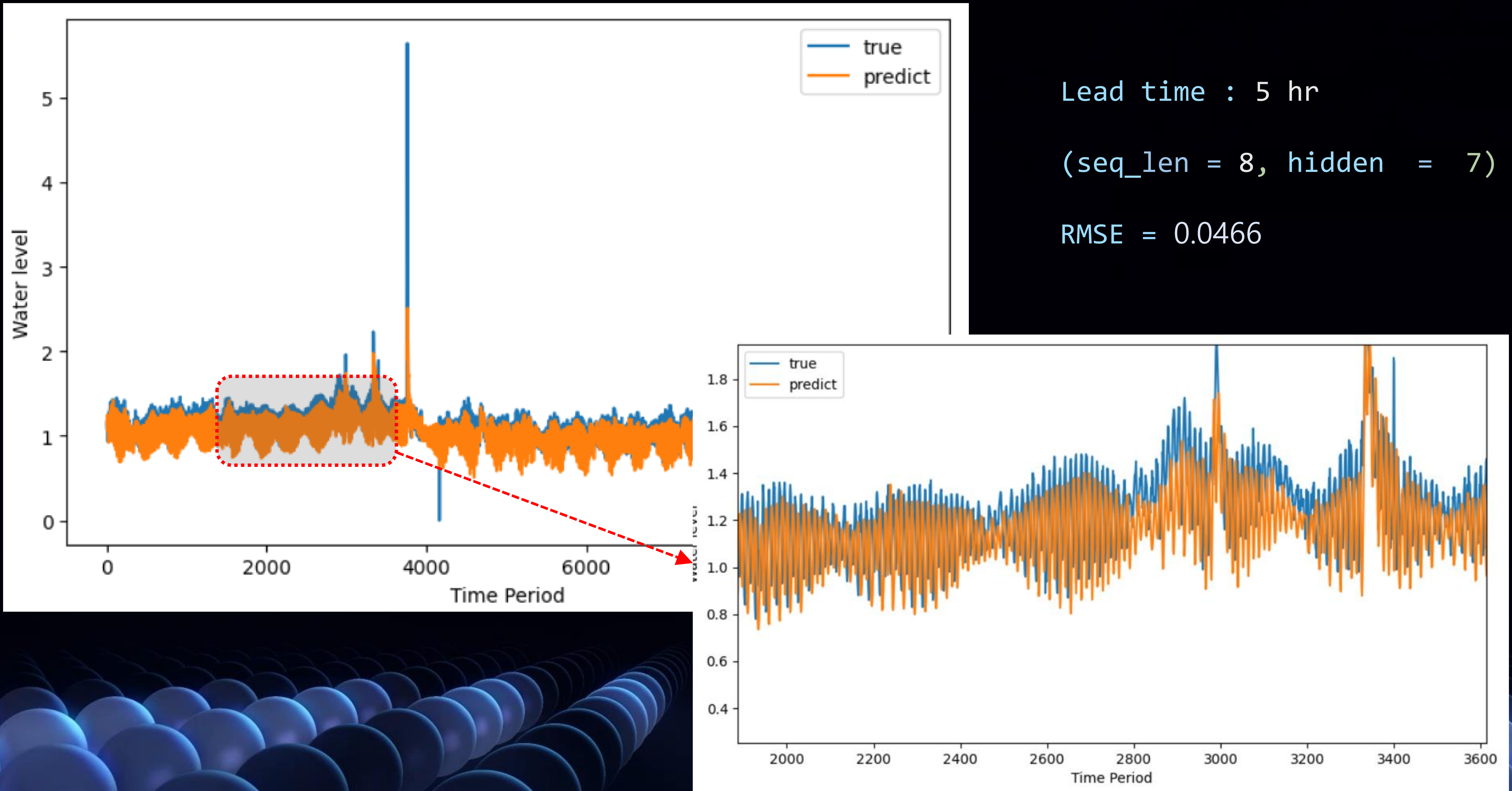
Lead time : 3 hr

(seq_len = 8, hidden = 7)

RMSE = 0.0321



(Part I-a) : ML application for river routing



Lead time : 5 hr

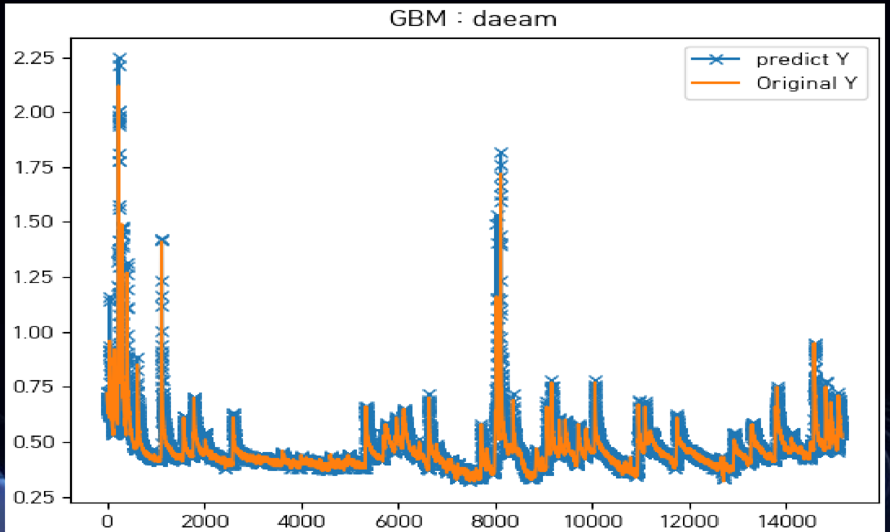
(seq_len = 8, hidden = 7)

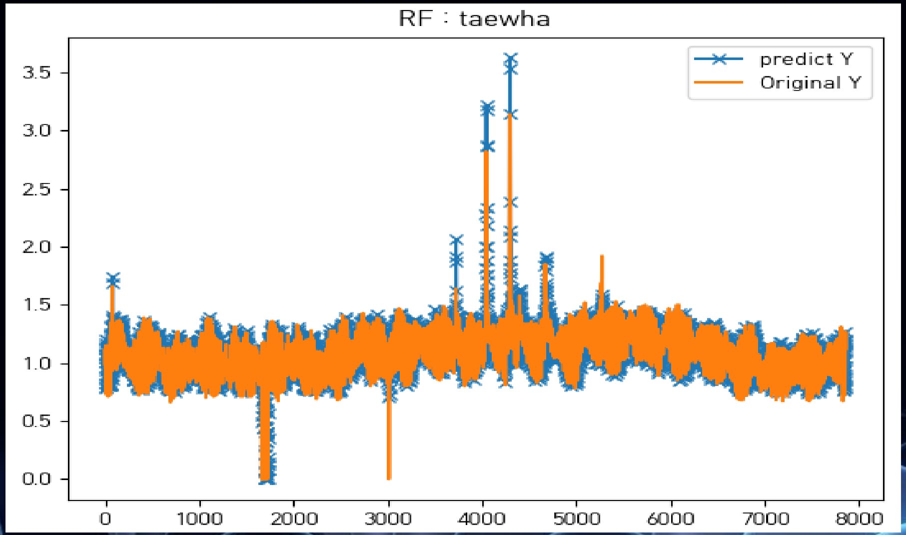
RMSE = 0.0466

(Part-I a): River Routing by using measured data

Automatic Hyper-parameter Optimization

Predicted for	Sequence length	Case	Station	Characteristics	Forecasting Results		Remarks
					RMSE	Deter. Coeff. (R^2)	Best ML Algorithms
3h	6h	various	DA	Upper	2.3cm	0.95	GBM > RF > BiL > LSTM
			SA	Middle	2.8cm	0.99	RF > GBM > BiL > LSTM
			GY	Middle	5.5cm	0.98	GBM > RF > BiL > LSTM
			SH	Middle	9.7cm	0.90	GBM~ > BiL > LSTM
			TH	Downstream (5km)	7.6cm	0.86	RF ~GBM > BiL > LSTM



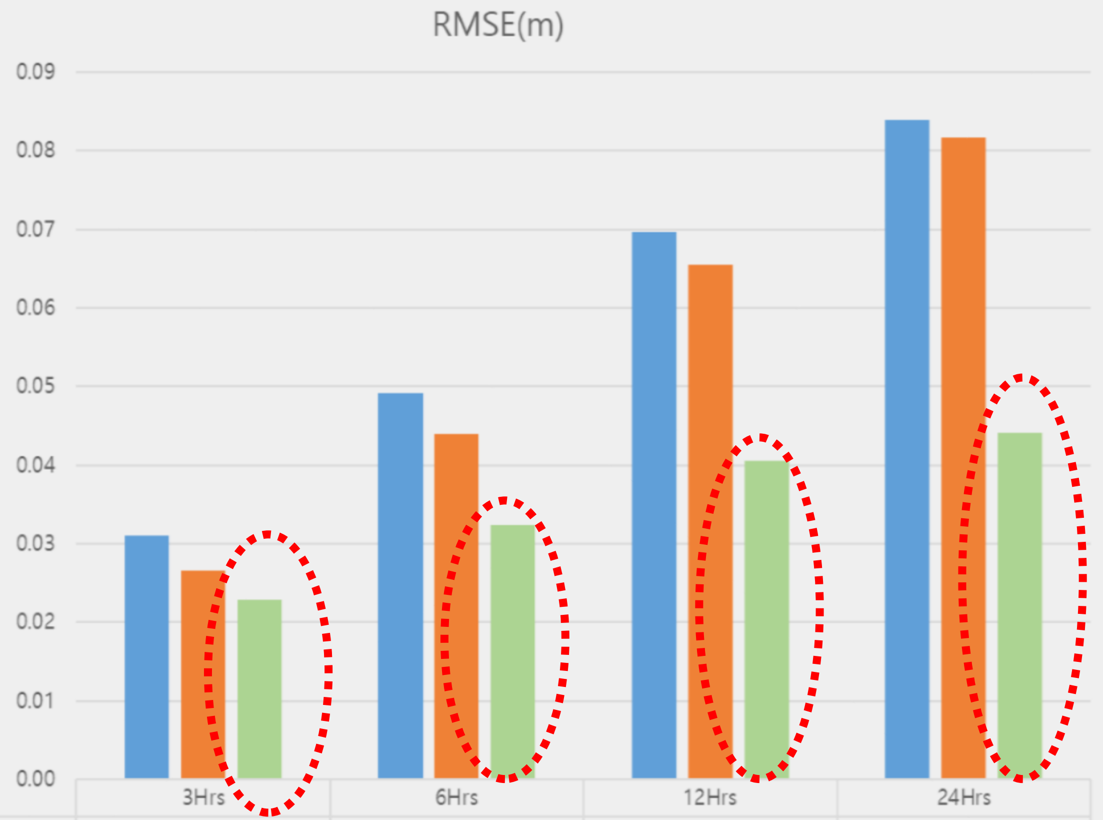


SO, what is Next Trial?

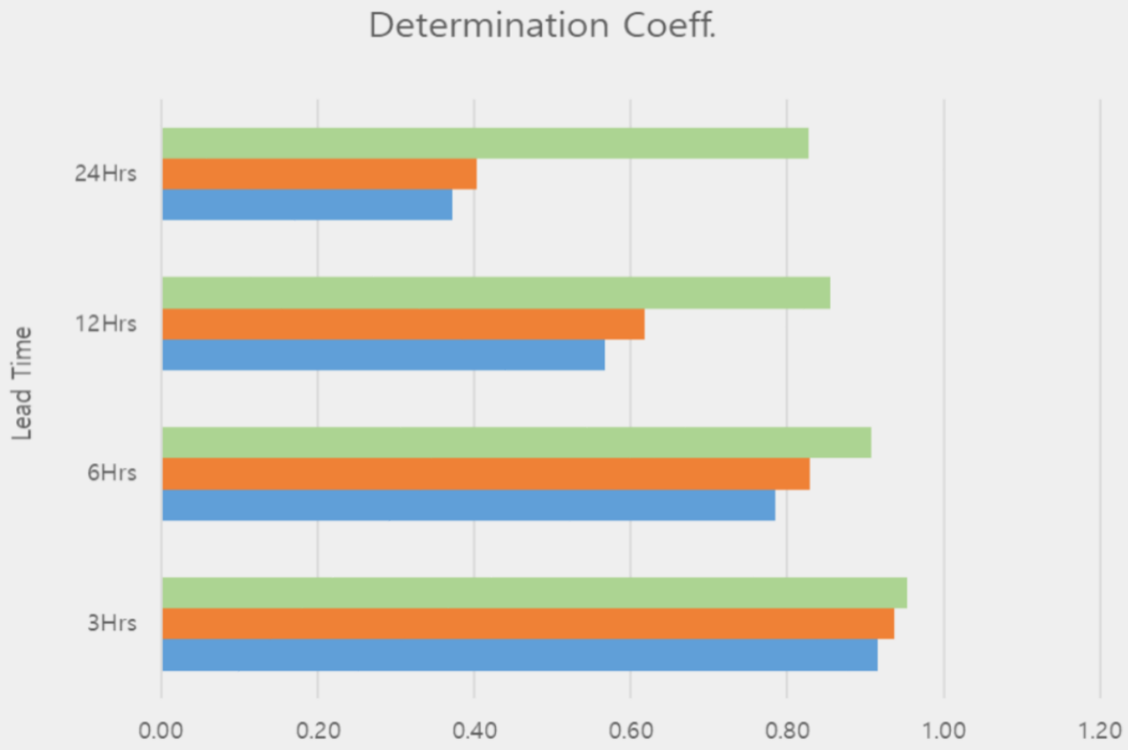
Combining weather forecast in M/L

(Part-I a): River Routing by using measured data

Lead-time Increase by adding rainfall forecast



	3Hrs	6Hrs	12Hrs	24Hrs
수위, 방류고려	0.03	0.05	0.07	0.08
실측강우 추가	0.03	0.04	0.07	0.08
예측강우 추가	0.02	0.03	0.04	0.04



	3Hrs	6Hrs	12Hrs	24Hrs
예측강우 추가	0.95	0.91	0.85	0.83
실측강우 추가	0.94	0.83	0.62	0.40
수위, 방류고려	0.91	0.78	0.57	0.37

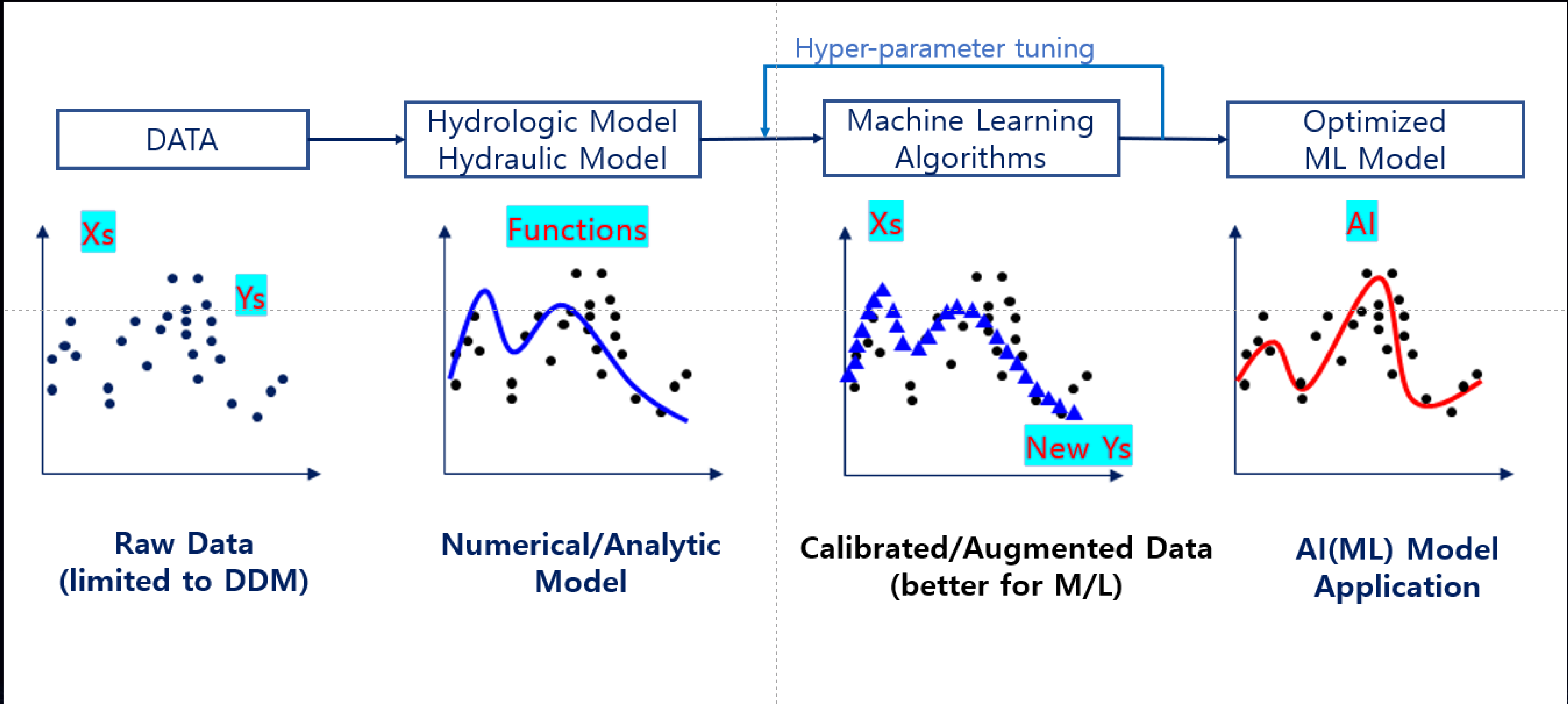
(Part-I b): Build the surrogate model for the river

K water

SO, what is Next Trial?

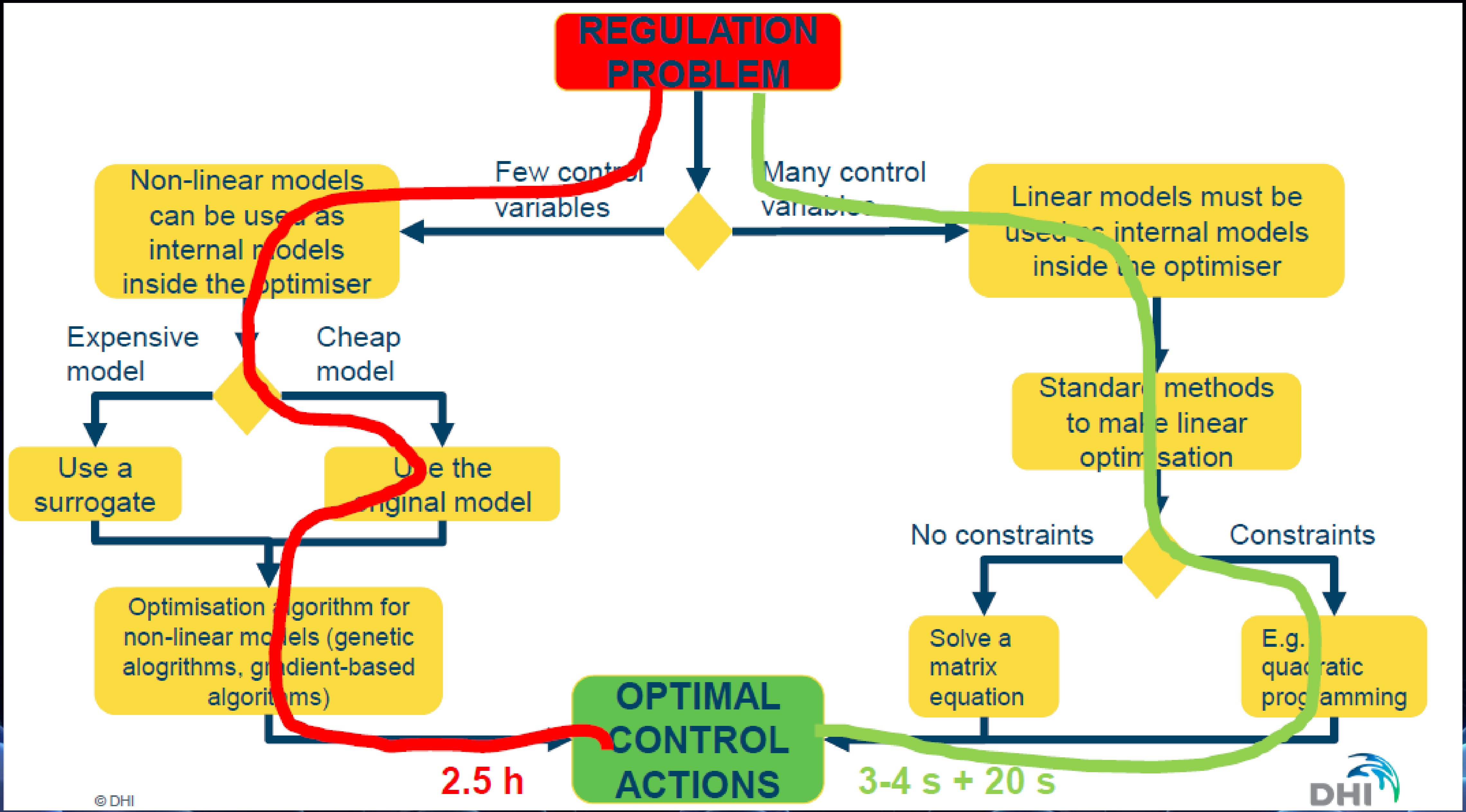
→ Physical Model +
Data Driven Model

Concept of Hybrid M/L (Physical Model + DDM)

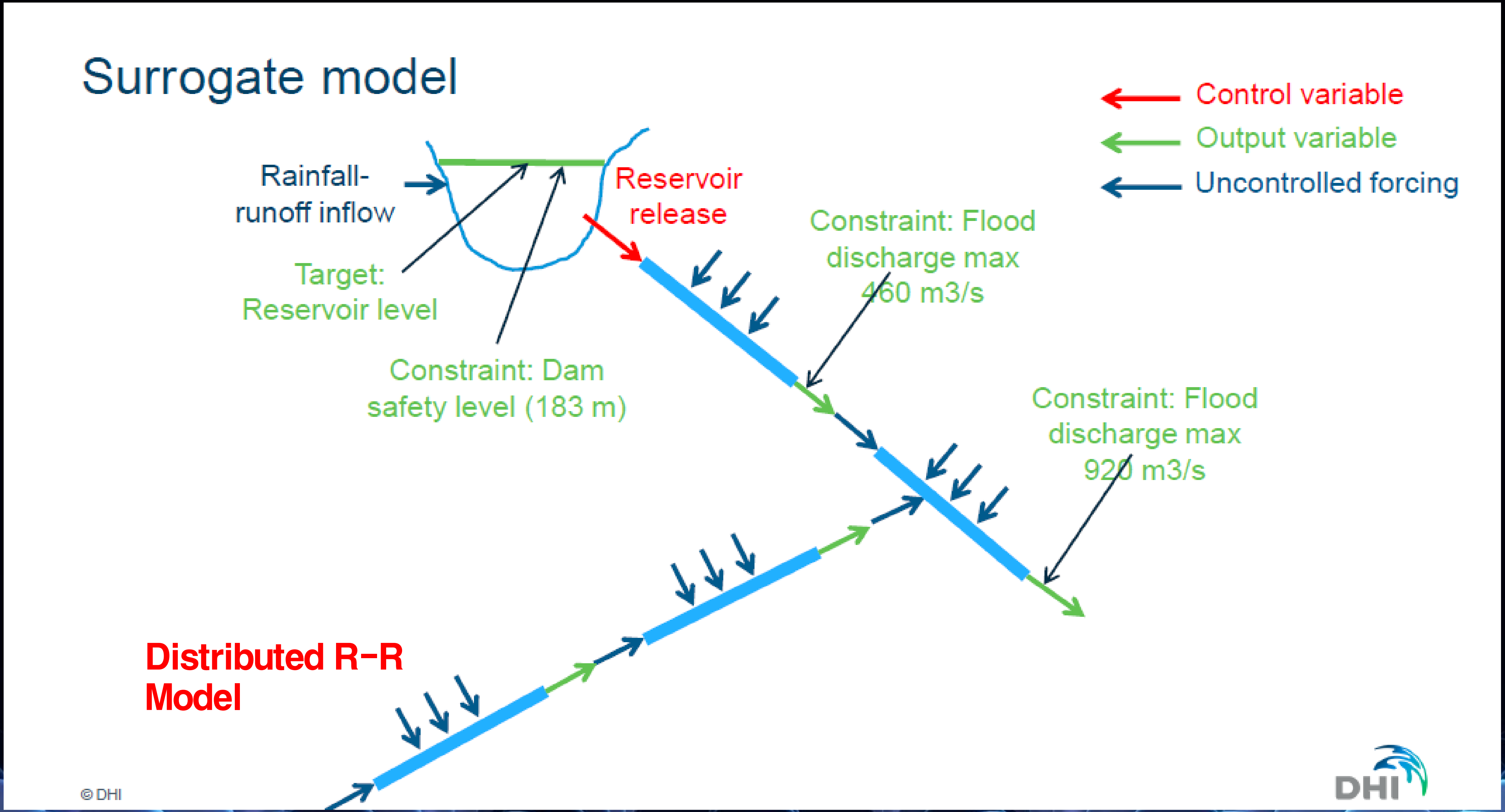


(Source) UNIST, 하이브리드(수리학-인공지능 연계) 유역 수질 예측모델 개발 및 영산강유역 수질관리 적용연구 (2021)

Challenges : surrogate model



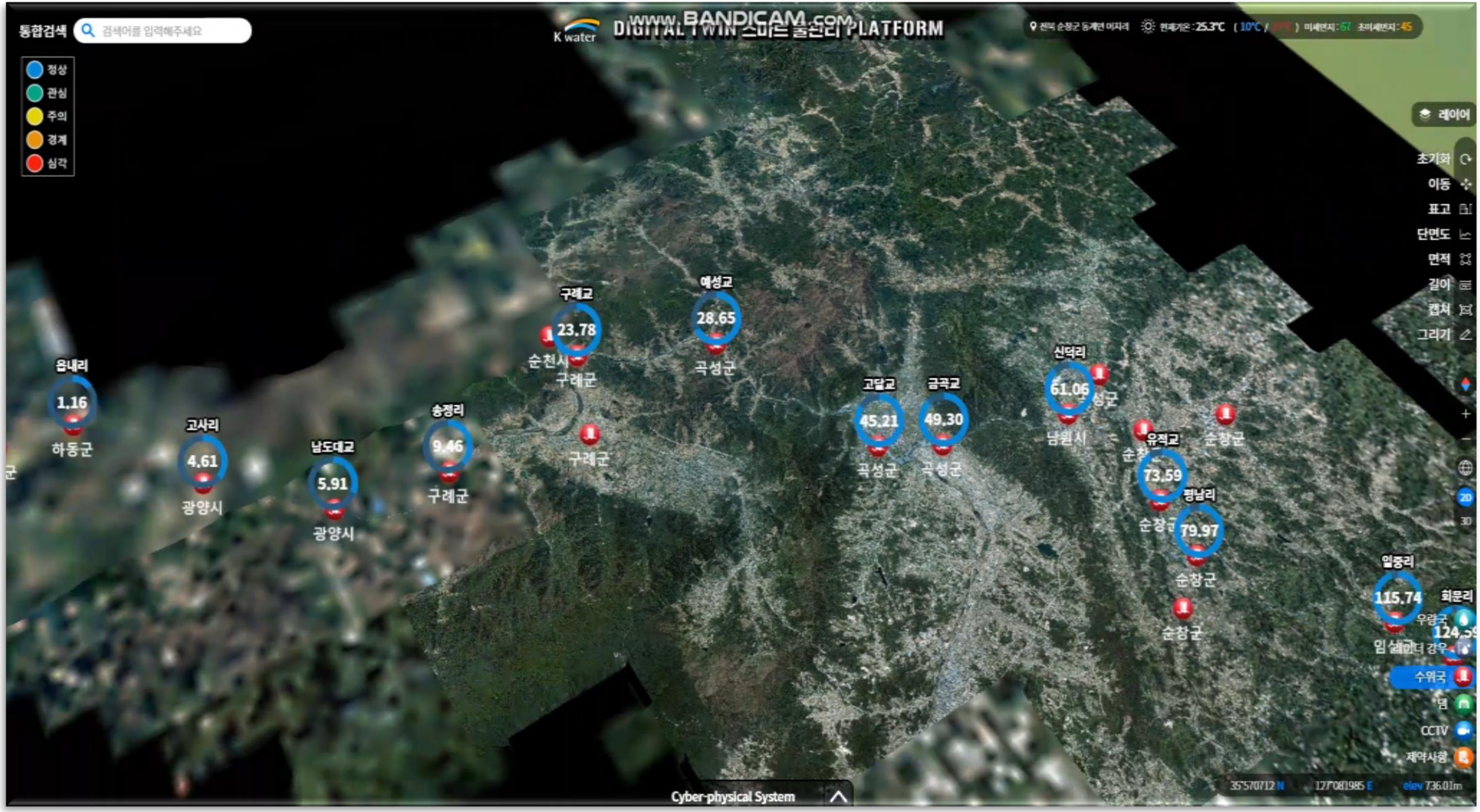
(Part-I b): Challenges : surrogate model



(Part-I b): Integrated Flood Simulation

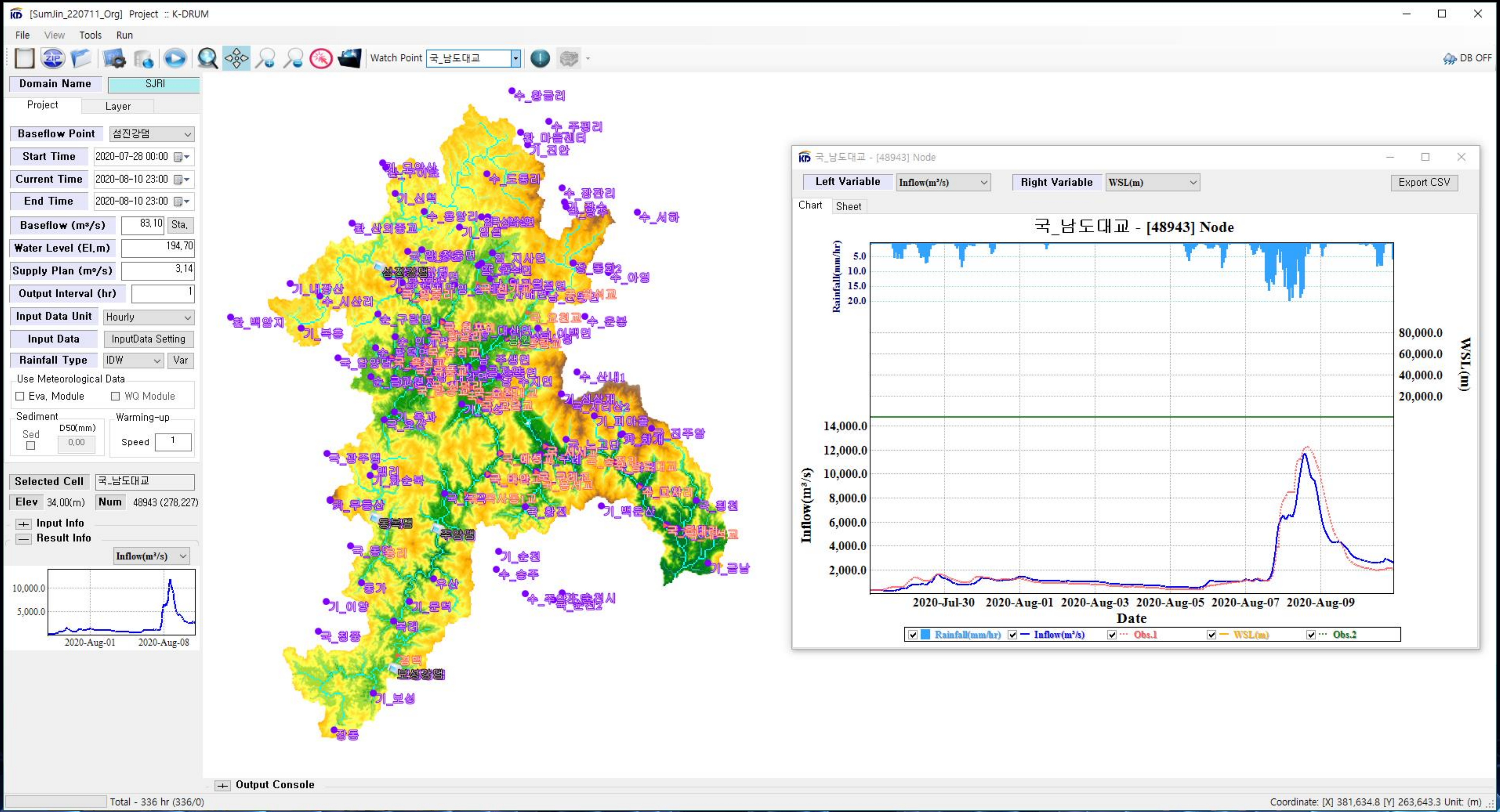
Target — Support Integrated Water Resources Management using the customized flood prediction system based on the developed SW

- Based on integrating the developed SW and systemization, support IWRM of each watershed



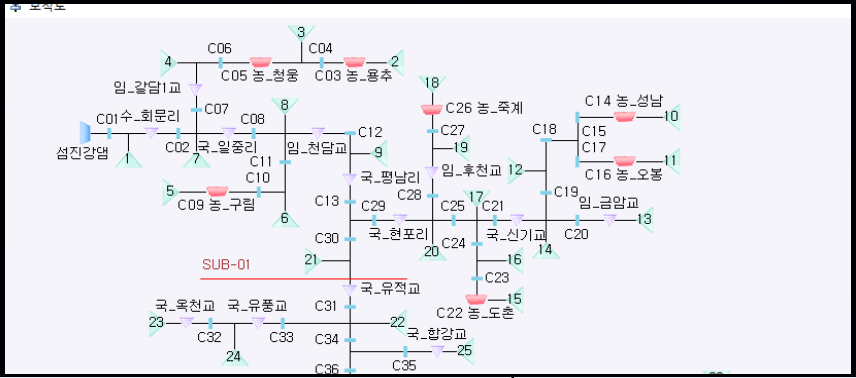
통합검색
검색어를 입력해주세요
www.BANDICAM.com
K water
DIGITAL TWIN 스마트 물관리 PLATFORM
정상
관심
주의
경계
심각
레이어
초기화
이동
표고
단면도
면적
길이
캡처
그리기
구례교
23.78
순천시
구례군
예성교
28.65
곡성군
읍내리
1.16
하동군
고사리
4.61
광양시
남도대교
5.91
광양시
송정리
9.46
구례군
구례군
고달교
45.21
곡성군
금곡교
49.30
곡성군
신덕리
61.06
남원시
순창군
유적교
73.59
순창군
평남리
79.97
순창군
순창군
일중리
115.74
회문리
수위국
댐
CCTV
제약사항
Cyber-physical System

Application of Distributed R-R Model

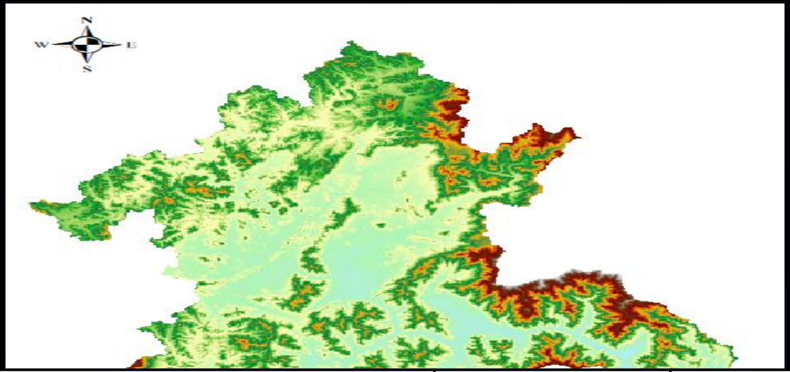


Application of Distributed R-R Model

Lumped (64) : 49km²



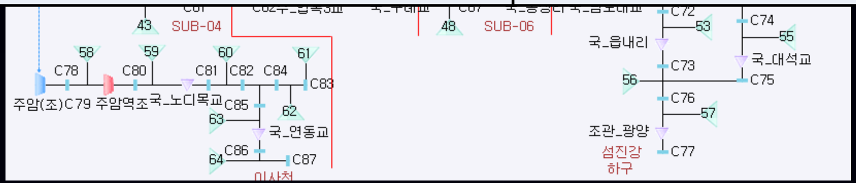
Distributed (50,250) : 0.0625km²

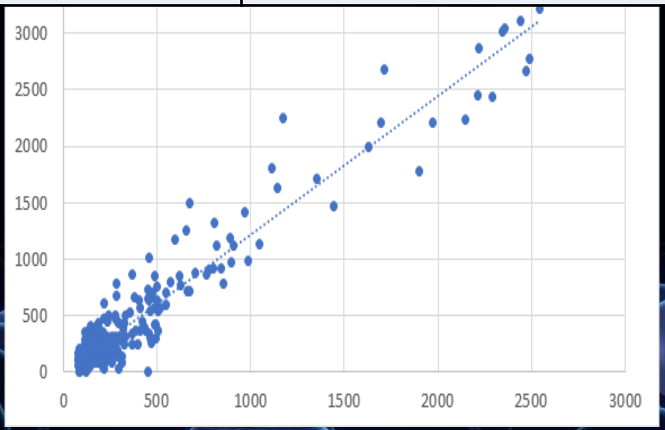


64 tributaries

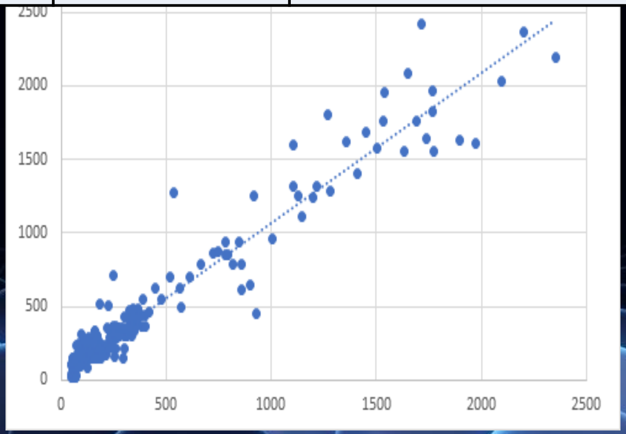
vs.

	Seojin DAM Inflow forecast	Juam DAM inflow forecast	Average	Period
NSE	0.8894	0.9290	0.9092	Flood Season (July and August in 2021)
Determination Coefficient(R^2)	0.9353	0.9426	0.9390	





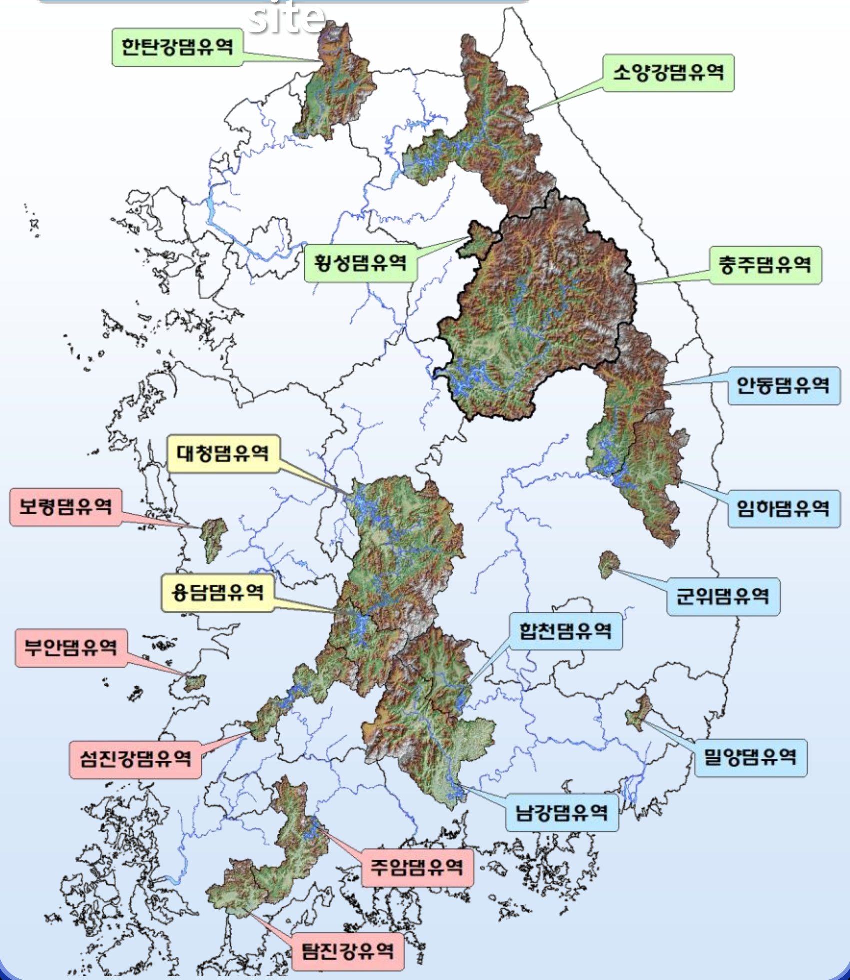
SJ DAM Inflow



JA DAM Inflow

Application of Distributed R-R Model

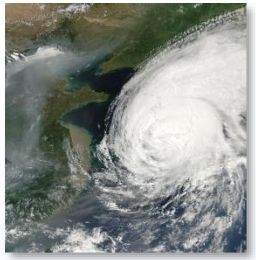
Application Dam site



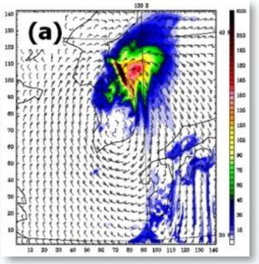
Dam site	Area	Grid size	Grid count	EFF	R^2
Cheungju	6,648㎢	500m	26,765	0.912	0.962
Soyang	2,703㎢	500m	11,099	0.859	0.936
Hoengseong	209㎢	100m	20,738	0.891	0.927
Hantan gang	1,279㎢	300m	14,201	0.921	0.972
Andong	1,584㎢	300m	18,085	0.879	0.953
Imha	1,361㎢	300m	15,144	0.890	0.962
Hapcheon	925㎢	200m	23,225	0.891	0.937
Namgang	2,285㎢	500m	9,163	0.953	0.984
Milyang	95.4㎢	100m	10,342	0.858	0.954
Gunwi	87.5㎢	60m	24,305	0.503	0.801
Daecheong	3,204㎢	500m	13,180	0.852	0.902
Youngdam	930㎢	250m	14,876	0.901	0.933
Seomjingang	763㎢	250m	12,208	0.803	0.875
Juam	1,010㎢	250m	16,160	0.815	0.893
Buan	59㎢	100m	5,900	-	-
Boryeong	163.6㎢	100m	16,360	0.793	0.845
Jangheung	193㎢	100m	19,300	0.831	0.882

Application of Distributed R-R Model

Typhoon RUSA



<Satellite image>



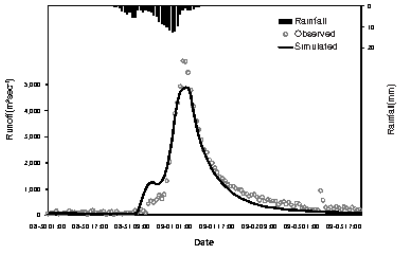
<Precipitation>

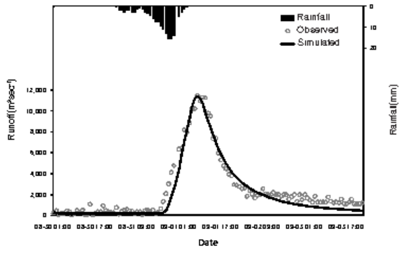
<Typhoon damage>

- 4 Grade typhoon
- Central Pressure : 950hPa
- Max. Wind speed : 40m/s
- Max. Precipitation : 870.5mm
- Period : 2002. 8. 23-2002. 9. 1

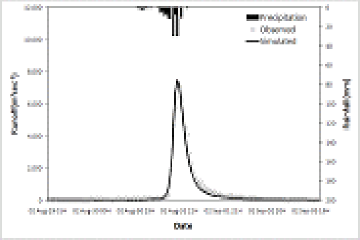
Simulation results

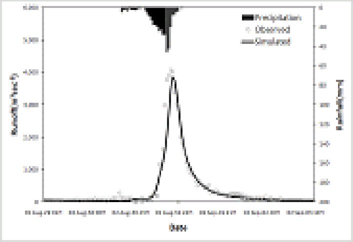
Han River

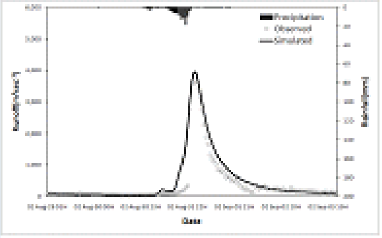




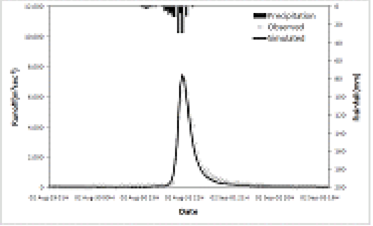
Nakdong

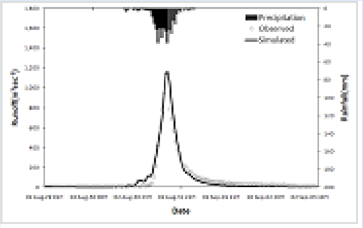




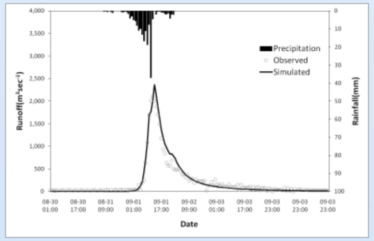


Gum River

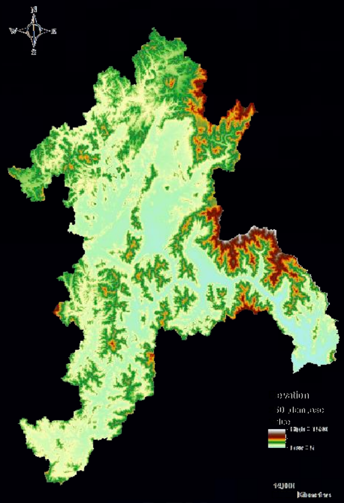


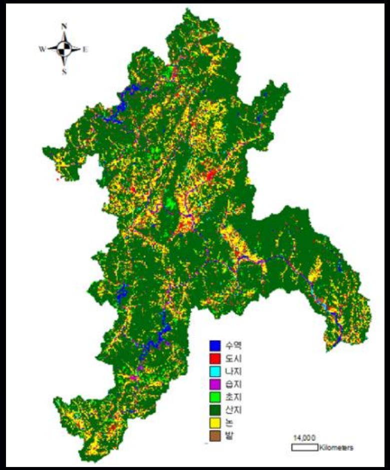


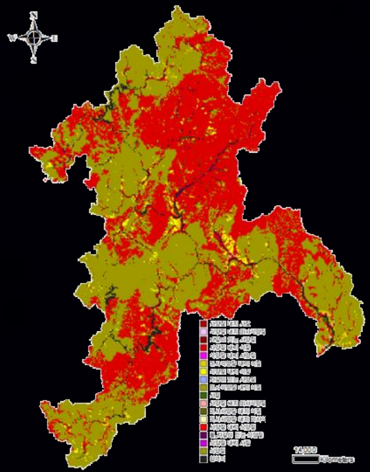
Other



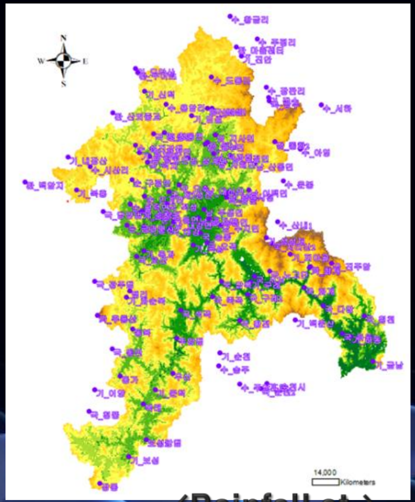
Application of Distributed R-R Model


<DEM>


<land use>

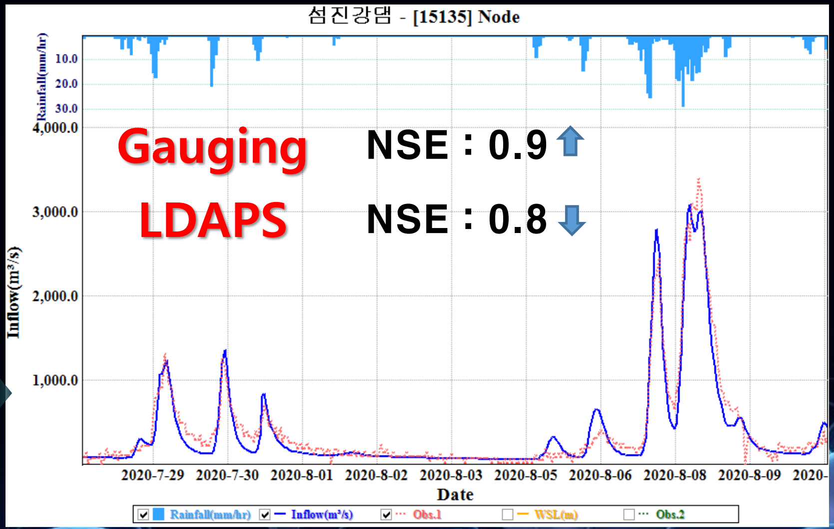

<Soil map>


<Soil depth>


<Rainfall st.>

- Rainfall : 88(144)

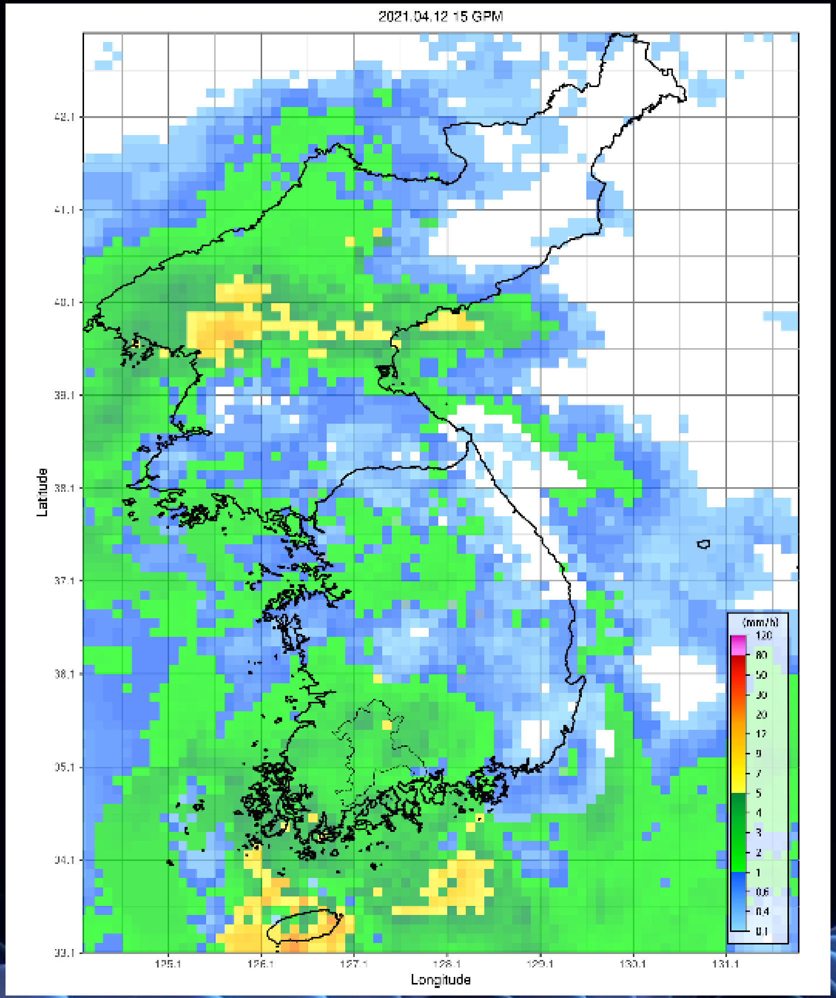
Seomjin dam



Application of Distributed R-R Model

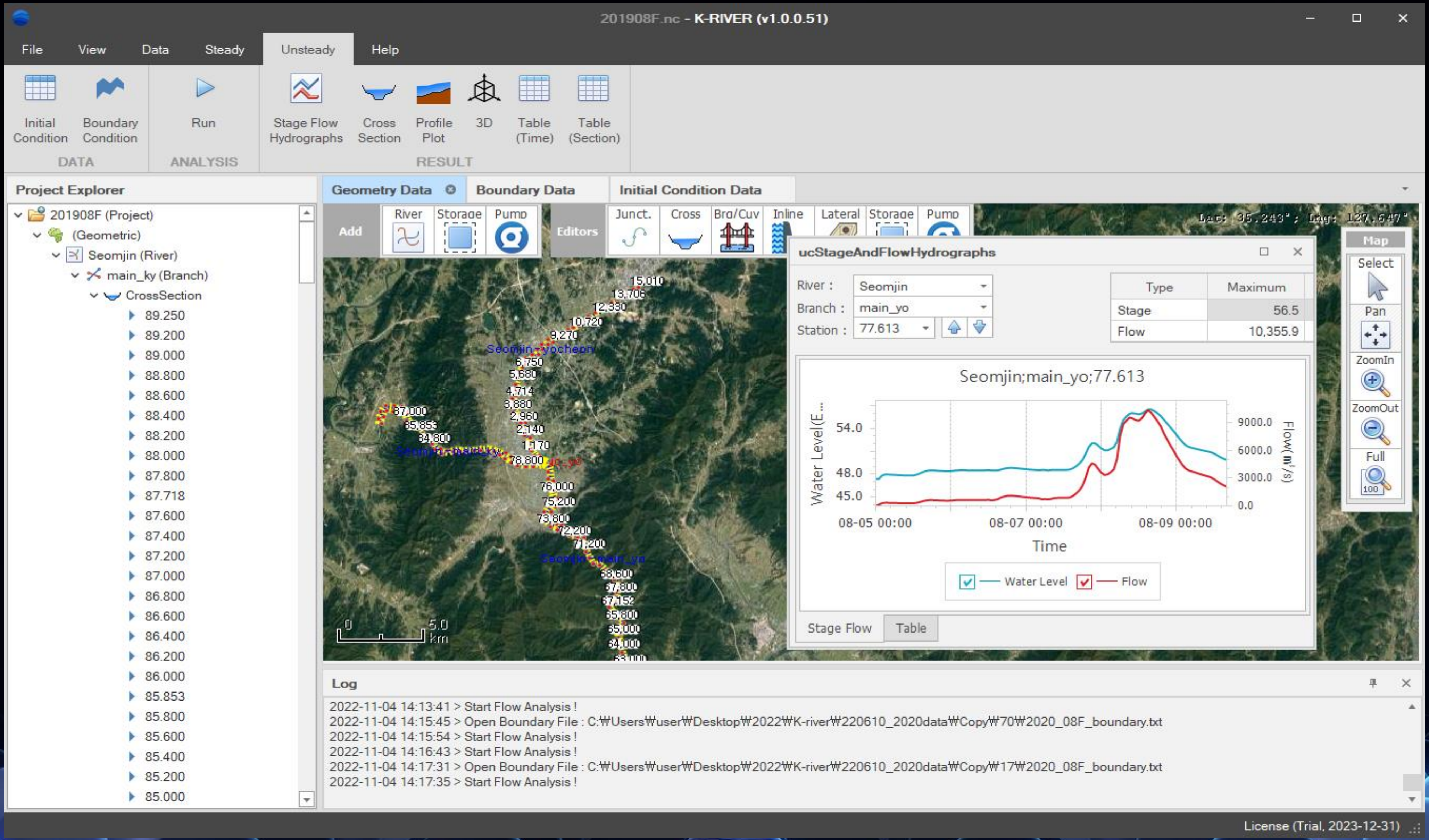
K water

LDAPS (Local Data Assimilation and Prediction System, KMA Rainfall prediction)



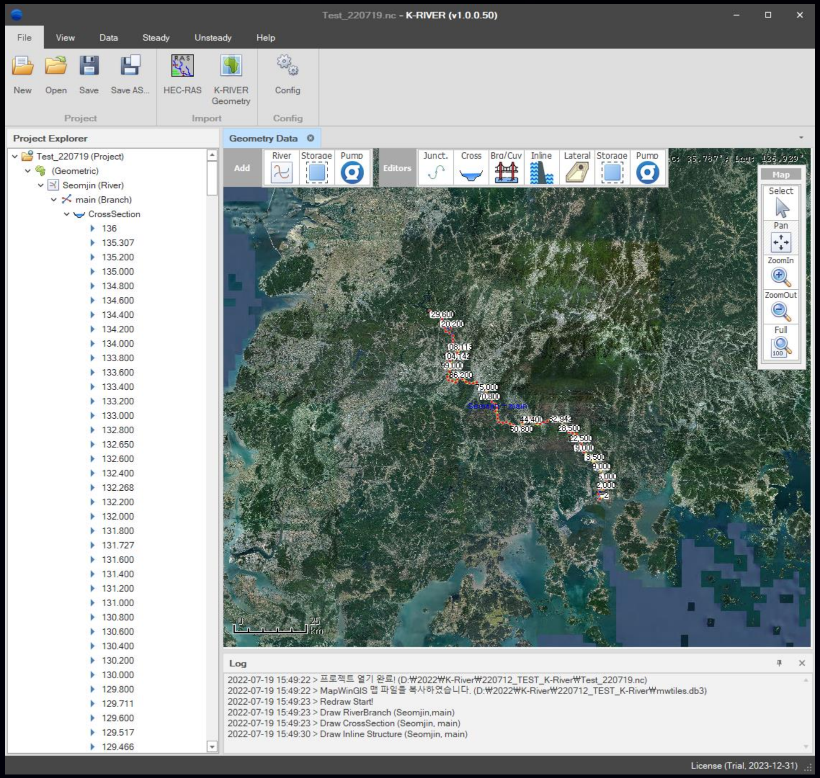
- Grid size : 1.5 km
- 48 hr prediction
- 1hr interval
- Prediction : 4 times / day
- Base model : UM

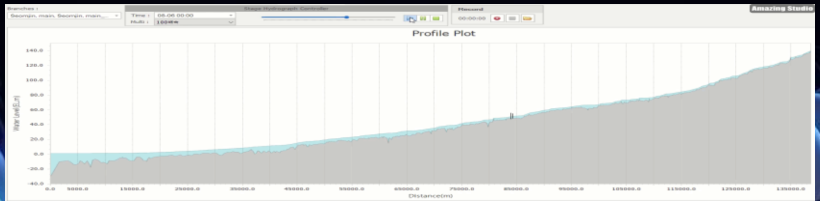
Linking to River and Flood Models



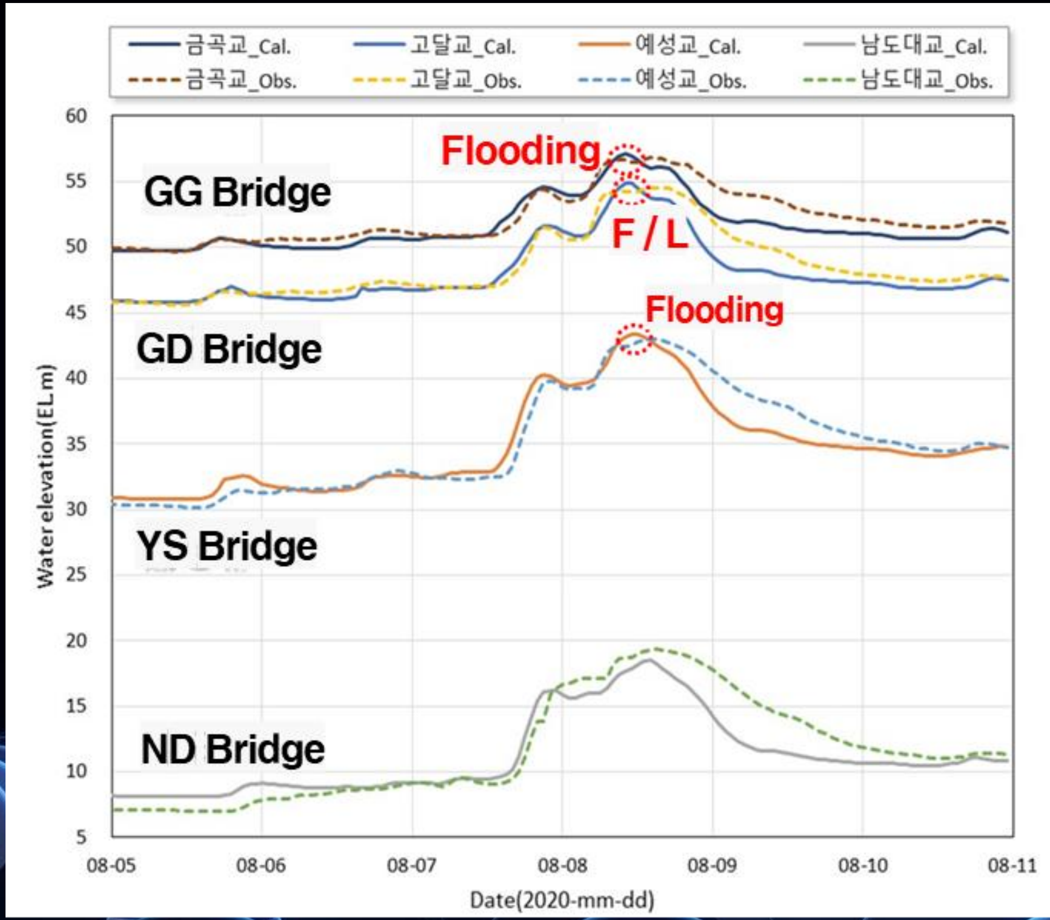
Linking to River and Flood Models

K-River (Water stage prediction along the river)



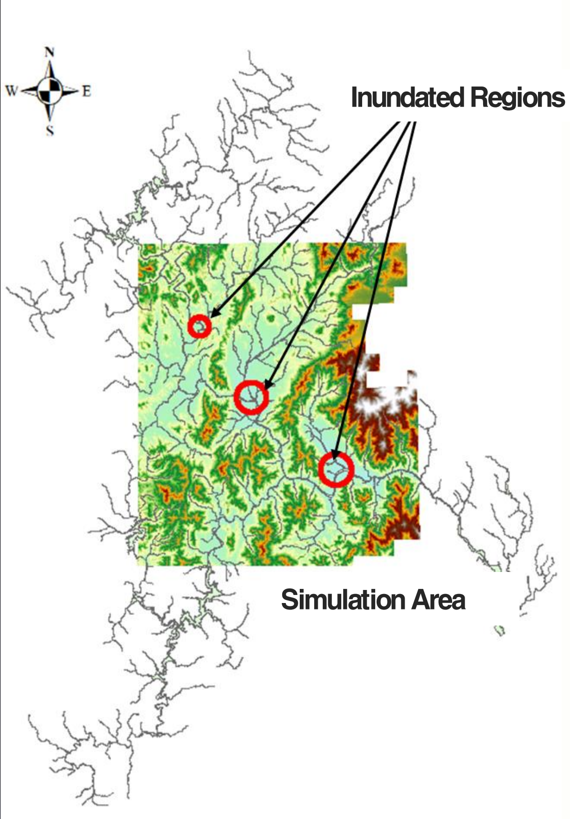


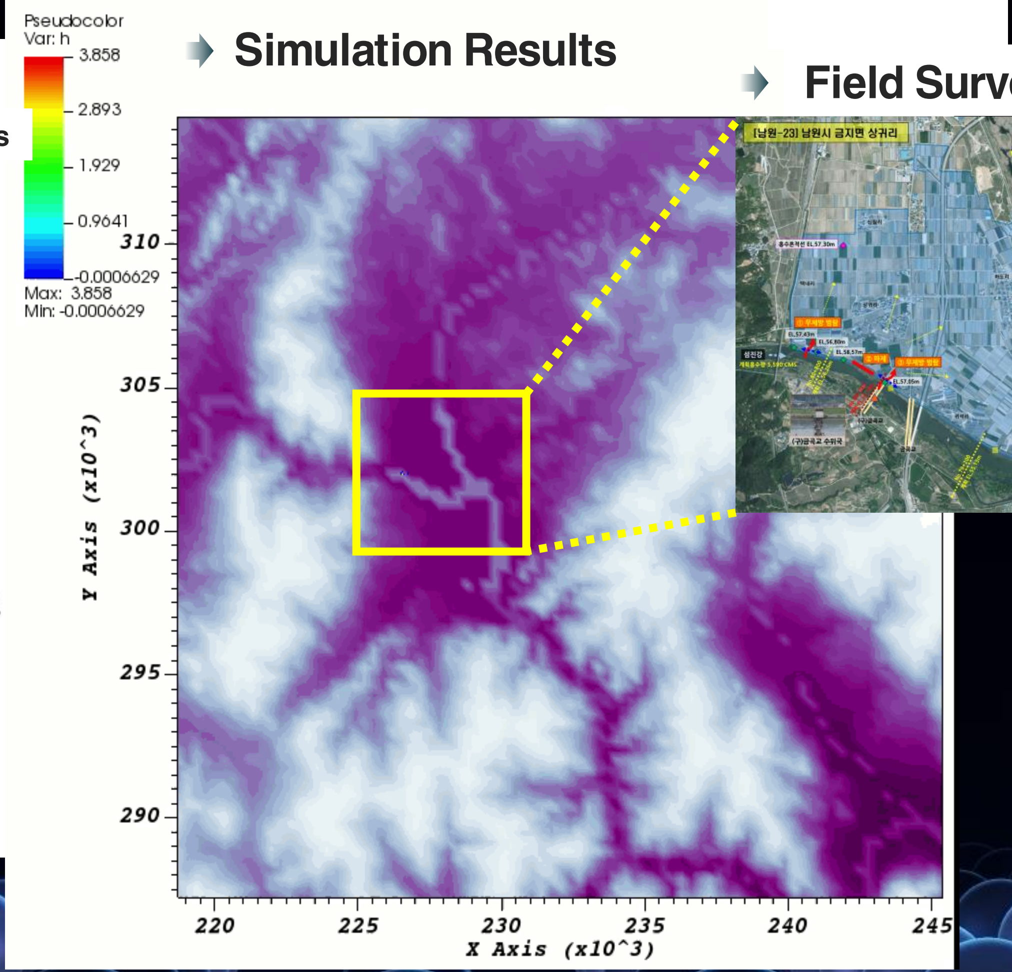
- K-DRUM – K-River linkage : 24 Laterals (Tributary) and Dam Discharge
- Max. Difference : 0.28m to 0.82m

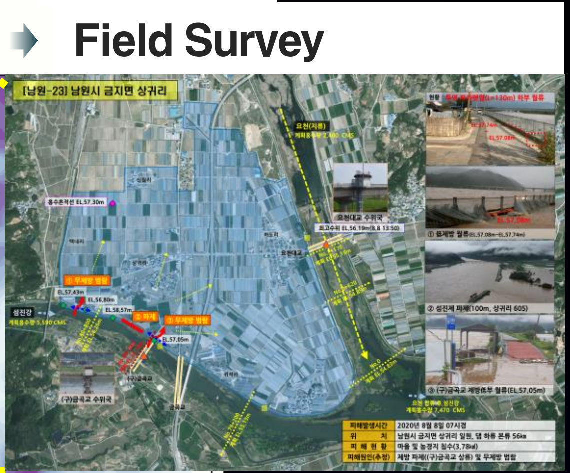


Linking to River and Flood Models

K-Flood (Flood inundation analysis)

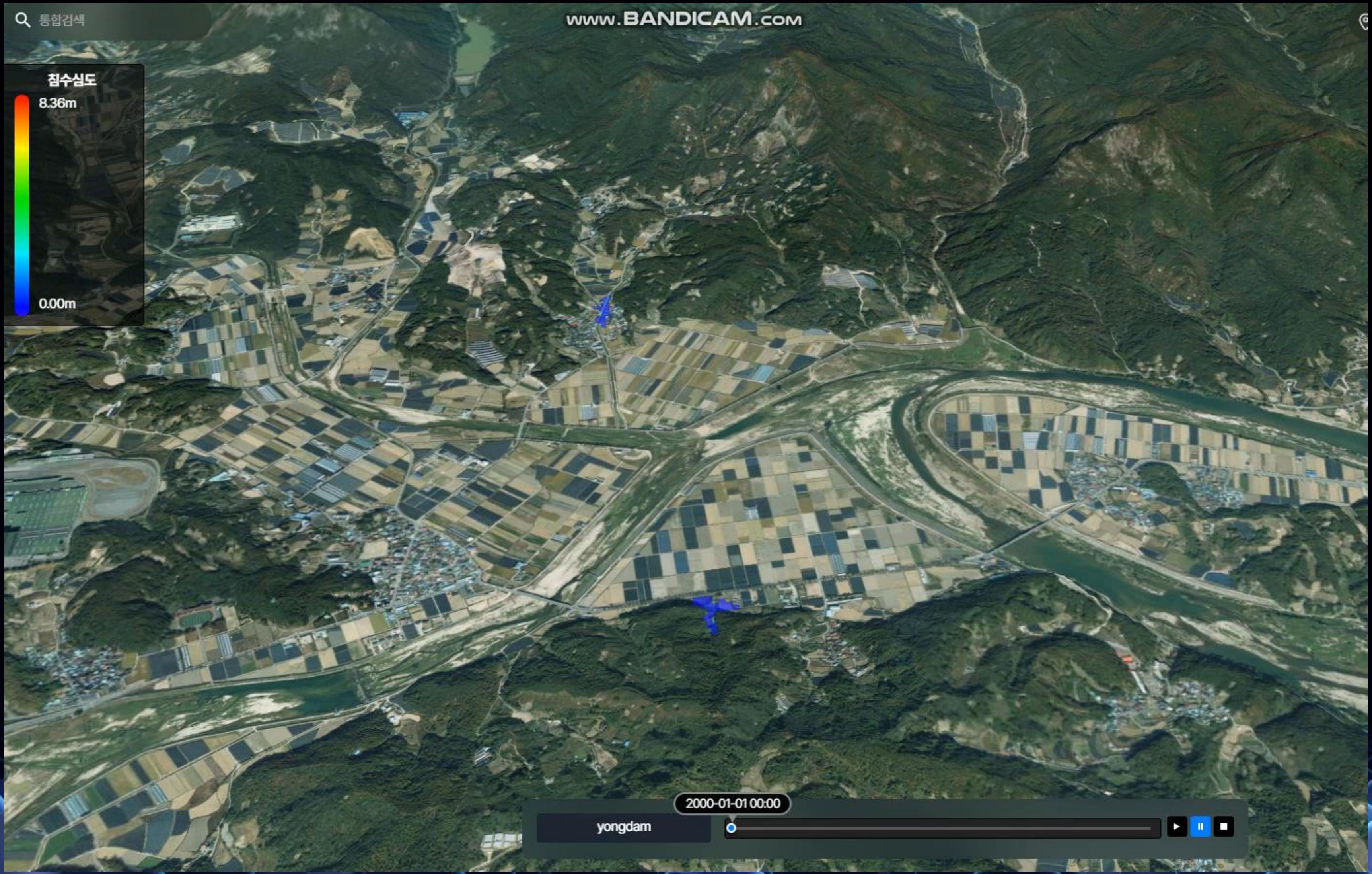






Linking to River and Flood Models

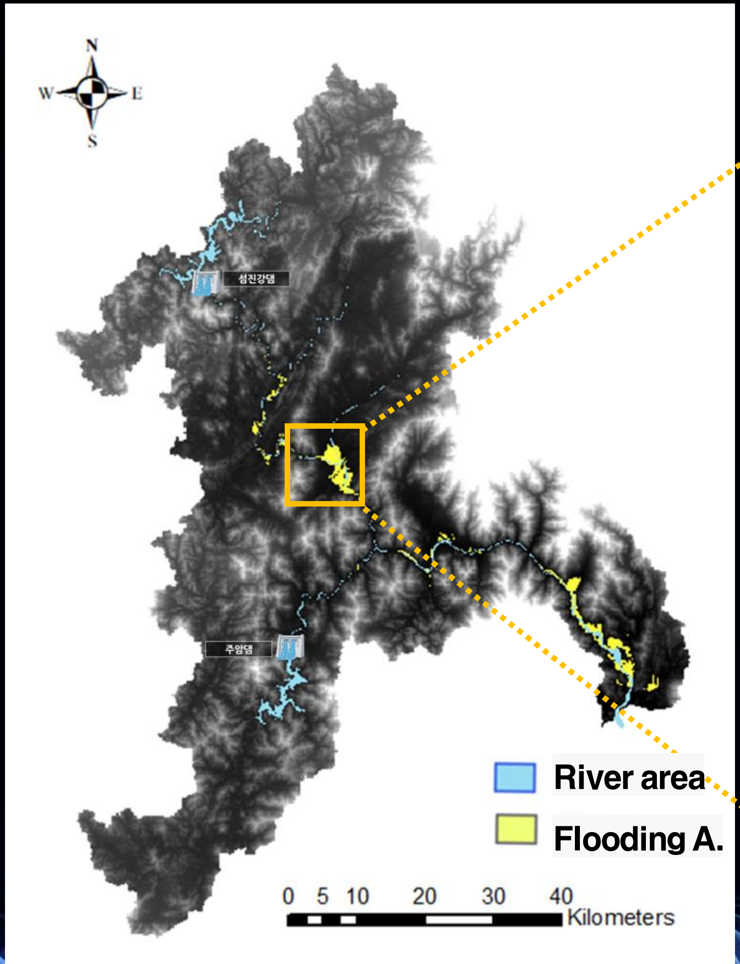
Downstream of Yongdam Dam With K-Flood

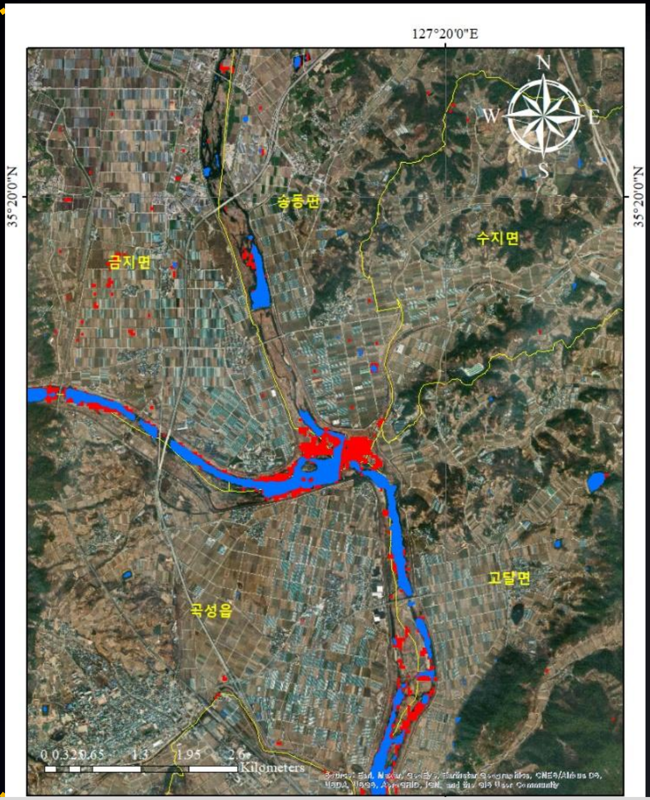


Linking to River and Flood Models

Downstream of Yongdam Dam With K-Flood

- Satellite Image at flooding in **August 2020**



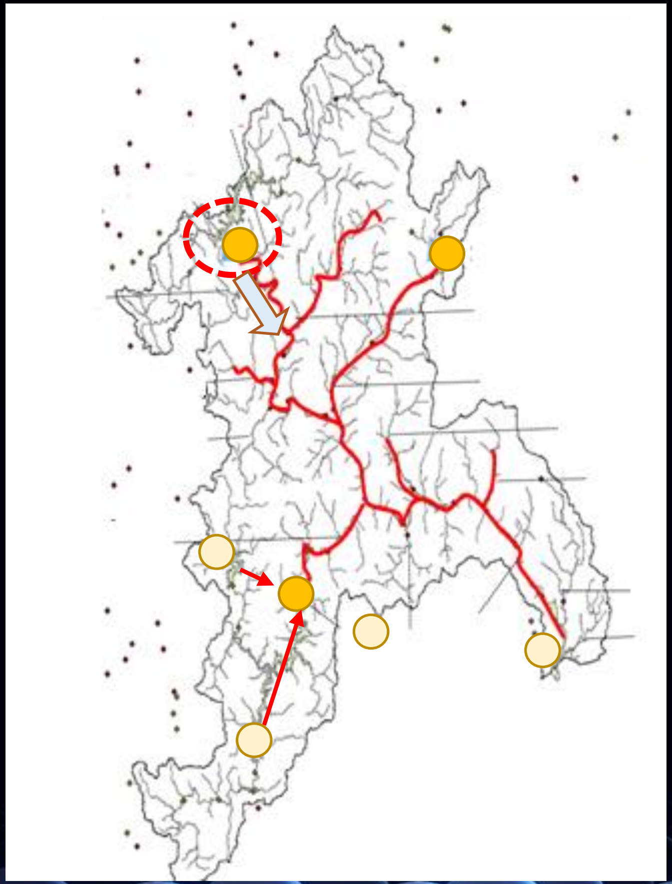


(Before) 8.2. 06:32 KST (Blue), Water Area : 4.8 ㎢
(After) 8.7. 18:31 KST (Red), Water Area: 7.9 ㎢ (+64.2%)

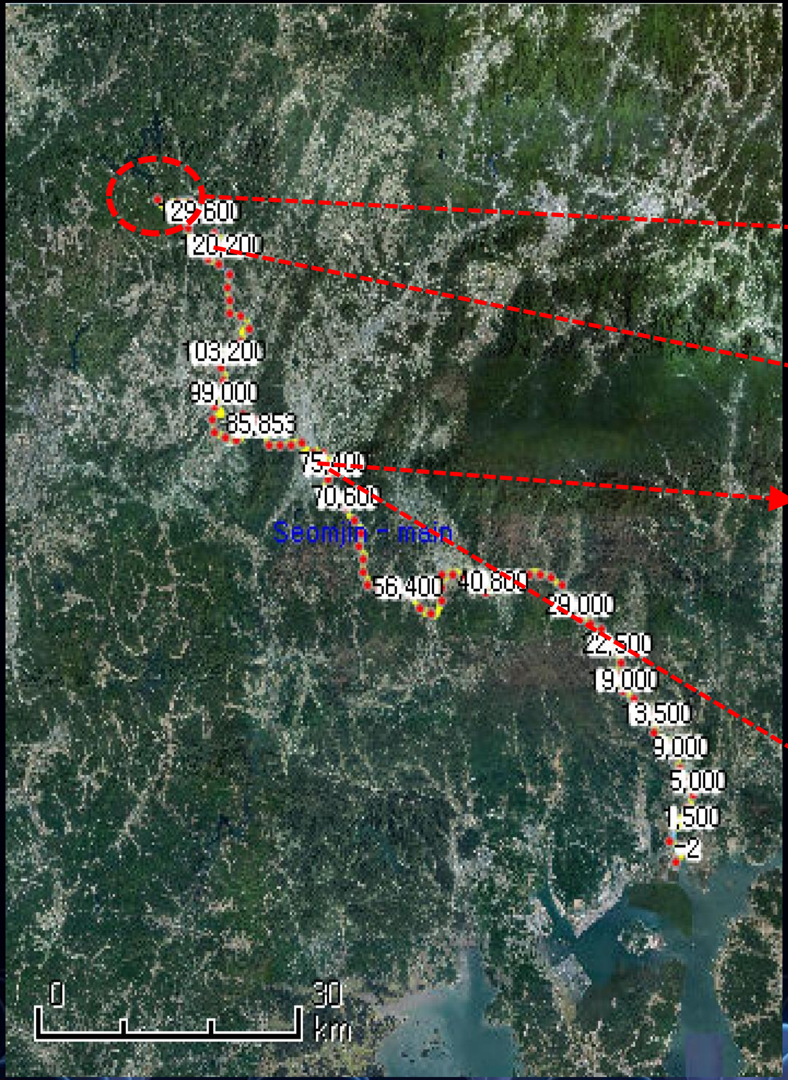
Performance of Surrogate River Model

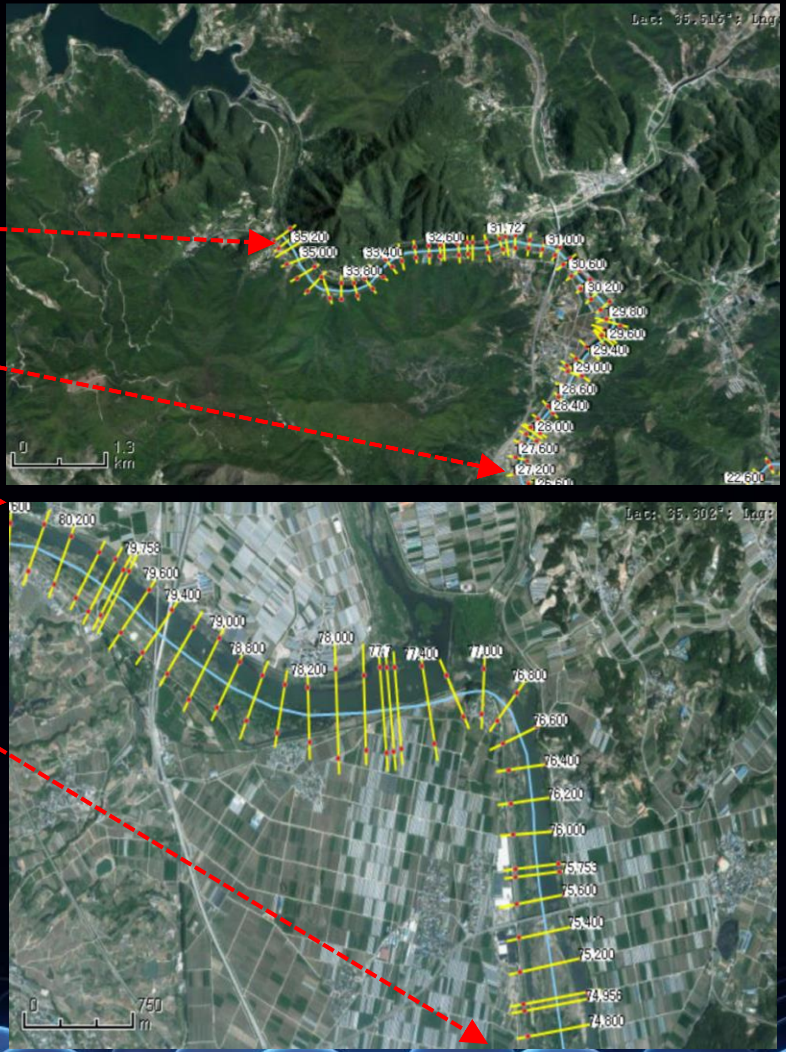
Performance of Surrogate River Model

Seomjin River
Watershed Area : 4,911 km^2
Main Channel Length : 173.5 km



Map of Seomjin River



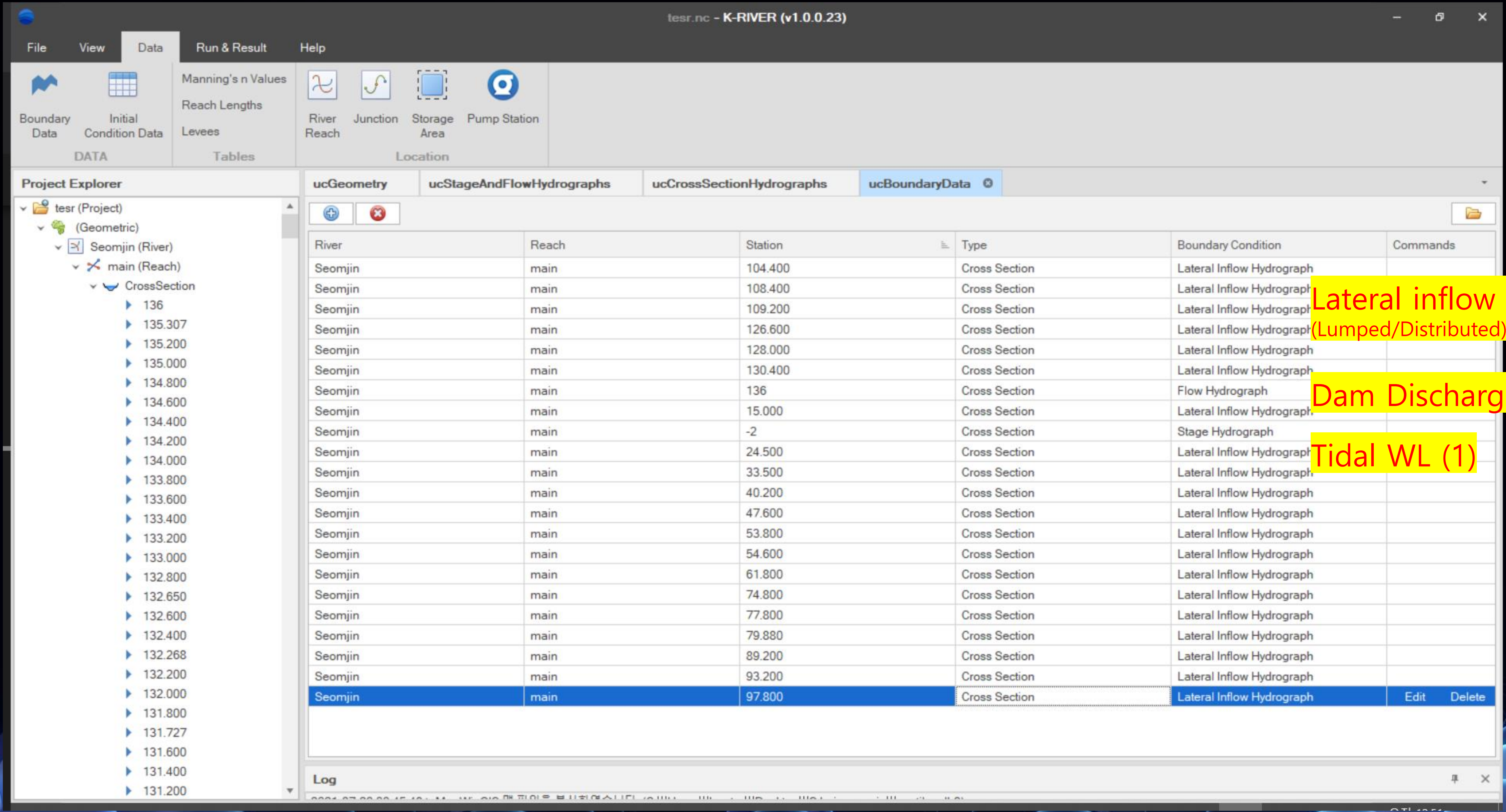


Main River Channel and Cross Sections

Material and Method : Hybrid Model

The 5th China-Japan-South Korea Water Science Research Forum

K water



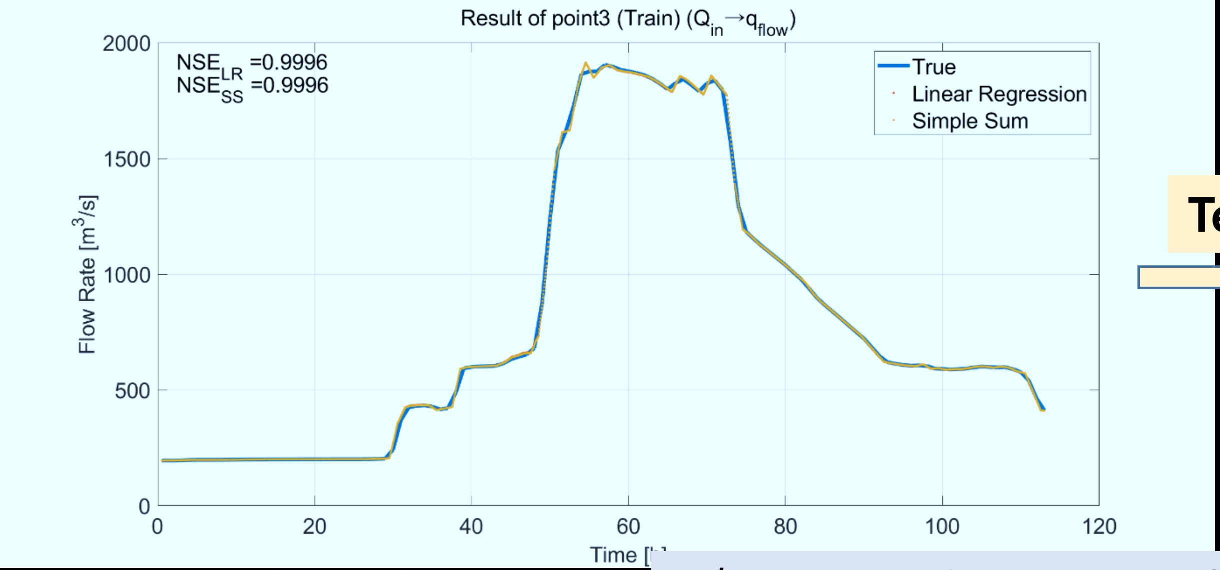
River	Reach	Station	Type	Boundary Condition	Commands
Seomjin	main	104.400	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	108.400	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	109.200	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	126.600	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	128.000	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	130.400	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	136	Cross Section	Flow Hydrograph	
Seomjin	main	15.000	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	-2	Cross Section	Stage Hydrograph	
Seomjin	main	24.500	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	33.500	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	40.200	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	47.600	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	53.800	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	54.600	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	61.800	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	74.800	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	77.800	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	79.880	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	89.200	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	93.200	Cross Section	Lateral Inflow Hydrograph	
Seomjin	main	97.800	Cross Section	Lateral Inflow Hydrograph	Edit Delete

Lateral inflow (20)
(Lumped/Distributed)

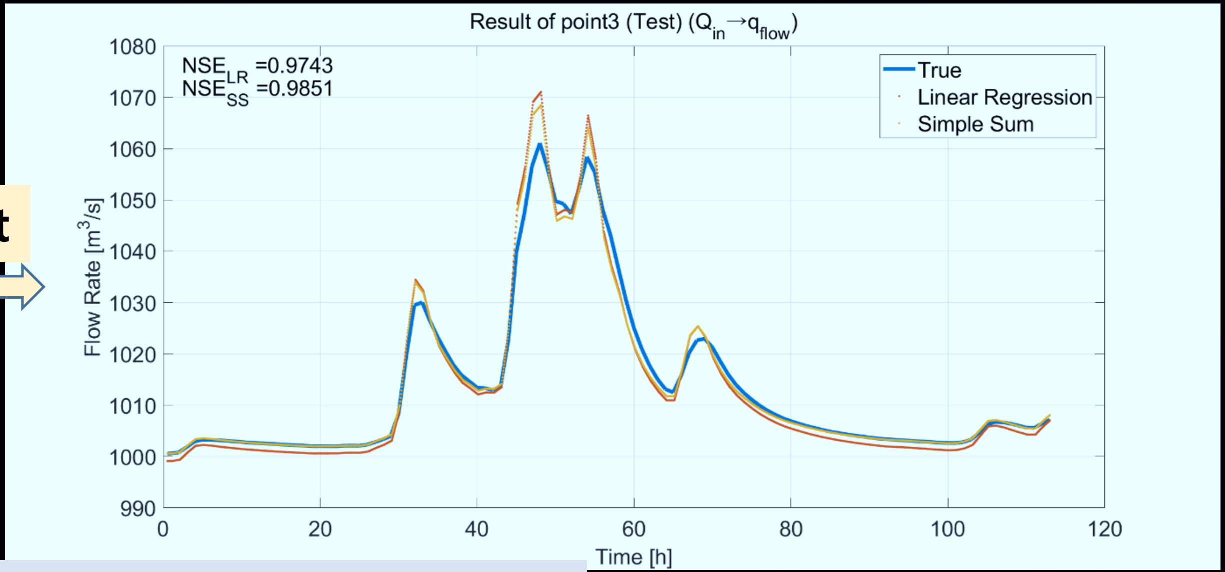
Dam Discharge (1)

Tidal WL (1)

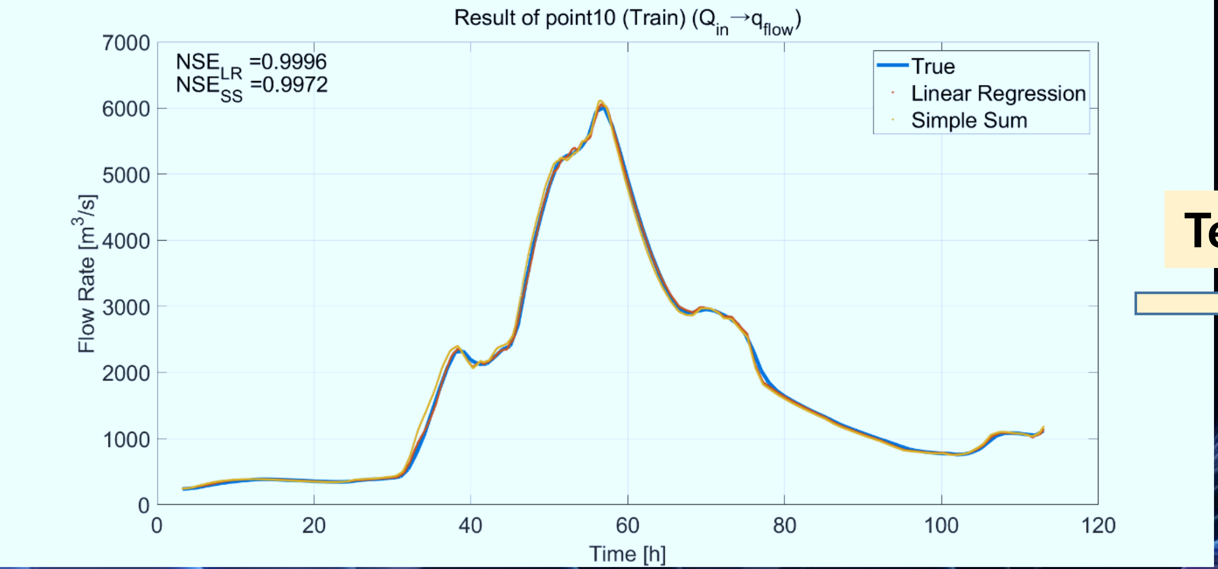
Performance of Surrogate River Model



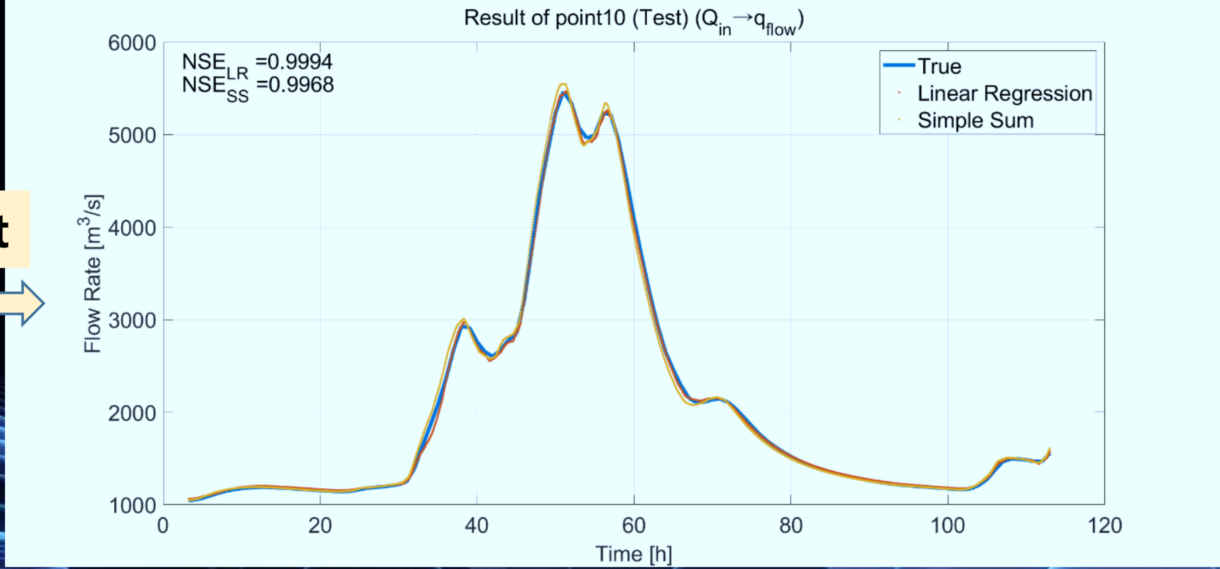
Test



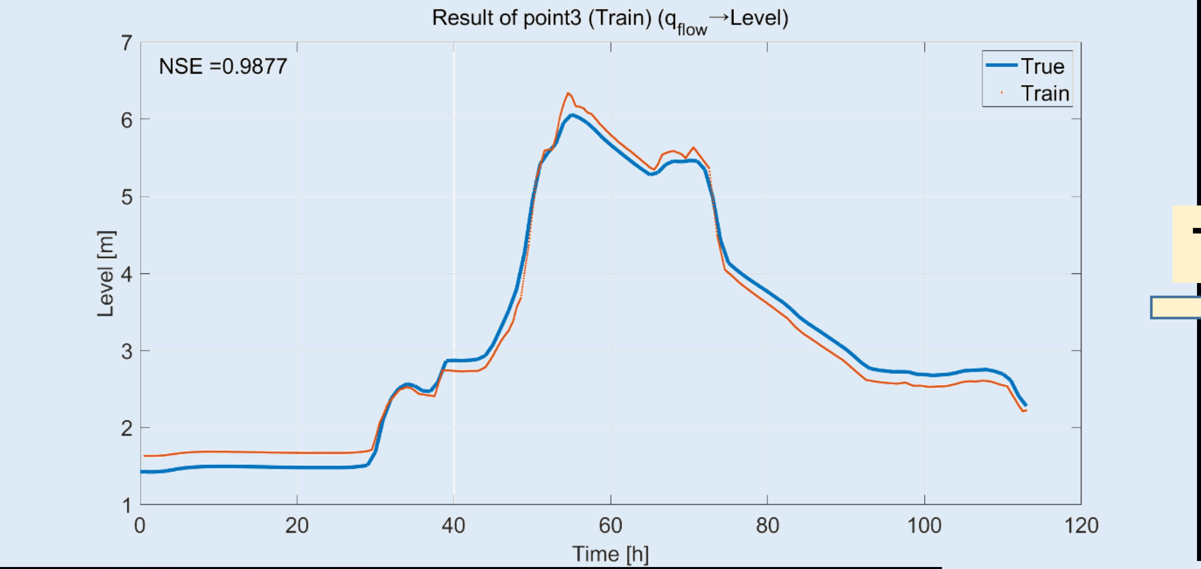
$$q^{i}_{flow} = a_1 Q^{1}_{in}(d_1) + a_2 Q^{2}_{in}(d_2) + \cdots a_{i-1} Q^{i-1}_{in}(d_3) + b_1$$



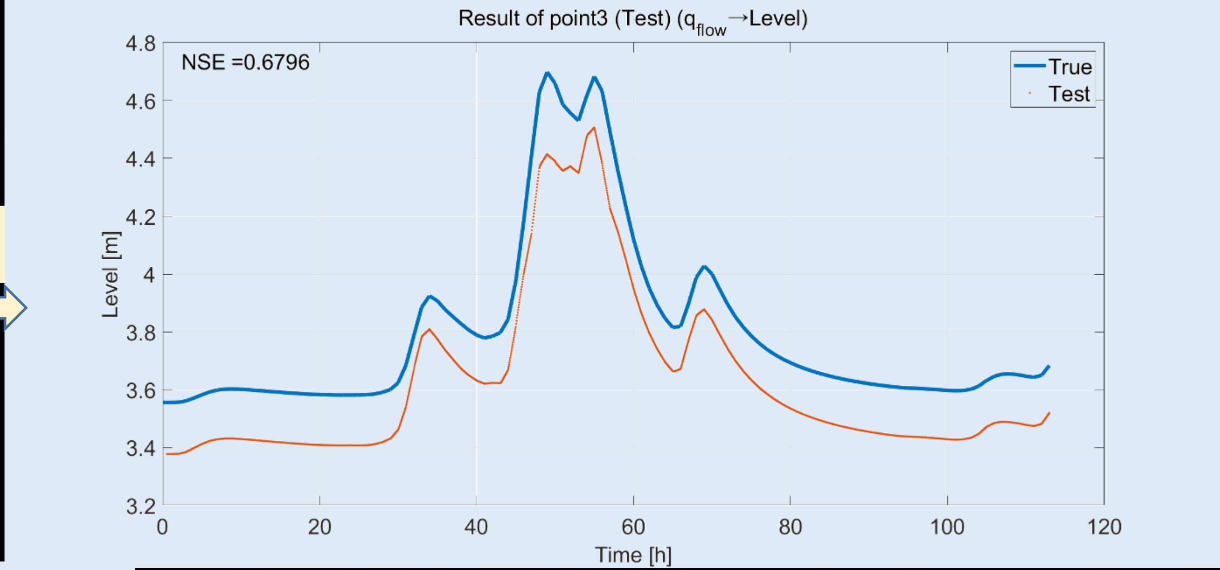
Test



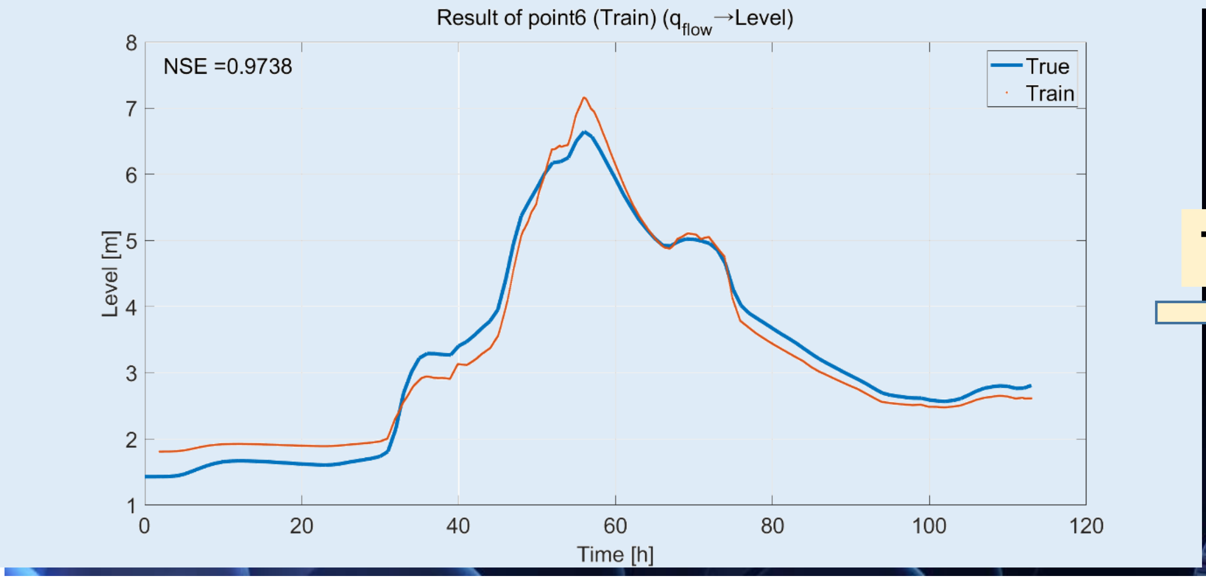
Performance of Surrogate River Model



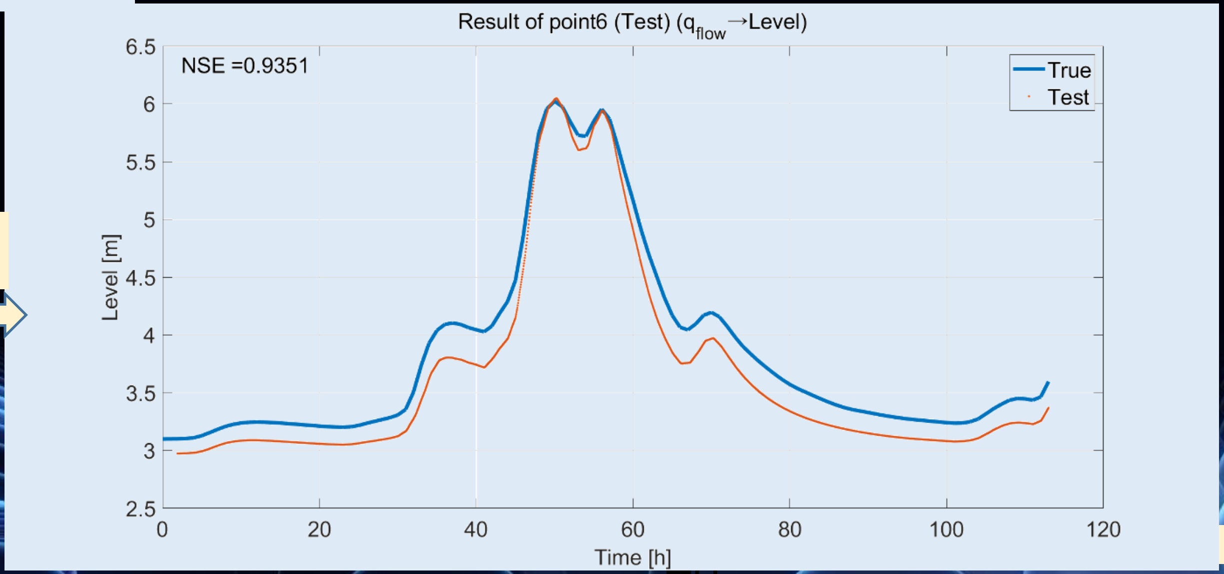
Test



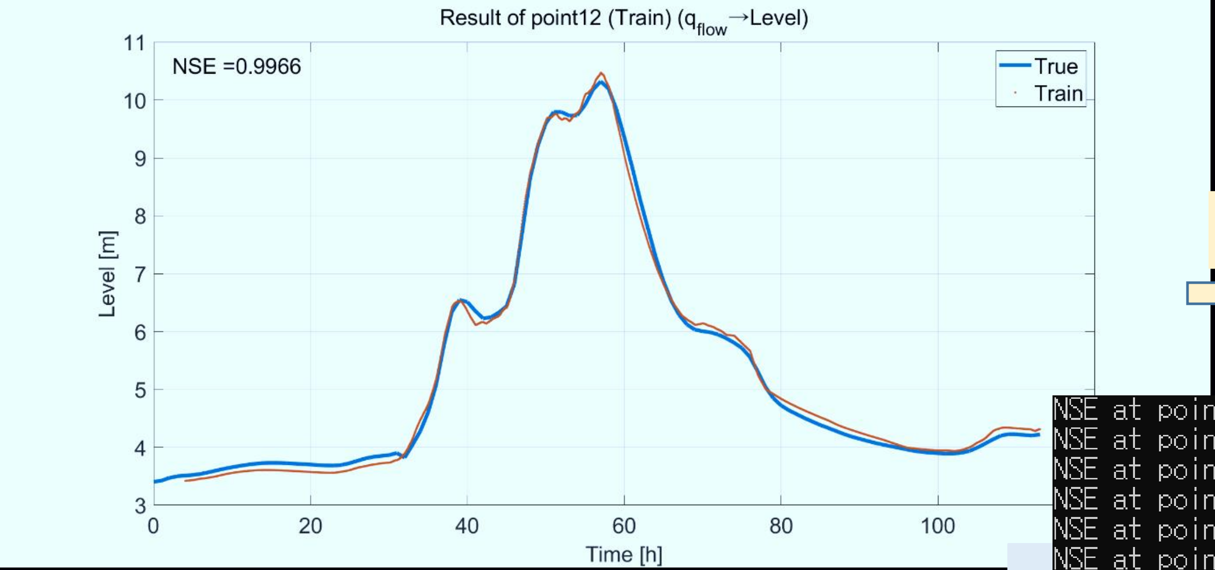
$$h = Aq + B$$



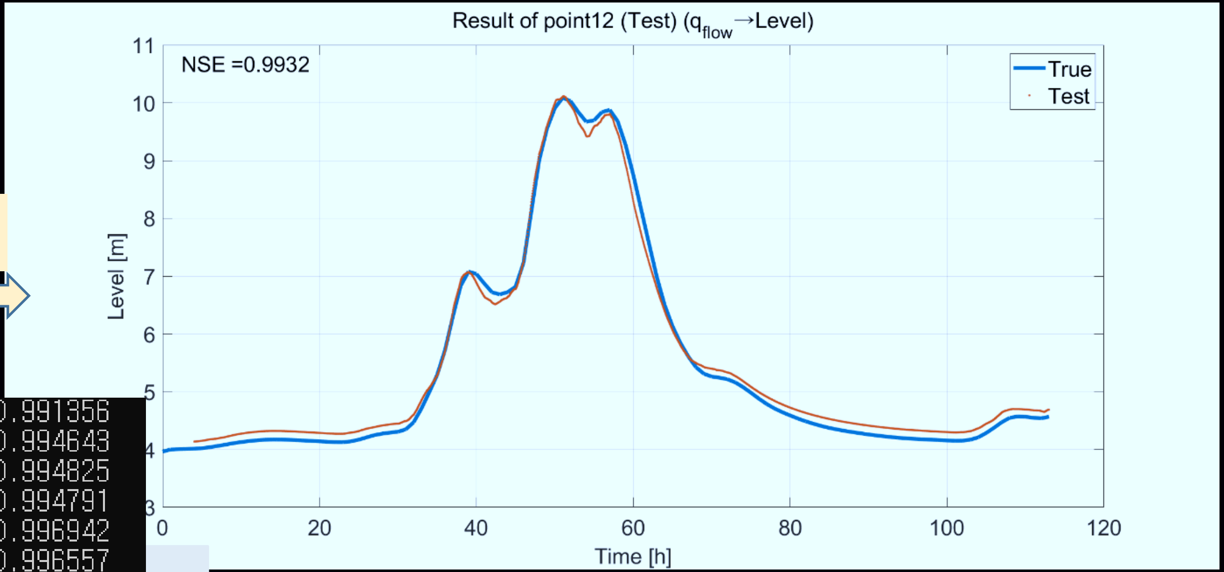
Test



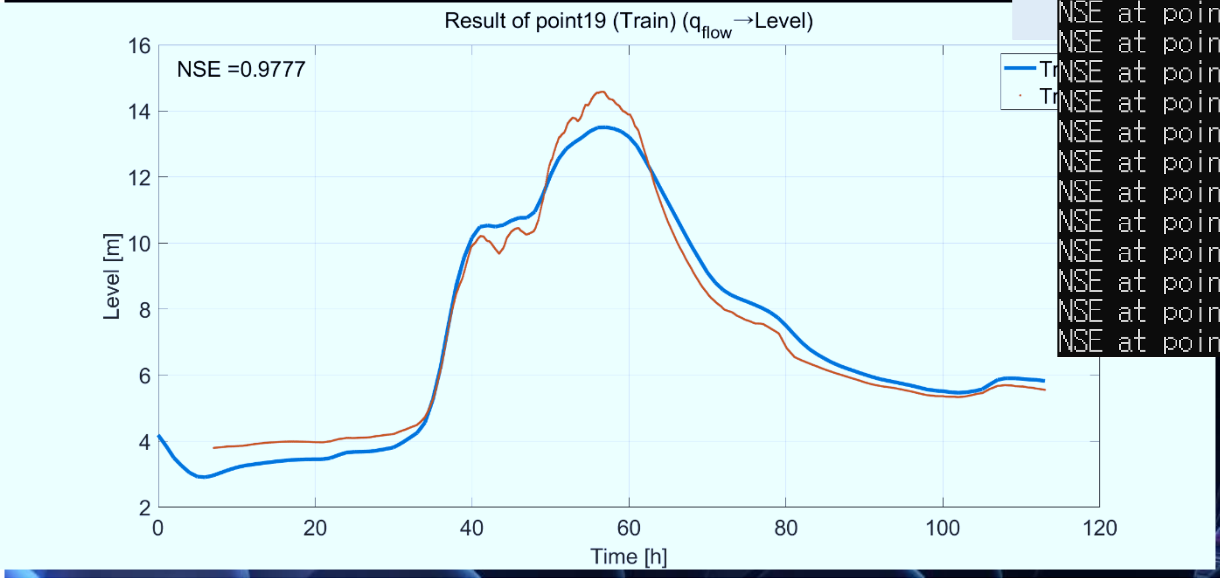
Performance of Surrogate River Model

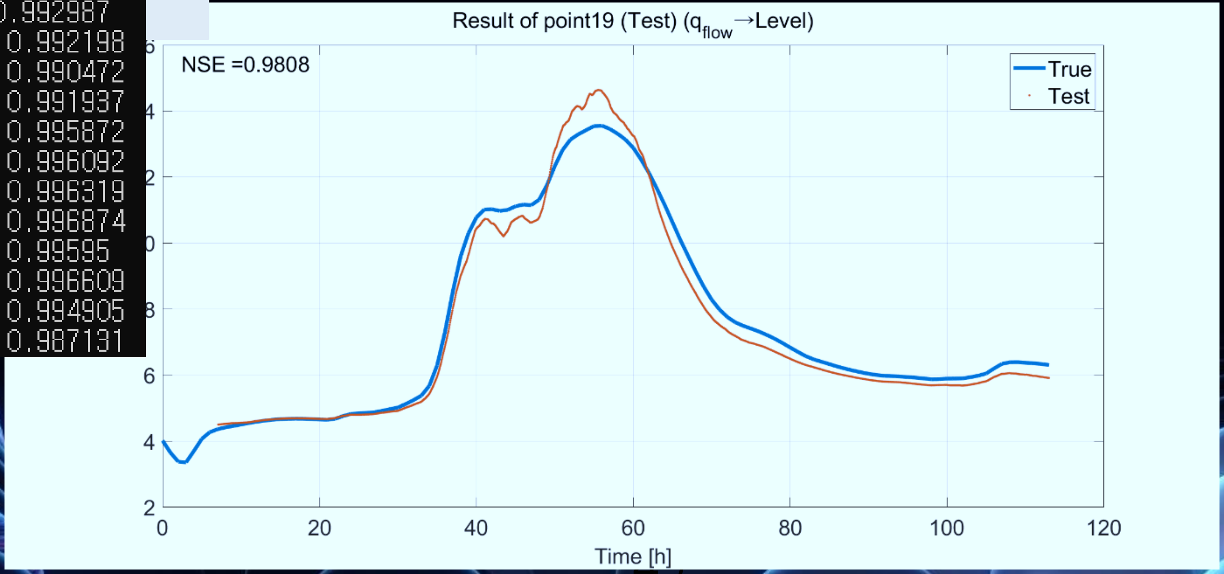


Test



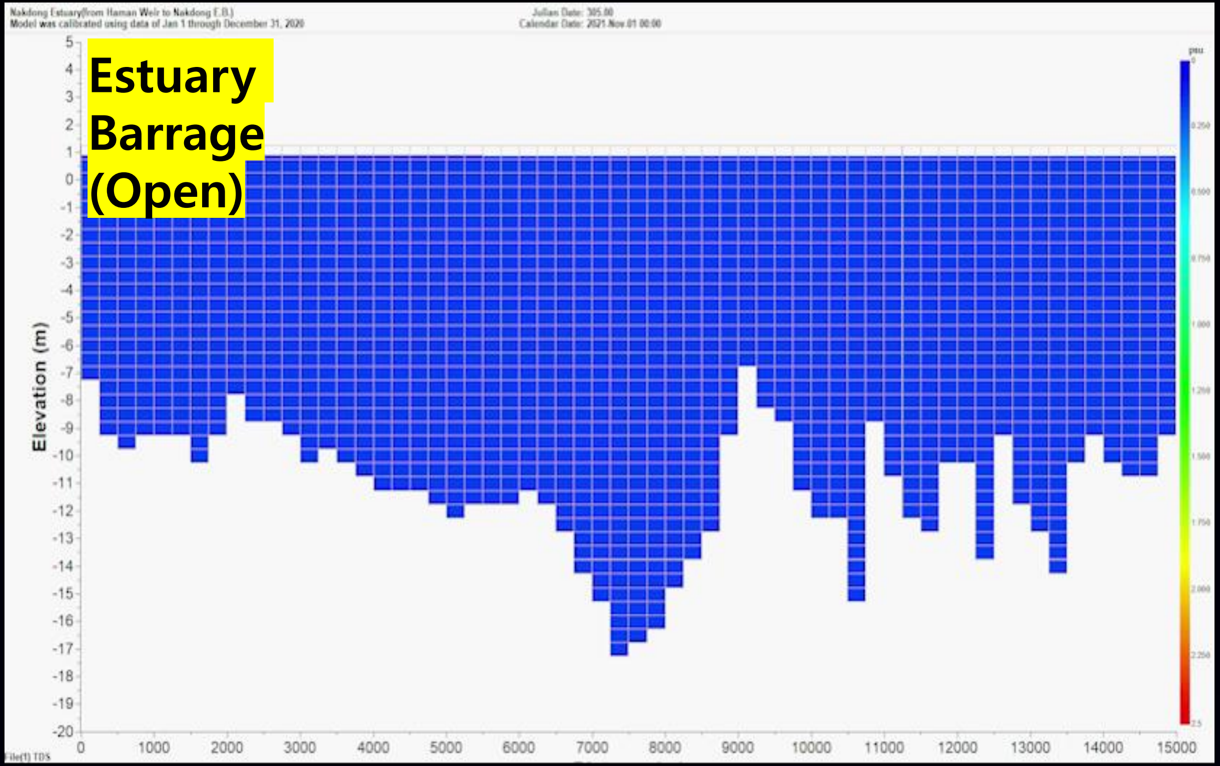
NSE at point 1	is	0.991356
NSE at point 2	is	0.994643
NSE at point 3	is	0.994825
NSE at point 4	is	0.994791
NSE at point 5	is	0.996942
NSE at point 6	is	0.996557
NSE at point 7	is	0.996031
NSE at point 8	is	0.995465
NSE at point 9	is	0.992987
NSE at point 10	is	0.992198
NSE at point 11	is	0.990472
NSE at point 12	is	0.991937
NSE at point 13	is	0.995872
NSE at point 14	is	0.996092
NSE at point 15	is	0.996319
NSE at point 16	is	0.996874
NSE at point 17	is	0.99595
NSE at point 18	is	0.996609
NSE at point 19	is	0.994905
NSE at point 20	is	0.987131



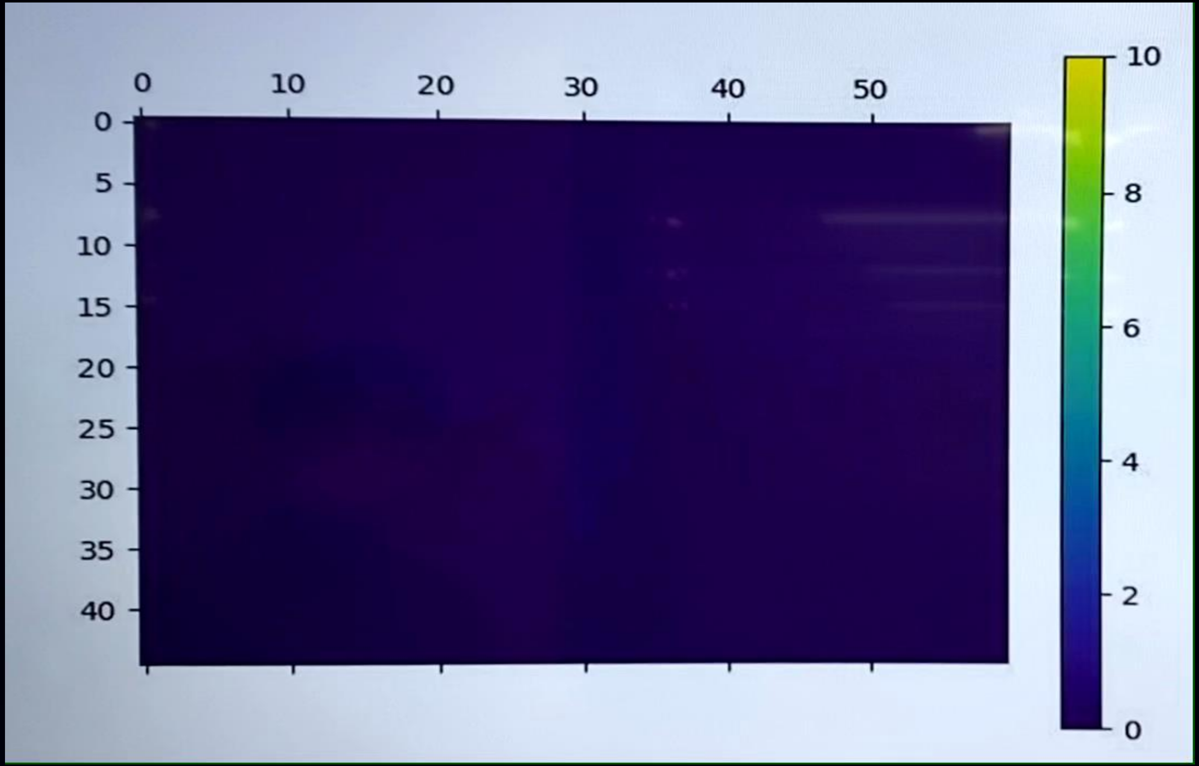


Performance of Surrogate River Model (Example)

CE-QUAL-W2 (x-z 2D) Model : Training 90 cases



AI (MLP-Mixer) Model (Test 18 cases)

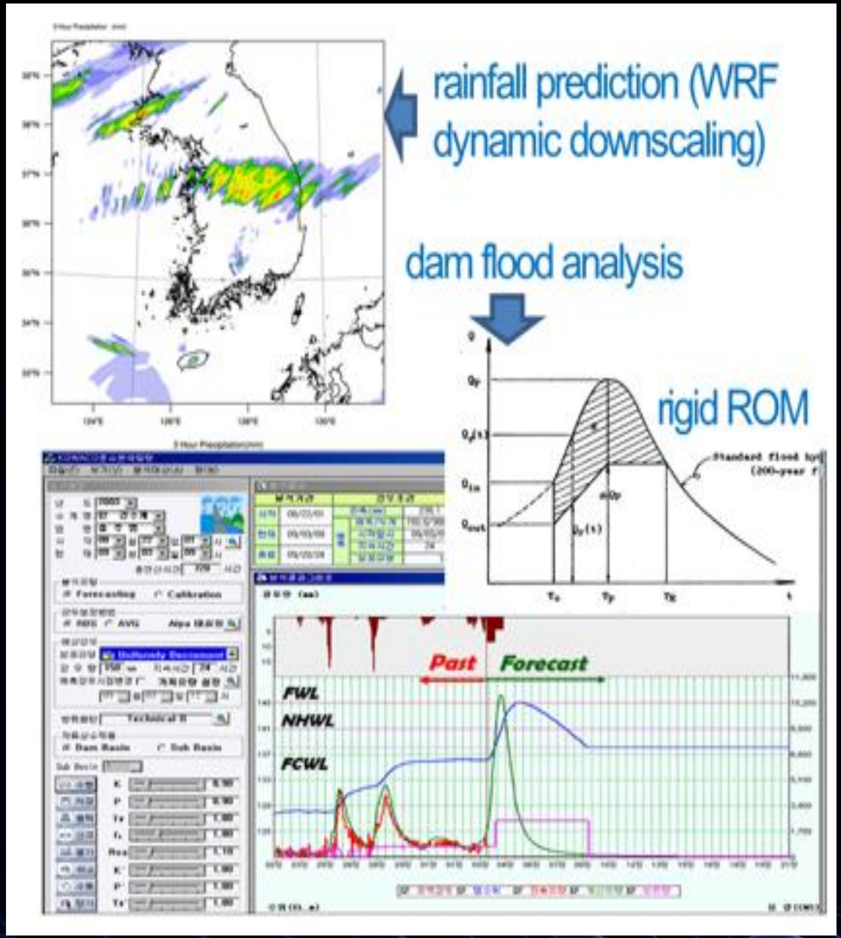


Salt Intrusion

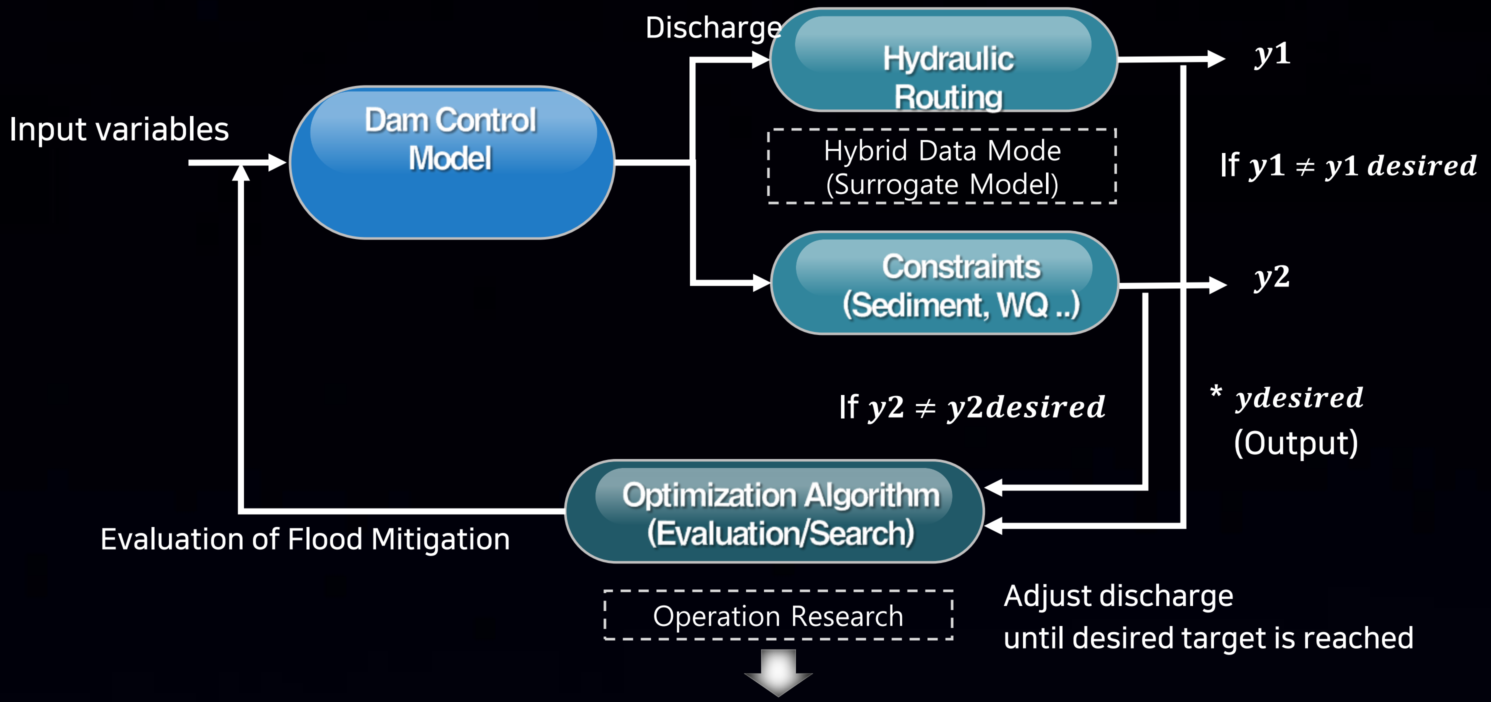
flood

Optimization w/ Hybrid Model

As-Is (in operation)



To-Be

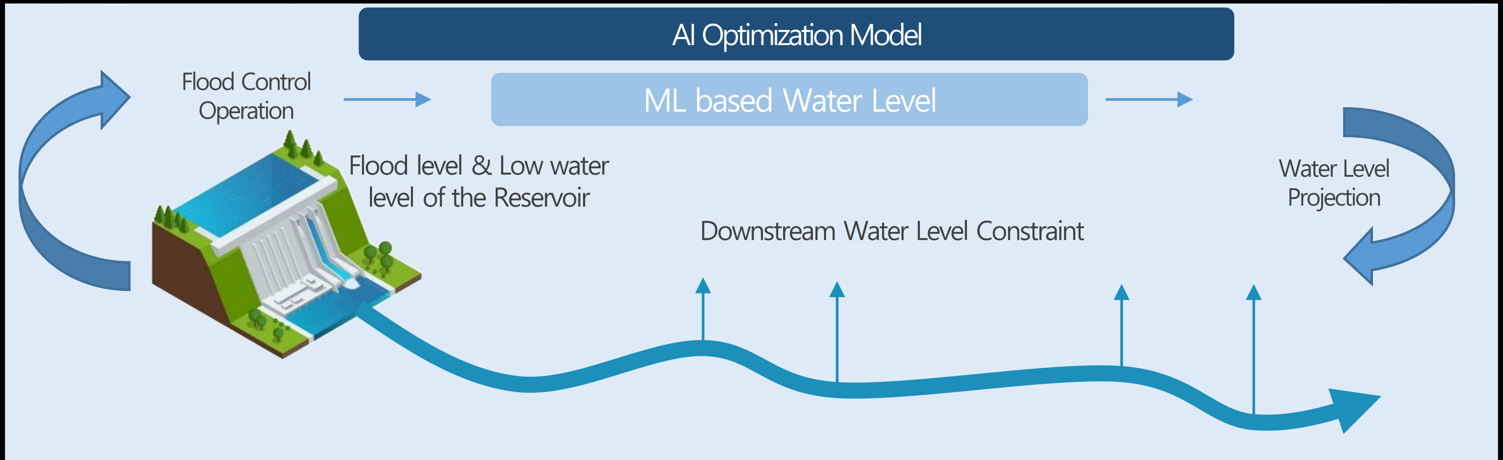


Cases	Flood Control(R)	Discharge(CMS)	y1	y2	Simulation (Flooding)	Simulation (Constraints)	Decision Making
1	0.455	1,750	OK	OK	X	X	Selected
2	0.515	2,000	OK	NG	O	X	counter plan
…	…	…	…	…	…	…	…
n	0.563	1,980	NG	OK	X	O	counter plan

$$Score(Risk) = \frac{Predicted\ WL - Level\ of\ Floodplain}{EL.\ of\ Levee - Level\ of\ Flood\ plain}$$

Optimization w/ Hybrid Model

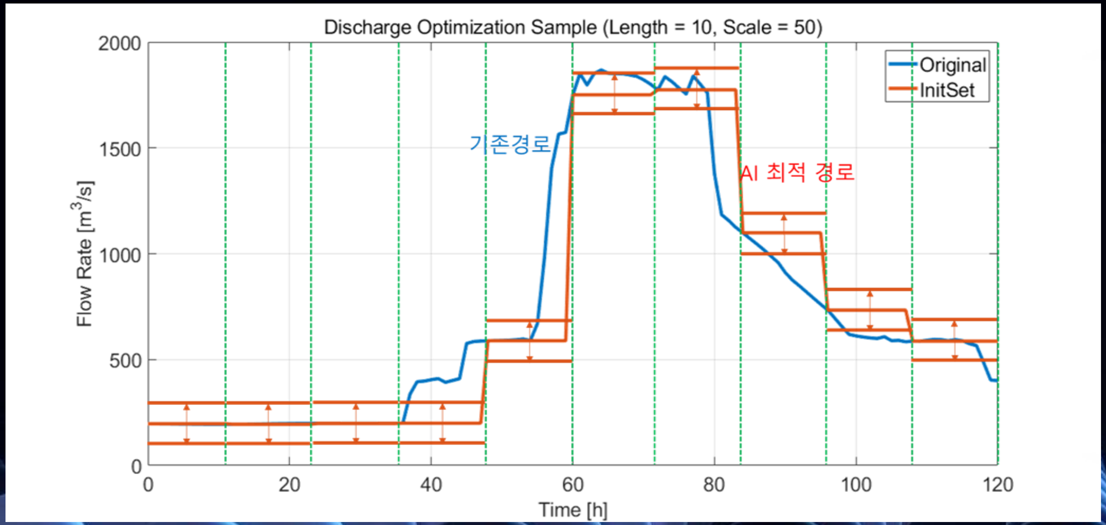
Lead-time Increase by adding rainfall forecast

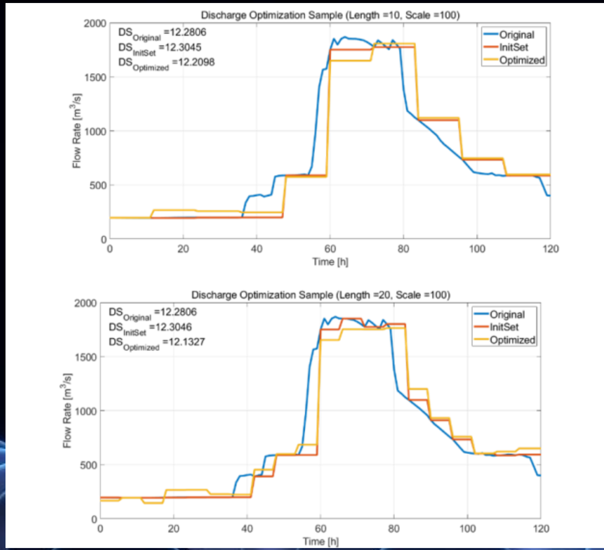


Aggressive

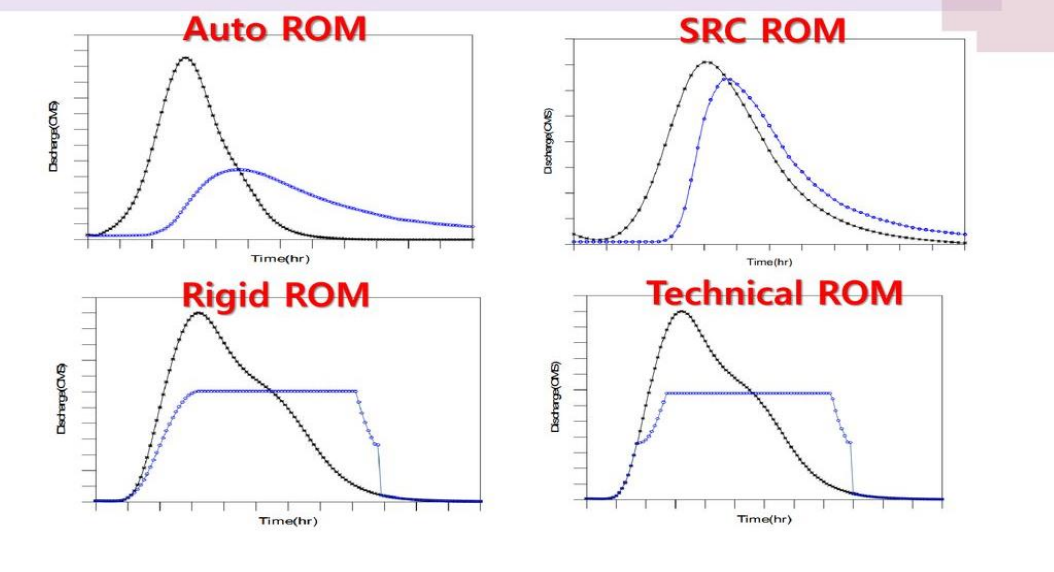
Moderate

Conservative

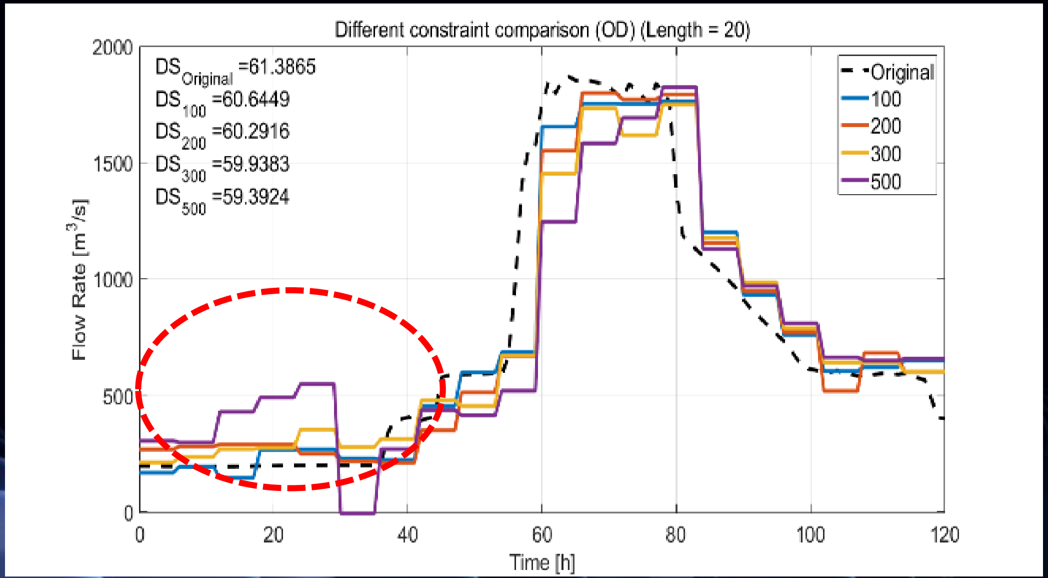


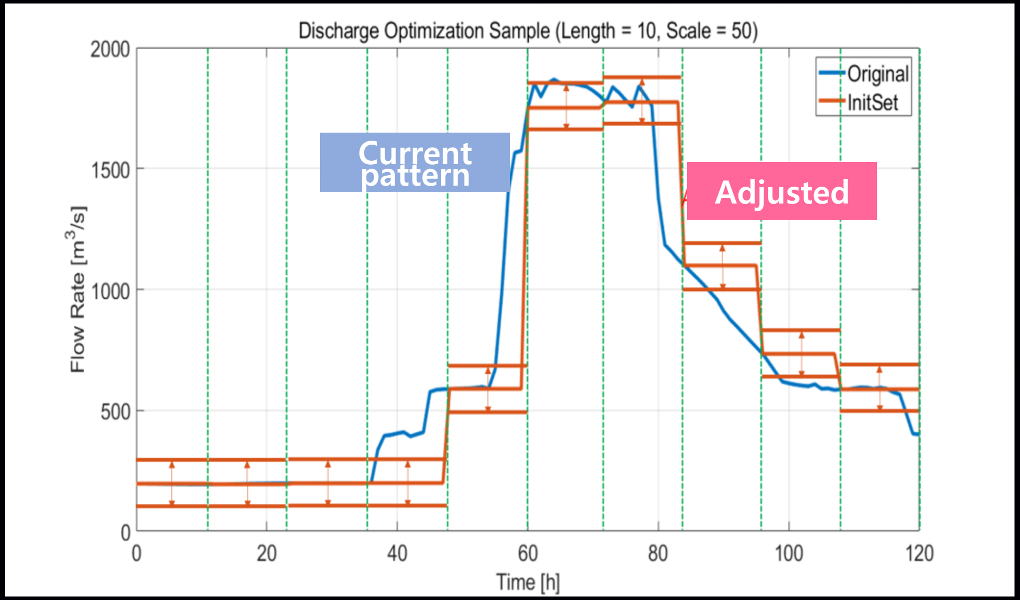


Optimization w/ Hybrid Model

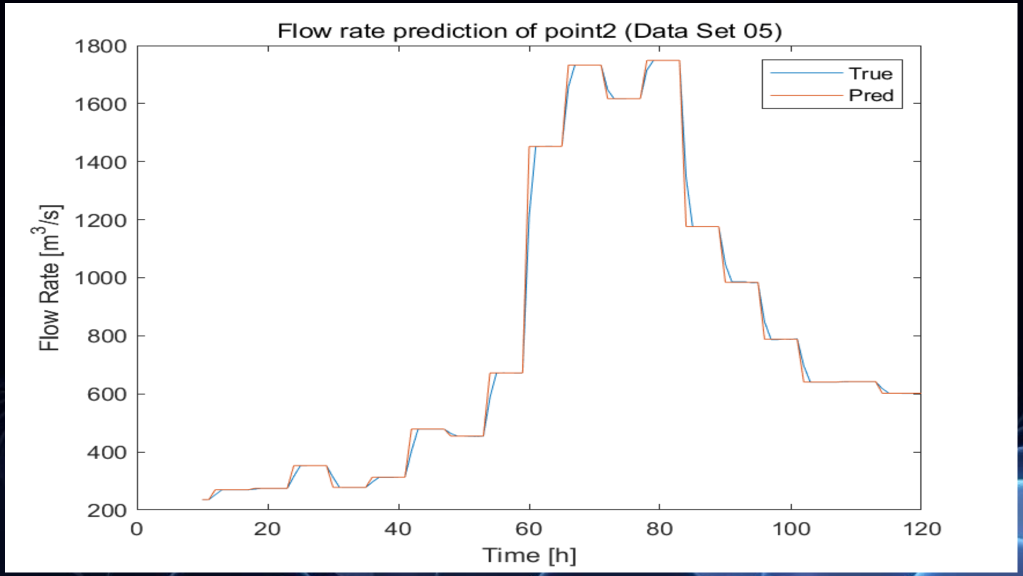


Possible dam discharge scenarios

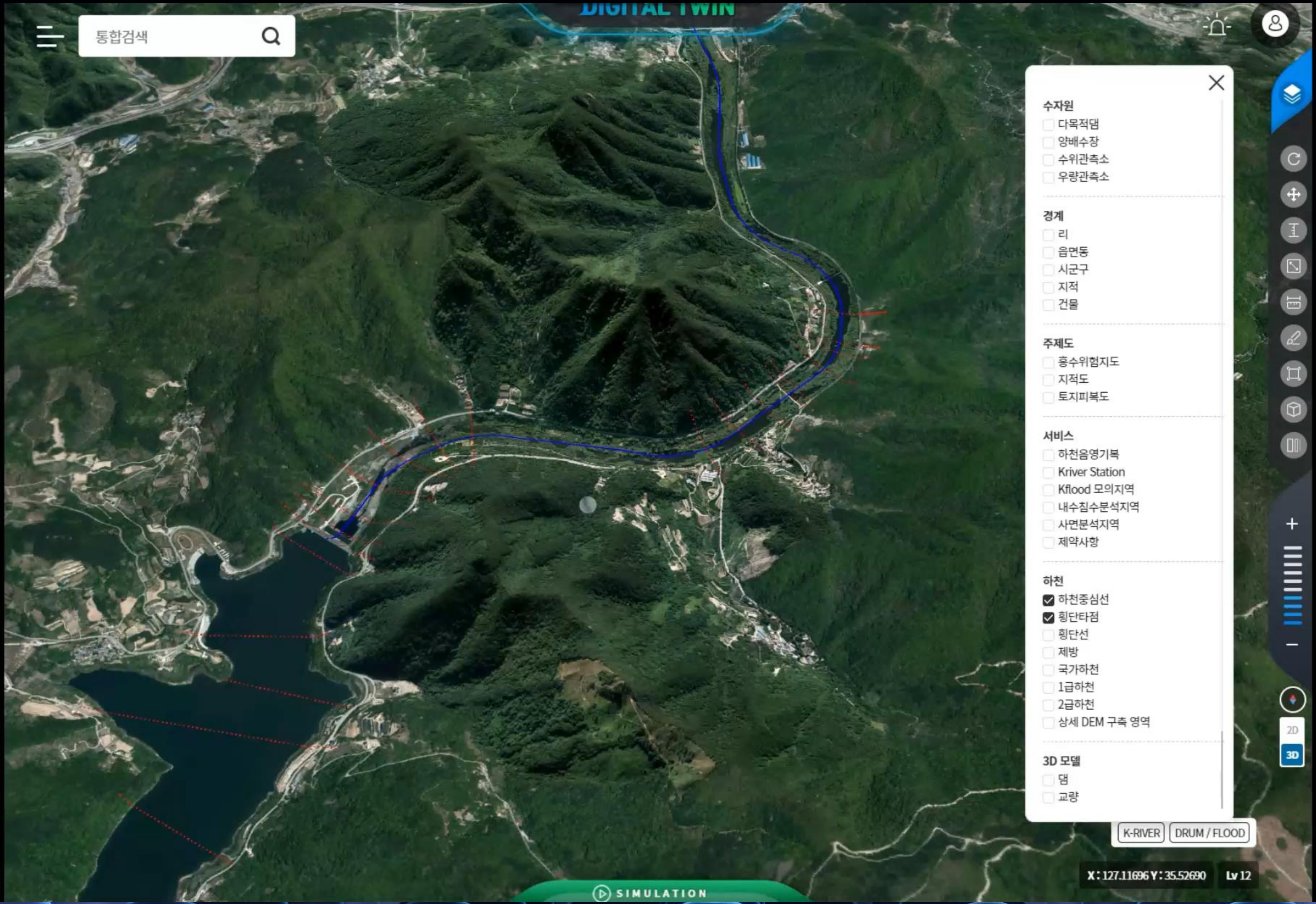




Optimized scenario at the current time

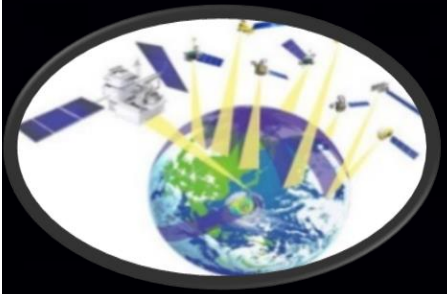


Optimization w/ Hybrid Model

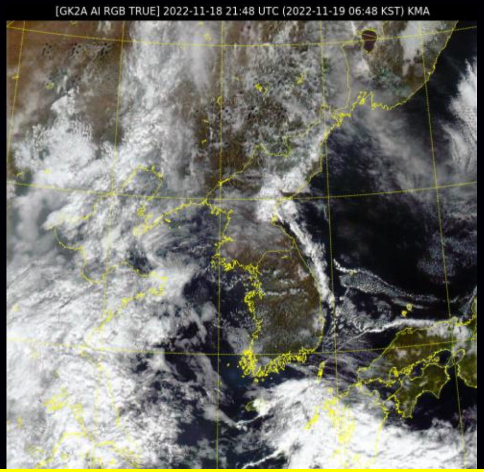


(Part II) Increasing accuracy of R-R model

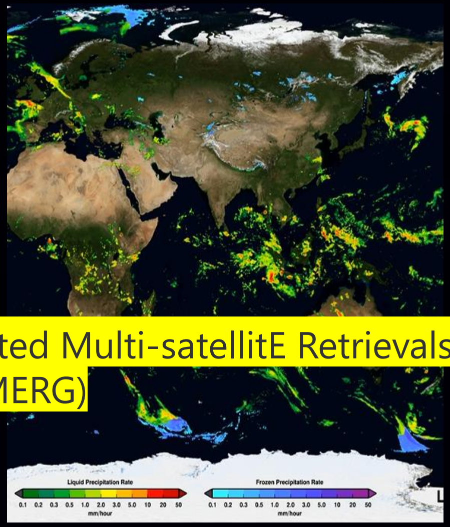
(Part-II) Remote Sensing Data Applications : Rainfall



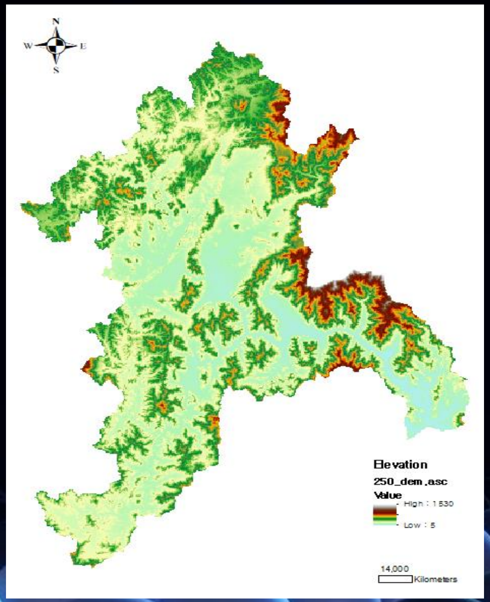
JAXA Global Rainfall Watch



Communication, Ocean and Meteorological, COMS

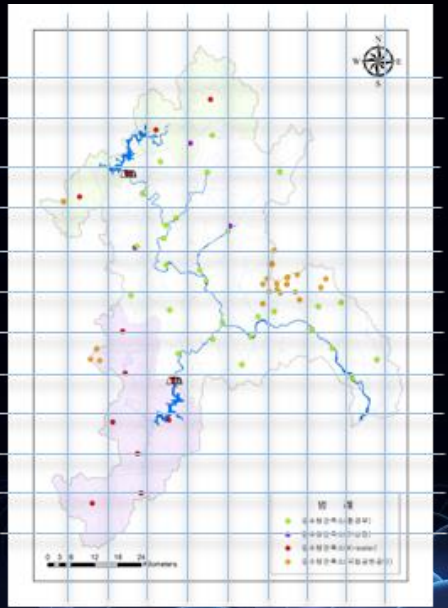


The Integrated Multi-satellitE Retrievals for GPM (IMERG)

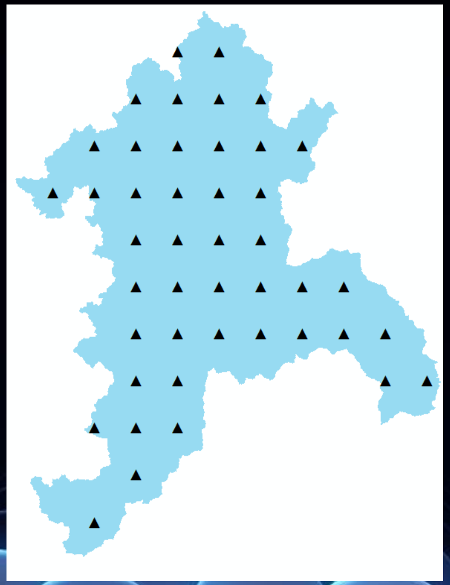


Seomjin River Basin

12 RF gauges

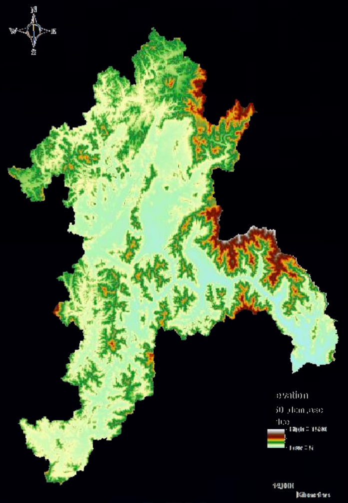


ML-based R/G value mapping

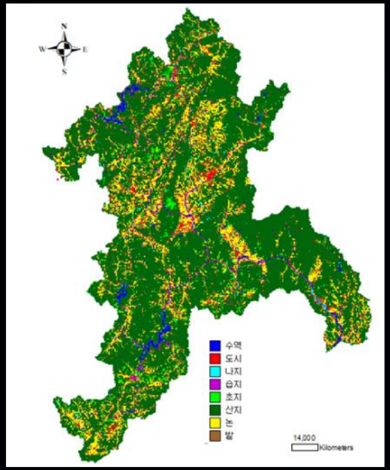


Rainfall distribution

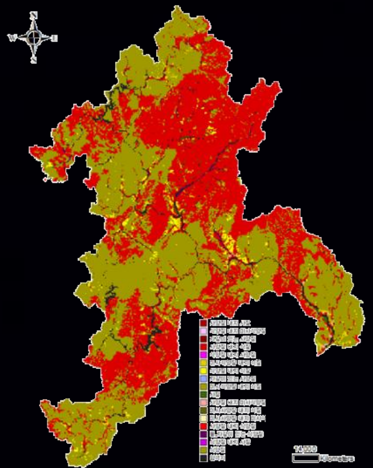
(Part-II) Remote Sensing Data Applications : for R-R Model



<DEM>

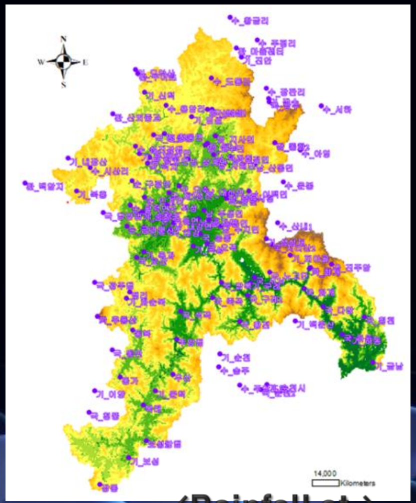


<land use>



<Soil map>

<Soil depth>



<Rainfall st.>

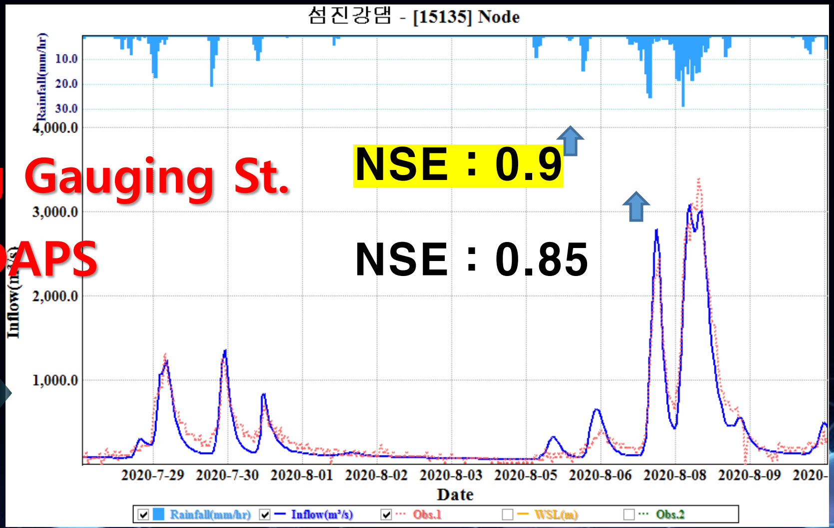
- Rainfall : 88(144)

Using Gauging St.

LDAPS

NSE : 0.9

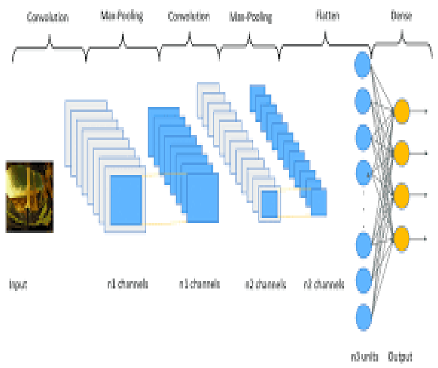
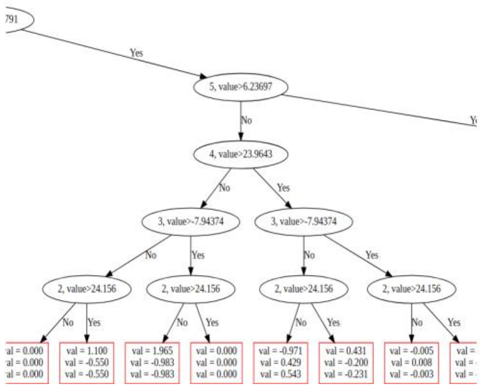
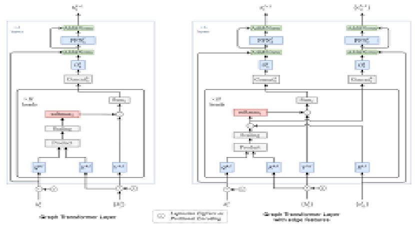
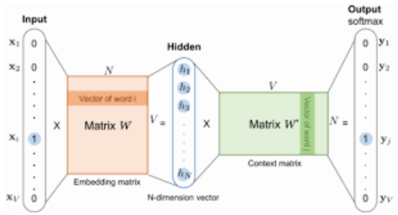
NSE : 0.85



Seomjin dam

(Part-II) Remote Sensing Data Applications : Rainfall

Global(Satellite Data) and gauge data(12) mapping (hourly)

	Conv2d (CNN)	CatboostRegressor (Decision Tree)	Sequence Modelling
	X (Images)	○	△ (Combined)
ML Libraries	Torch	Geopandas(for Grid) Xarray	Torch(Network structuring)
Algorithms			
	Prediction as a whole	-	One-hot-encoding for 12 Rainfall gauges
Metric (RMSE)	1.313	1.354	1.410

(Part-II) Remote Sensing Data Applications : Rainfall

Expanding monitoring to ungagged area (rainfall distribution) Seomjin River Basin (on-going)

① Collecting Data

- IMERG, GSMaP, PERSIANN-CCS https://chrsdata.eng.uci.edu/

- IMERG Final

* **Final (3months, research)**, Late (14hr, forecasting), Early (4hr, flash flooding)

Spatial Resolution: 10×10km (0.1×0.1°), Temporal Resolution: 30min

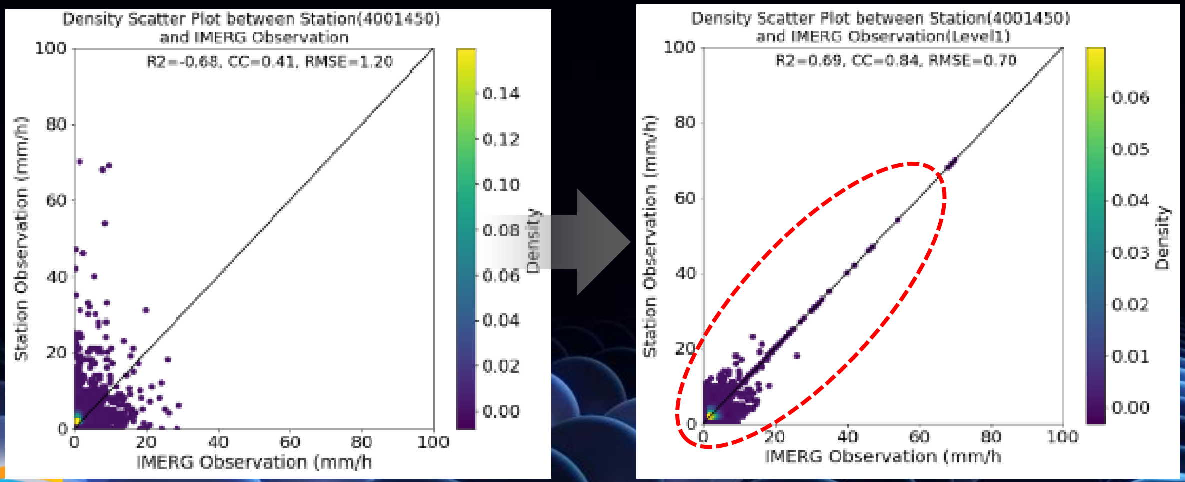
- GSMaP (Global Satellite Mapping of Precipitation, JAXA)

Spatial Resolution: 10×10km (0.1×0.1°), Temporal Resolution: 1hr

- PERSIANN-CCS (Precipitation Estimation form Remotely Sensed Information using ANN

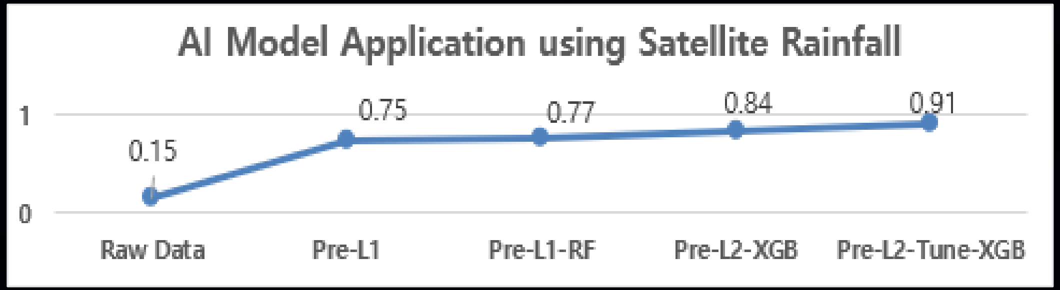
Spatial Resolution: 4×4km (0.04×0.04°), Temporal Resolution: 1hr

② Preprocessing

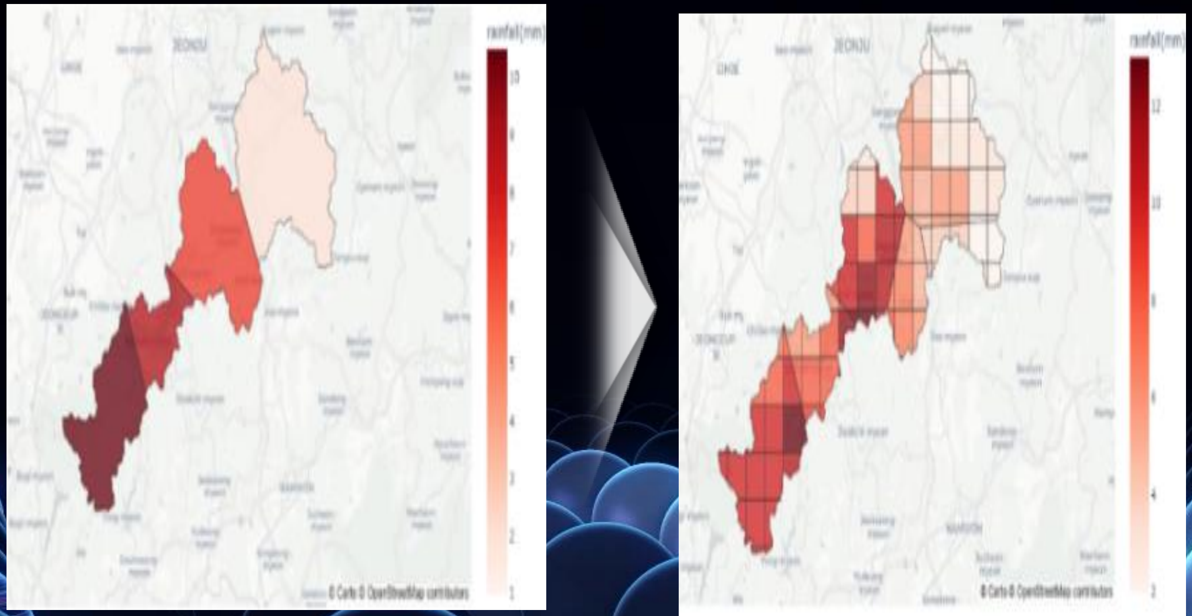


③ Application of ML

- Random Forest, XGBoost (w/ radar, Thiessen, ...)

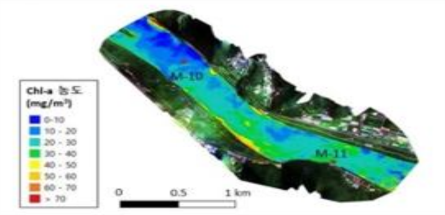


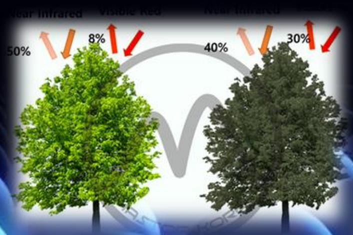
④ Rainfall Distribution of SJ Basin (upstream)



(Part-II) Remote Sensing Data Applications : Algae Monitoring

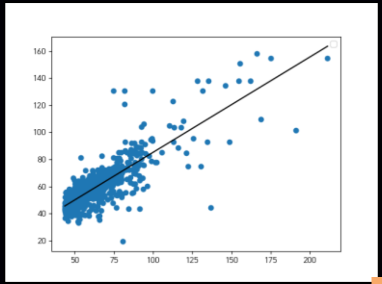
- R,G,B, NDVI, NDRE, GNDVI,GRVI
- NDBE, FR, FB, FG



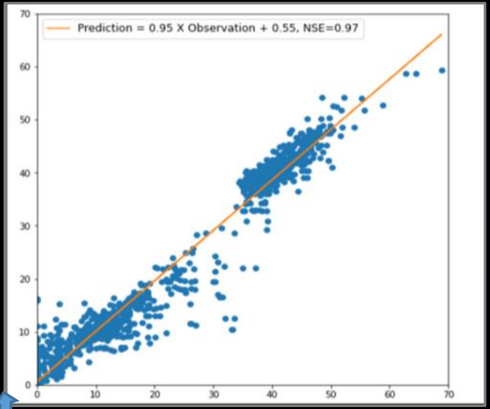


Sentinel-2 bands	Central wavelength (μm)	Resolution (m)
Band 1 – Coastal aerosol	0.443	60
Band 2 – Blue	0.490	10
Band 3 – Green	0.560	10
Band 4 – Red	0.665	10
Band 5 – Vegetation red edge	0.705	20
Band 6 – Vegetation red edge	0.740	20
Band 7 – Vegetation red edge	0.783	20
Band 8 – NIR	0.842	10
Band 8A – Vegetation red edge	0.865	20
Band 9 – Water vapour	0.945	60
Band 10 – SWIR – Cirrius	1.375	60
Band 11 – SWIR	1.610	20
Band 12 – SWIR	2.190	20

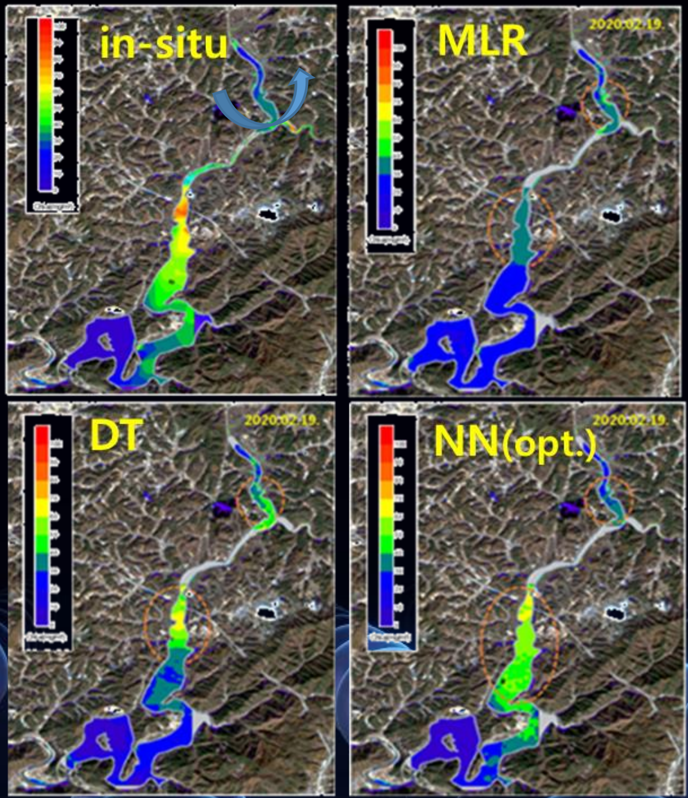
R^2 = 0.43 ~ 0.69

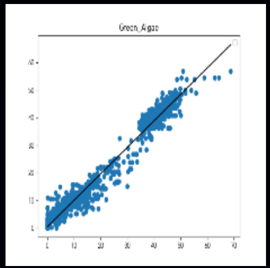


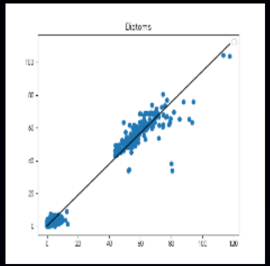
30% ↑



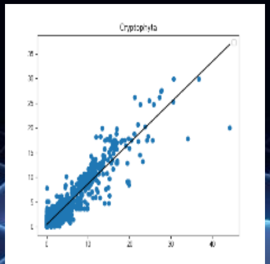
R^2 = 0.43 ~ 0.69

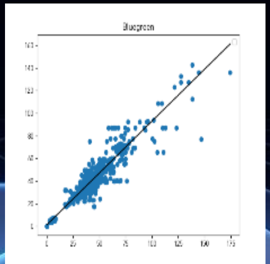






Green
Green





Diatoms

Dinoflagellates

Summary

- Feasibility of Hybrid ML techniques to partly surrogate physical models with predicted flow boundary conditions was studied.

  - Proposed model has shown a good performance under limited conditions so far.

  - If complementary models are established, it is expected to greatly support to optimize real-time dam operation system, and so on.

  - And machine learning-aided research is also continuing to improve the evaluation of rainfall for the specific basin by using satellite data.

Thank You!