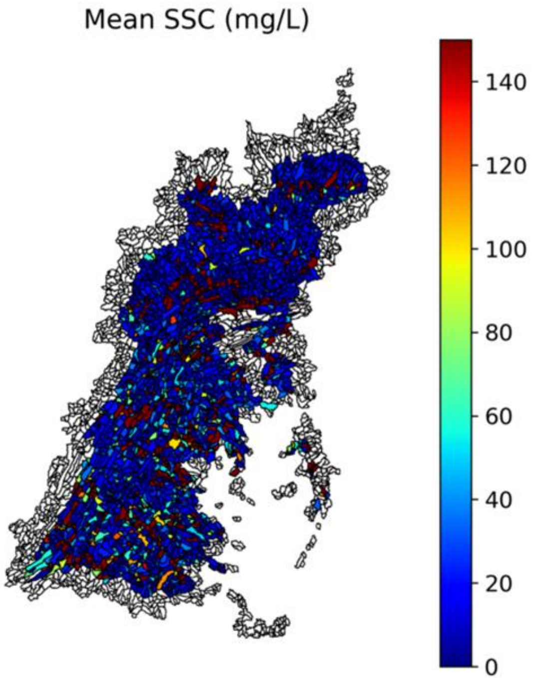
# Multiscale machine-learning sediment prediction update

Xueting Pu, Chaopeng Shen, Yalan Song
Civil & Environmental Engineering, Penn State



# Purely data-driven ML in water

- Examined comparison with in-situ data & long-term projections

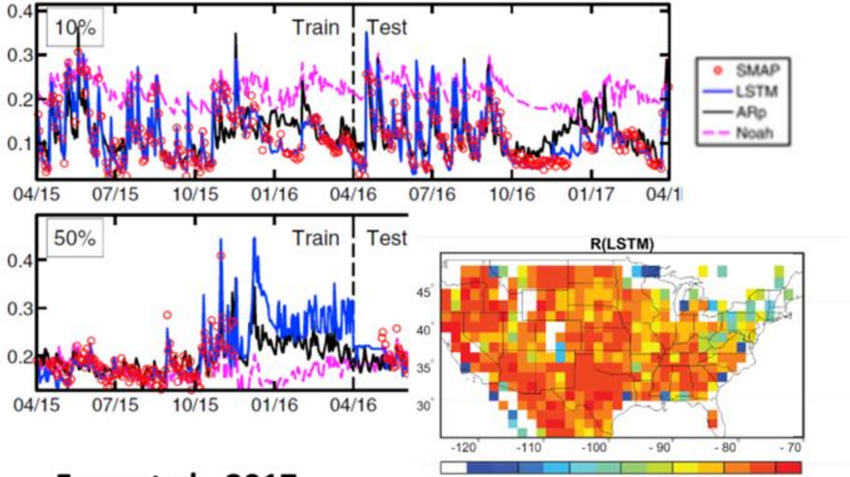
**Geophysical Research Letters**

Research Letter | Full Access

**Prolongation of SMAP to Spatiotemporally Seamless Coverage of Continental U.S. Using a Deep Learning Neural Network**

Kuai Fang, Chaopeng Shen, Daniel Kifer, Xiao Yang

First published: 16 October 2017 | https://doi.org/10.1002/2017GL075619 | Cited by: 3




**Fang et al., 2017**
doi: 10.1002/2017gl075619

**Water Resources Research**

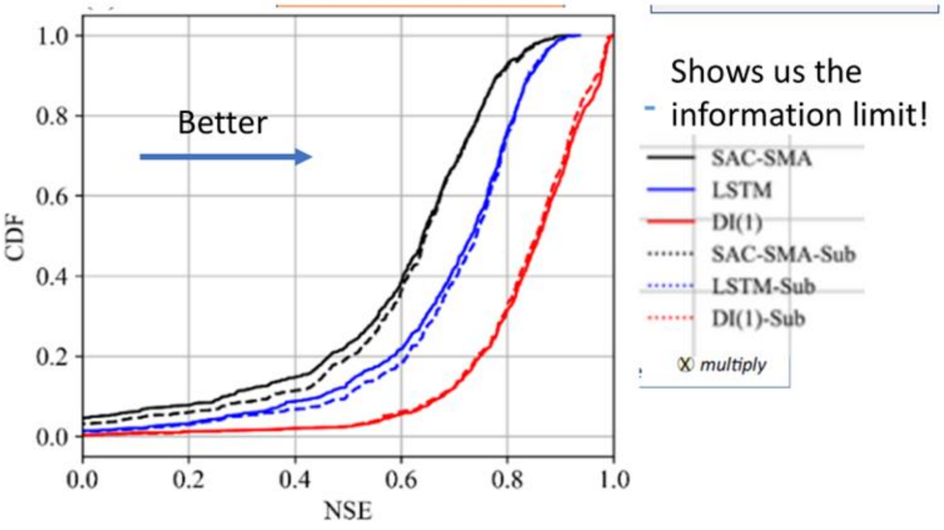
RESEARCH ARTICLE
10.1029/2019WR026793

**Special Section:**
Big Data & Machine Learning in Water Sciences: Recent Progress and Their Use in Advancing Science

**Enhancing Streamflow Forecast and Extracting Insights Using Long-Short Term Memory Networks With Data Integration at Continental Scales**

Dapeng Feng[1], Kuai Fang[1,2], and Chaopeng Shen[1]

[1]Civil and Environmental Engineering, Pennsylvania State University, State College, PA, USA, [2]Now at: Earth System Science, Stanford University, Stanford, CA, USA

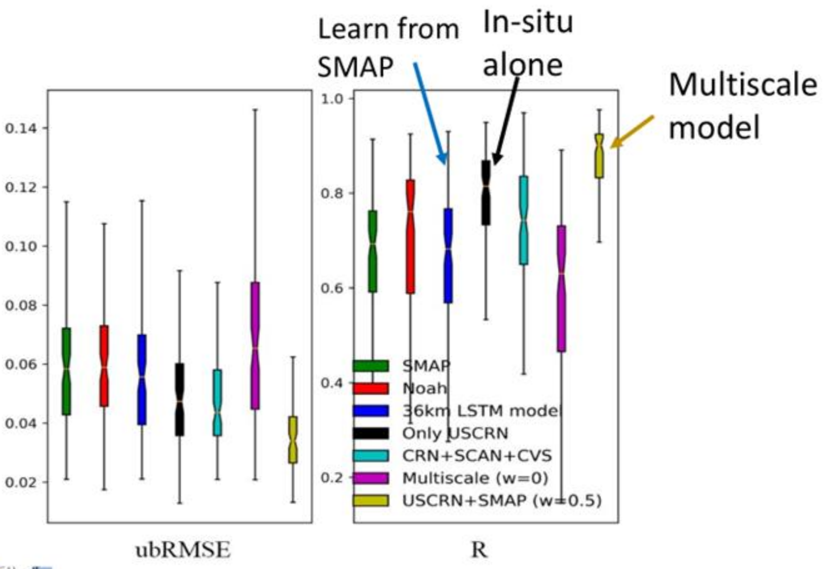


# Multiscale soil moisture – learning from two teachers

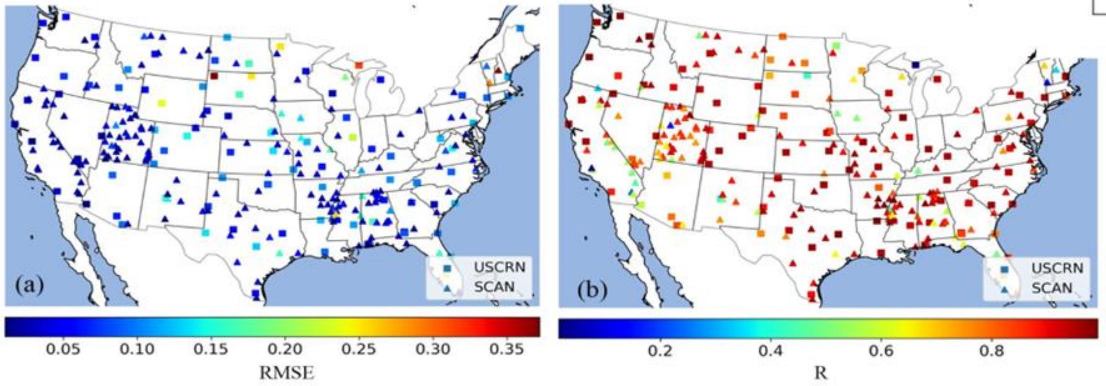
High-res LSTM - - - - - In-situ data

Coarsen ↓

Low-res sim - - - - - Satellite data




Test period: 2015-04-01 to 2020-03-31




**Geophysical Research Letters**

Research Letter | Full Access

A multiscale deep learning model for soil moisture integrating satellite and in-situ data

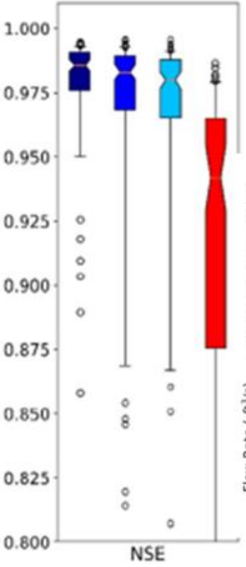
Jiangtao Liu, Farshid Rahmani, Kathryn Lawson, Chaopeng Shen

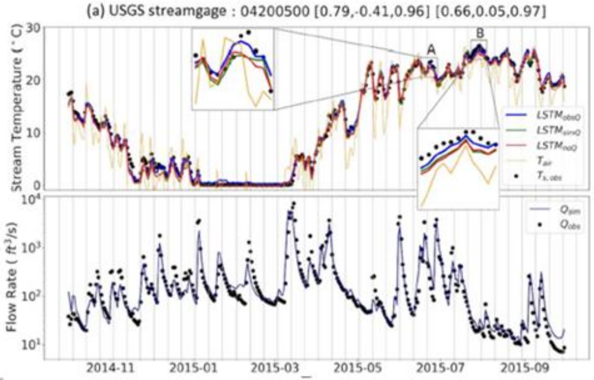
First published: 14 March 2022 | https://doi.org/10.1029/2021GL096847

# LSTM applications in water quality

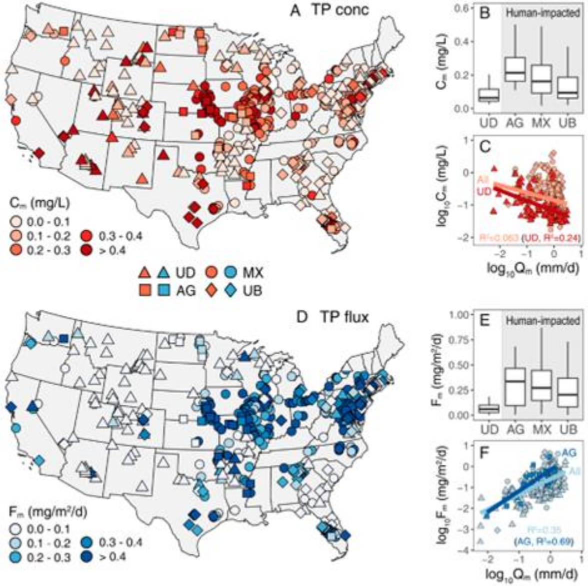
Water temperature

Rahmani et al., 2021, ERL




Dissolved Oxygen & Total Phosphorous (Wei Zhi & Li Li)




Journal of Hydrology
Volume 639, August 2024, 131573

Suspendid Sediment Conc

Research papers

## Deep learning insights into suspended sediment concentrations across the conterminous United States: Strengths and limitations

Yalan Song [a], Piyaphat Chaemchuen [a], Farshid Rahmani [a], Wei Zhi [a], Li Li [a], Xiaofeng Liu [a], Elizabeth Boyer [b], Tadd Bindas [a], Kathryn Lawson [a], Chaopeng Shen [a]

Nitrate time series

Contents lists available at ScienceDirect
Science of the Total Environment
journal homepage: www.elsevier.com/locate/scitotenv

## A deep learning-based novel approach to generate continuous daily stream nitrate concentration for nitrate data-sparse watersheds

Gourab Kumer Saha [a], Farshid Rahmani [b], Chaopeng Shen [b], Li Li [b], Raj Cibin [a,b,*]

PNAS RESEARCH ARTICLE | ENVIRONMENTAL SCIENCES

## Increasing phosphorus loss despite widespread concentration decline in US rivers

Wei Zhi, Hubert Baniecki, Jiangtao Liu, Elizabeth Boyer, Chaopeng Shen, Gary Shenk, Xiaofeng Liu, and Li Li
Affiliations are included on p. 8.

Edited by Nils Stenseth, Universitetet i Oslo, Oslo, Norway; received January 30, 2024; accepted October 10, 2024

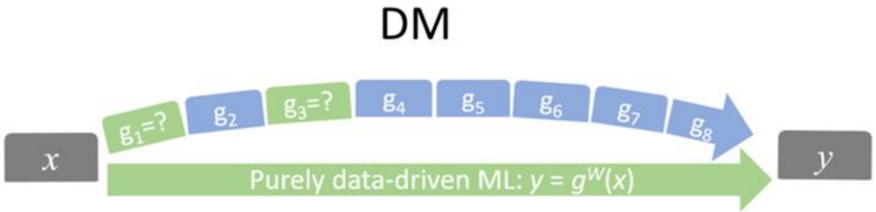
nature water
Explore content About the journal Publish with us Subscribe
nature > nature water > articles > article
Article | Published: 09 March 2023

## Temperature outweighs light and flow as the predominant driver of dissolved oxygen in US rivers

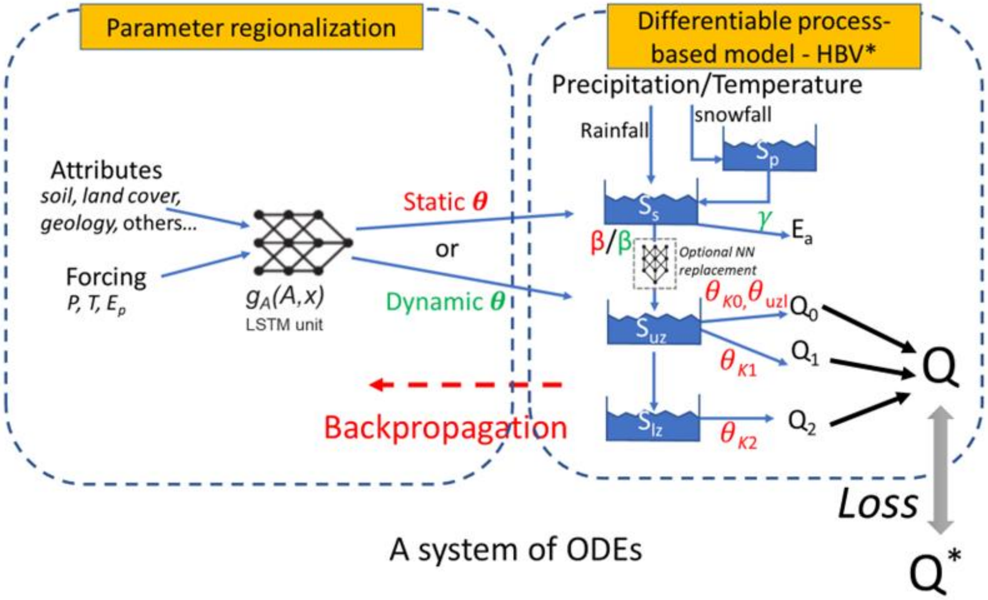
Wei Zhi, Wenyu Ouyang, Chaopeng Shen & Li Li

Nature Water 1, 249–260 (2023) | Cite this article

# What is Differentiable Modeling?

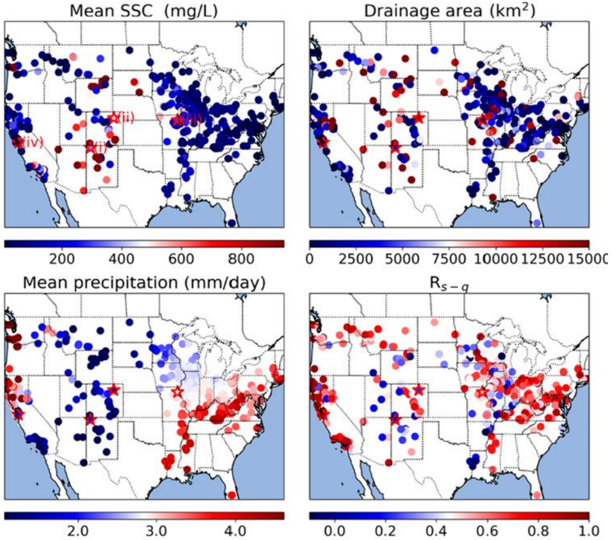



- NNs ("?") mixed w/ process-based equations (priors)
- Breaks a problem into parts, with some as priors
- "end-to-end" training on big data
- The priors constrain the learning to an interpretable scope.
- Can be used a forward simulator as well as




A system of ODEs

This is NOT physics-informed neural network (PINN)!




Research papers

# Deep learning insights into suspended sediment concentrations across the conterminous United States: Strengths and limitations

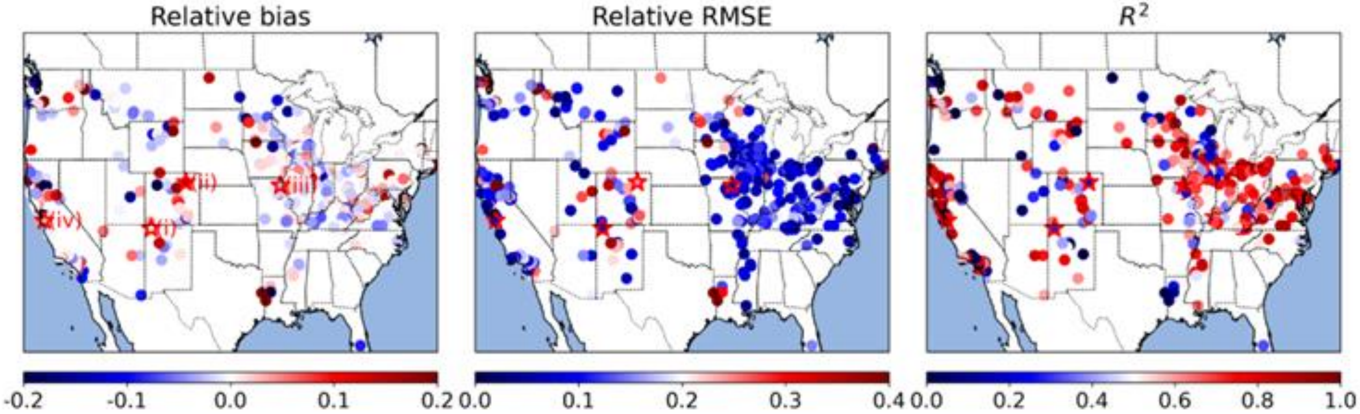
Yalan Song [a], Piyaphat Chaemchuen [a], Farshid Rahmani [a], Wei Zhi [a], Li Li [a], Xiaofeng Liu [a], Elizabeth Boyer [b], Tadd Bindas [a], Kathryn Lawson [a], Chaopeng Shen [a]

Show more



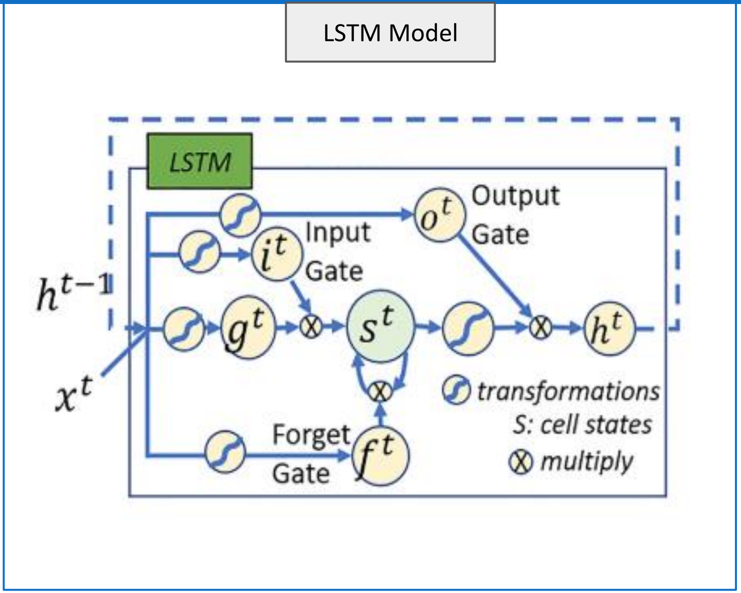
https://doi.org/10.1016/j.jhydrol.2024.131573

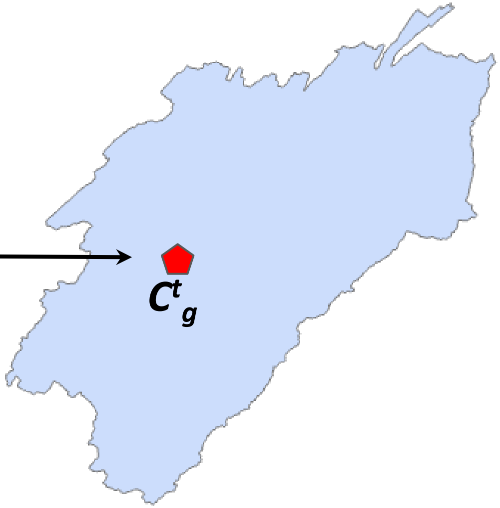




# Lumped LSTM

Basin averaged inputs $X_g^t$, $A_g$




A=[‘ETPOT_Hargr’, ‘FW’, ‘HWSD_clay’, ‘HWSD_gravel’, ‘HWSD_sand’, ‘HWSD_silt’, ‘NDVI’, ‘Porosity’, ‘SoilGrids1km_clay’, ‘SoilGrids1km_sand’, ‘SoilGrids1km_silt’, ‘T_clay’, ‘T_gravel’, ‘T_sand’, ‘T_silt’, ‘aridity’, ‘glaciers’, ‘meanP’, ‘meanTa’, ‘meanelevation’, ‘meanslope’, ‘permafrost’, ‘permeability’, ‘seasonality_P’, ‘seasonality_PET’, ‘snow_fraction’, ‘snowfall_fraction’,‘area’]

x = [P, PET, Temp]

$SSC_g^t = LSTM(x_g^t, A_g)$

# Multiscale LSTM

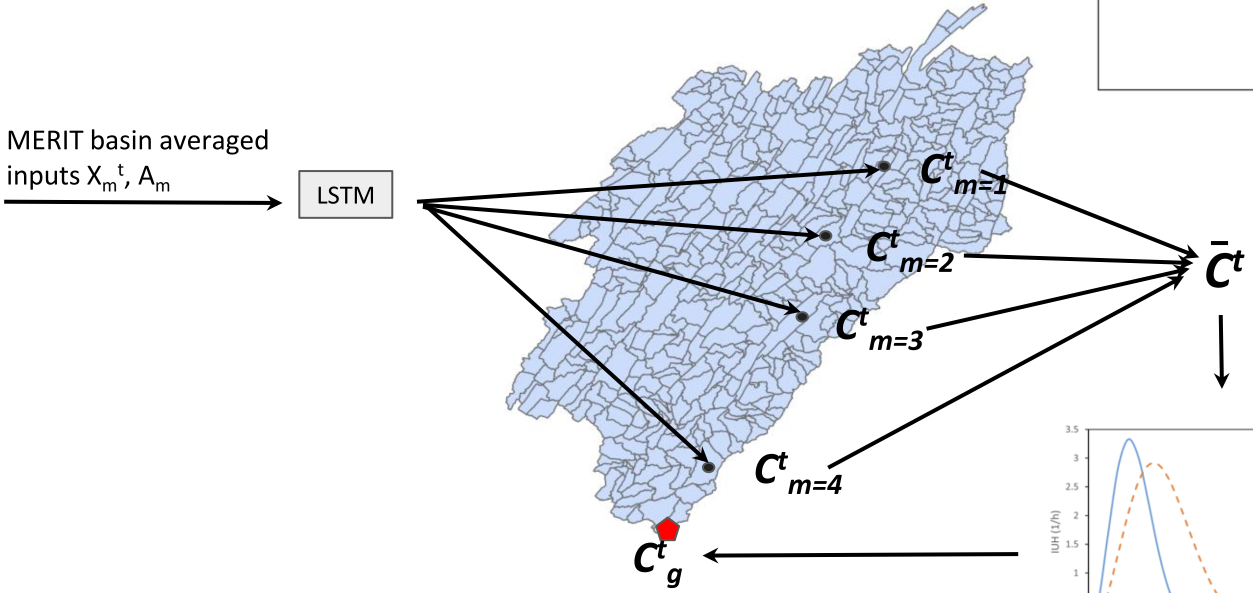
$SSC_m^t = LSTM(x_m^t, A_m)$ # m is a merit basin.

$\underline{SSC_g}^t = average(SSC_m^t)$

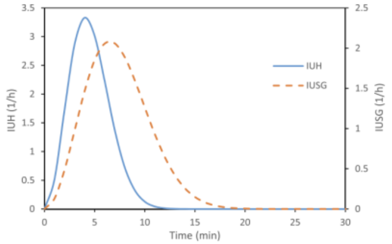
$SSC_G = Routing(\underline{SSC_g}^t, A_{r,\,G})$ # G is a gage containing one or more merit basins

# Routing0: average the SSCm belong to a gage

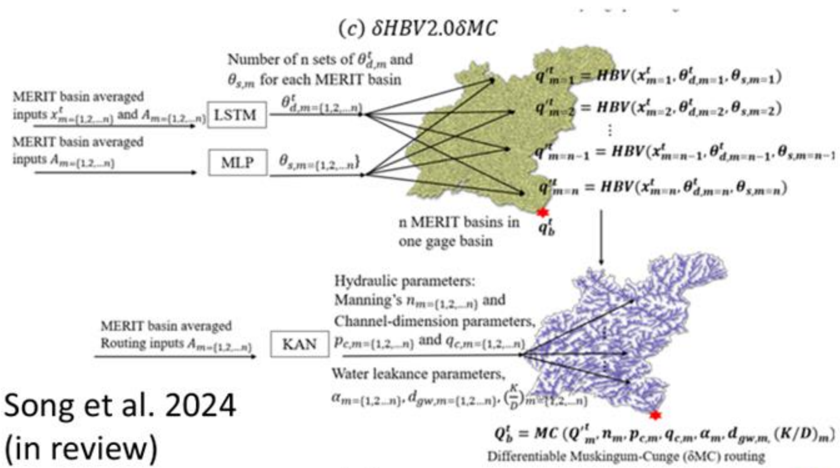
# Routing1: unit hydrograph




Unit hydrograph

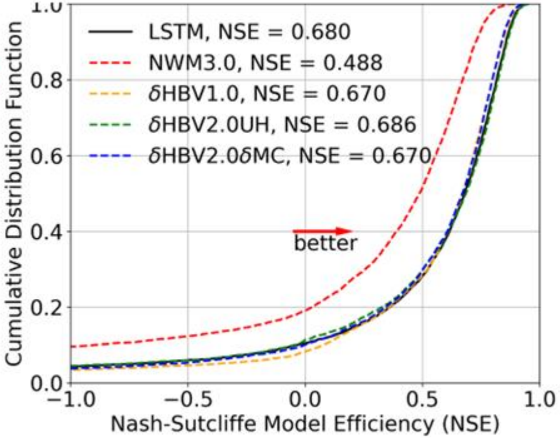


# Large-scale, operational




Song et al. 2024
(in review)

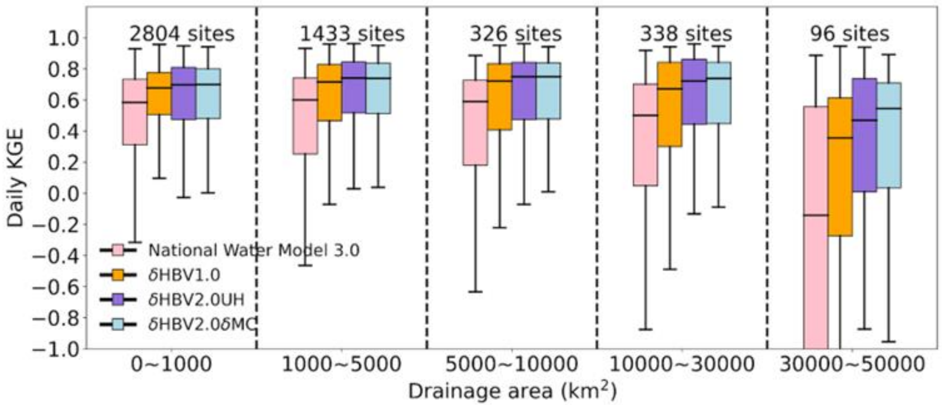
Paper:
https://bit.ly/3NnqDNB



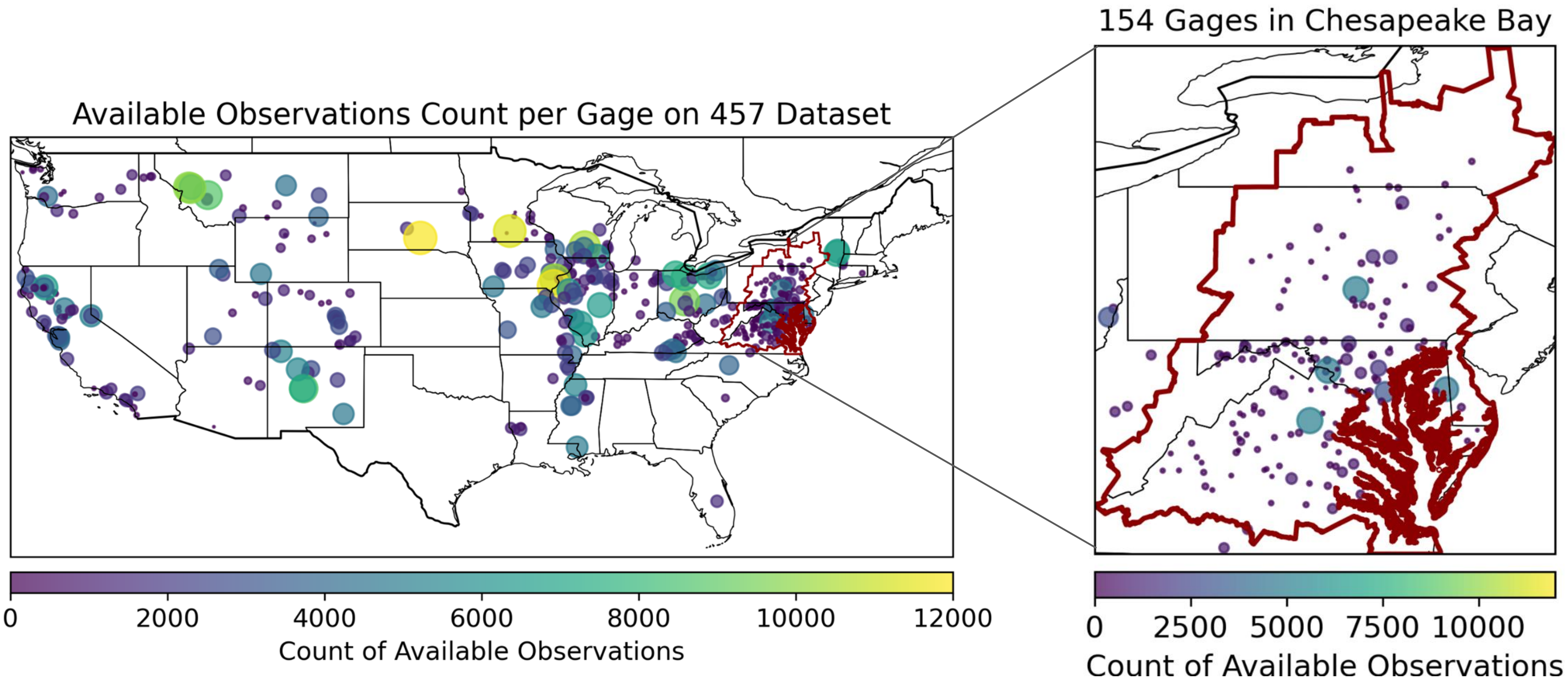

The mean seamless daily streamflow of 1995WY
on the MERIT river network


Data:

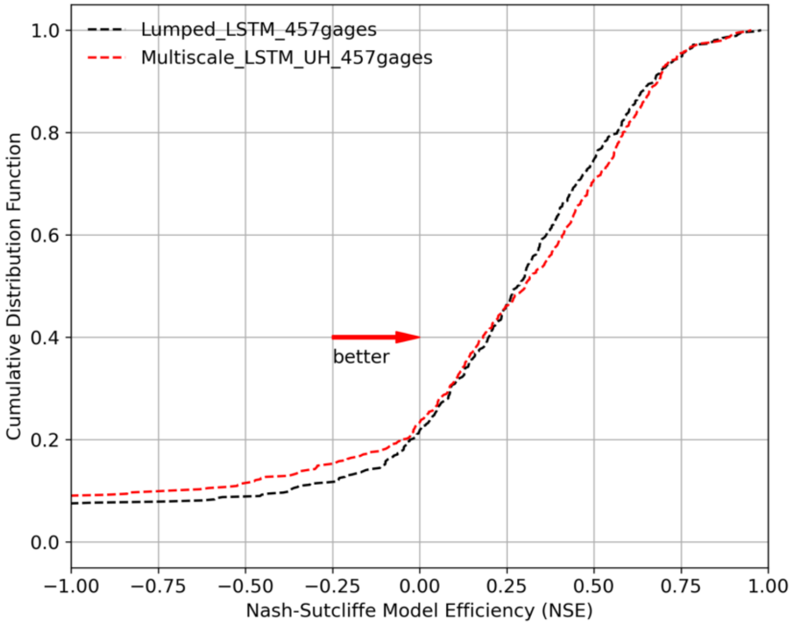
Tested in 1980-2019 on 4,997 GAGES-II stations



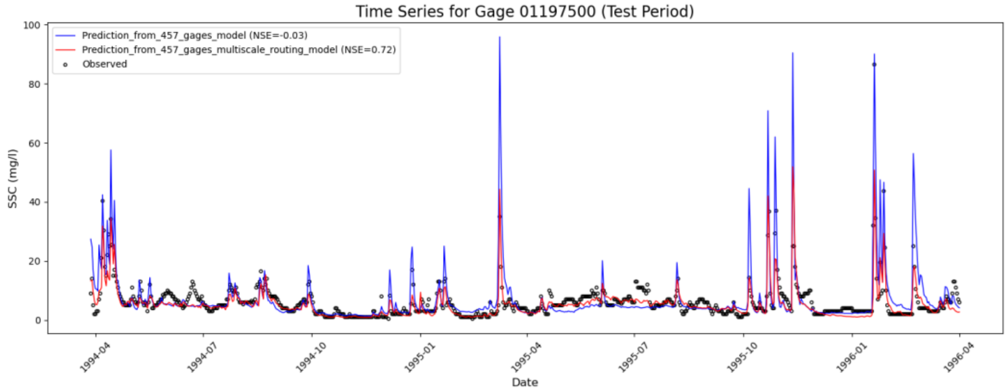
# Dataset


Available Observations Count per Gage on 457 Dataset
0
2000
4000
6000
8000
10000
12000
Count of Available Observations
154 Gages in Chesapeake Bay
0
2500
5000
7500
10000
Count of Available Observations

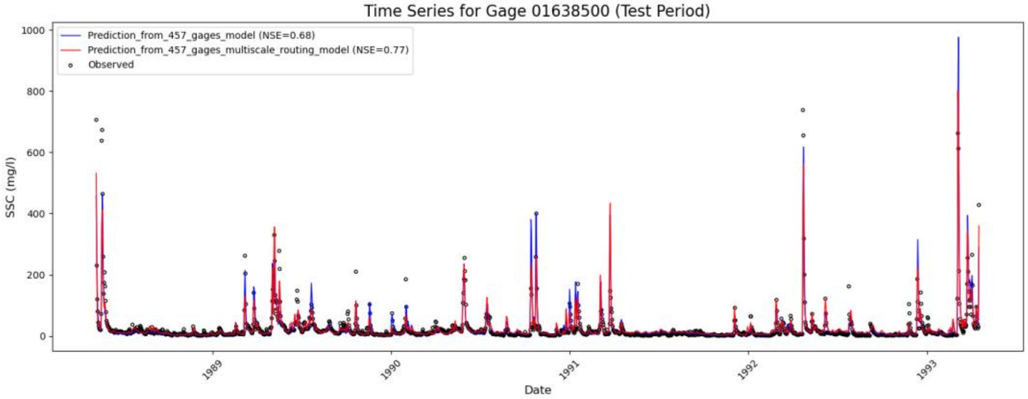
# Results



Lumped_LSTM_457gages
Multiscale_LSTM_UH_457gages
better
Cumulative Distribution Function
Nash-Sutcliffe Model Efficiency (NSE)

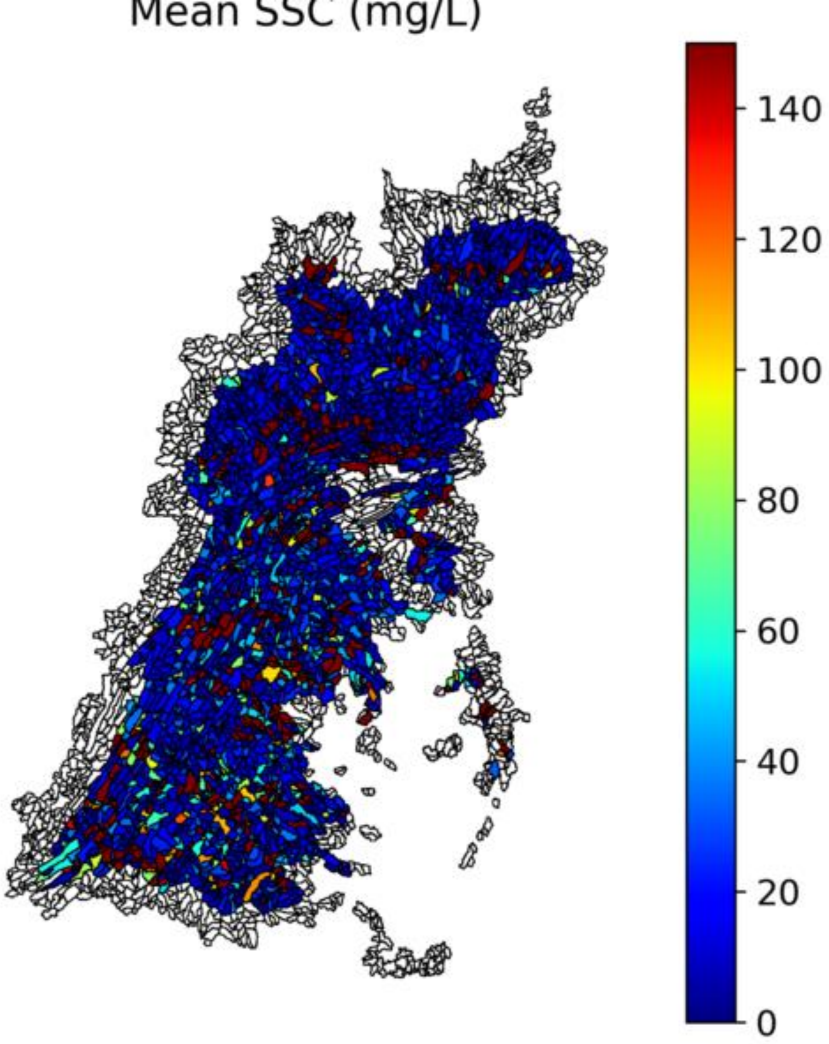



Time Series for Gage 01197500 (Test Period)
Prediction_from_457_gages_model (NSE=-0.03)
Prediction_from_457_gages_multiscale_routing_model (NSE=0.72)
Observed
SSC (mg/l)
Date

Time Series for Gage 01638500 (Test Period)
Prediction_from_457_gages_model (NSE=0.68)
Prediction_from_457_gages_multiscale_routing_model (NSE=0.77)
Observed
SSC (mg/l)
Date

# Results



Next steps:

- Incorporate more input descriptors (connectivity? Buffer strips? Aggregated Geomorphic units? Management? Intersection(soil, slope, LULC)) and forcings (maximum rainfall?)
- Include more sites for training, small and large
- Add physical constraints (sensitivity, variance, known relationships)