Peishi Jiang

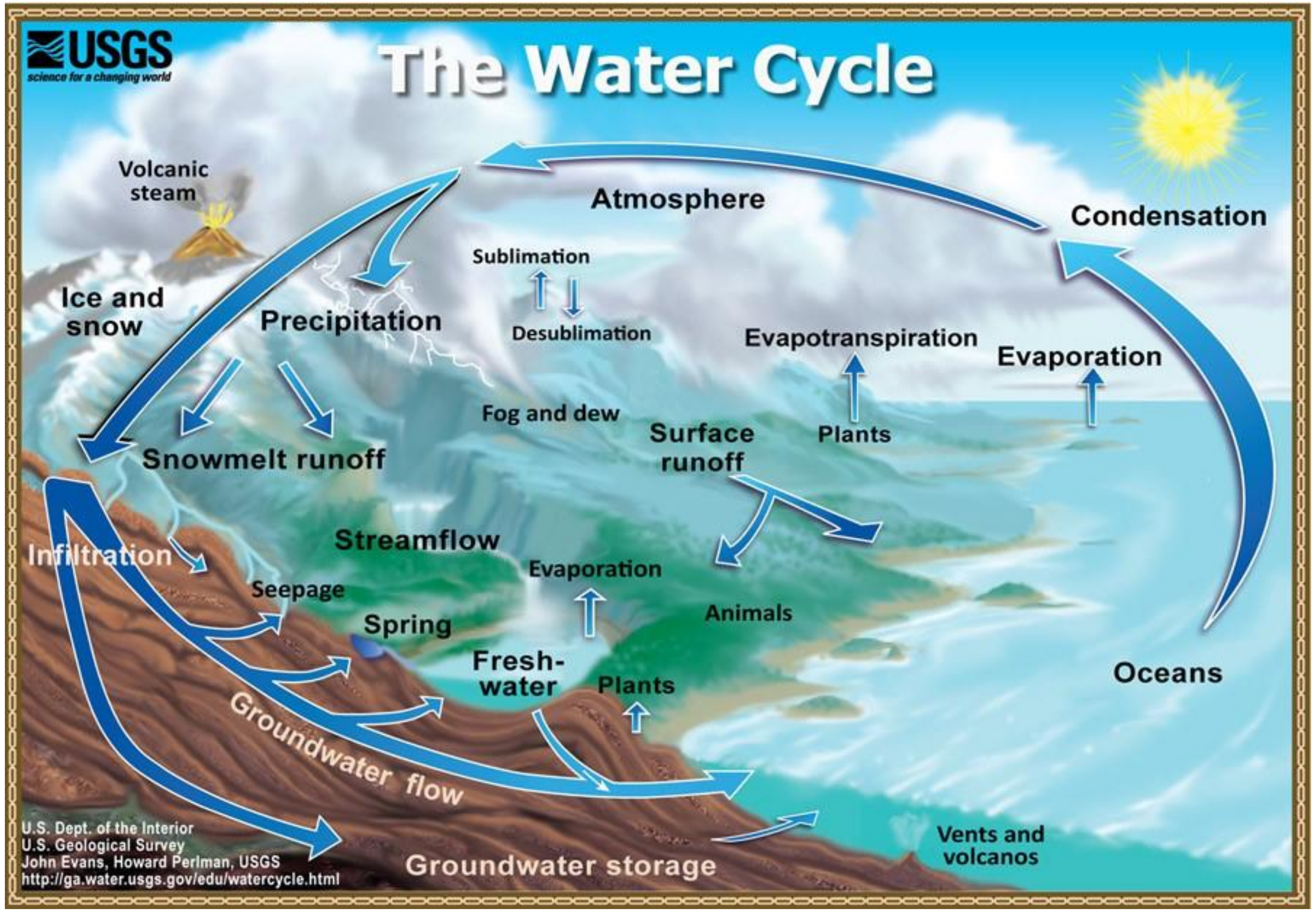
Pacific Northwest National Laboratory

Machine Learning (ML) Applications in Hydrology

Hydrology: the study of water

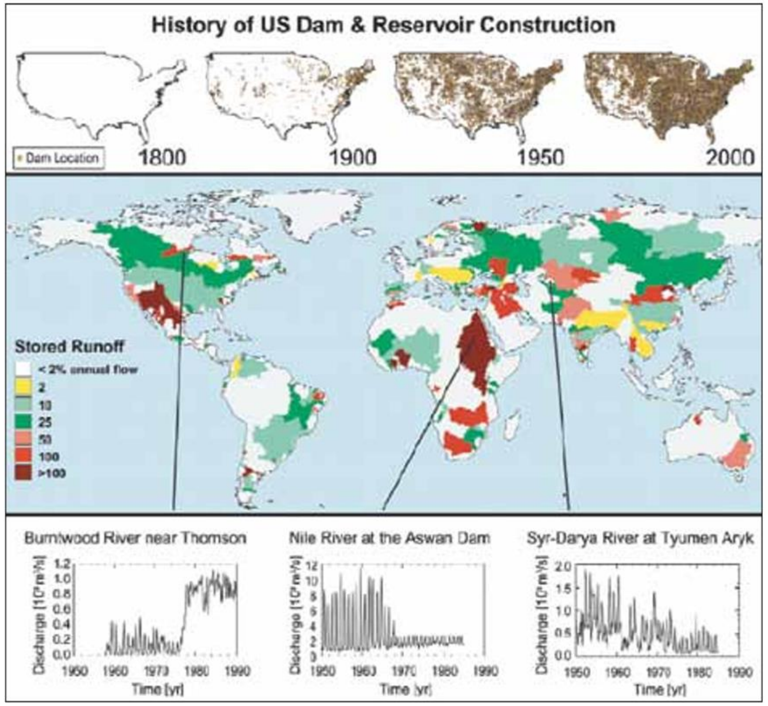
"Hydrology is the science that encompasses the occurrence, distribution, movement and properties of the waters of the earth and their relationship with the environment within each phase of the hydrologic cycle."

by United States Geological Survey



What are the grand challenges in Hydrology?

History of US Dam & Reservoir Construction



SOURCE: Vörösmarty et al. (2004) American Geophysical Union.

The water cycle: an agent of change

Water and life

Clean water for people and ecosystems

National Research Council, et al.
"Challenges and opportunities in the hydrologic sciences." (2012).

Flowing meltwater on the surface of the Greenland ice sheet

SOURCE: Roger Braithwaite, University of Manchester and Specialiststock.com.

How have Machine Learning (ML) been applied in Hydrology?

machine learning in hydrology

About 171,000 results from Google search

Machine learning for hydrologic sciences: An introductory overview — [PDF] wiley.com, Full View
T Xu, F Liang - Wiley Interdisciplinary Reviews: Water, 2021 - Wiley Online Library
… **machine learning** algorithms and deep **learning** … **machine learning** in **hydrologic** sciences. Finally, we conclude with challenges associated with applying **machine learning** for **hydrologic** …
Save Cite Cited by 116 Related articles All 11 versions

Machine learning applications in hydrology — [PDF] researchgate.net
H Lange, S Sippel - Forest-water interactions, 2020 - Springer
… In this section, we provide a short overview on some important **machine learning** techniques used in **hydrology**. Given the pace of development in this field, the overview is necessarily …
Save Cite Cited by 125 Related articles All 5 versions

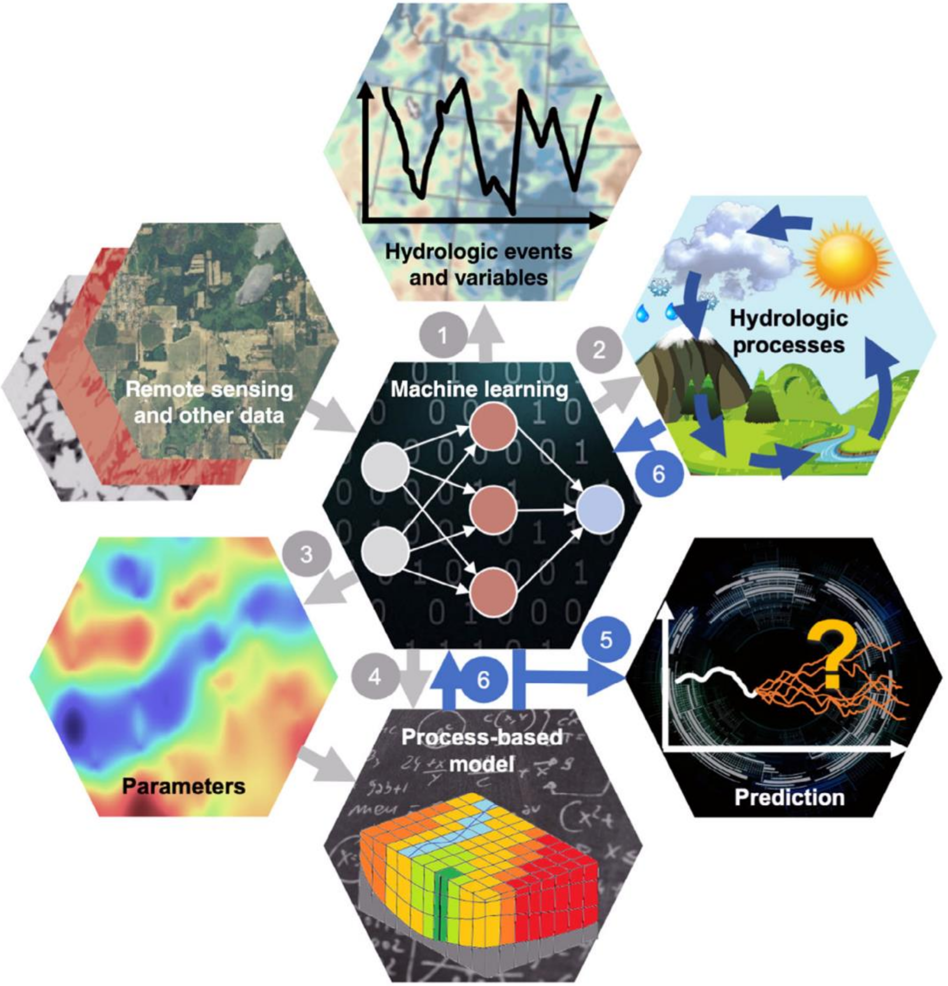
[HTML] **Ensemble machine learning paradigms in hydrology: A review** — [HTML] sciencedirect.com, ViewIt@pnnl.gov
M Zounemat-Kermani, O Batelaan, M Fadaee… - … of Hydrology, 2021 - Elsevier
… generalized stacked, in different application fields of **hydrology**. The main **hydrological** topics in this review study cover subjects such as surface **hydrology**, river water quality, rainfall-…
Save Cite Cited by 383 Related articles All 7 versions

Application of machine learning algorithms in hydrology
H Mosaffa, M Sadeghi, I Mallakpour… - Computers in earth and …, 2022 - Elsevier
… in **hydrology**. The Science Direct and Springer databases were used to investigate the number of studies in the subfield of **hydrology** with **machine learning** and **hydrology** keywords for …
Save Cite Cited by 49 Related articles

Deep learning and machine learning in hydrological processes climate change and earth systems a systematic review — [PDF] preprints.org
S Ardabili, A Mosavi, M Dehghani… - … for Sustainable Future …, 2020 - Springer
… of **machine learning** and deep **learning** methods … **machine learning** and deep **learning** are presented through a novel classification of methods. The paper concludes that deep **learning** …
Save Cite Cited by 170 Related articles All 8 versions

Broadening the use of machine learning in hydrology — [PDF] frontiersin.org, ViewIt@pnnl.gov
C Shen, X Chen, E Laloy - Frontiers in Water, 2021 - frontiersin.org
… **machine learning** in various subdomains of **hydrology**. In this Research Topic, we sought to broaden the use of **machine learning** (ML) in **hydrology** … applications of **machine learning** in …
Save Cite Cited by 70 Related articles All 9 versions

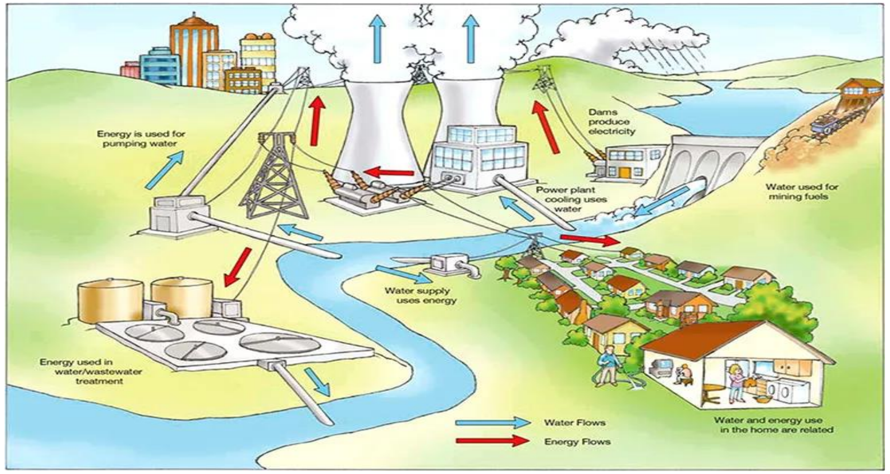
What role does hydrological science play in the age of machine learning? — [PDF] wiley.com, Full View
GS Nearing, F Kratzert, AK Sampson… - Water Resources …, 2021 - Wiley Online Library
… **learning** to rainfall-runoff simulation indicate that there is significantly more information in large-scale **hydrological** … interest in **machine learning** in the **hydrological** sciences community, …



Xu, Tianfang, and Feng Liang. "Machine learning for hydrologic sciences: An introductory overview." Wiley Interdisciplinary Reviews: Water 8.5 (2021): e1533.

Streamflow prediction

Water resources



Source: medium

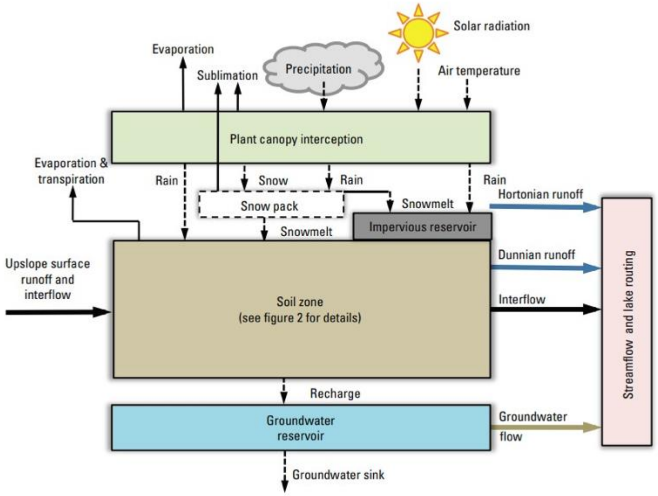
Drought

Source: Phys.org

Flooding

Source: CNN

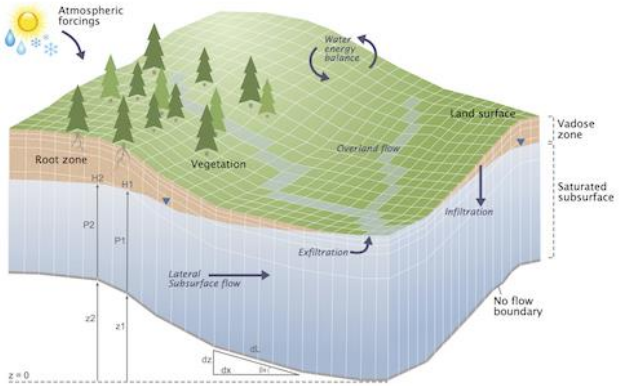
Process-based hydrological models



Source: PRMS model

<u>Bucket model:</u>

- Simplified process
- Lack of accuracy

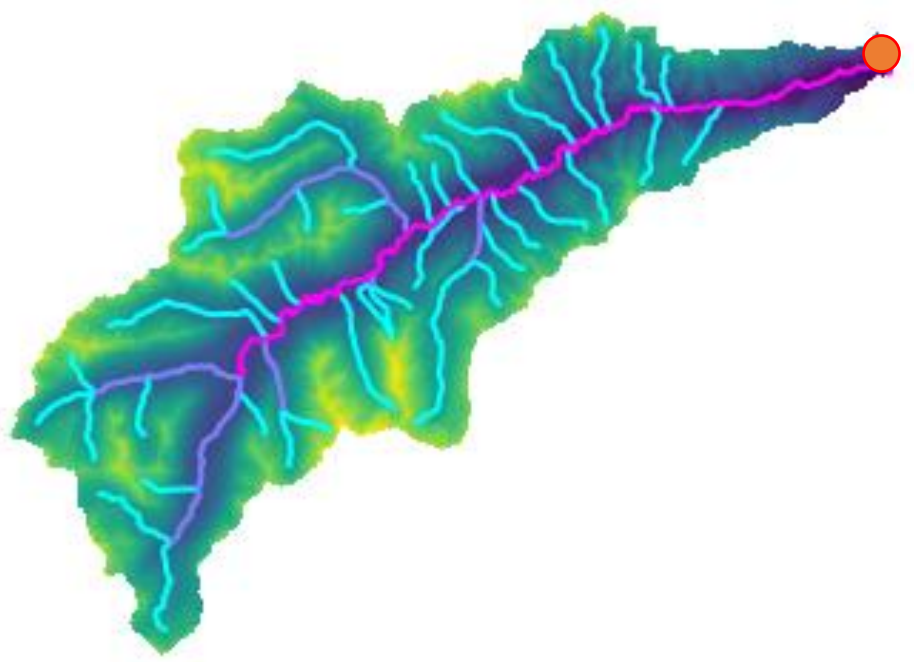


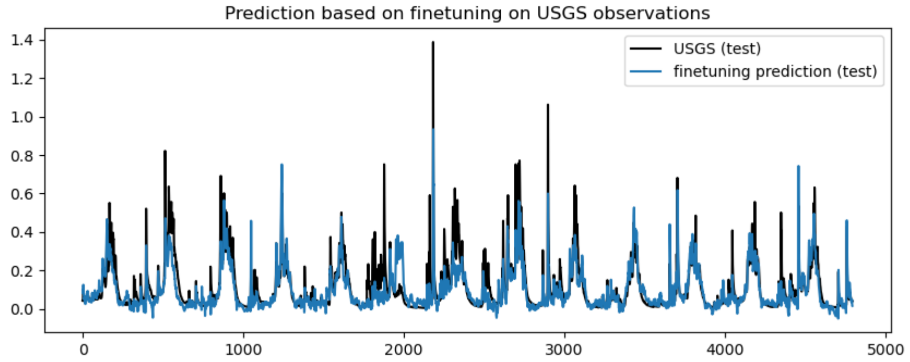
Source: ParFlow model

<u>Fully integrated model:</u>

- Physically accurate
- Computationally expensive

ML application in streamflow prediction





One location (point-based)

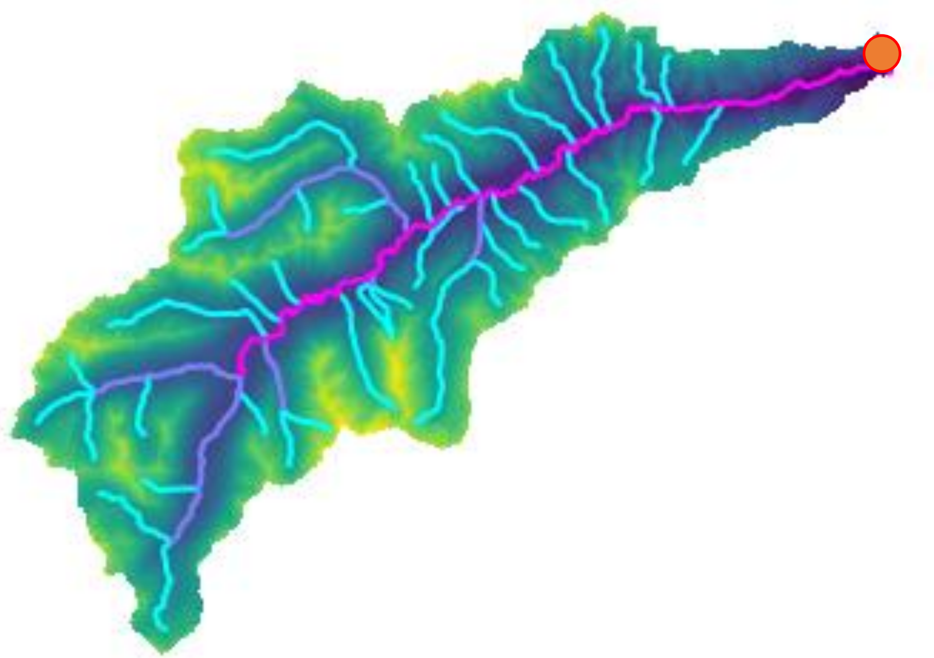
Multiple locations (point-based)

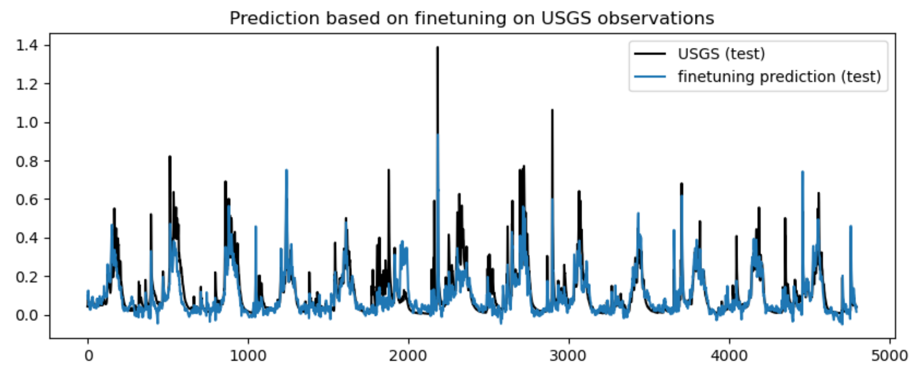
Streamflow prediction

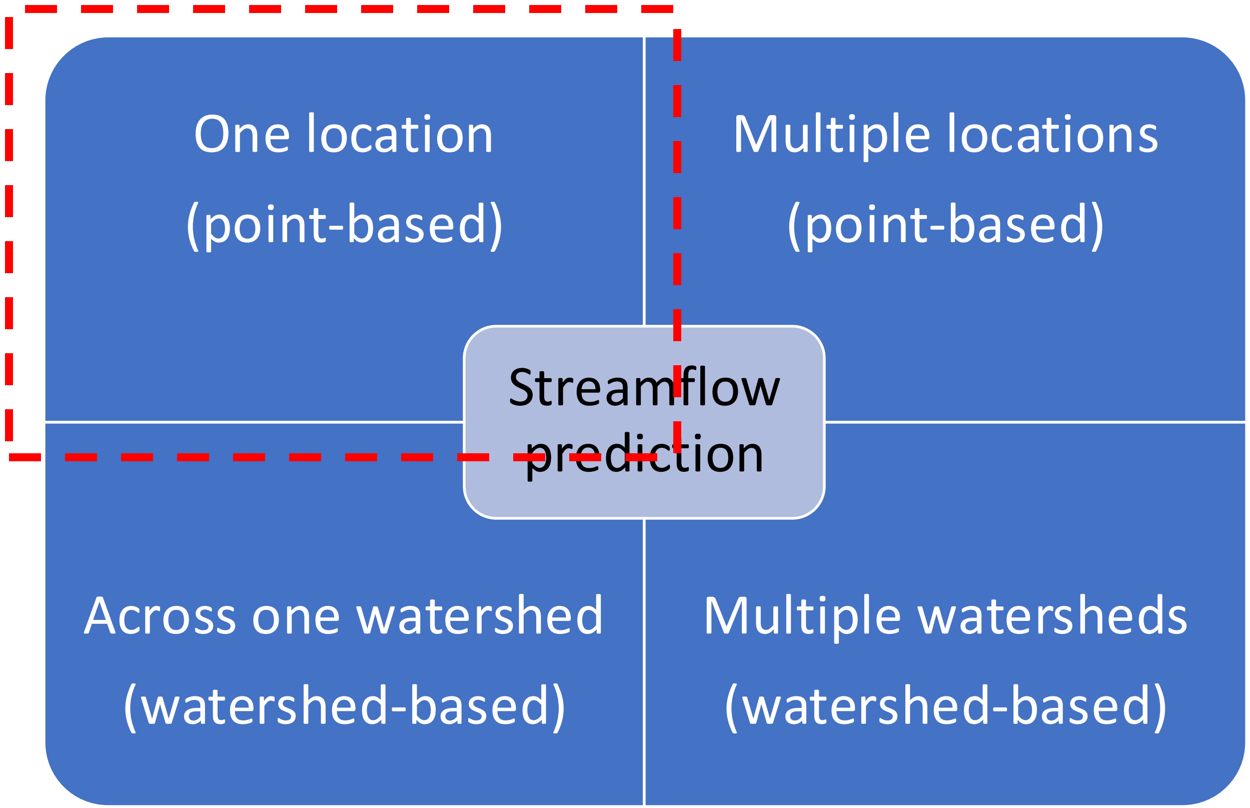
Across one watershed (watershed-based)

Multiple watersheds (watershed-based)

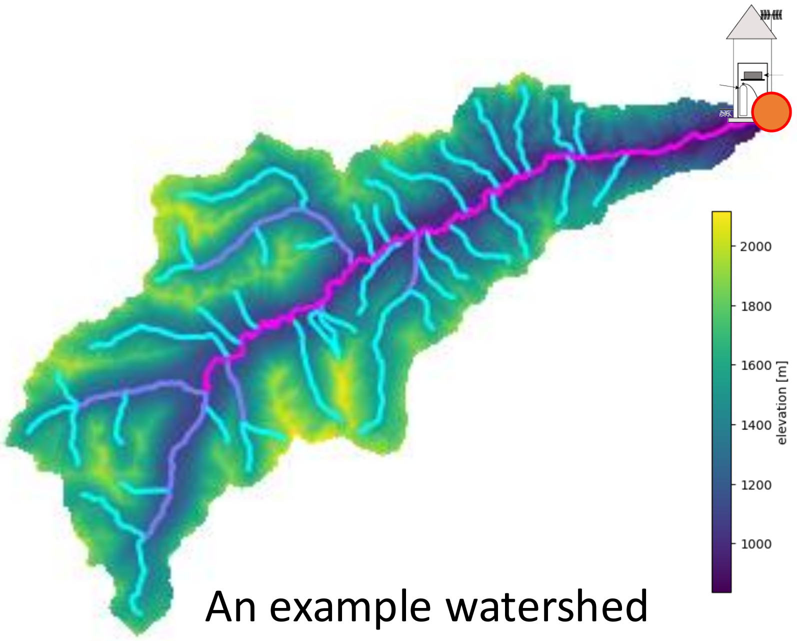
ML application in streamflow prediction







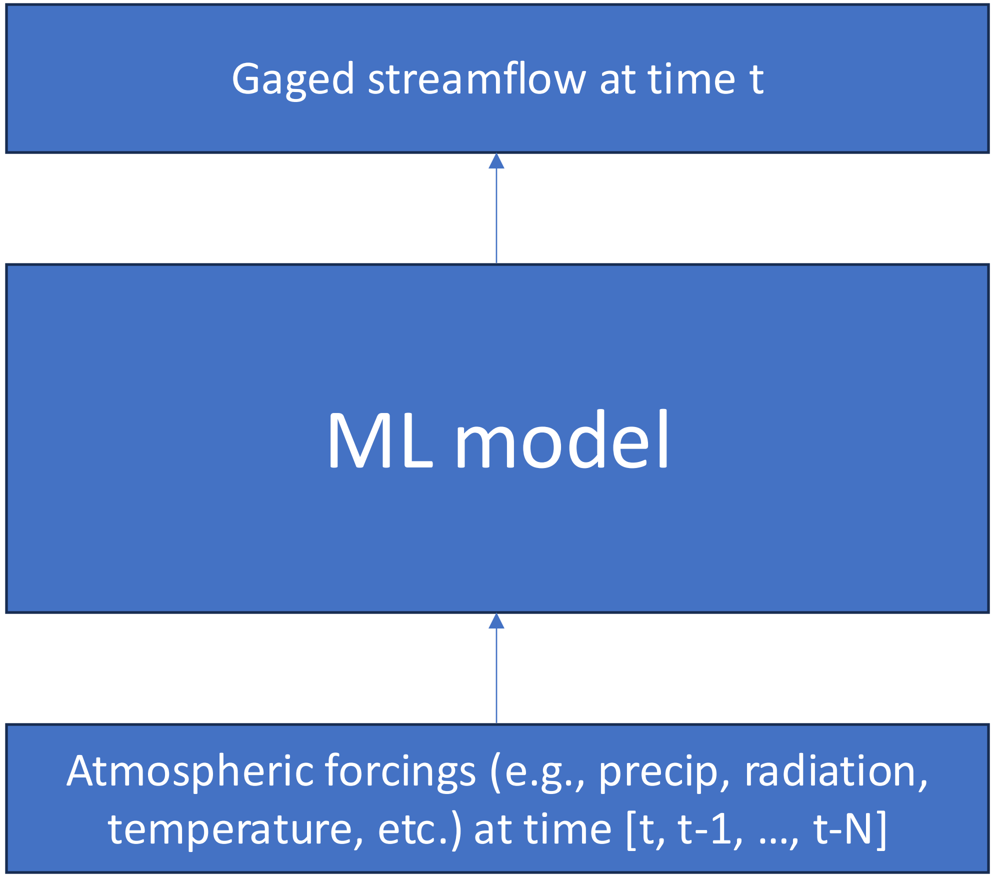
Streamflow prediction at one gauged location:



An example watershed

Location of interest

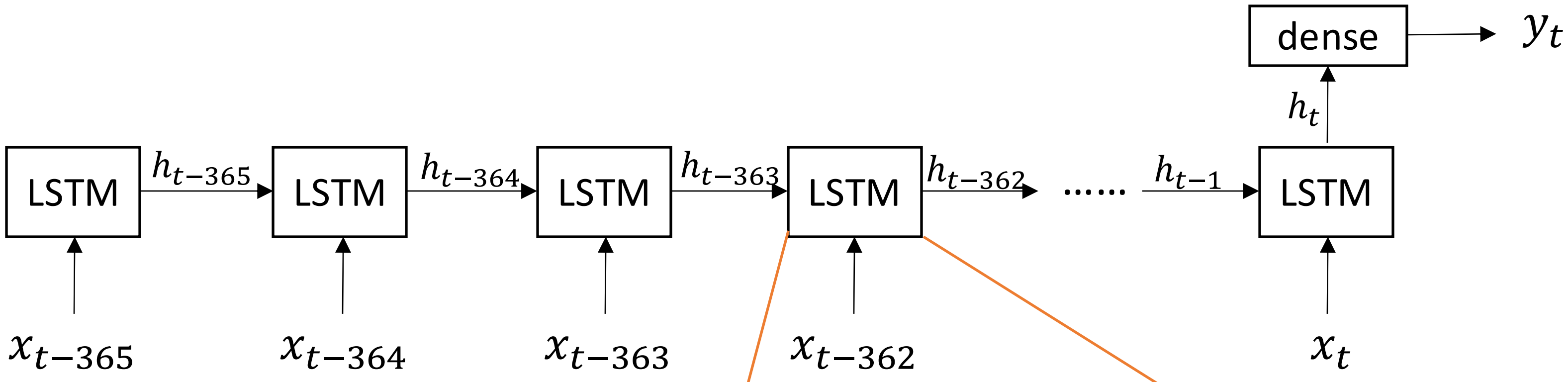
gaging station



Streamflow prediction at one gauged location: Long Short-Term Memory model (LSTM)

Input x:

- precipitation
- temperature
- solar radiation
- vapor pressure



Output y: runoff

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

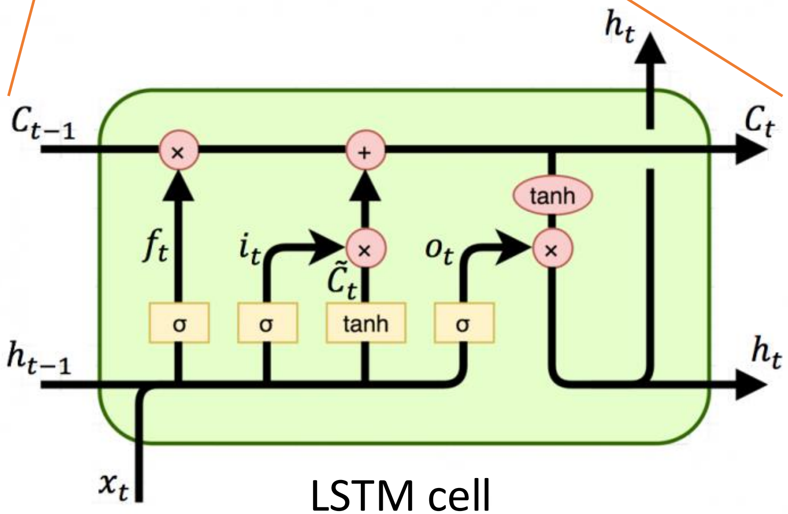
Input gate $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$

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

Cell input $\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$

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

Hidden state $h_t = o_t \odot \tanh(C_t)$



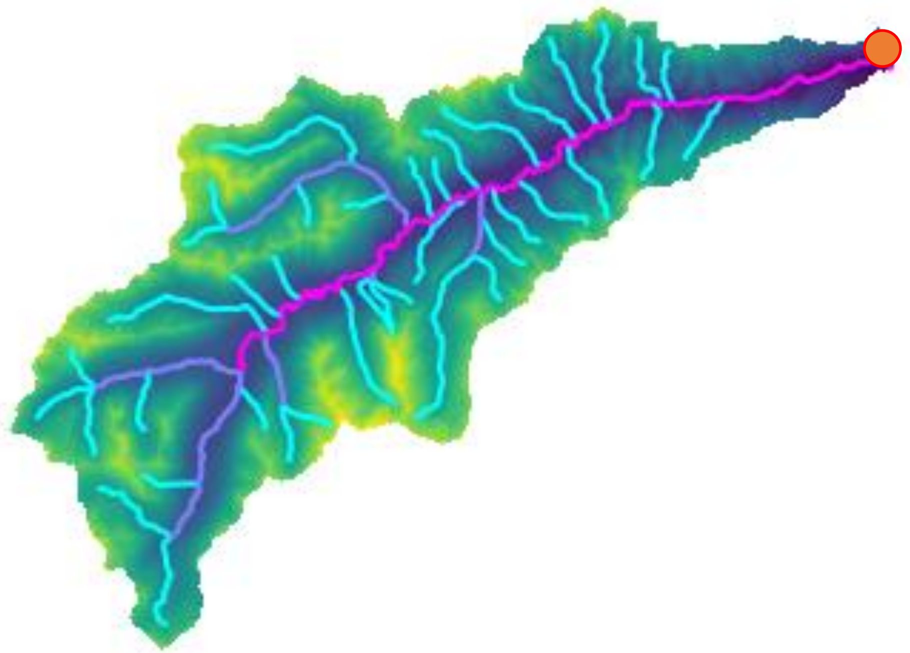
LSTM cell

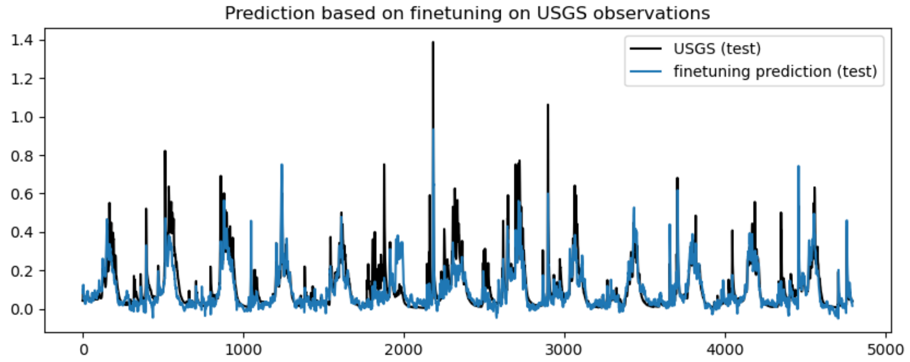
Kratzert, Frederik, et al. "Rainfall–runoff modelling using long short-term memory (LSTM) networks." Hydrology and Earth System Sciences 22.11 (2018): 6005-6022.

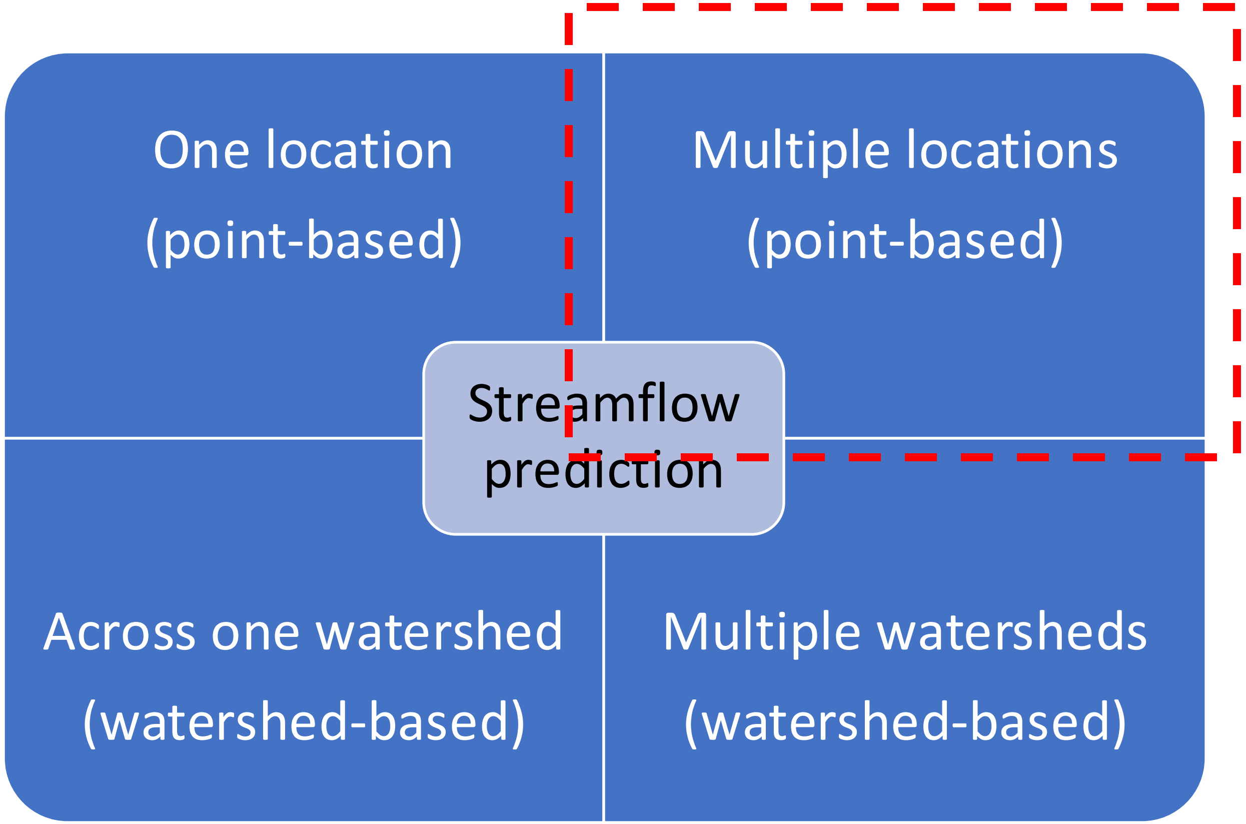
Streamflow prediction at one gauged location: Long Short-Term Memory model (LSTM)

Jupyter Notebook

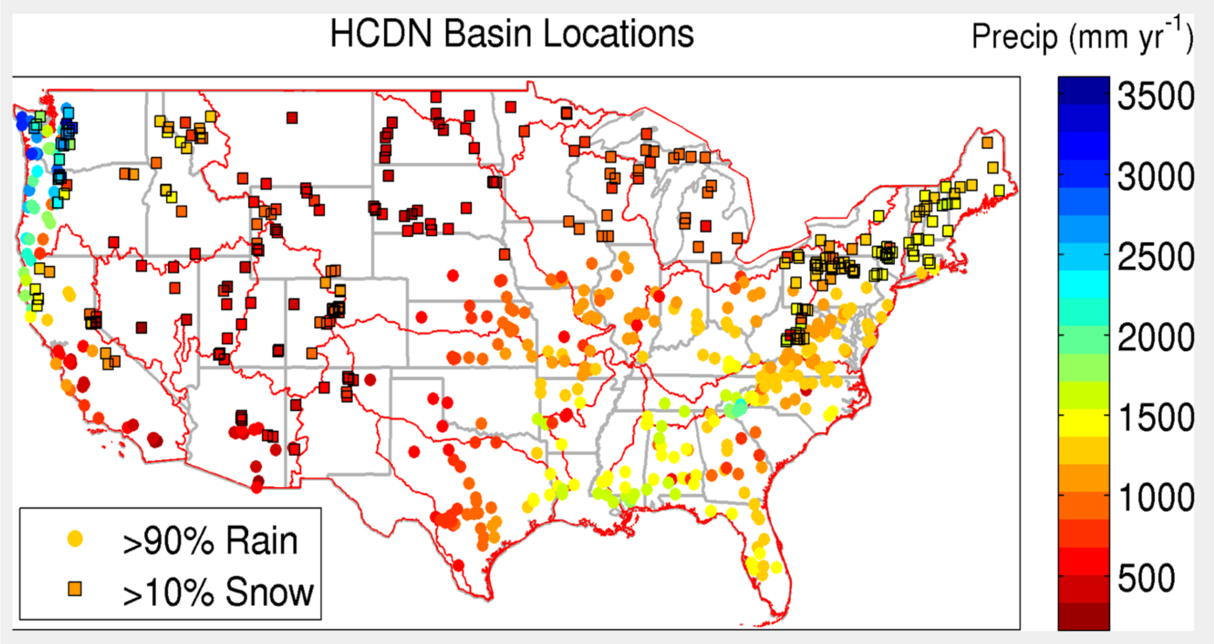
ML application in streamflow prediction



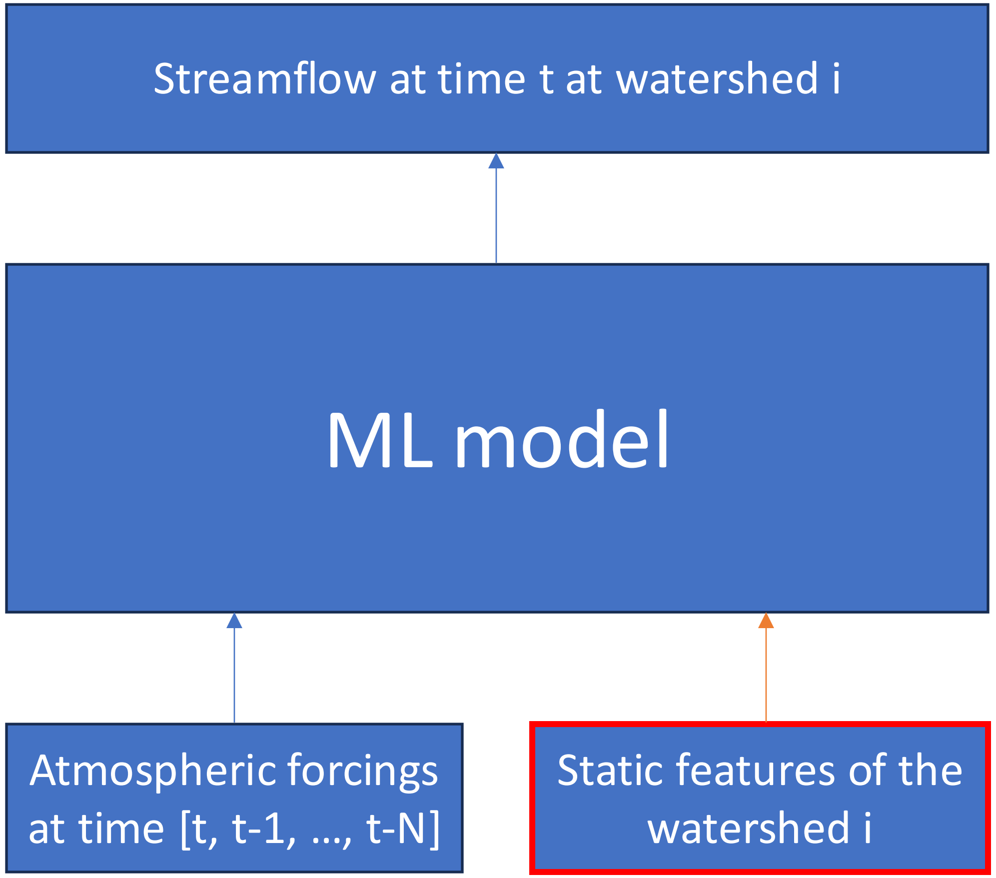




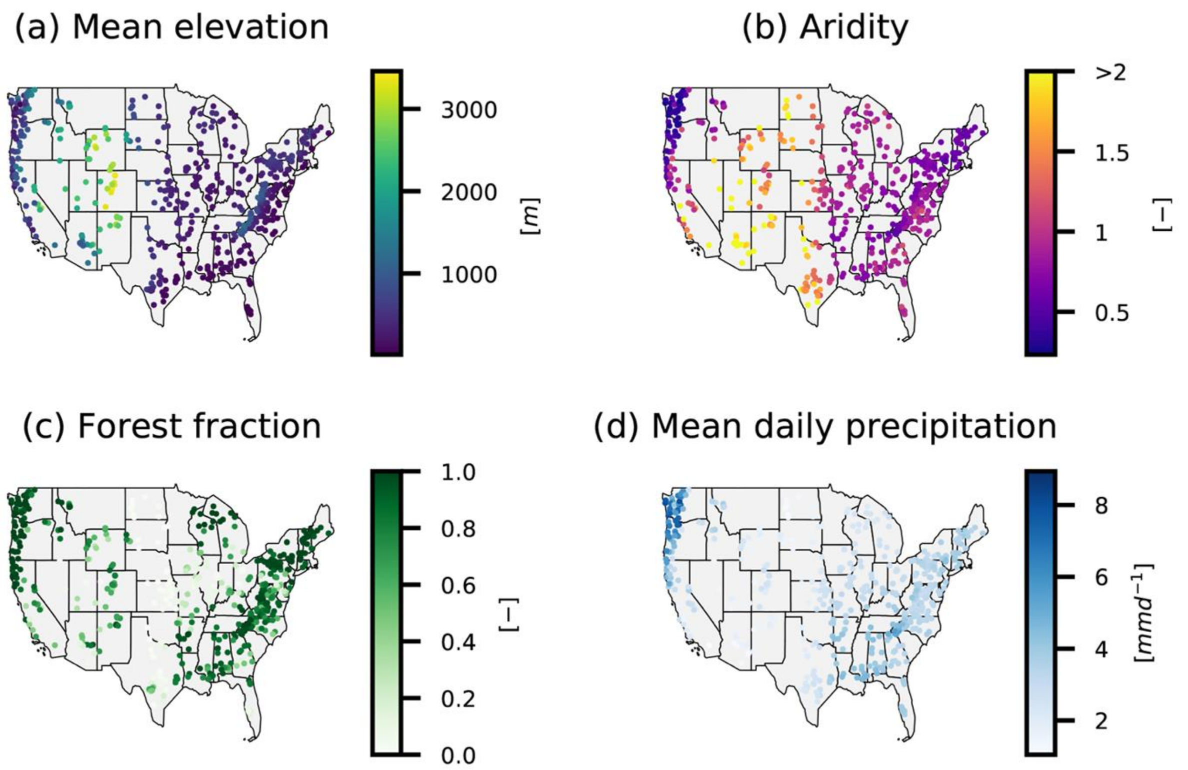
Streamflow prediction at multiple gauged locations: LSTM

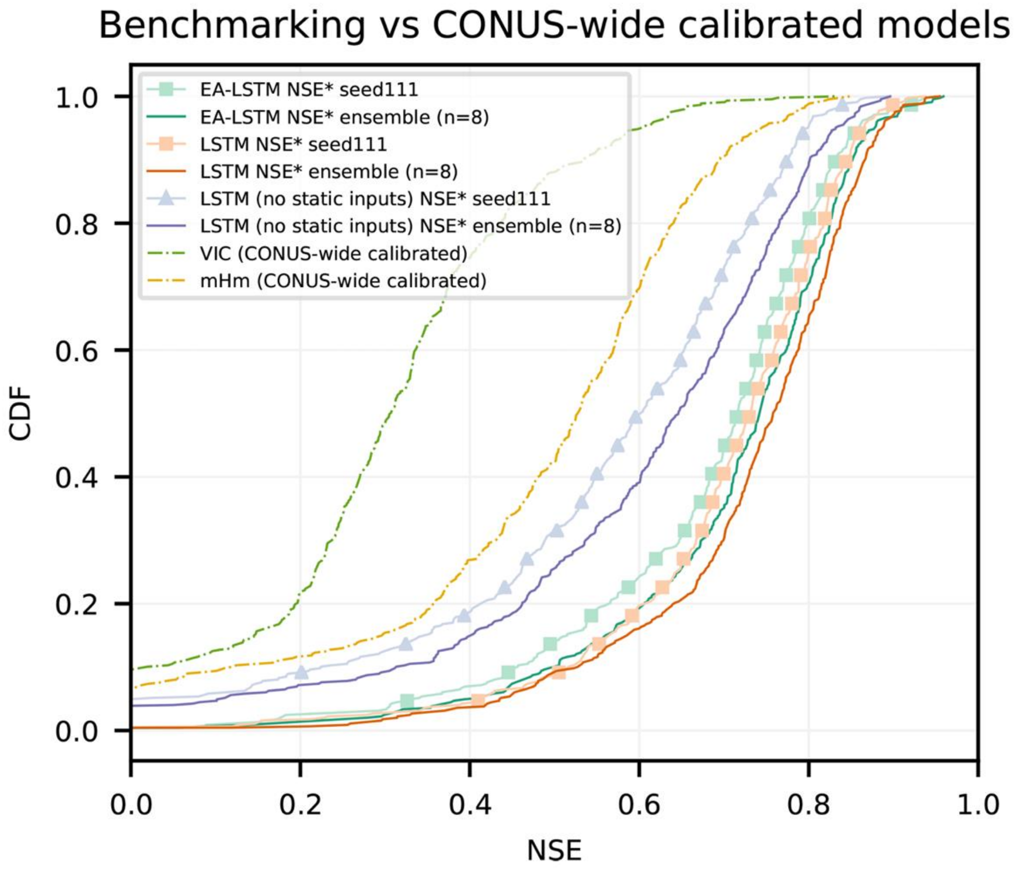


Source: UCAR - CAMELS: Large-Sample Hydrometeorological Dataset



Streamflow prediction at multiple gauged locations: LSTM

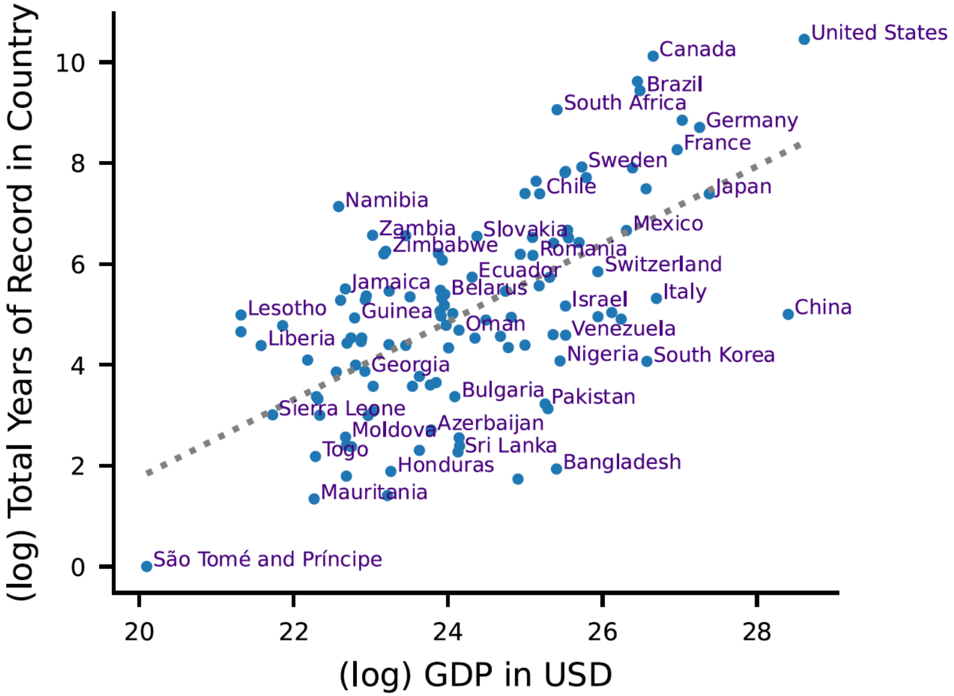




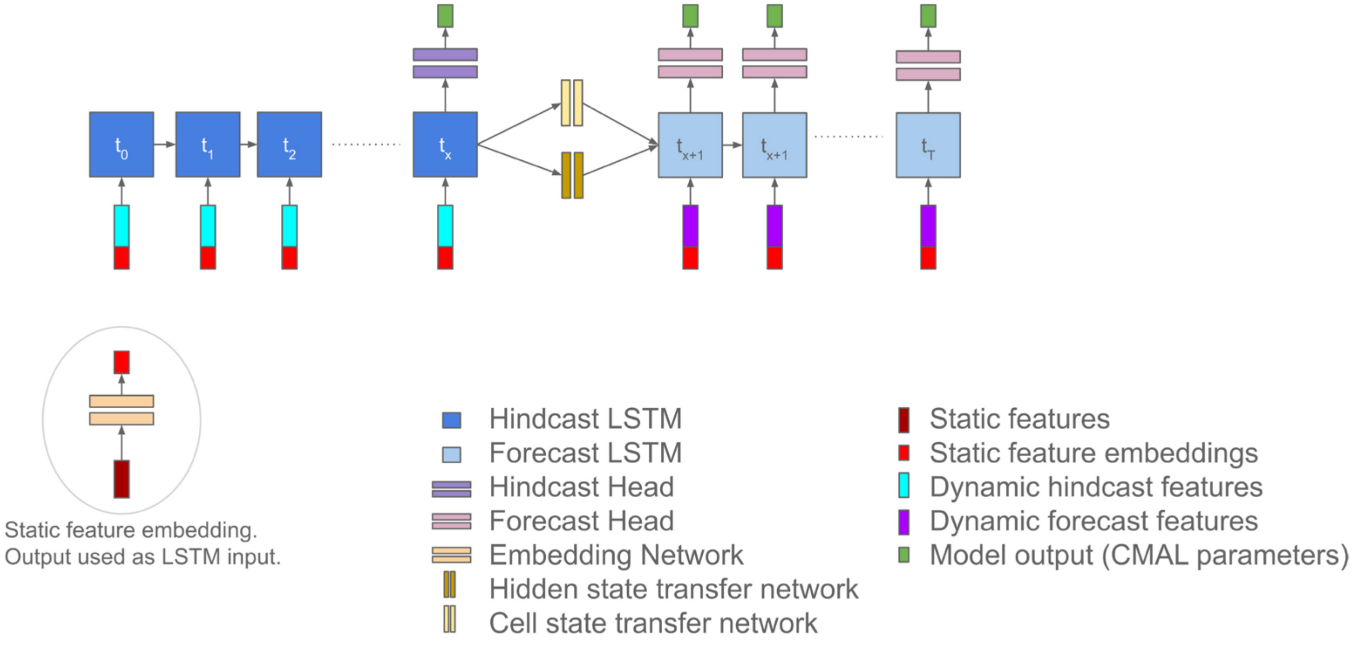
VIC, mHm: two hydrological models
NSE: Nash–Sutcliffe model efficiency coefficient
CDF: Cumulative distribution function

Kratzert, Frederik, et al. "Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets." Hydrology and Earth System Sciences 23.12 (2019): 5089-5110.

Streamflow prediction at ungauged locations: LSTM & transfer learning



Extended Data Fig. 1 | Streamflow data availability correlates with national GDP. There is a log-log correlation (r=0.611; N=117) between national Gross Domestic Product (GDP) and the total number of years worth of daily streamflow data available in a country from the Global Runoff Data Center. GDP data are sourced from The World Bank[46].



Architecture of the LSTM-based forecast model developed for this project

Nearing, Grey, et al. "Global prediction of extreme floods in ungauged watersheds." *Nature* 627.8004 (2024): 559-563.

Streamflow prediction at ungauged locations: LSTM & transfer learning

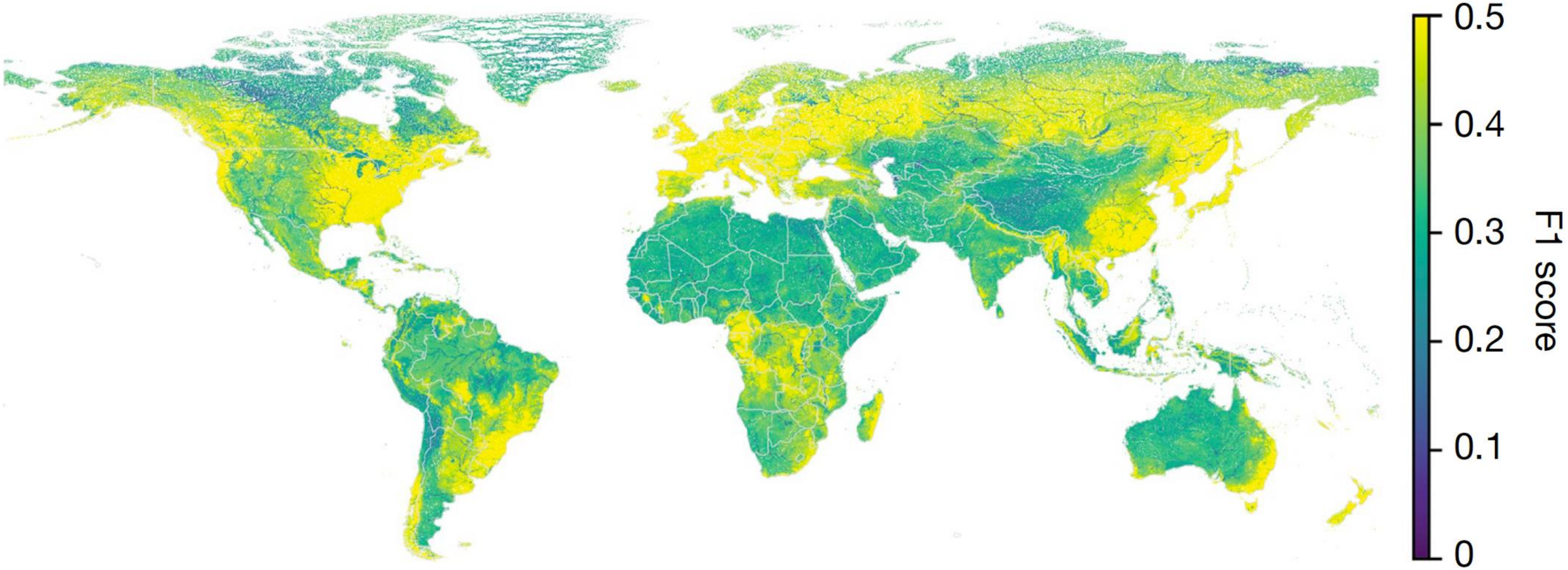
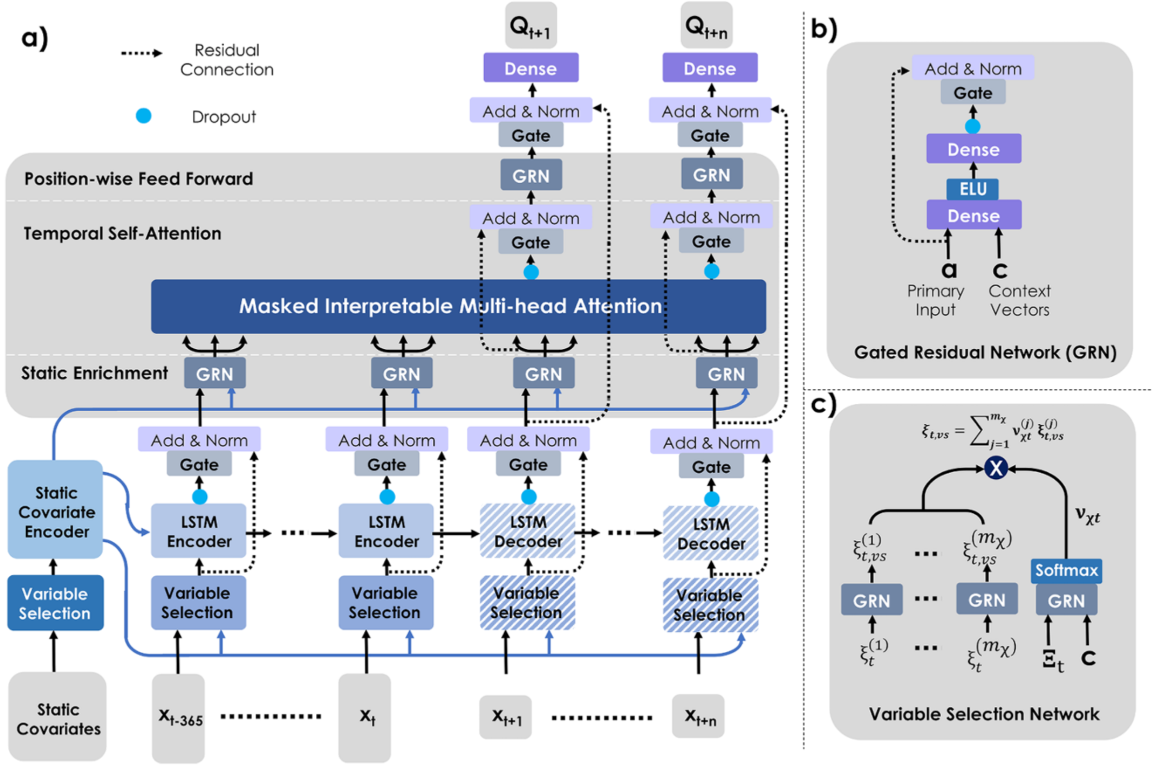


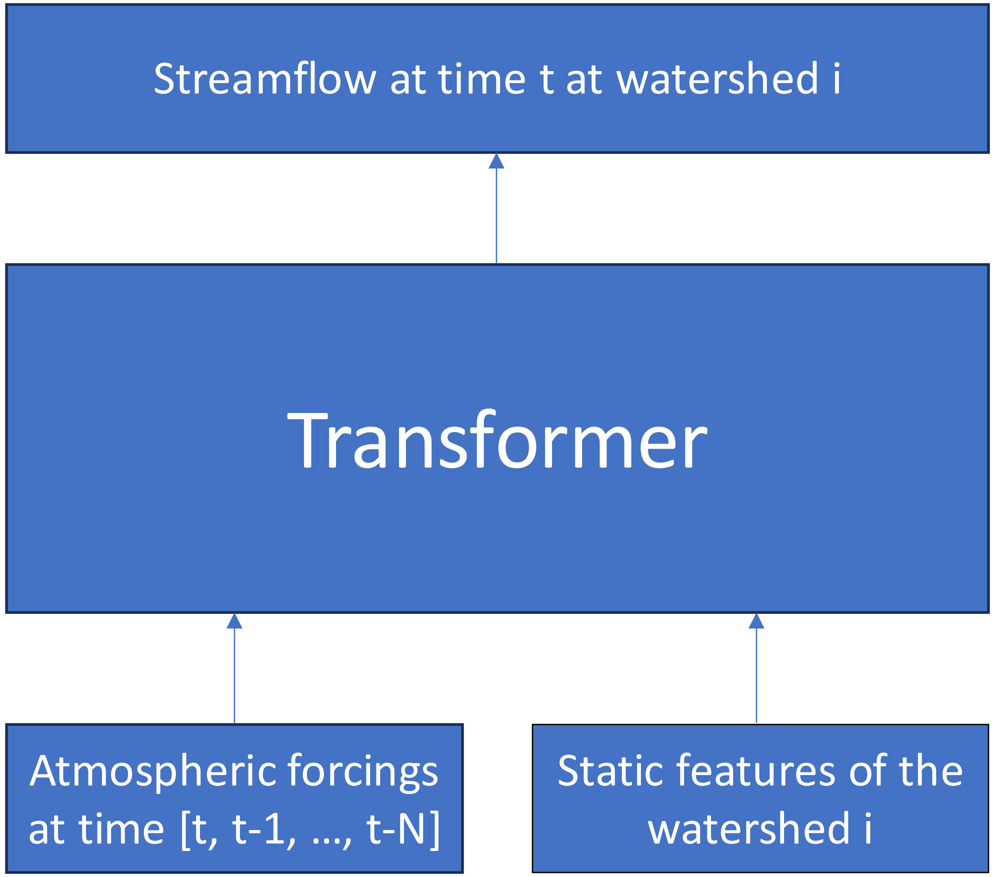
Fig. 6 | Global predicted skill. This map shows predictions of 2-year return period F1 scores over 1.03 million HydroBASINS level-12 watersheds for the AI forecast model. Basemap from GeoPandas[34].

F1 score: the harmonic mean of precision and recall

Nearing, Grey, et al. "Global prediction of extreme floods in ungauged watersheds." *Nature* 627.8004 (2024): 559-563.

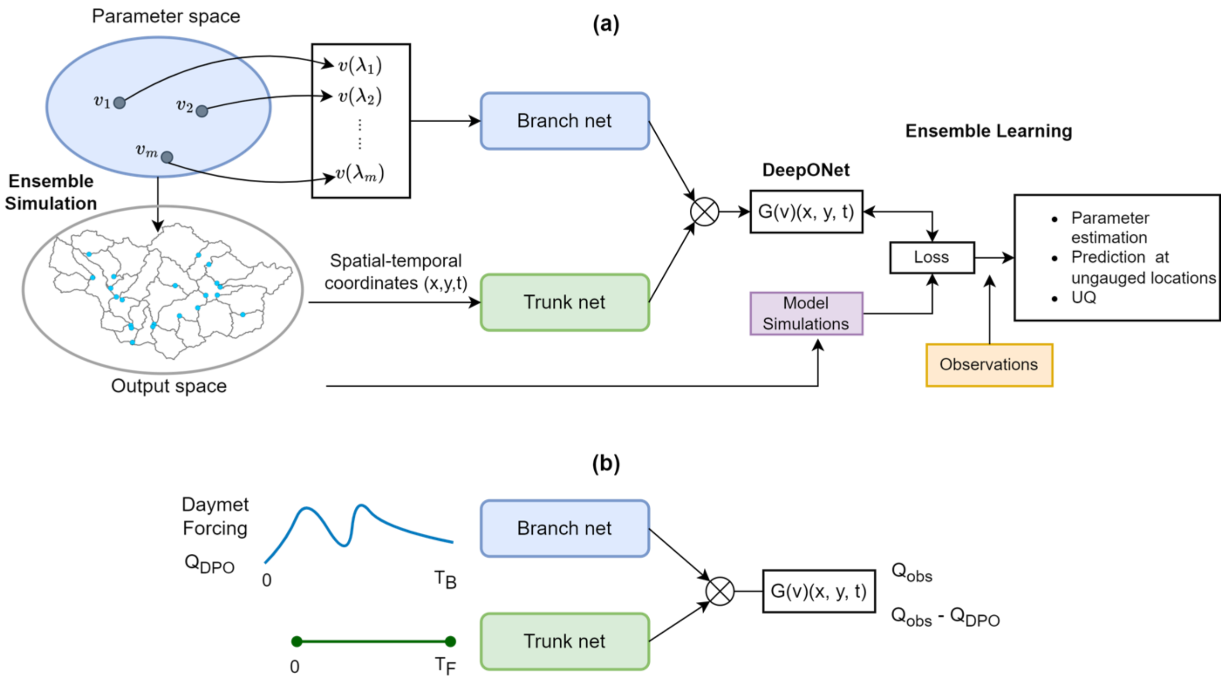
Streamflow prediction at multiple locations: Other ML models

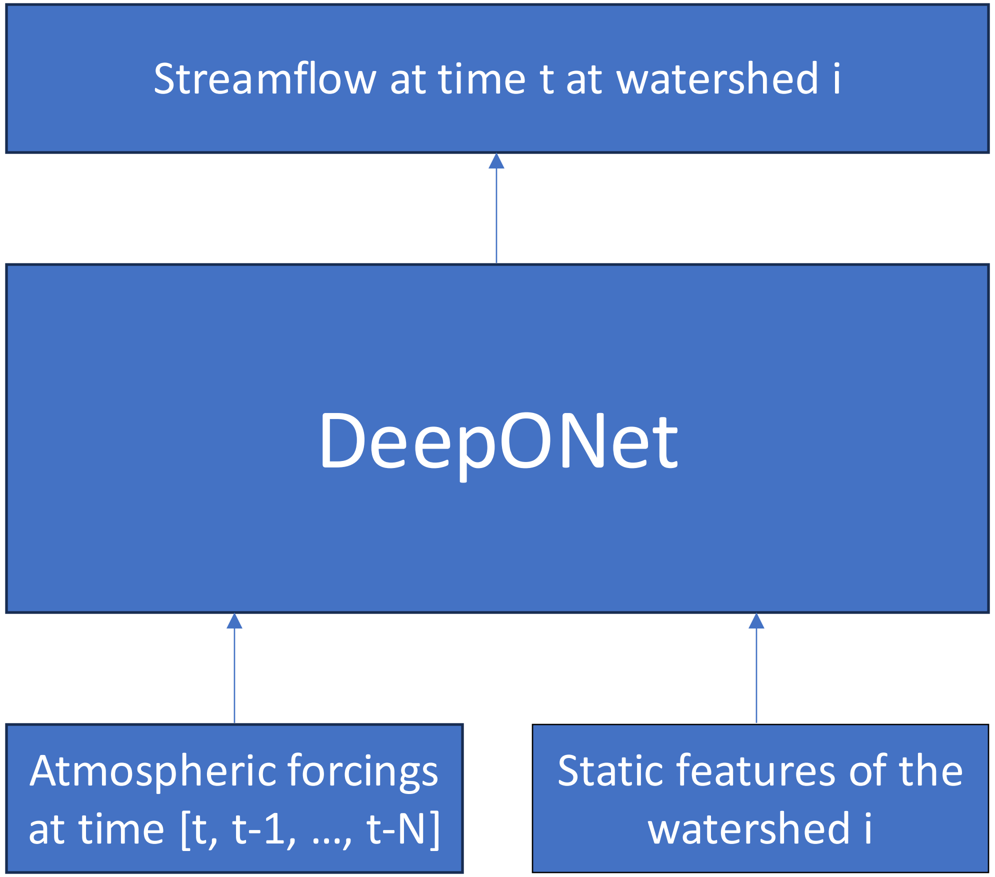




Koya, Sinan Rasiya, and Tirthankar Roy. "Temporal Fusion Transformers for streamflow Prediction: Value of combining attention with recurrence." *Journal of Hydrology* 637 (2024): 131301.

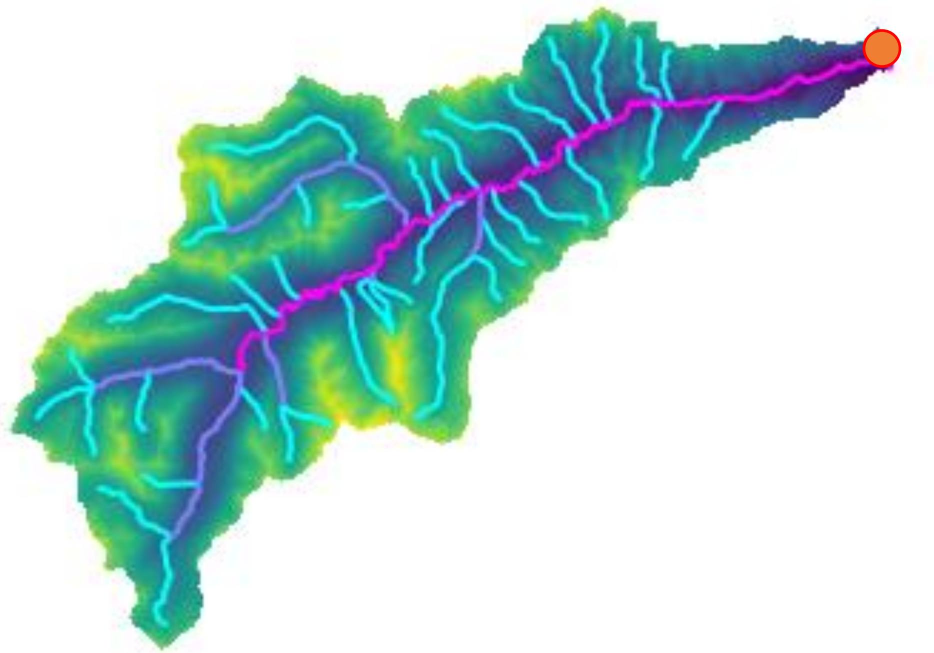
Streamflow prediction at multiple locations: Other ML models

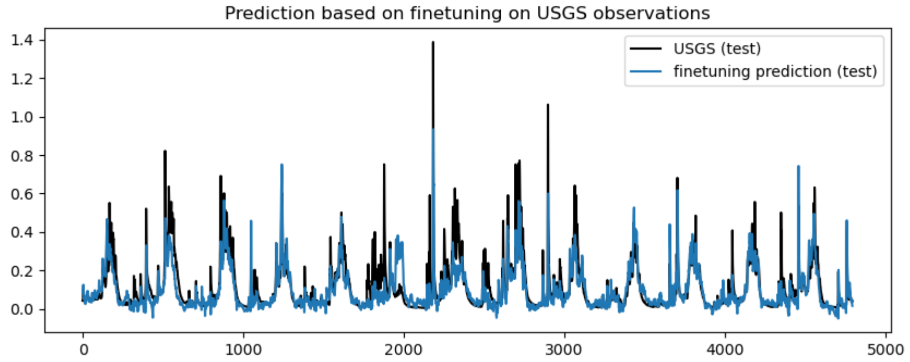


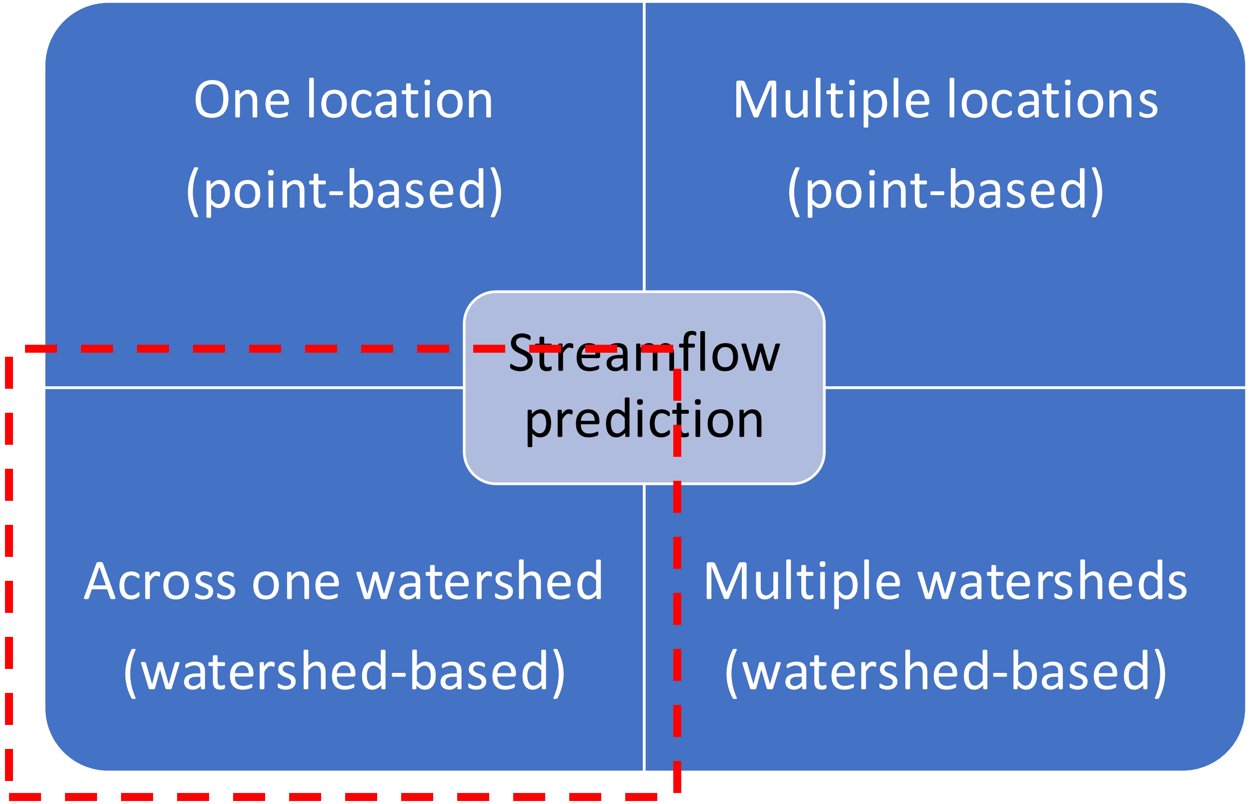


Sun, Alexander, Peishi Jiang, Pin Shuai, and Xingyuan Chen "Bridging hydrological ensemble simulation and learning using deep neural operators." *Water Resources Research* (2024).

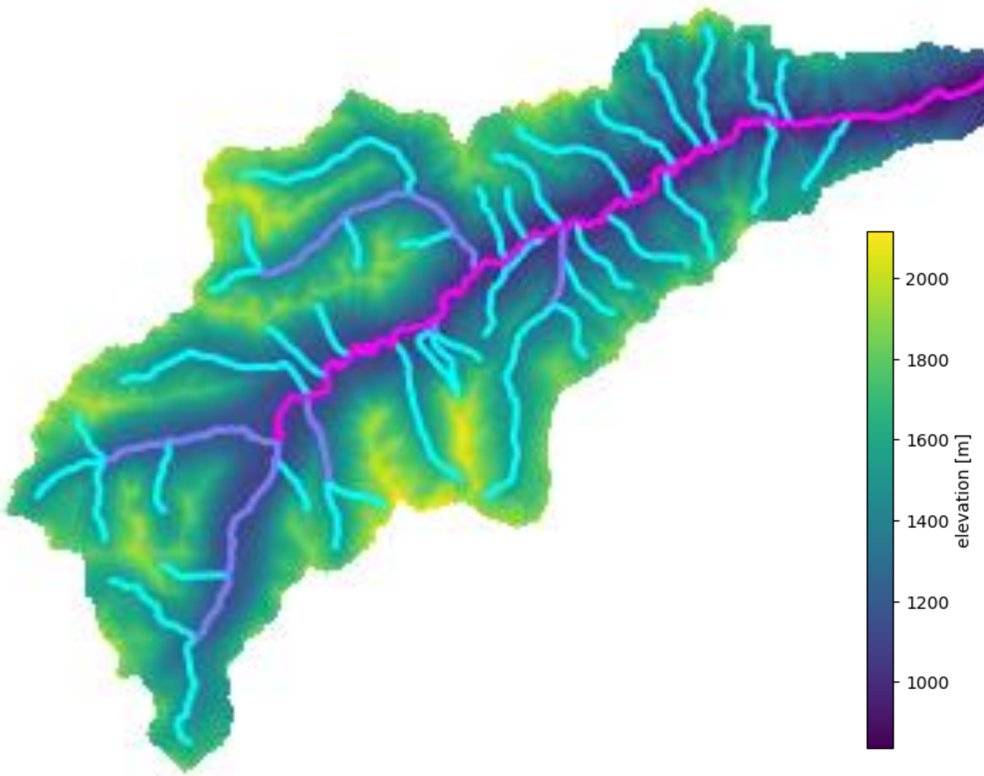
ML application in streamflow prediction





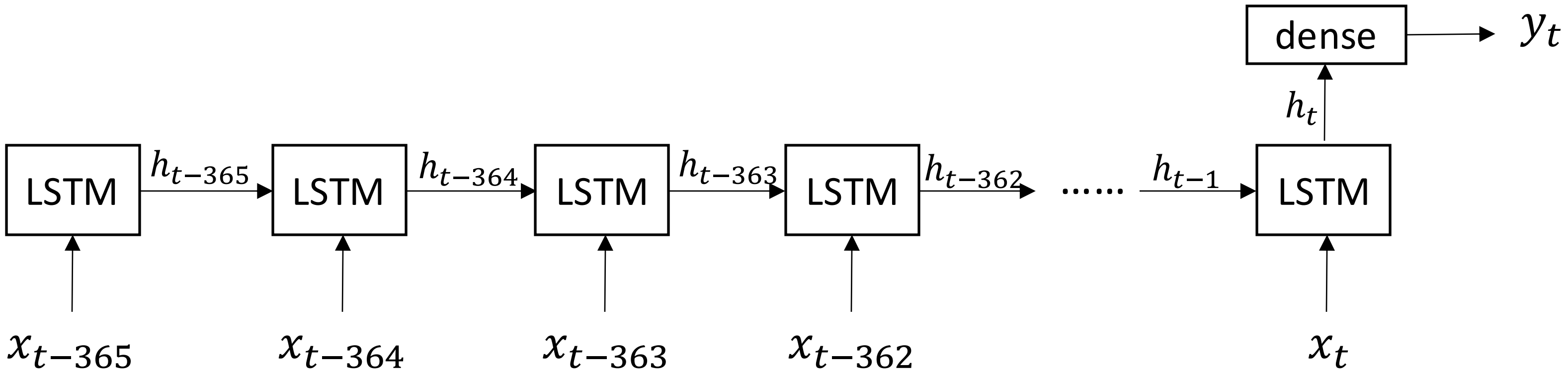


Streamflow prediction across a watershed



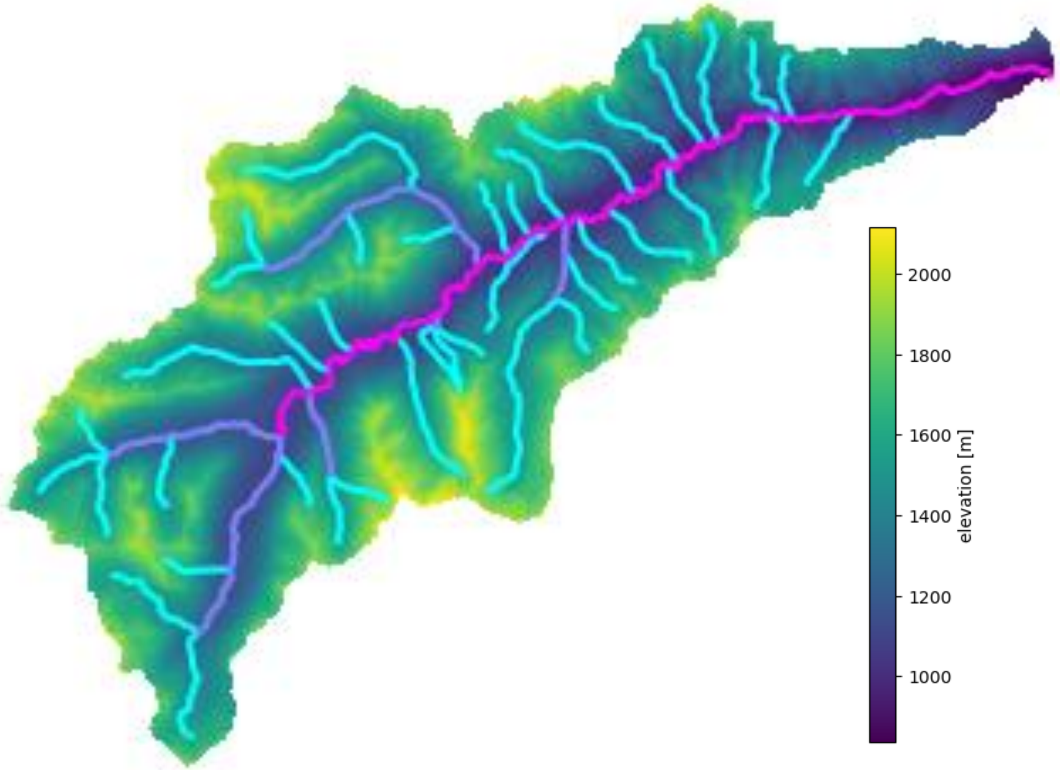
An example watershed

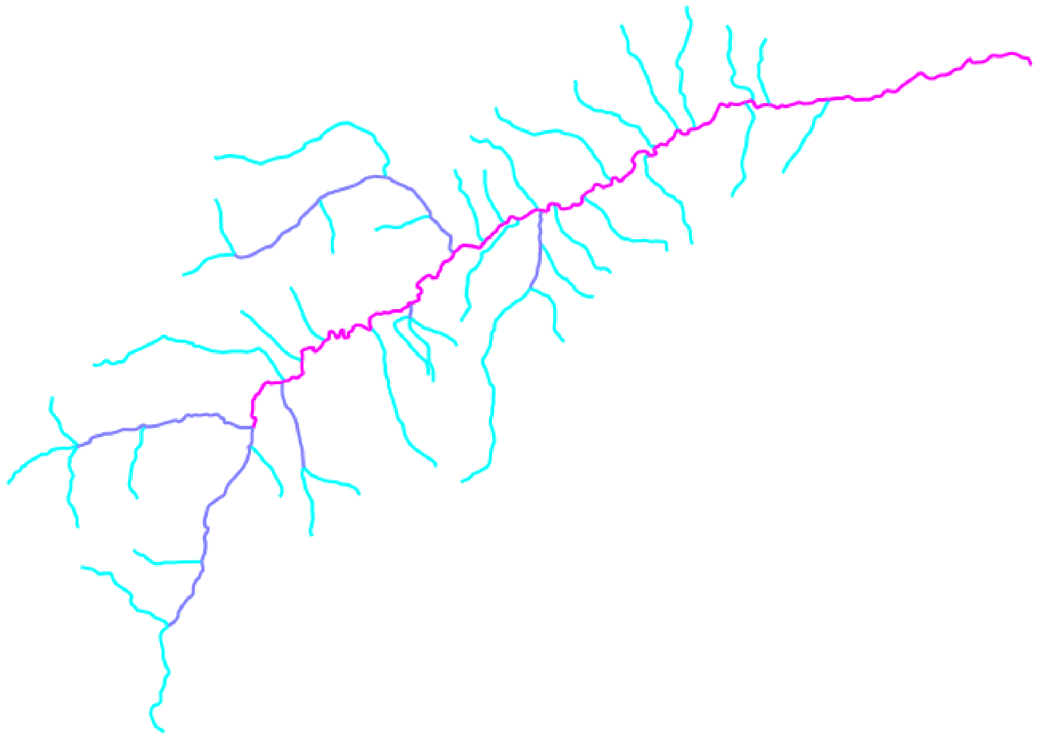
Streamflow prediction across a watershed: Issues with LSTM



LSTM does not explicitly capture the physical connections between river reaches

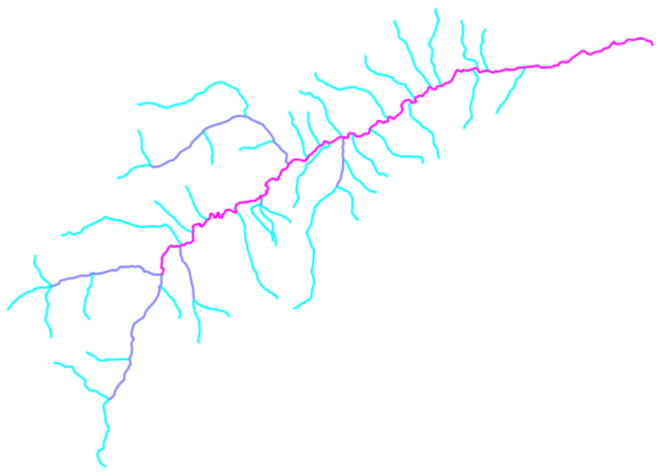
Streamflow prediction across a watershed: river network as a graph

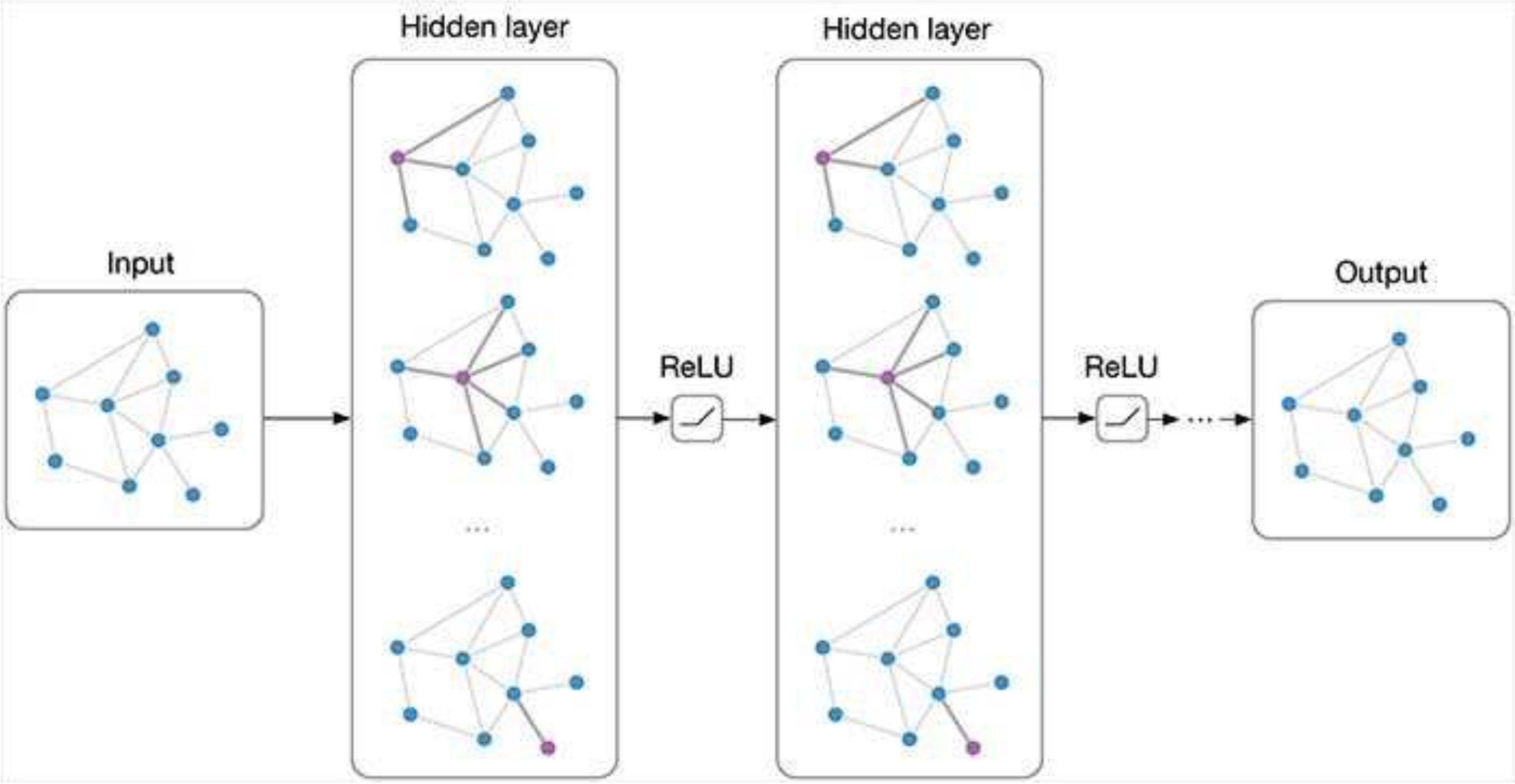




An example watershed

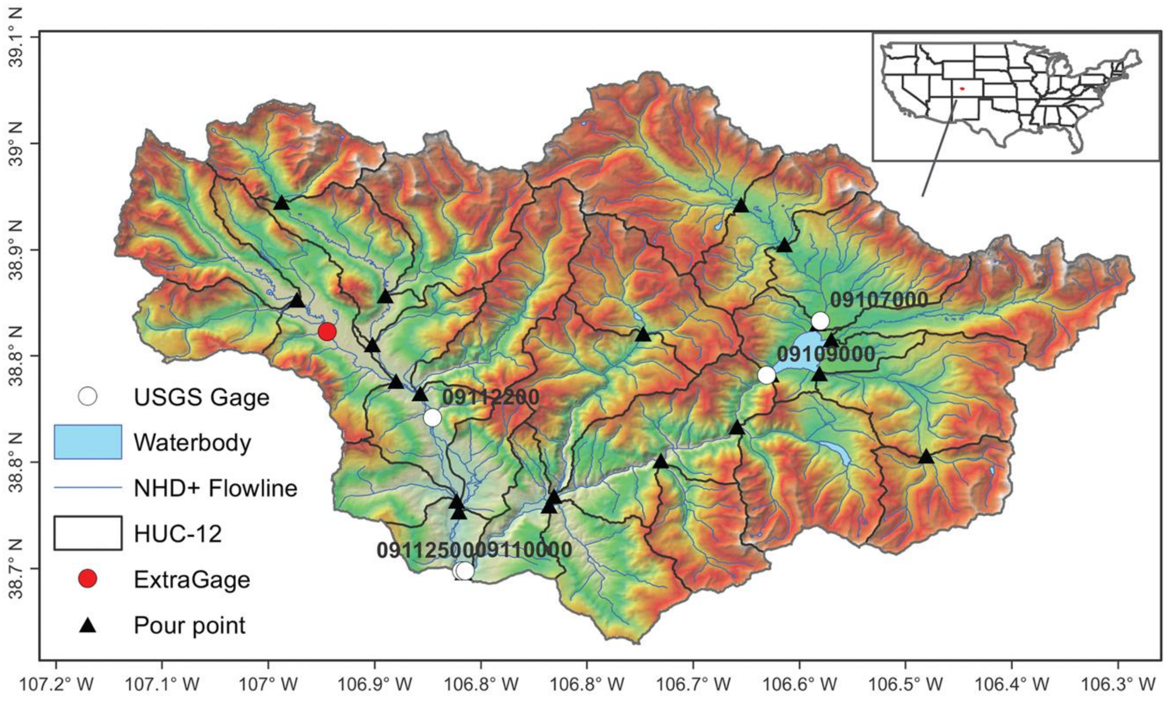
Streamflow prediction across a watershed: Graph Neural Network

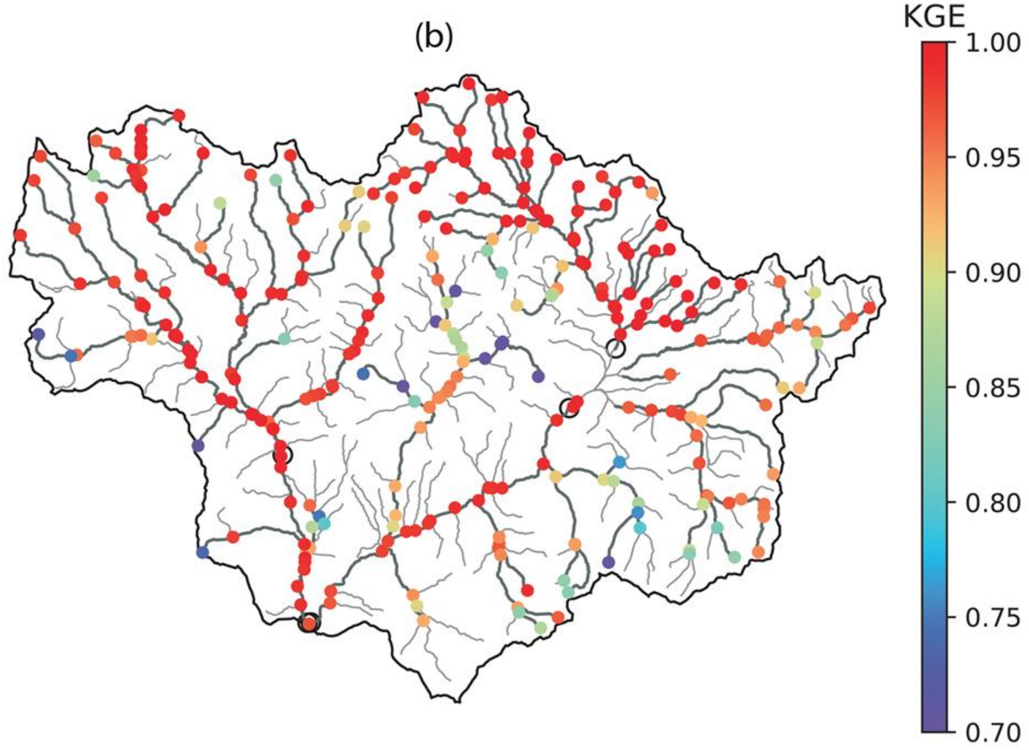




Kipf, Thomas N., and Max Welling. "Semi-supervised classification with graph convolutional networks." *arXiv preprint arXiv:1609.02907* (2016).

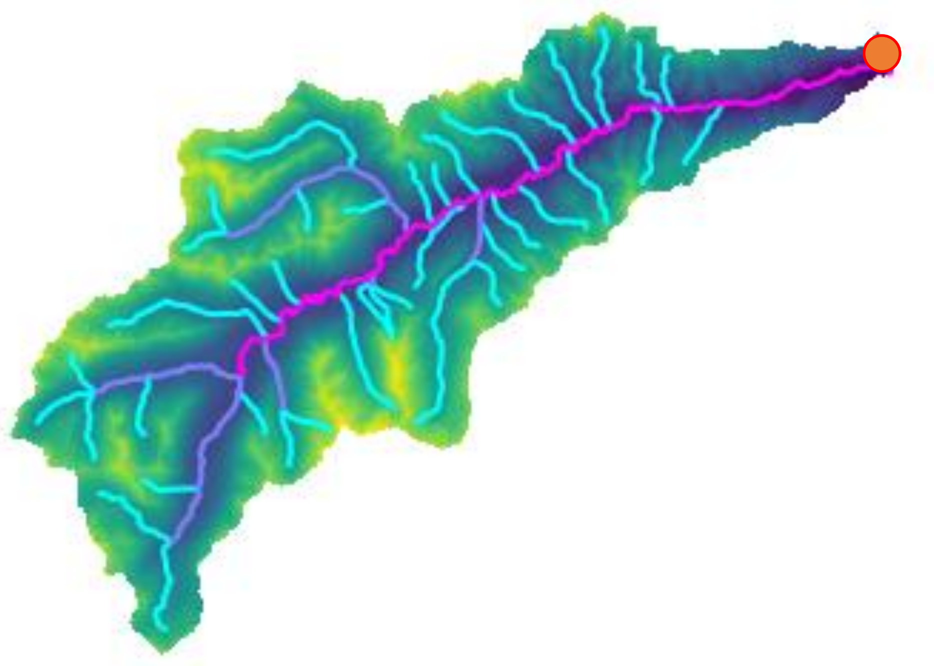
Streamflow prediction across a watershed: Graph Neural Network

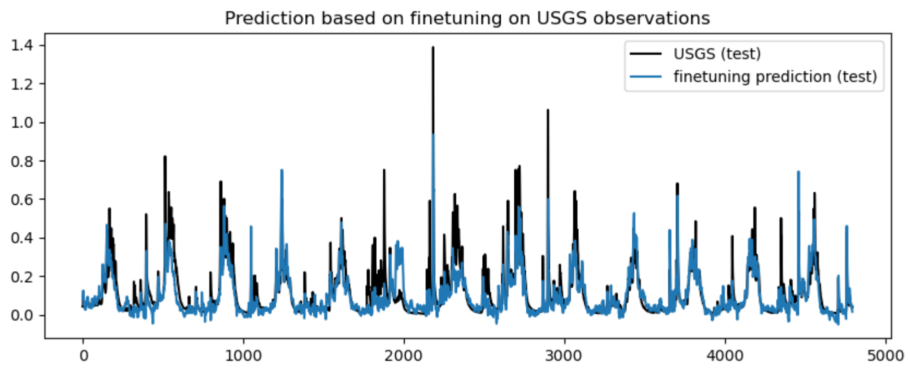


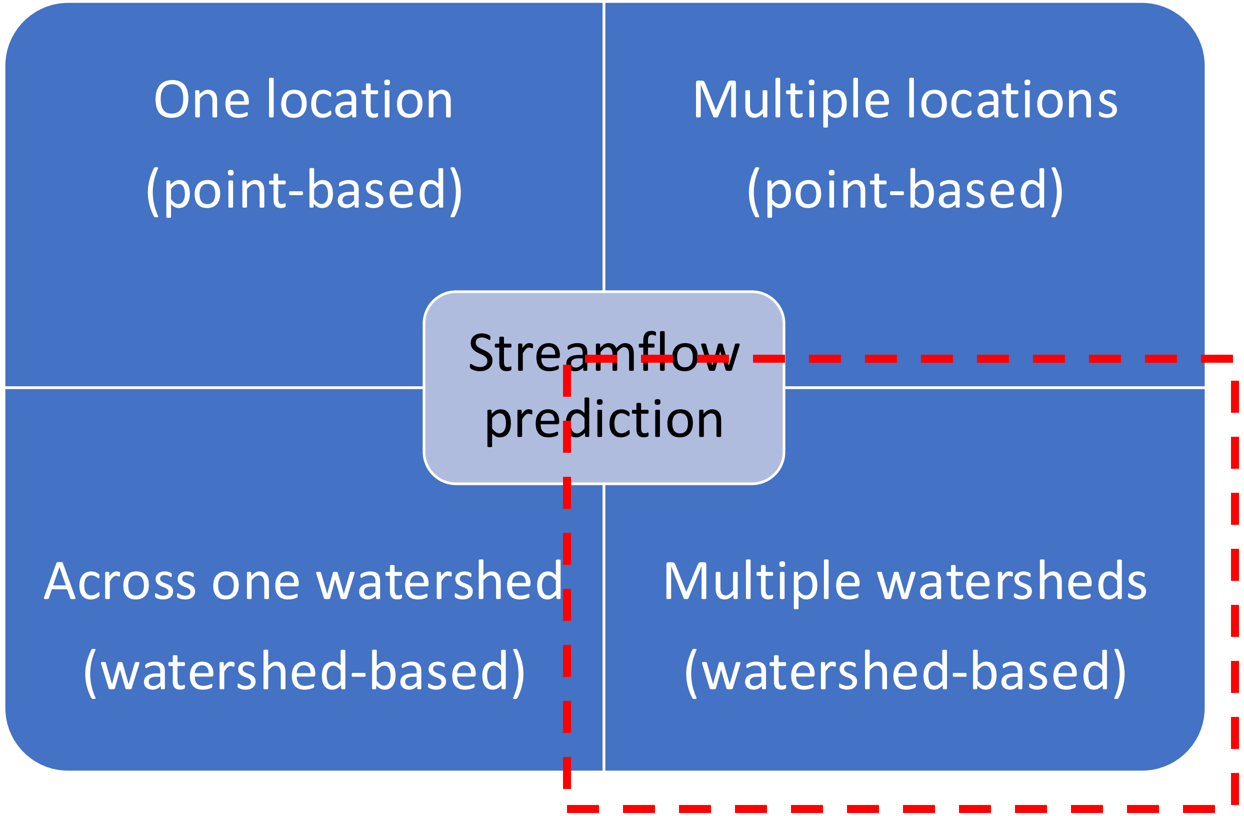


Sun, Alexander Y., Peishi Jiang, Zong-Liang Yang, Yangxinyu Xie, and Xingyuan Chen. "A graph neural network (GNN) approach to basin-scale river network learning: the role of physics-based connectivity and data fusion." *Hydrology and Earth System Sciences* 26, no. 19 (2022): 5163-5184.

ML application in streamflow prediction

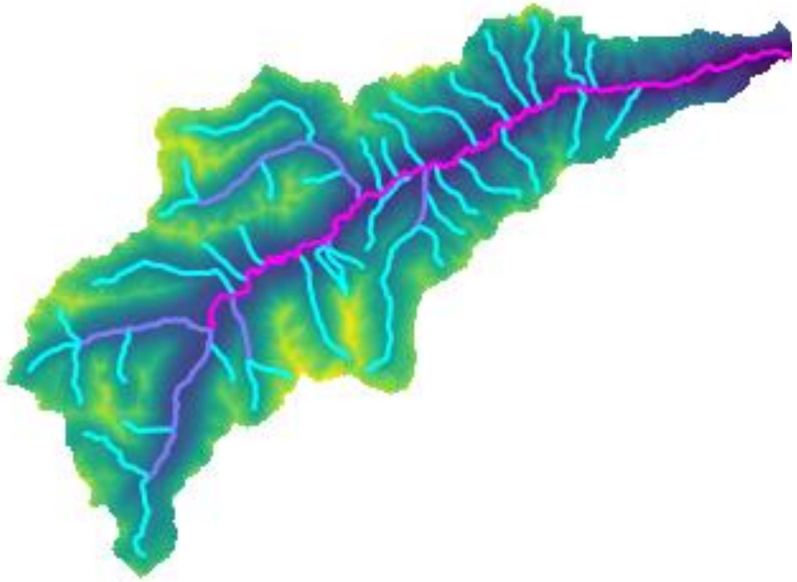


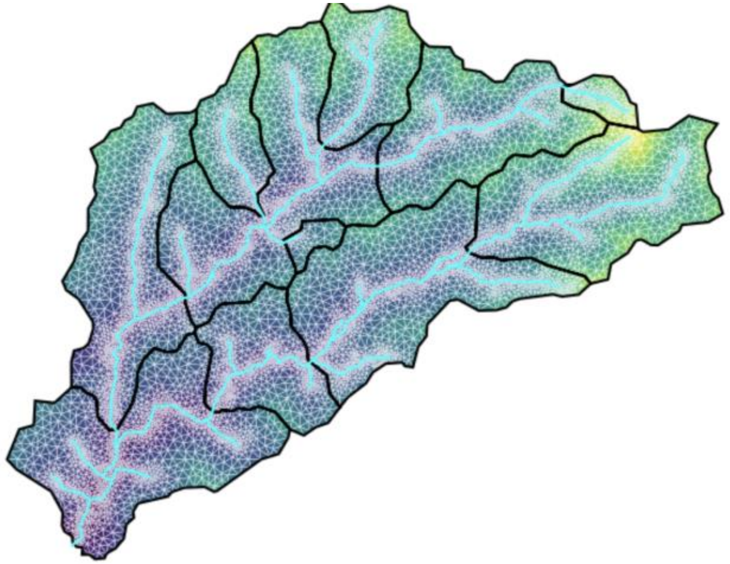


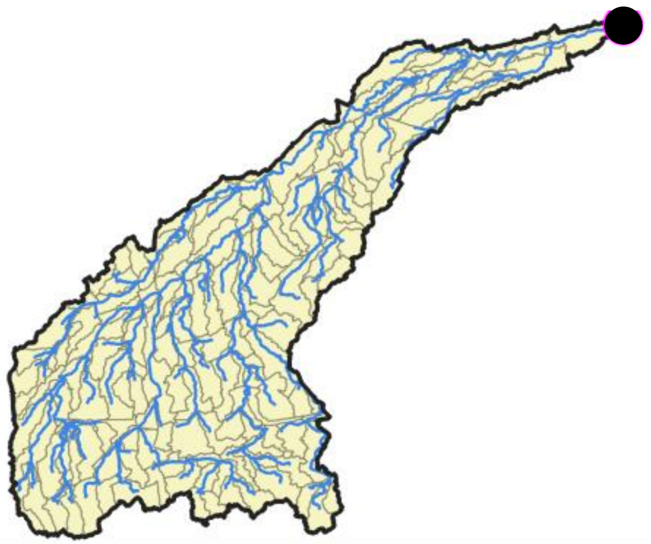


Streamflow prediction across ungauged watersheds: Graph Neural Network & Hybrid Modeling

What if the goal is to predict streamflow at reach scale across watersheds?

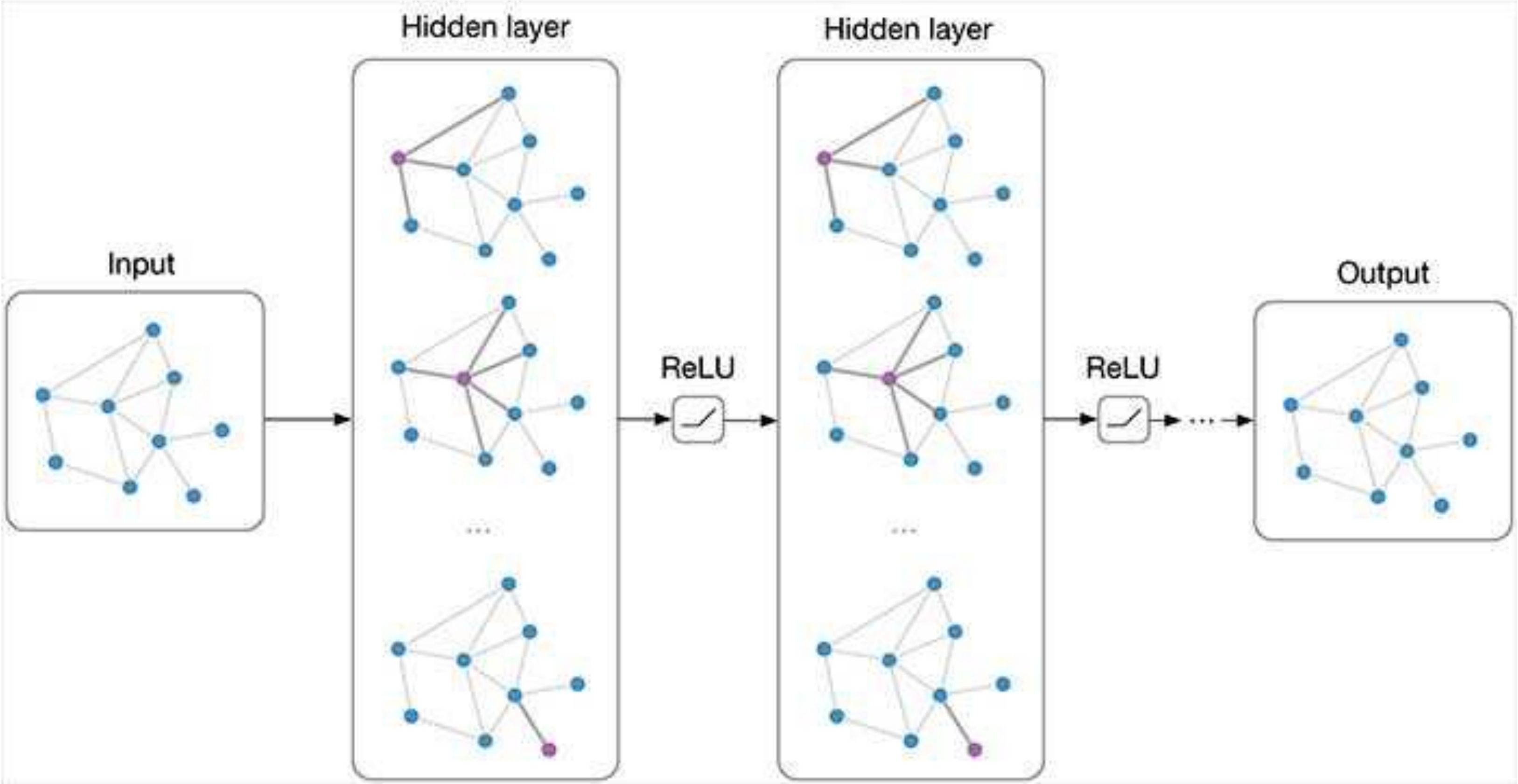






Streamflow prediction across ungauged watersheds: Graph Neural Network & Hybrid Modeling

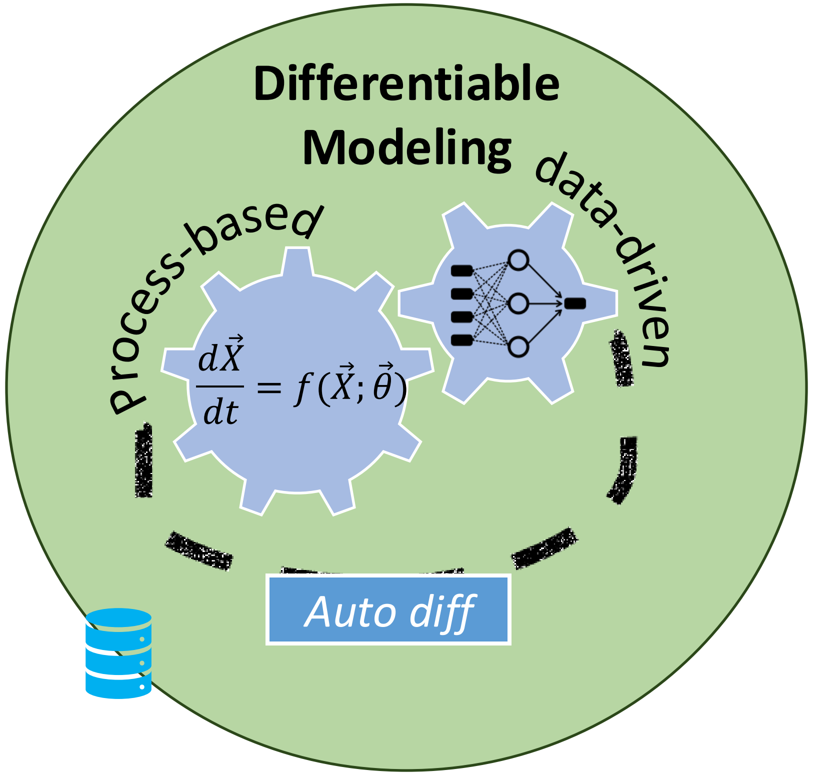
The vanilla Graph Neural Network won't work any more



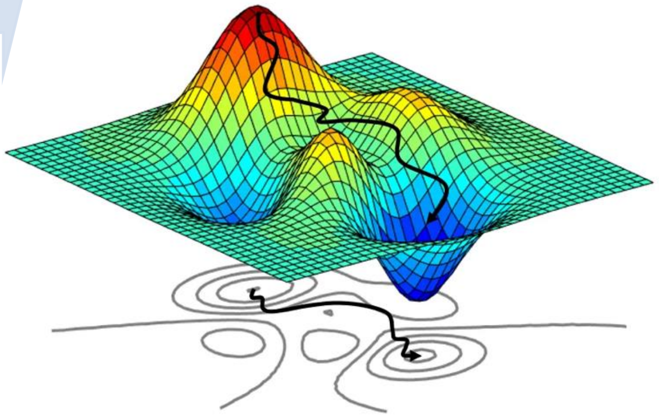
Streamflow prediction across ungauged watersheds: Graph Neural Network & Hybrid Modeling

Ongoing work but there is a promise in using differentiable modeling

explainable

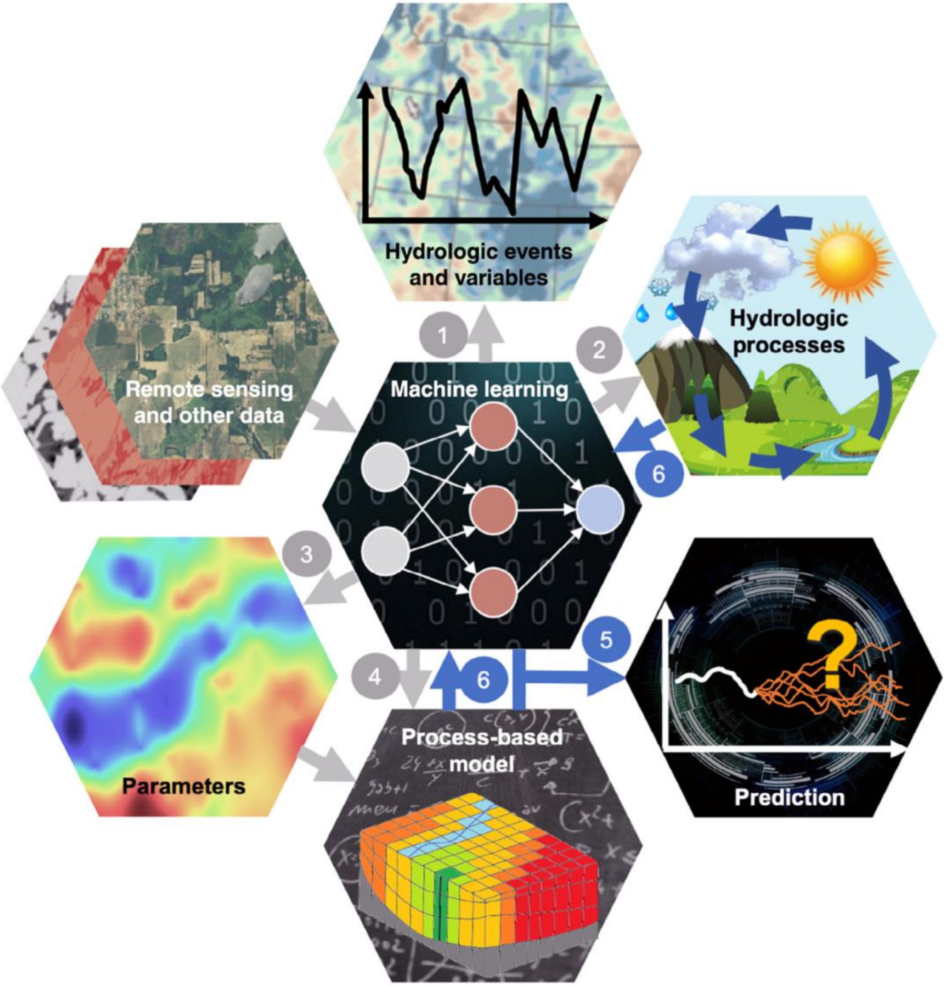


optimizable



Other ML applications in Hydrology

- ~~Surrogate modeling~~
- Inverse modeling
- Hybrid modeling

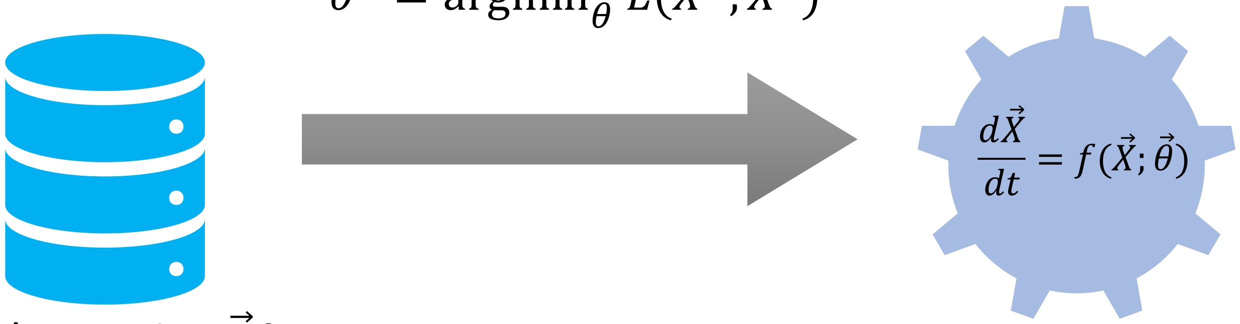


Xu, Tianfang, and Feng Liang. "Machine learning for hydrologic sciences: An introductory overview." Wiley Interdisciplinary Reviews: Water 8.5 (2021): e1533.

Inverse modeling: How to estimate the parameters of process-based models from observations?

$$\vec{\theta}^e = \mathrm{argmin}_{\vec{\theta}}\, L(\vec{X}^o, \vec{X}^s)$$

$\vec{\theta}^e$: the optimal parameters
$\vec{X}^o$: the observations
$\vec{X}^s \in \vec{X}$: the observed model states
L: the loss function



observation $\vec{X}^o$

process-based model

Traditional approaches

Needs a separate workflow for each model

Jiang et al., EMS, 2021

Model forward runs

Update parameters $\vec{\theta}^e$

EnKF

GLUE

coupled iterations

Alternative approach?

Model ensemble runs

Estimate parameters $\vec{\theta}^e$

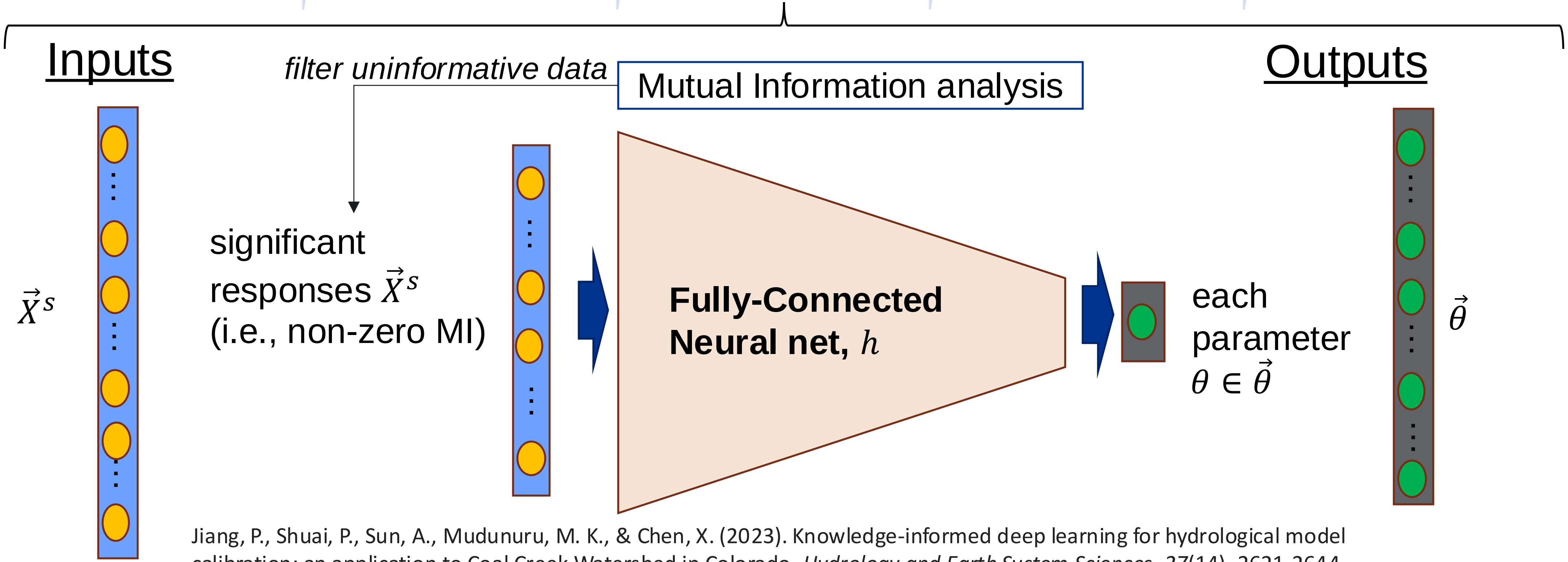
can be separate

Knowledge-Informed deep learning approach hydrological model calibration

Inverse mapping $\vec{\theta}^e = h(\vec{X}^o; \vec{w}^e)$ with

$$\vec{w}^e = \text{argmin}_{\vec{w}}\ L(\vec{\theta}\ \ ; h(\vec{X}^s; \vec{w}))$$

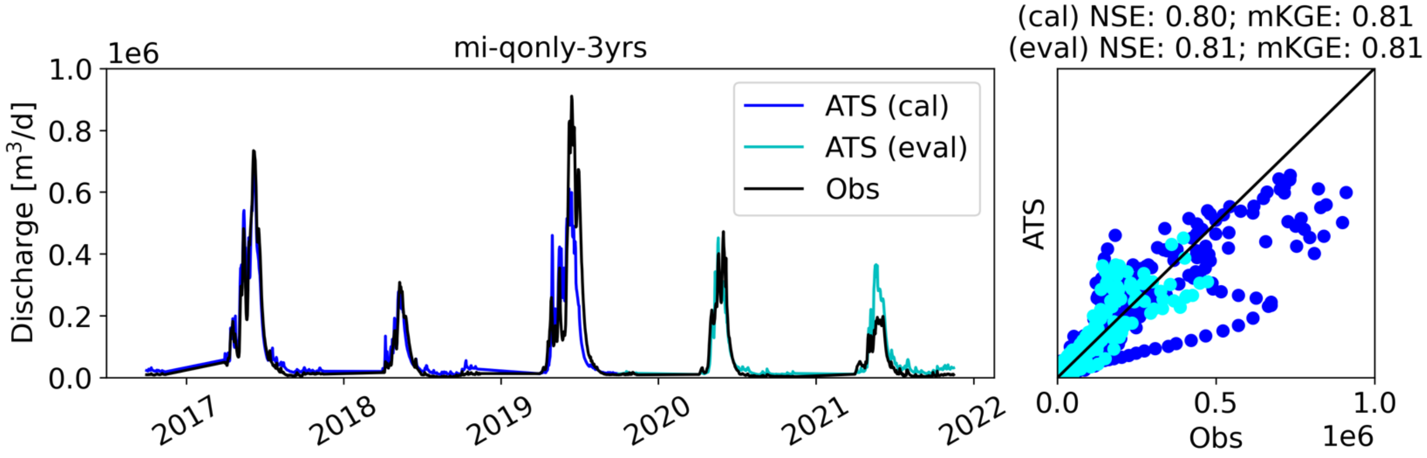
h: a deep learning model ;
$\vec{w}$: the trainable parameters of h;
L: the loss function.



Jiang, P., Shuai, P., Sun, A., Mudunuru, M. K., & Chen, X. (2023). Knowledge-informed deep learning for hydrological model calibration: an application to Coal Creek Watershed in Colorado. *Hydrology and Earth System Sciences*, *27*(14), 2621-2644.

An application at Coal Creek Watershed in Colorado

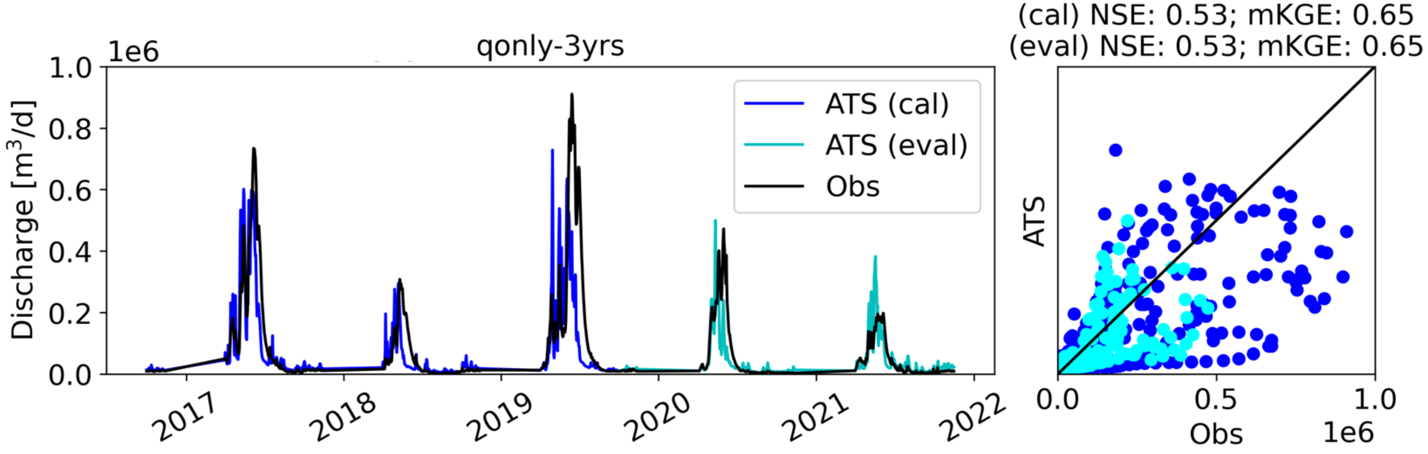
knowledge-informed inverse mapping



NSE: the Nash–Sutcliffe model efficiency coefficient

mKGE: the modified the Kling-Gupta efficiency metric

original inverse mapping (w/o MI analysis)



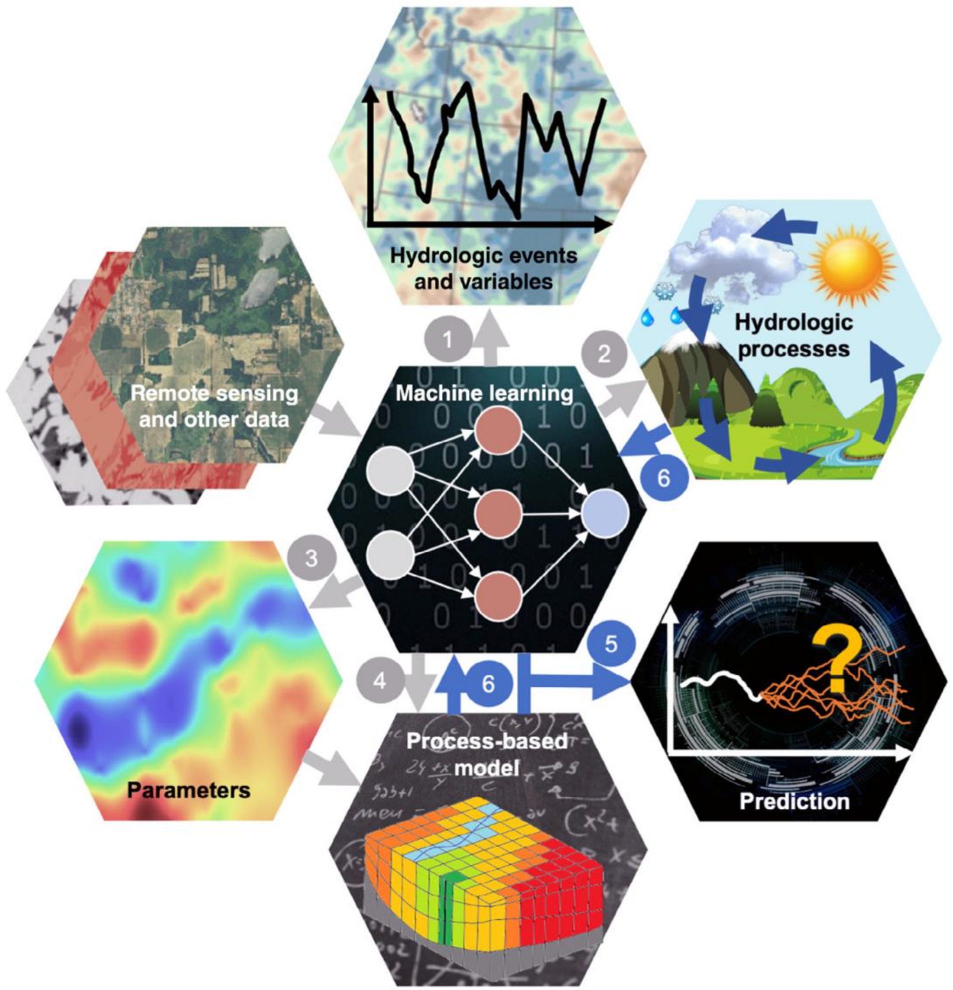
Jiang, P., Shuai, P., Sun, A., Mudunuru, M. K., & Chen, X. (2023). Knowledge-informed deep learning for hydrological model calibration: an application to Coal Creek Watershed in Colorado. *Hydrology and Earth System Sciences*, *27*(14), 2621-2644.

Other ML applications in Hydrology

- ~~Surrogate modeling~~

- ~~Inverse modeling~~

- Hybrid modeling



Xu, Tianfang, and Feng Liang. "Machine learning for hydrologic sciences: An introductory overview." Wiley Interdisciplinary Reviews: Water 8.5 (2021): e1533.

Differentiable hybrid models for ecohydrological processes

Differentiable hybrid modeling

$$\frac{d\vec{X}}{dt} = f\left(\vec{X}; \vec{\theta}^{e} = h(\vec{X}; \vec{w}^{e})\right) \quad \text{with}$$
$$\vec{w}^{e} = \text{argmin}_{\vec{w}} \, L(\vec{X}^{o}, \vec{X}^{S})$$

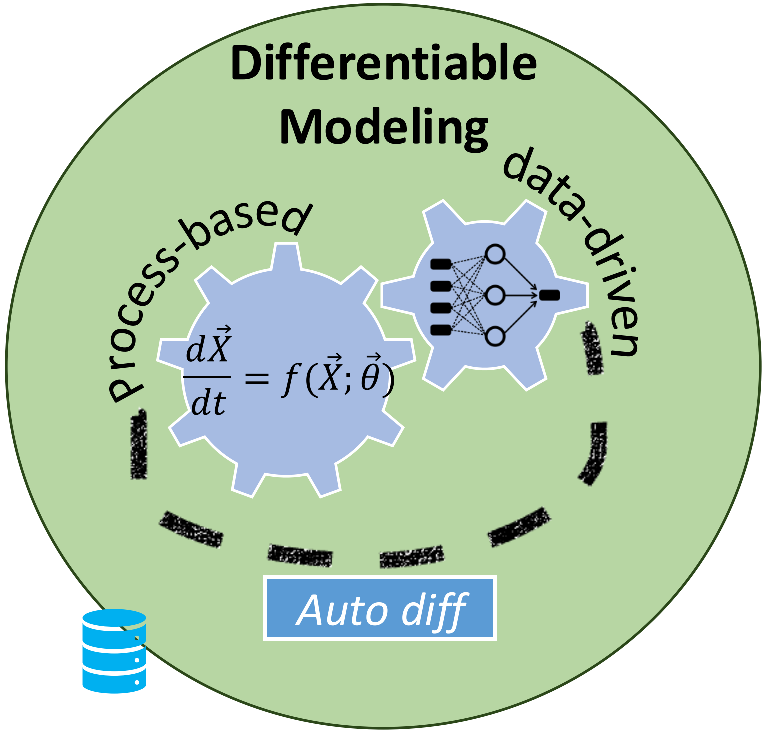
h: a deep learning model

$$\neq$$

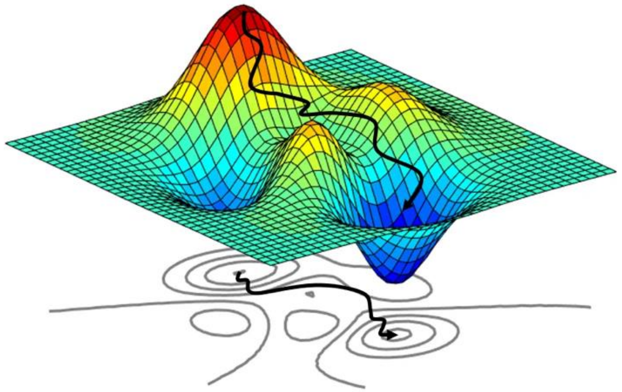
Inverse mapping for parameter estimation

$$\vec{\theta}^{e} = h(\vec{X}^{o}; \vec{w}^{e}) \quad \text{with}$$
$$\vec{w}^{e} = \text{argmin}_{\vec{w}} \, L(\vec{\theta}^{prior}; h(\vec{X}^{s}; \vec{w}))$$

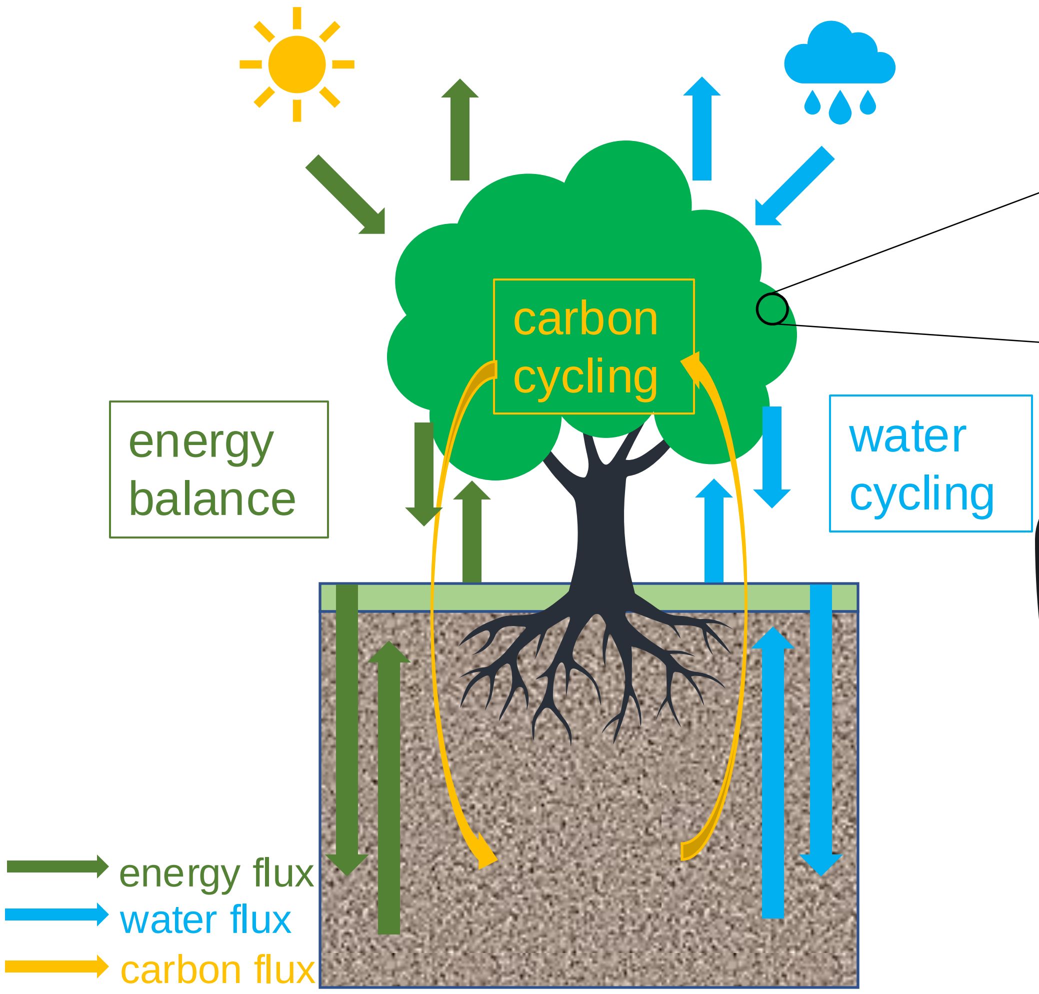
explainable

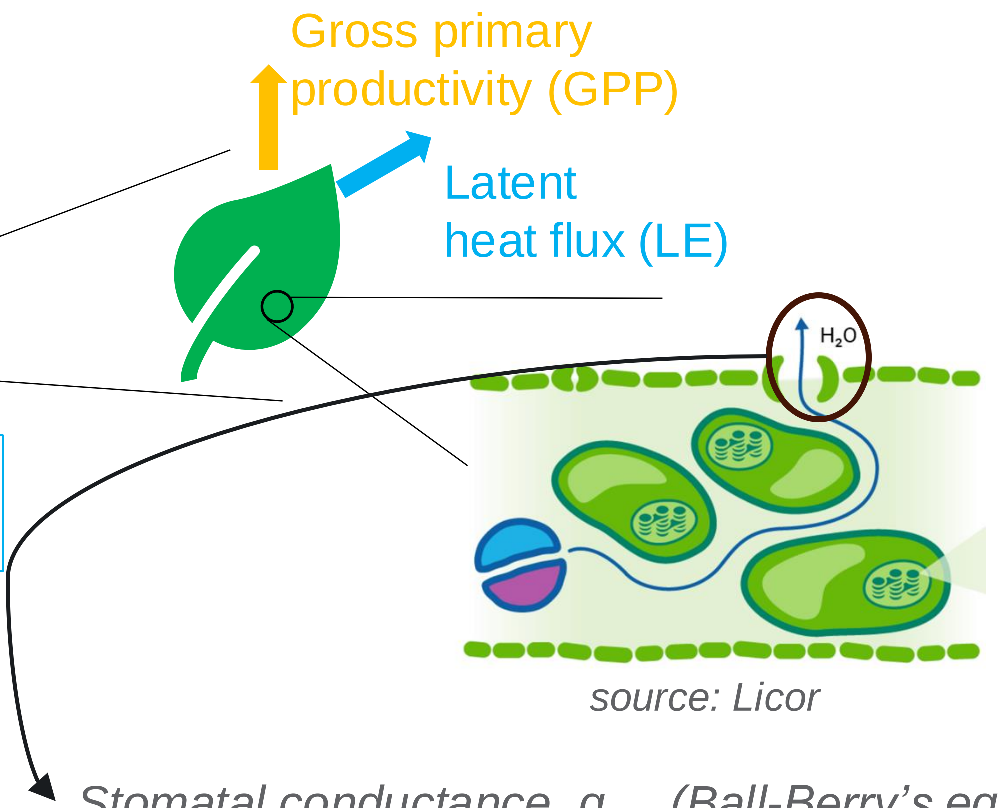


optimizable



Current land surface models are facing uncertainties in model parameters and mechanisms





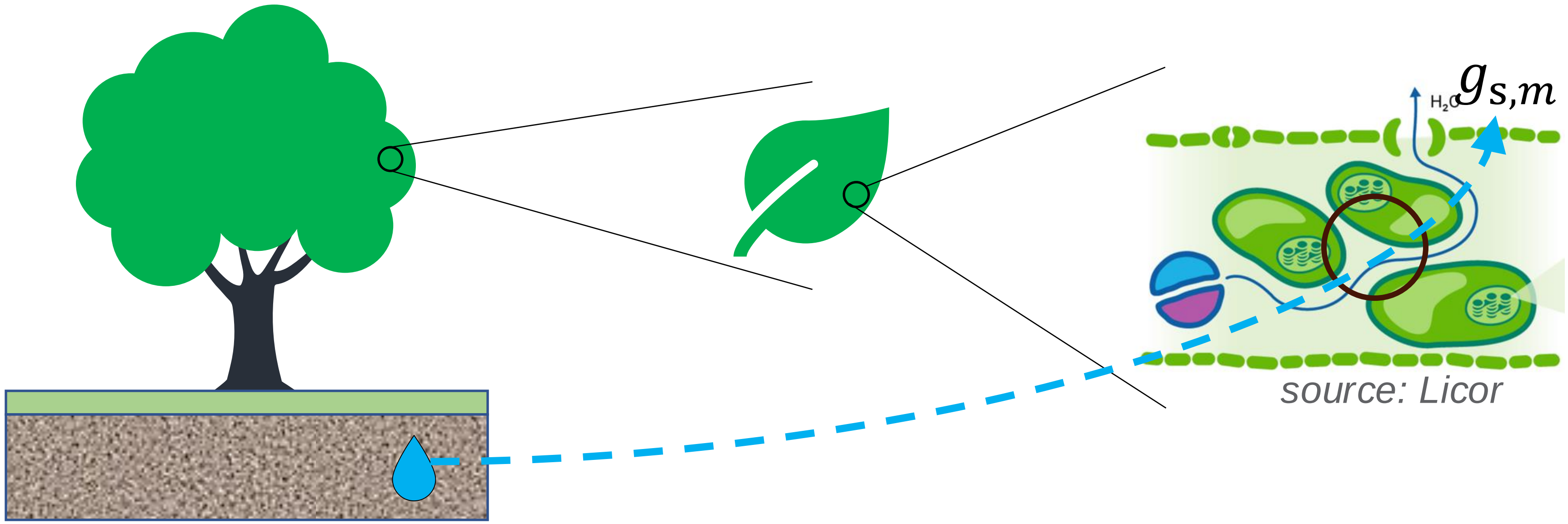
Stomatal conductance, $g_{s,m}$ (Ball-Berry's eqn)

$$g_{s,m} = \frac{\boldsymbol{mA\theta_l}}{C_s} + \boldsymbol{b}$$

- m, b: unknown parameters
- $\boldsymbol{\theta_l}$: leaf water content without plant water deficit accounted

Jiang, Peishi, et al. (2024). JAX-CanVeg: A Differentiable Land Surface Model. *Water Resources Research*, in review.

We proposed a hybrid version of Ball-Berry eqn calculating stomatal conductance $g_{s,m}$ to account for plant water stress impact



The original Ball-Berry equation

$$g_{s,m} = \frac{mA\theta_l}{C_s} + b$$

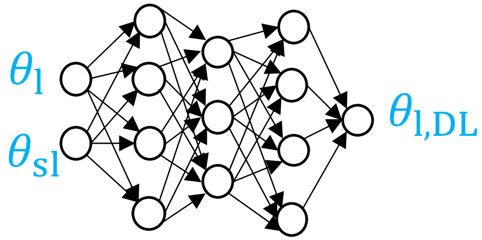
$$\theta_l = \frac{e_a}{e_*(T_l)}$$

- Not accounting for plant water stress

The hybrid Ball-Berry equation approximates soil water stress impact on $g_{s,m}$ using soil water content

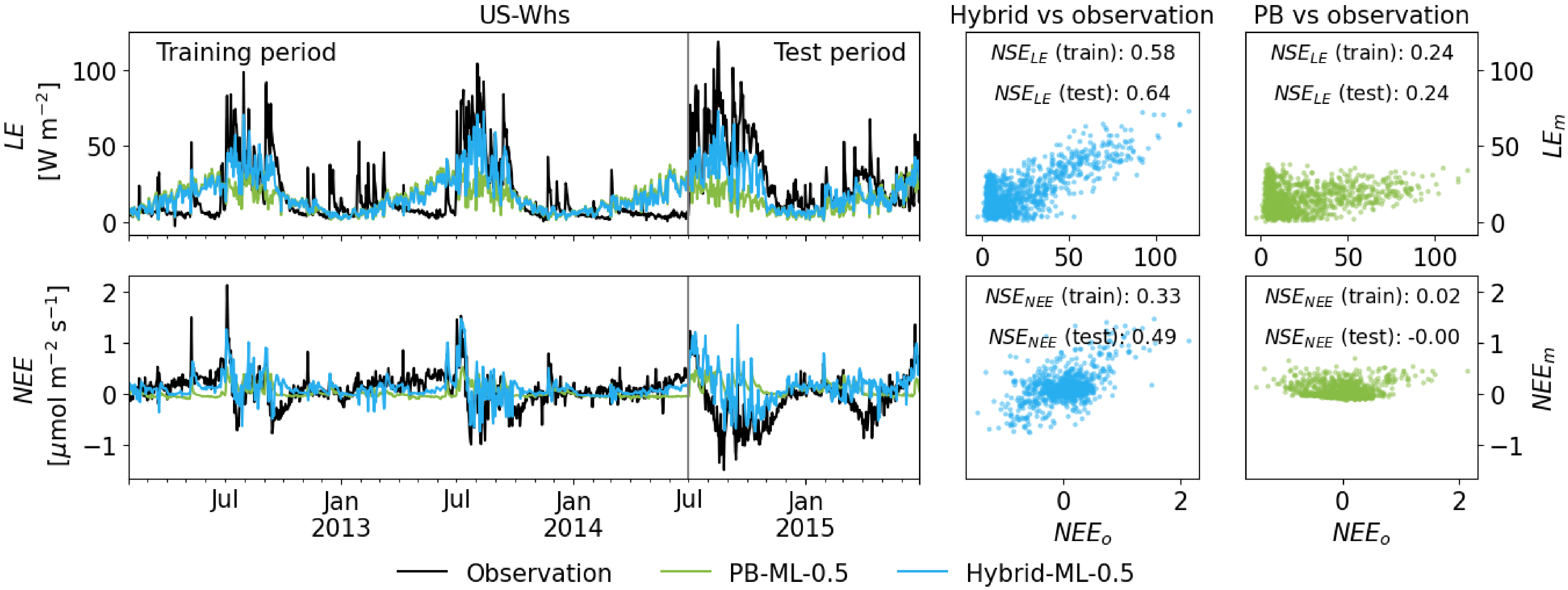
$$g_{s,m,\mathrm{DL}} = \frac{mA\theta_{l,\mathrm{DL}}}{C_s} + g_0$$

$$\theta_{l,\mathrm{DL}} = h(\theta_l, \theta_{l,\mathrm{DL}})$$



Jiang, Peishi, et al. (2024). JAX-CanVeg: A Differentiable Land Surface Model. *Water Resources Research*, in review.

Hybrid modeling improves both the latent heat fluxes (LE) and net ecosystem exchange (NEE) simulations at a semi-arid ecosystem



Jiang, Peishi, et al. (2024). JAX-CanVeg: A Differentiable Land Surface Model. *Water Resources Research*, in review.

Conclusion

- Overview of ML application in Hydrology
- ML application in streamflow prediction
- ML application in parameter estimations of hydrological models
- Differentiable hybrid modeling in ecohydrology