J. Korea Water Resour. Assoc. Vol. 51, No. 6 (2018), pp. 503-514

Comparison of physics-based and data-driven models for streamflow simulation of the Mekong River

2019. 1. 23

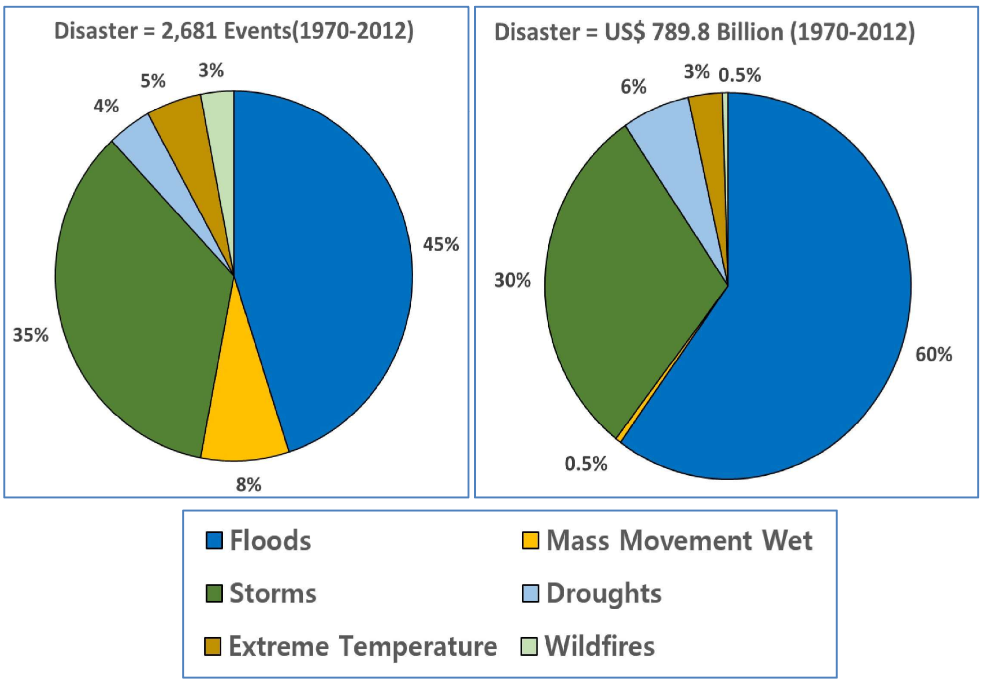
Giha Lee

leegiha@knu.ac.kr / http://leegiha.knu.ac.kr

KNU KYUNGPOOK NATIONAL UNIVERSITY

I. Background

Global disasters from 1970 to 2012 were reported that there were 8,835 disaster events, 1.94 million deaths, and US$ 2.4 trillion of economic losses (WMO, 2014).



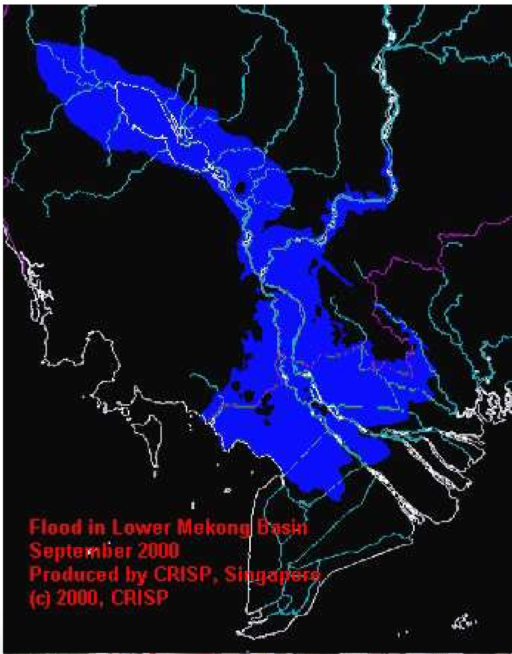
The number of Disaster and Total Economic losses by hazard type in Asia (1970-2012) (WMO, 2014)

I. Background

Flood happens every year in the Lower Mekong River Basin causing huge economic damage, numerous of affected people, and destroying large area of agriculture.

Mekong River Flood in Sep. 2000

- Flood in 50-year return period with flood volume 500 km^3 (MRC, 2005)
- More than 90 people died in Mekong Delta
- 600,000 people were homeless
- At lease 140,000 homes were inundated
- Widespread 800,000 km^2 in **Cambodia**, Lao, Thailand and Vietnam

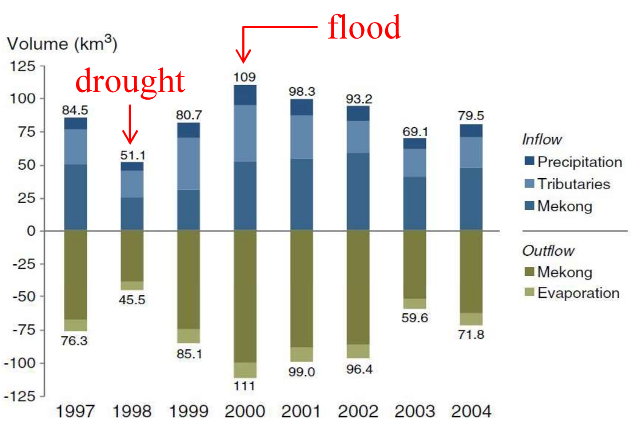


I. Background

Normal Flow

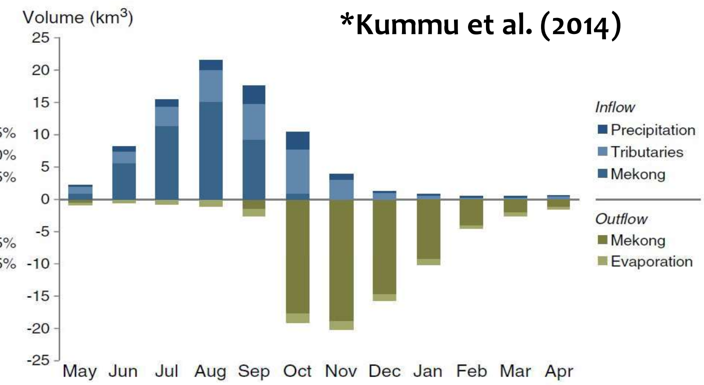
Wet Season (Reverse flow into the Tônlé Sap)

Receding Flow During Dry Season



Annual Water Balance

*Kummu et al. (2014)

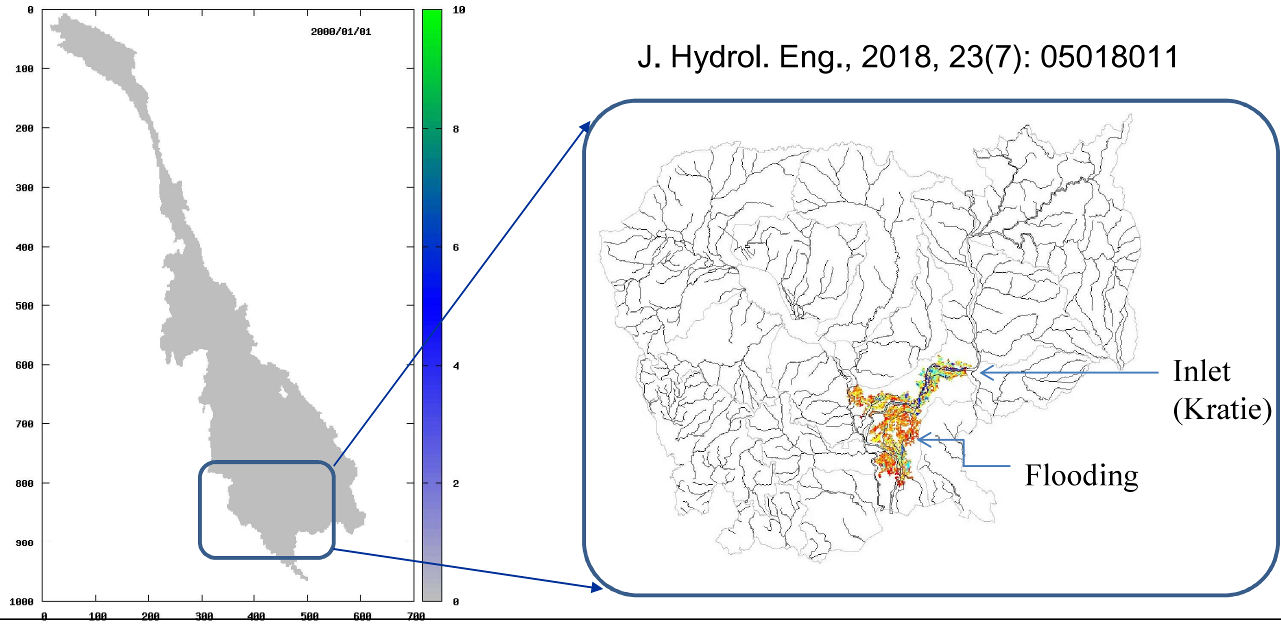


Monthly Water Balance

- **Major Inflow Sources of the Tonle Sap Lake: (1) Tonle Sap River – Mekong influence (50.3%), (2) Tributaries (34.1%), (3) Rainfall (12.4%), (4) Overland flow via floodplain (3.2%)**

II. Research Project (IHP, 2018 ~ 2021)

- **Project Title:** Countermeasures against water-related problems due to urbanization and land development in a developing country
- **One of the Goals:** To develop a **hydrologic modelling system (rainfall-runoff-inundation)** for water disaster reduction and damage mitigation **in the lower Mekong River basin** (in particular, **Cambodia**) **under basin development & climate change scenarios**

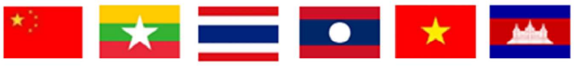


II. Research Project (IHP, 2018 ~ 2021)

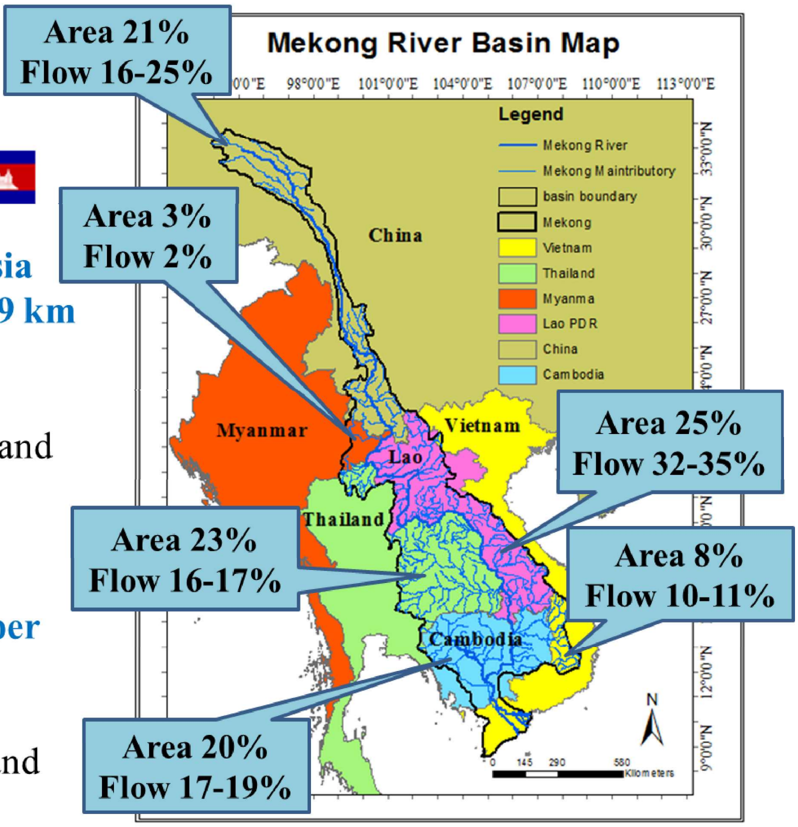
- The rainfall-runoff-inundation modelling process **needs huge computational time** (especially, CC scenarios) to apply to **the whole Mekong River basin**.
- To make modelling **system more compact and efficient**, independent rainfall-runoff model is fundamental to **simulate the streamflow of the inlet point** (Kratie – key station for flood warning in the LMB)
- Then, the system should be **combined with a 2D inundation** model for modelling flooding patterns of the Lower Mekong River basin (ex. Tonle Sap lake)
- To simulate discharge of the Kartie by **two models: SWAT(quasi physics-based) & LSTM(deep neural network)** were tested using globally-available or local database. → **Objective of this presentation**

III. The Mekong River

70 Million people: 6 countries

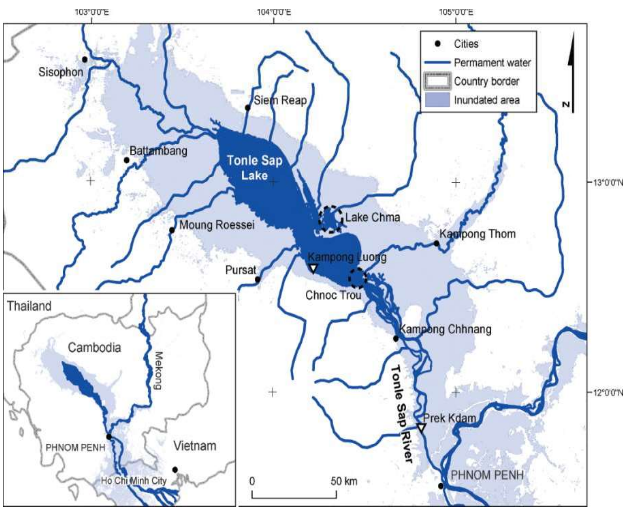


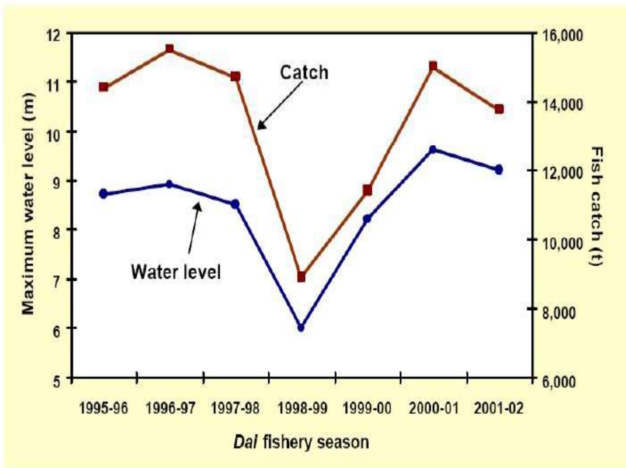
- Location: **The South-Eastern Asia**
- River length: approximately **4,909 km**
- Area: **795,000 km^2**
- Mean discharge: 15,000 m^3/s
- Ranked as the **12th longest** river, and **8th largest** water discharge in the world.
- Containing two major parts: **Upper Basin** and **Lower Basin**.
- Climate: tropical monsoon (**wet** and **dry** season).



III. The Tonle Sap Lake

The largest freshwater Lake in Southeast Asia





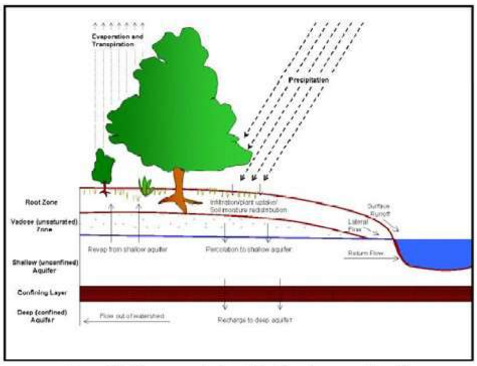
*Keskinen and Kummu (2013)

- More than **50% of Cambodian GDP** depends on the **primary industries** such as agriculture, fishing, and forestry, and the **Tonle Sap Lake** plays an **important role** to support the **national economy** in Cambodia
- In addition, **Cambodian people** usually take **nourishment** from the fish and aquatic animals of Tonle Sap Lake.

8

IV. Model – SWAT & SWAT-CUP

- **SWAT(Soil Water Assessment Tool)** is **a process-based hydrological model**, developed by USDA-ARS to evaluate alternative management strategies on water resources and non-point source pollution in large river basins. → **Basic tool of the MRC**



Land Phase of the Hydrologic Cycle

- Controls the amount of water, sediment, nutrient and pesticide loading to the main channel
- Use water balance equation:

$SW_t = SW + \Sigma(R_{day}-Q_i-E_a-P_i-QR_i)$

SW, soil water content

t, time

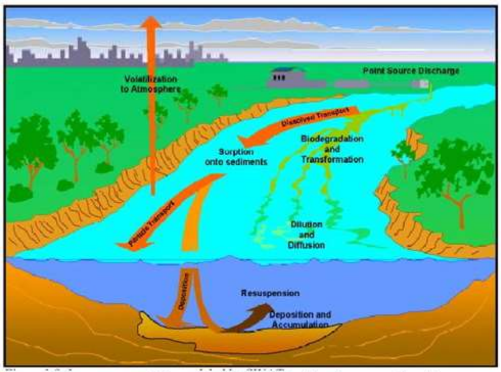
R_{day}, amount of precipitation

Q_i, amount of surface runoff

E_a, amount of evapotranspiration

P_i, amount of percolation

QR_i, amount of return flow



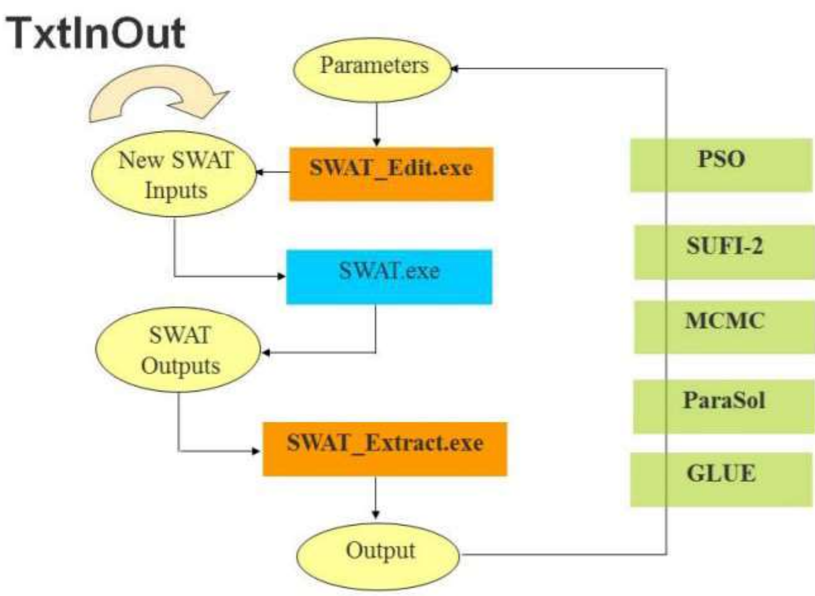
Routing Phase of the Hydrologic Cycle

- The movement of water, sediments, etc. through the channel network of the watershed to the outlet
- Keeps track of mass flow in the channel
- Models the transformation of chemicals in the stream and streambed
- Inputs for Routing in the Main Channel or Reach: Flood Routing, Sediment Routing, Nutrient Routing, Channel Pesticide Routing
- Inputs for Routing in the Reservoir: Reservoir Outflow, Sediment Routing,

http://www.brc.tamus.edu/swat/

IV. Model – SWAT & SWAT-CUP

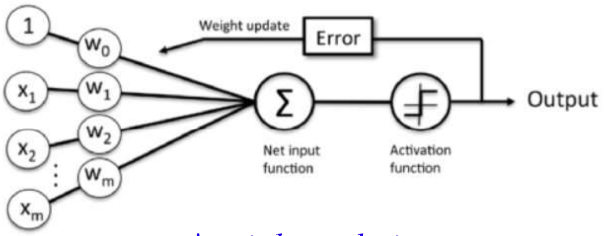
- **SWAT-CUP(SWAT Calibration and Uncertainty Programs)** is computer program for SWAT calibration.
- It contains **various optimization programs such as SUFI2, PSO, GLUE, ParaSol, and MCMC** algorithms for sensitivity analysis, calibration, validation as well as uncertainty analysis of the SWAT.



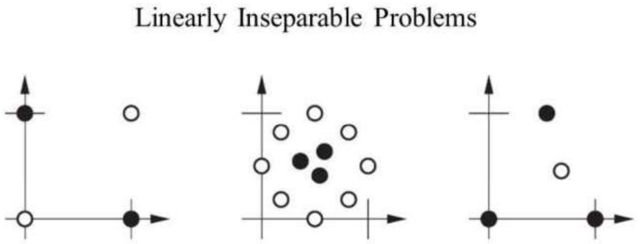
https://swat.tamu.edu/media/114860/usermanual_swatcup.pdf

IV. Model – LSTM (Deep Learning)

Perceptron model

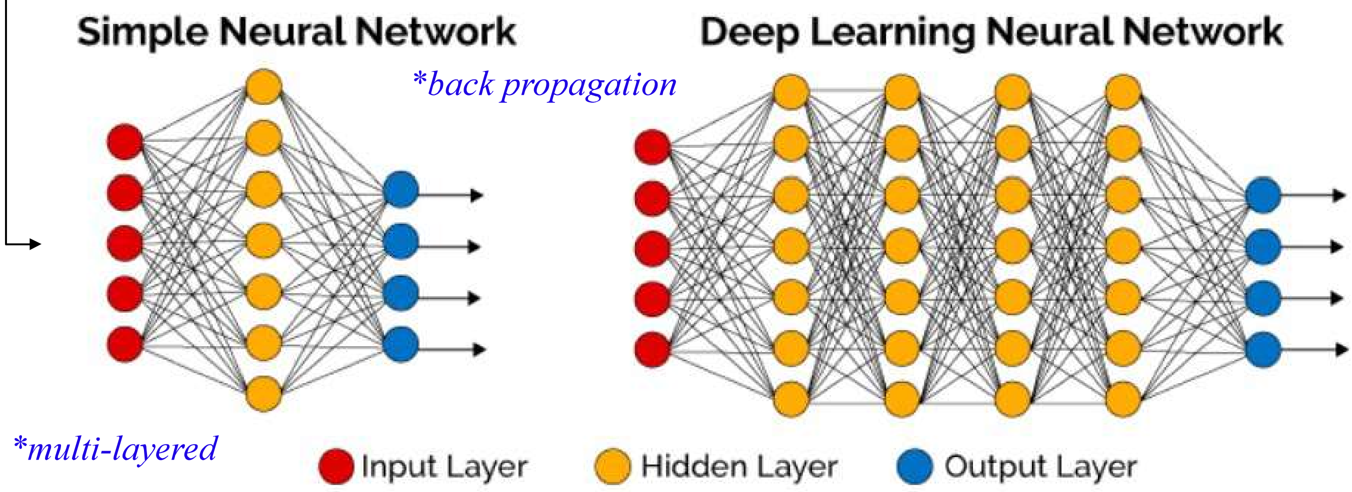


*weight updating



30 years

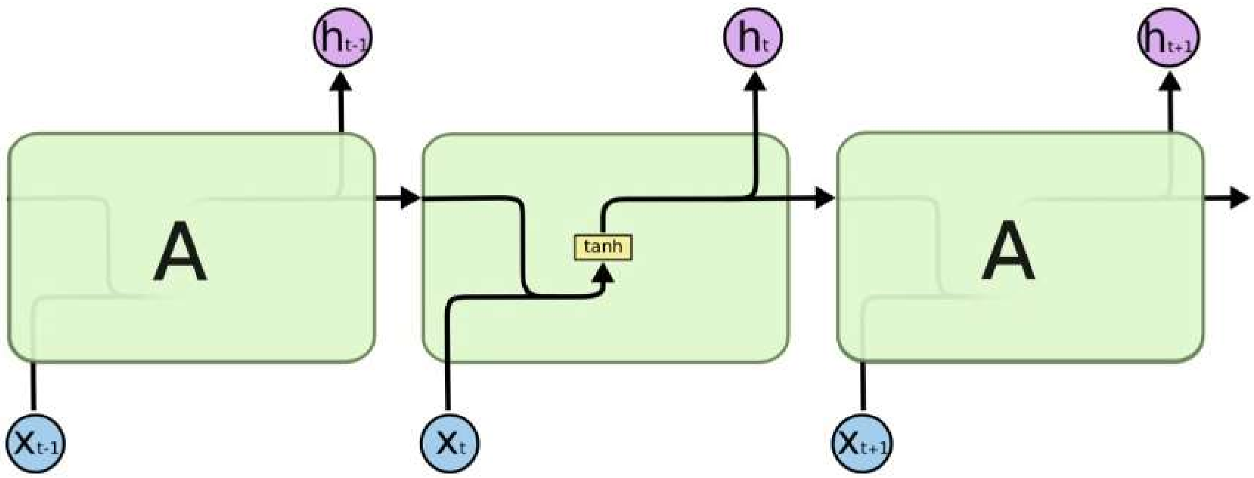
*back propagation



*multi-layered

IV. Model – LSTM (Deep Learning)

- **Long Short Term Memory networks** – usually just called “LSTMs” – are a **special kind of RNN(Recurrent Neural Network)**, capable of learning long-term dependencies.
- In standard RNNs, **a chain of repeating module** will have a very simple structure, such as a single tanh layer.

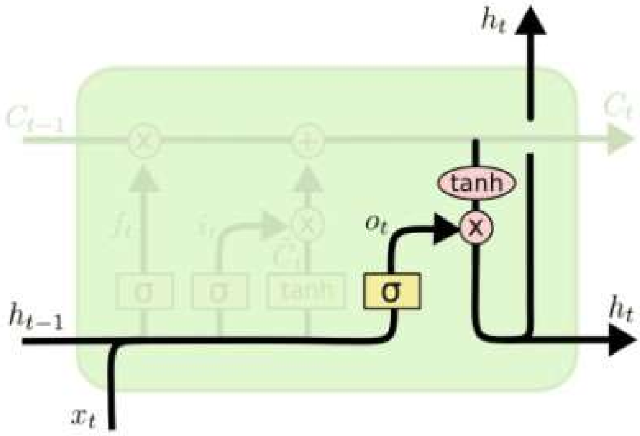


The repeating module in a standard RNN contains a single layer.

http://colah.github.io/posts/2015-08-Understanding-LSTMs/

IV. Model – LSTM (Deep Learning)

- LSTMs also have this chain like structure, but the repeating module has a different structure. **Instead of having a single neural network layer, there are four, interacting in a very special way**



Output gate

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

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

The repeating module in an LSTM contains four interacting layers.

http://colah.github.io/posts/2015-08-Understanding-LSTMs/

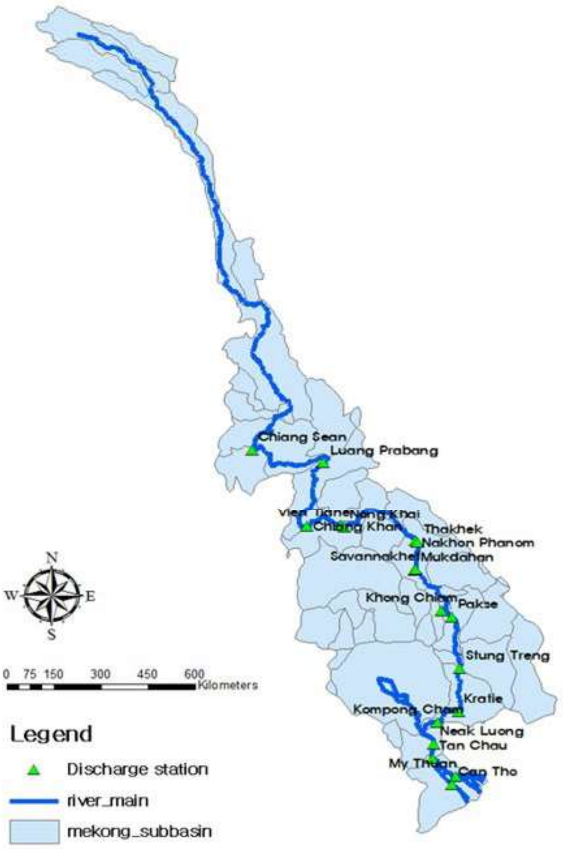
IV. Model – LSTM (Deep Learning)

- **TensorFlow by Google** is a popular **deep learning open source library** providing various deep neural network algorithms including LSTM
- In this study, the Mekong streamflow time series of the Kratie station was simulated by the **TensorFlow-LSTM**

Library	API	Platforms	Started by	Year
Caffe	Python, C++, Matlab	Linux, macOS, Windows	Y. Jia, UC Berkeley (BVLC)	2013
Deeplearning4j	Java, Scala, Clojure	Linux, macOS, Windows, Android	A. Gibson, J.Patterson	2014
H2O	Python, R	Linux, macOS, Windows	H2O.ai	2014
MXNet	Python, C++, others	Linux, macOS, Windows, iOS, Android	DMLC	2015
TensorFlow	Python, C++	Linux, macOS, Windows, iOS, Android	Google	2015
Theano	Python	Linux, macOS, iOS	University of Montreal	2010
Torch	C++, Lua	Linux, macOS, iOS, Android	R. Collobert, K. Kavukcuoglu, C. Farabet	2002

V. Model Applications

- **Modeling Domain –** Lower Mekong River Basin from the Chiang Saen

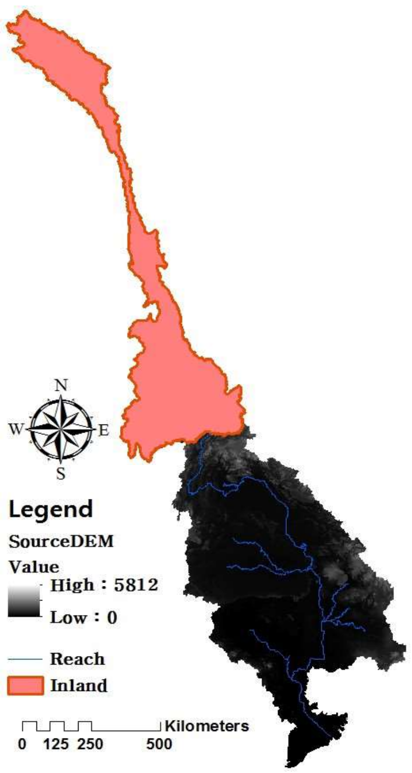


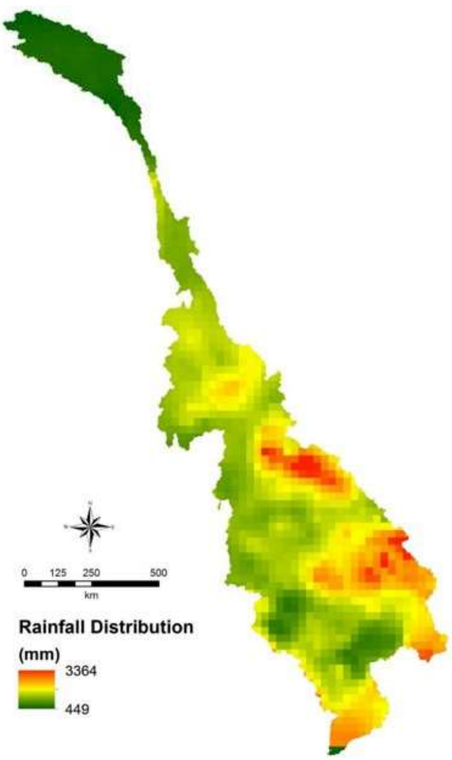
Mean annual discharge at main water level stations

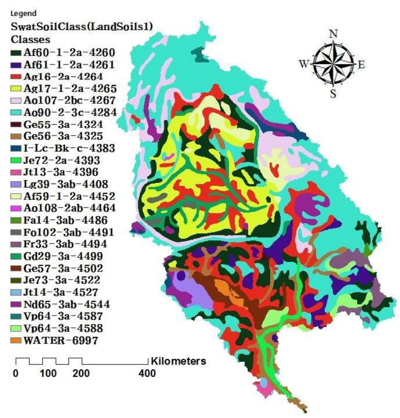
	Station Name	Catchment area (km^2)	Mean annual discharge (m^3/s)	as % total Mekong
1	Chiang Saen	189,000	2,700	18
2	Luang Prabang	268,000	3,900	26
3	Chiang Khan	292,000	4,200	28
4	Vientiane	299,000	4,400	29
5	Nong Khai	302,000	4,500	30
6	Nakhon Phanom	373,000	7,100	47
7	Mukdahan	391,000	7,600	51
8	Pakse	545,000	9,700	65
9	Stung Treng	635,000	13,100	87
10	Kratie	646,000	13,200	88
	Basin total	795,000	15,000	100

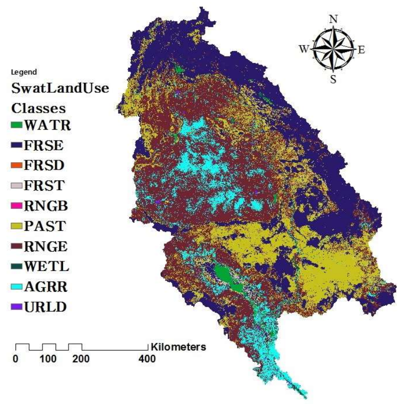
V-1. SWAT Application

- **Model Input Data –** DEM (USGS-HydroSHED), Landuse (GLCF-MODIS LandCover), Geology (FAO), Rainfall (APHRODITE) / Resolution (30 arc-seconds: Approx. 1km)



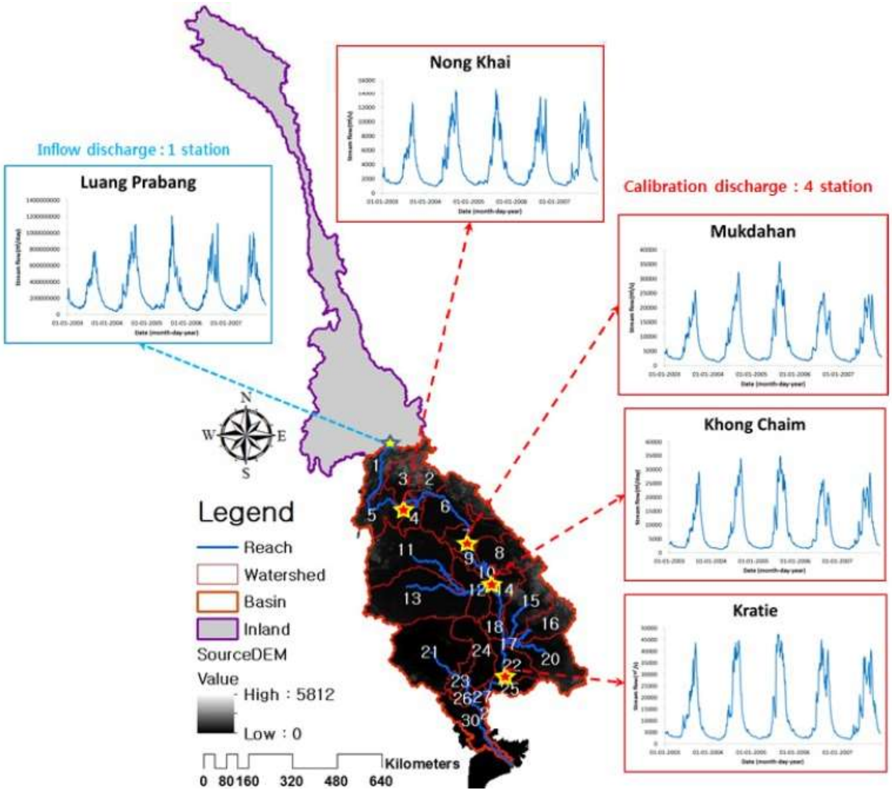






V-1. SWAT Application

- **Model Calibration –** 4 stations, Hydrologic Response Unit(No. 30) / Period (2003~2007, warming-up: 2000~2002) / 11 selected sensitive parameters (2000 iteration)



V-1. SWAT Application

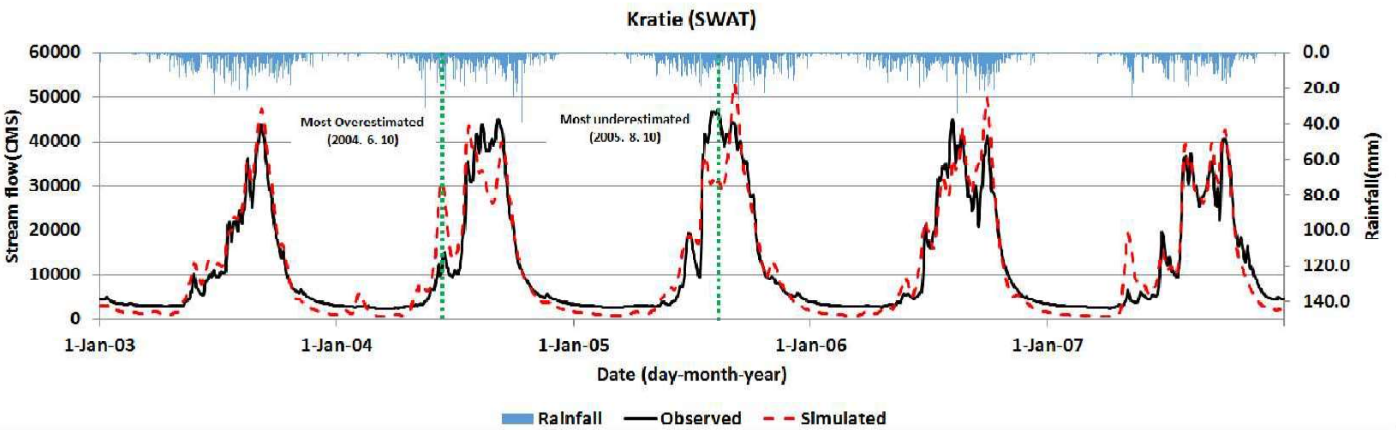
- **Model Calibration –** 4 stations, Hydrologic Response Unit(No. 30) / Period (2003~2007, warming-up: 2000~2002) / 11 selected sensitive parameters (2000 iteration)

Parameter	Description and units	Value	t-stat
ALPHA_BF.gw	Baseflow alpha factor (–)	0.540	29.500
RCHRG_DP.gw	Deep aquifer percolation fraction (–)	0.247	-1.086
CN2.mgt	Curve number (–)	60.913	0.667
CH_K2.rte	Channel effective hydraulic conductivity (mm/hr)	281.443	-1.631
SOL_AWC(..).sol	Available water capacity (mm/mm)	0.372	-0.624
CH_N2.rte	Manning's n-value for main channel (–)	0.124	-0.152
SURLAG.bsn	Surface runoff lag (days)	18.280	-0.983
ESCO.hru	Soil evaporation compensation factor (–)	0.994	-0.488
SOL_K(..).sol	Saturated hydraulic conductivity (mm/hr)	1938.421	0.503
GW_DELAY.gw	Groundwater delay time (days)	190.658	-0.398
CANMX.hru	Canopy storage (mm)	54.132	-0.719

V-1. SWAT Application

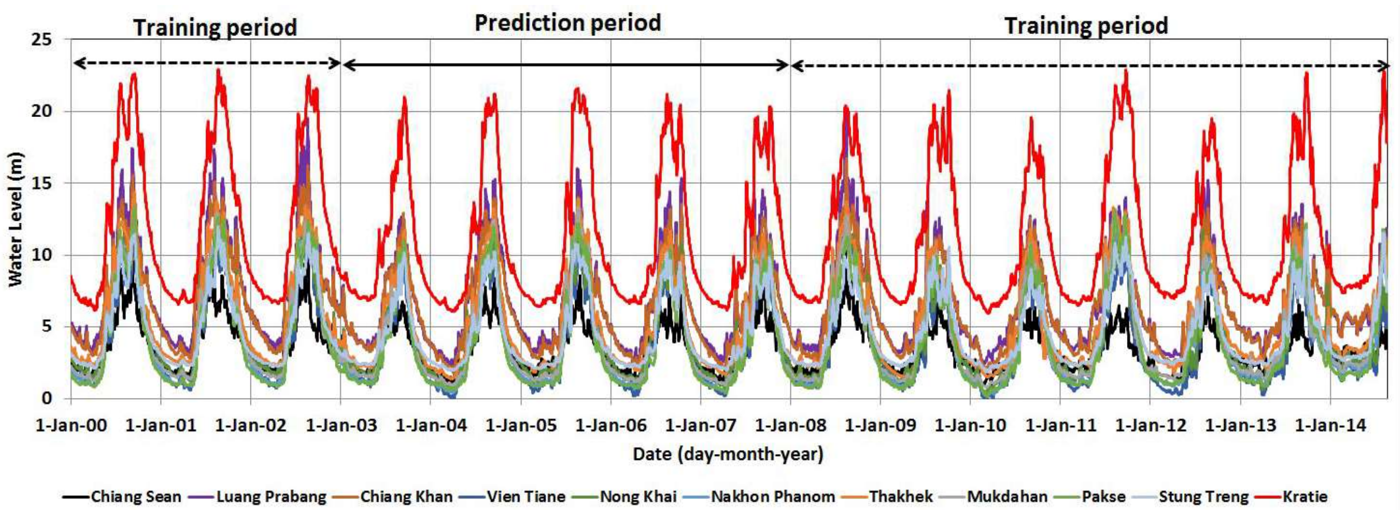
- **SWAT Simulation Results** – RMSE ($3941.71 m^3/s$), NSE(0.9)

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(O_t - P_t)^2} \qquad NSE = 1 - \frac{\sum_{t=1}^{N}(O_t - P_t)^2}{\sum_{t=1}^{N}(O_t - \overline{O})^2}$$



V-2. LSTM Application

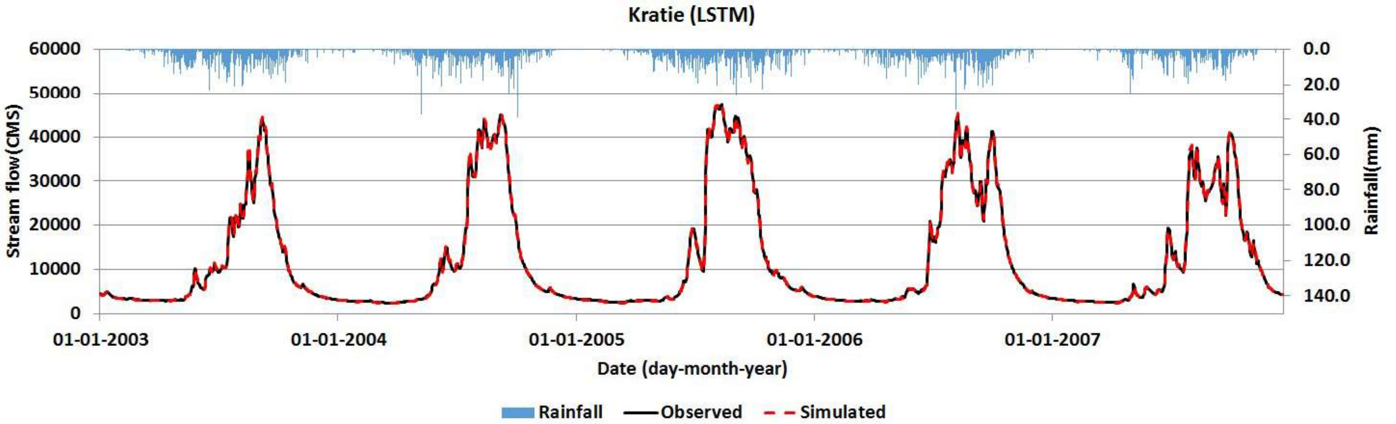
- **Model Training & Prediction** – Use of water level at 10 stations from 2000 to 2014. 8 (training: 2000 ~ 2003, 2008 ~ 2014 / prediction: 2003 ~ 2007) → convert into discharge using the Kratie rating curve, Hyper parameters (Learning rate: 0.01, No. hidden layers: 22, No. epoch: 5000, sequence length: 3, 5, 7 days)



V-2. LSTM Application

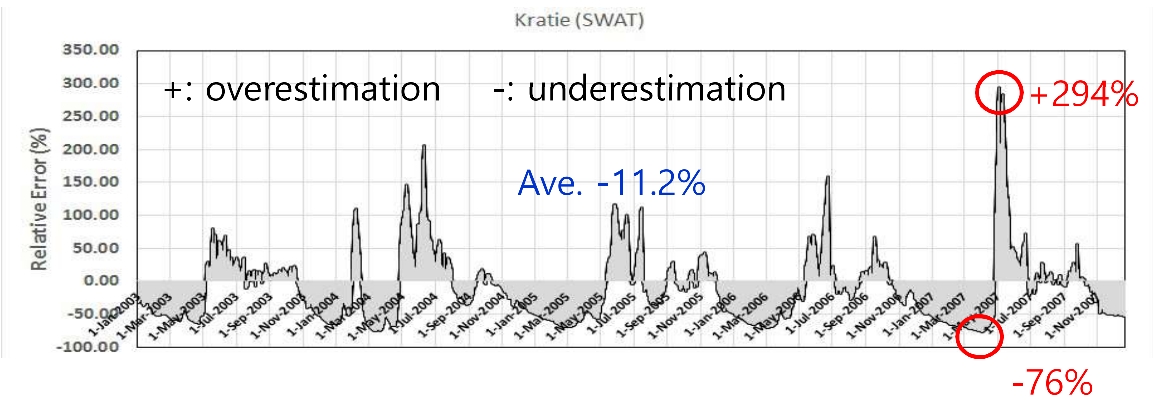
- **LSTM Simulation Results –** provide highly accurate model performances despite sequence length

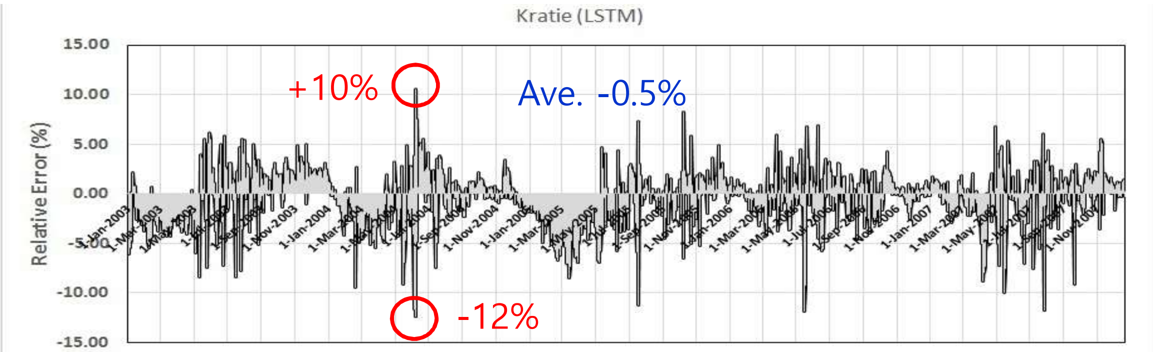
Performance \ Sequence Length	3 days	5 days	7 days
RMSE [㎥/s]	345.24	334.70	322.21
NSE [-]	0.99	0.99	0.99
R^2	0.99	0.99	0.99



V-2. Comparison of two models

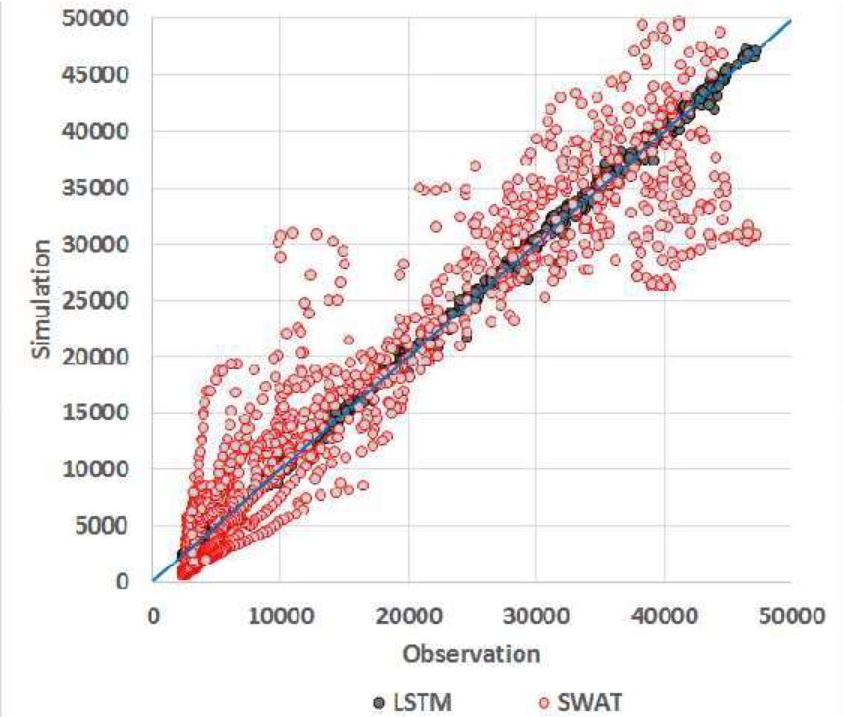
- **LSTM** leads to **more stable and accurate** hydrographs (i.e. both low and high flow) at the Kratie station **than SWAT**





V-2. Comparison of two models

- **LSTM** leads to **more stable and accurate** hydrographs (i.e. both low and high flow) at the Kratie station **than SWAT**



Summary & Concluding Remarks

- In this study, **runoff simulations at the Kratie station of the lower Mekong river** are performed using **SWAT, a physics-based hydrologic model, and LSTM, a data-driven deep learning algorithm**.

- The simulation results show that **Nash-Sutcliffe Efficiency (NSE)** of each model were calculated at **0.9(SWAT) and 0.99(LSTM)**, respectively.

- In order to **simply simulate hydrological time series of a ungauged large watershed**, data-driven model like the **LSTM method is more applicable** than the physics-based hydrological model having complexity due to various database pressure because the LSTM is able to **memorize the preceding time series sequences** and reflect them to prediction.

Thank You

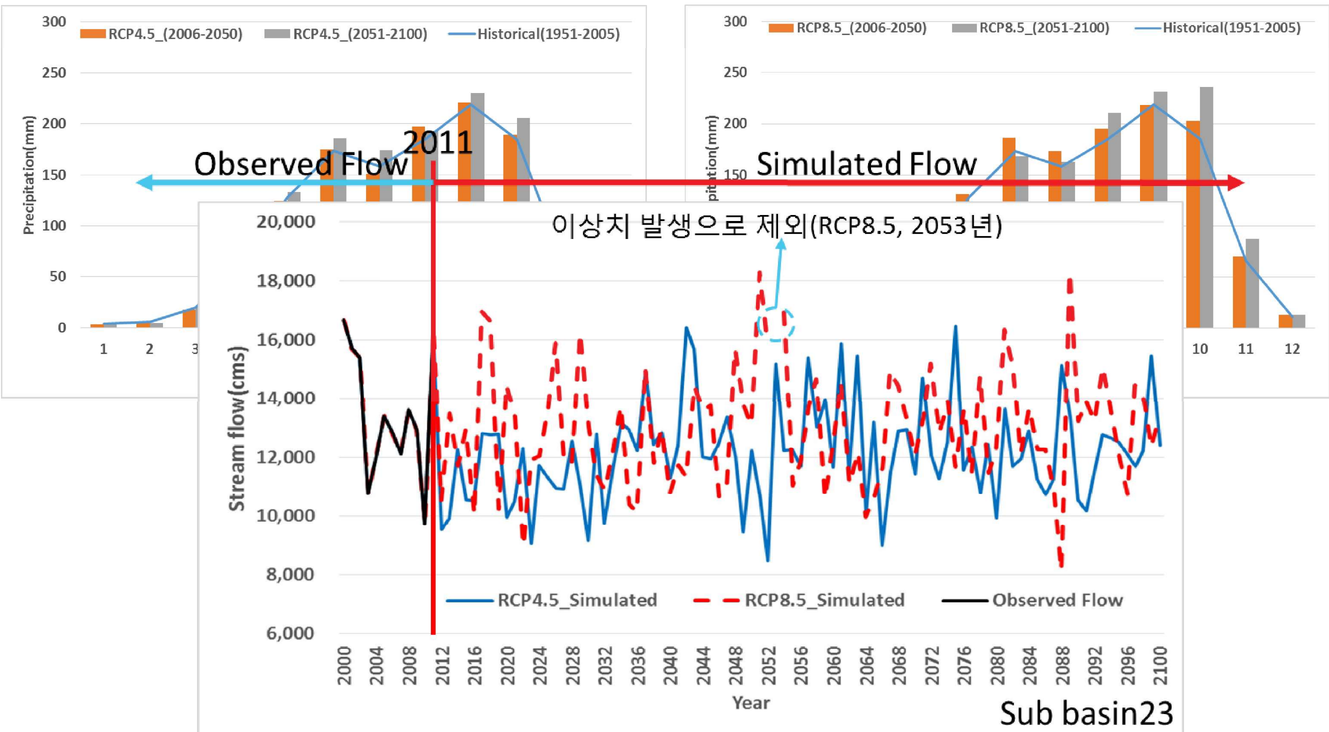
Q&A

25

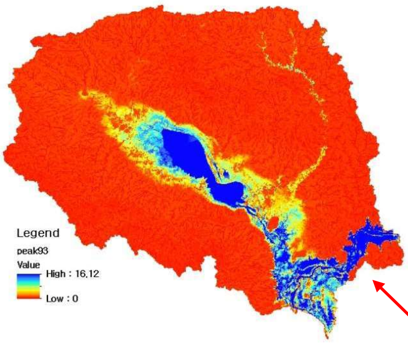
Future Research

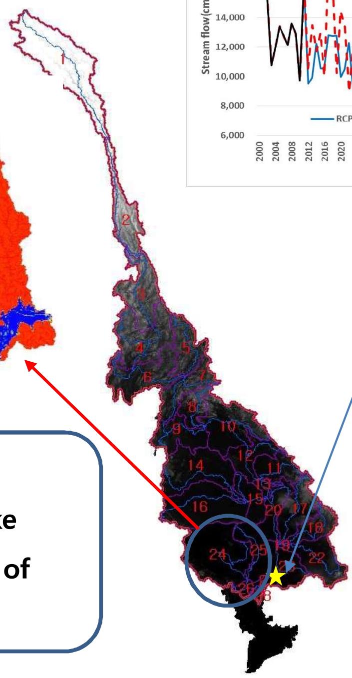
- Climate Change Impact on Hydrologic Regime in the Tonle Sap Basin, Cambodia

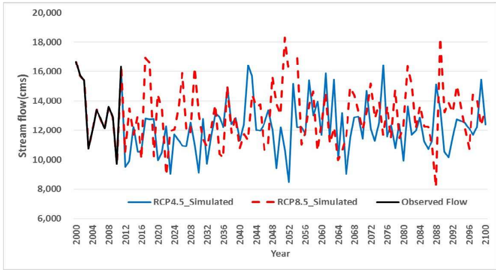
HadGEM3(Hadley Center Global Environmental Model version3)-RA: RCP 4.5 & RCP 8.5



Future Research







- **Generate inlet discharge data from CC scenarios**
- **Consideration of land use change**

- **Alternation of hydrologic regime of the Tonle Sap lake**
- **Socioeconomic assessment of the Tonle Sap basin**