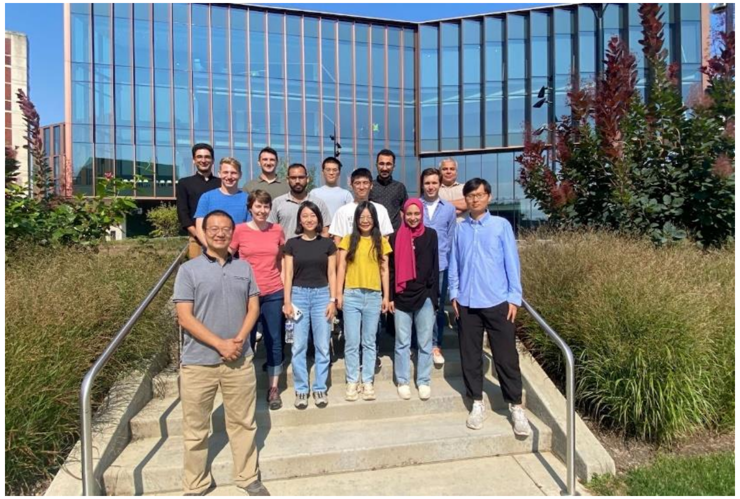
State-of-the-Art AI and Physics-informed ML in Hydrology & Water Quality: Insights and Synergies

Chaopeng Shen

[1]Civil and Environmental Engineering
Penn State University

cshen@engr.psu.edu



Hydroml.org
HydroML Symposium, May 22-26, 2022, Penn State
HydroML 2, May 2023, Berkeley, CA
HydroML 3, May 2024, near PNNL, WA

Our code collection: https://mhpi.github.io/
New framwork: https://mhpi.github.io/frameworks/

Yalan Song, Farshid Rahmani, Tadd Bindas, Jiangtao Liu, Doaa Aboelyazeed, Kamlesh Sawadekar, Dapeng Feng,

Overview

I. ***Data-Driven Machine Learning (ML)***
II. ***Physics-informed** ("Differentiable") modeling (DM)*
III. ***Future Outlook***

nature reviews earth & environment https://doi.org/10.1038/s43017-023-00450-9

Perspective Check for updates

Differentiable modelling to unify machine learning and physical models for geosciences

Chaopeng Shen[1], Alison P. Appling[2], Pierre Gentine[3], Toshiyuki Bandai[4], Hoshin Gupta[5], Alexandre Tartakovsky[6], Marco Baity-Jesi[7], Fabrizio Fenicia[7], Daniel Kifer[8], Li Li[1], Xiaofeng Liu[1], Wei Ren[9], Yi Zheng[10], Ciaran J. Harman[11], Martyn Clark[12], Matthew Farthing[13], Dapeng Feng[1], Praveen Kumar[6,14], Doaa Aboelyazeed[1], Farshid Rahmani[1], Yalan Song[1], Hylke E. Beck[15], Tadd Bindas[1], Dipankar Dwivedi[16], Kuai Fang[17], Marvin Höge[7], Chris Rackauckas[18], Binayak Mohanty[19], Tirthankar Roy[20], Chonggang Xu[21] & Kathryn Lawson[1]

nature COMMUNICATIONS

ARTICLE Check for updates

https://doi.org/10.1038/s41467-021-26107-z OPEN

From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling

Wen-Ping Tsai[1], Dapeng Feng[1], Ming Pan[2,3], Hylke Beck[4], Kathryn Lawson[1,5], Yuan Yang[6,7], Jiangtao Liu[1] & Chaopeng Shen[1,5]

ICLR

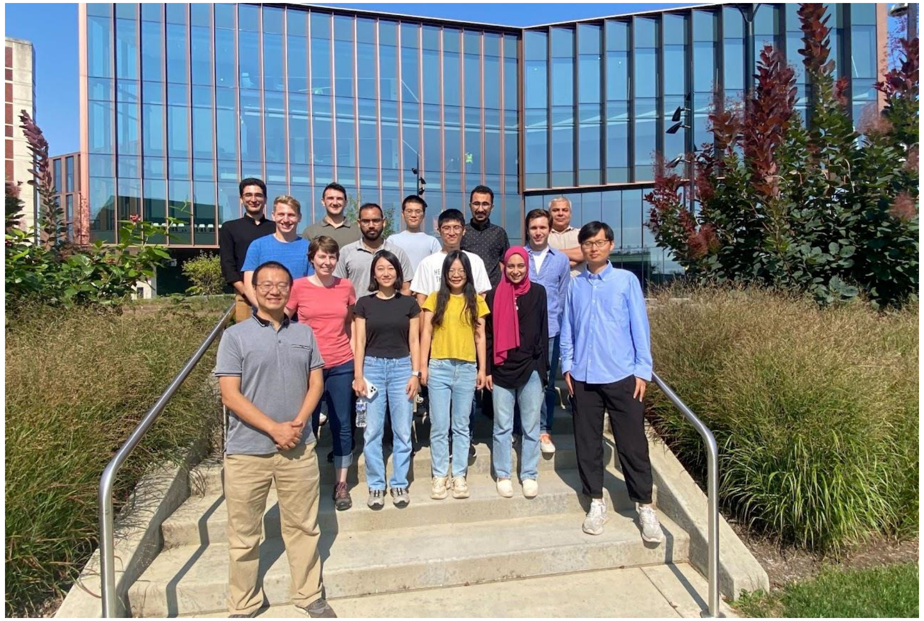
Sensitivity-Constrained Fourier Neural Operators for Forward and Inverse Problems in Parametric Differential Equations

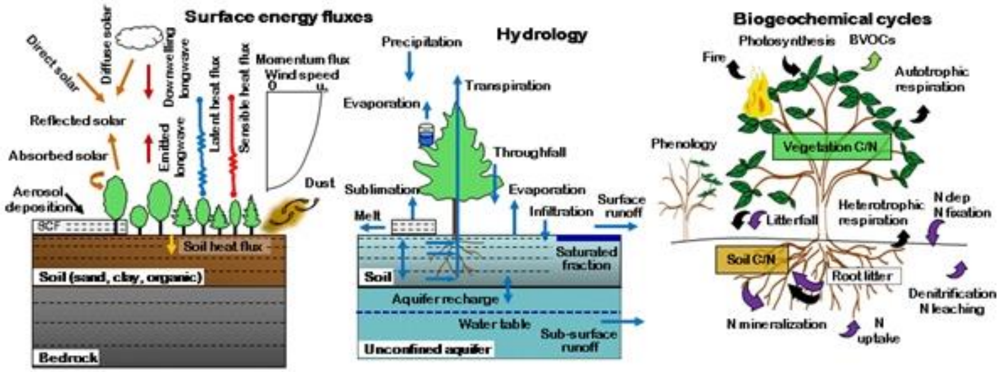
Abdolmehdi Behroozi, Chaopeng Shen, Daniel Kifer

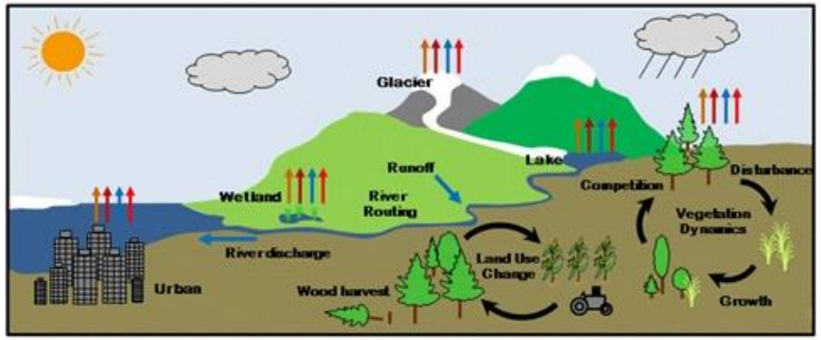
Published: 22 Jan 2025, Last Modified: 05 Feb 2025 ICLR 2025 Everyone Revisions BibTeX CC BY 4.0

About me

- Ph.D. Michigan State in Env. Engr.
- Postdoc Lawrence Berkeley National Lab
- 12 current group members + 3 incoming – 1 graduating.
- “Grew up” as a process-based modeler, solving PDEs. See both sides of the story.
- Working with ML since 2016.







Purely data-driven ML in water

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

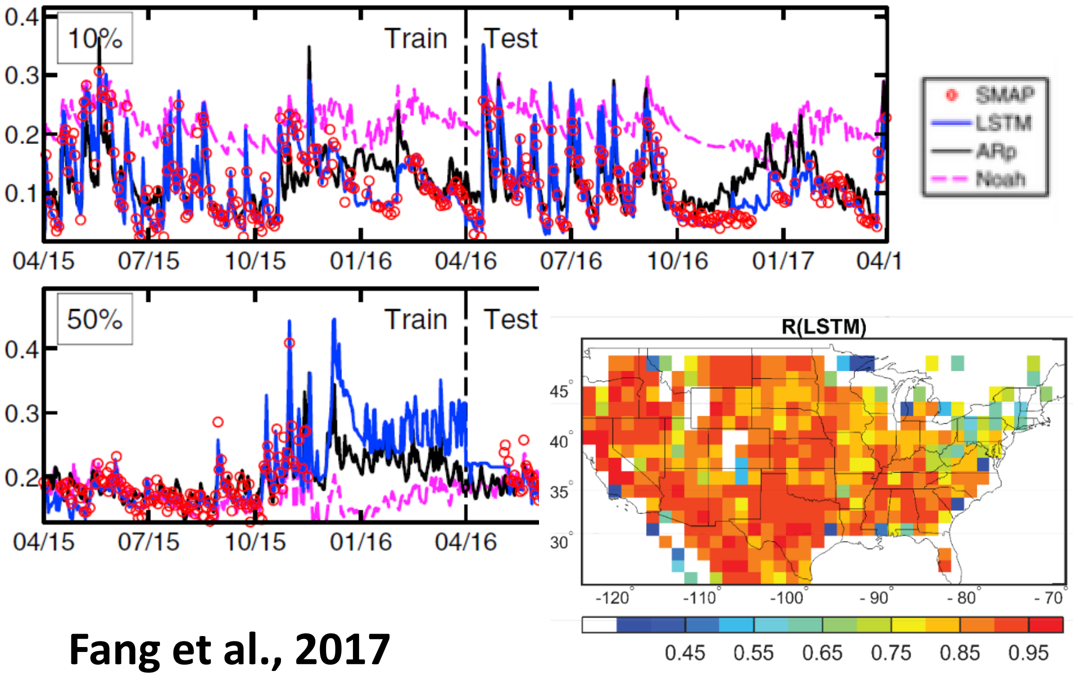
Geophysical Research Letters

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

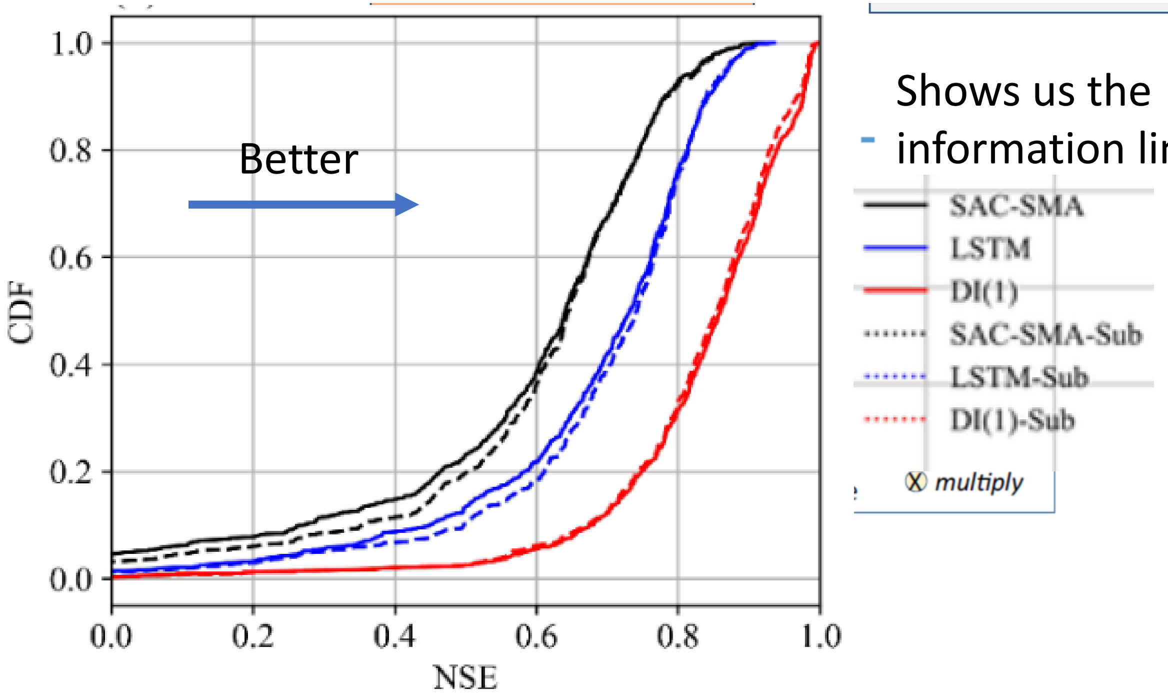
RESEARCH ARTICLE
10.1029/2019WR026793

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



Shows us the information limit!

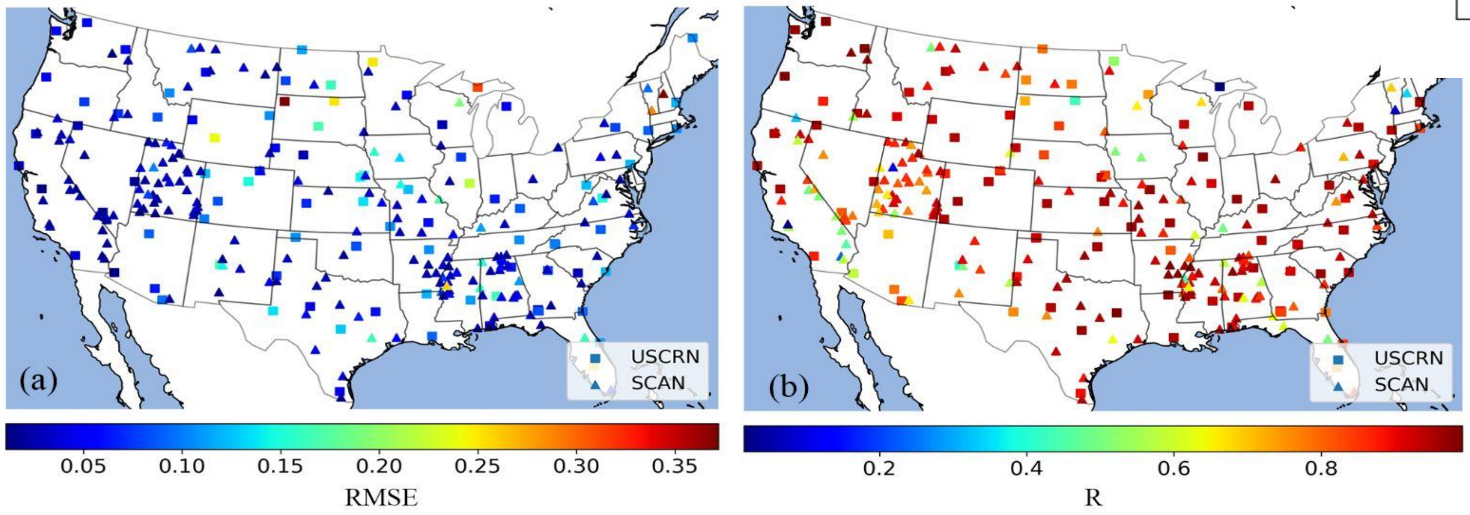
Multiscale soil moisture – learning from two teachers

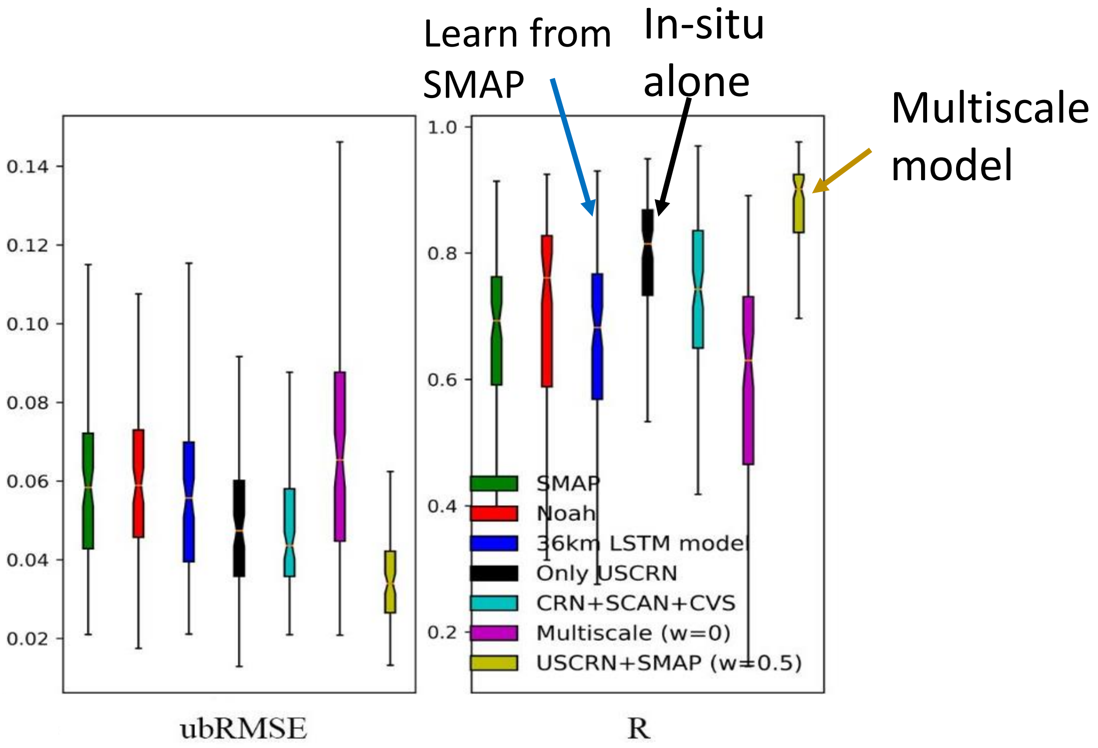
High-res LSTM - - - - - In-situ data

Coarsen ↓

Low-res sim - - - - - Satellite data

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





Geophysical Research Letters

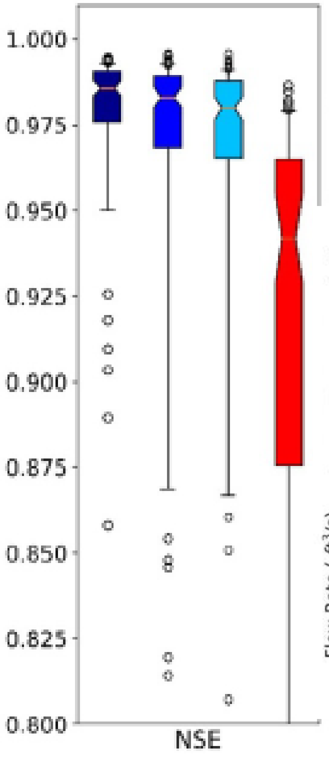
Research Letter | Full Access

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

Jiangtao Liu, Farshid Rahmani, Kathryn Lawson, Chaopeng Shen

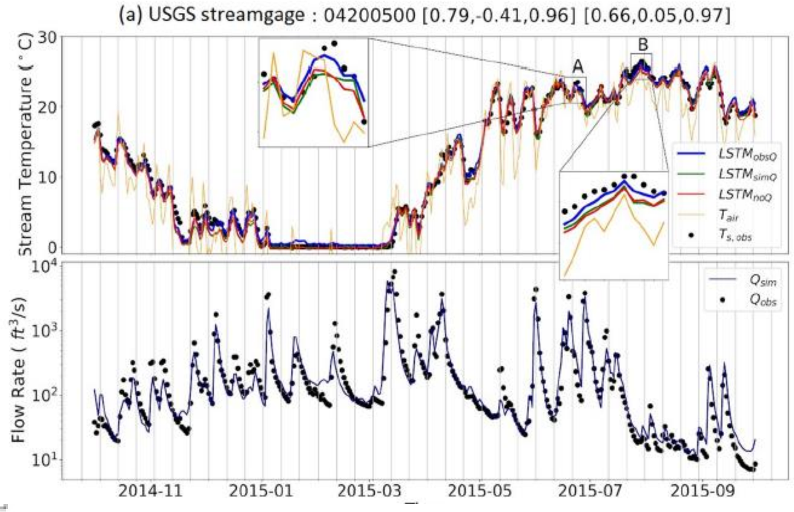
First published: 14 March 2022 | https://doi.org/10.1029/2021GL096847

LSTM applications in water quality

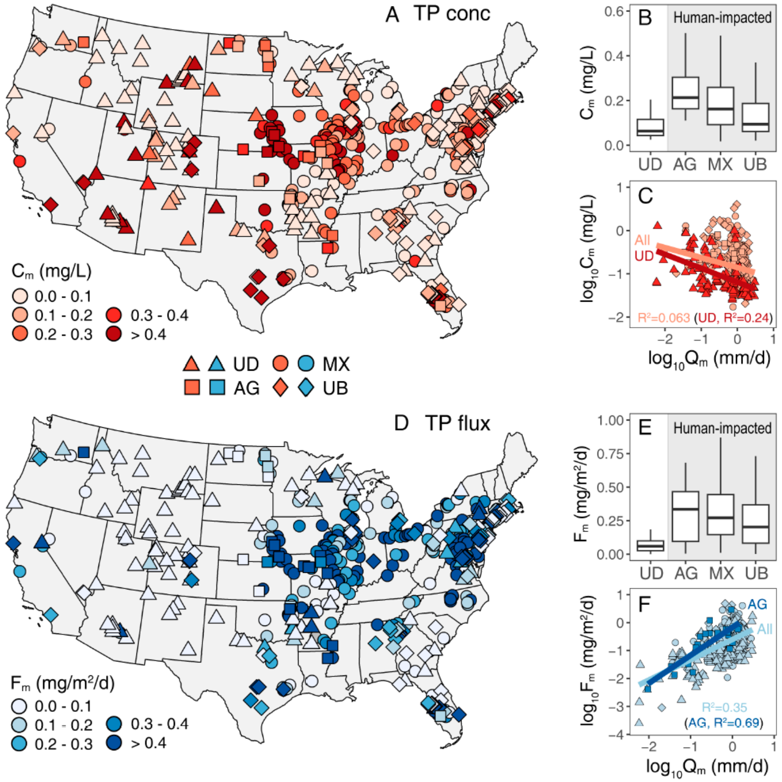


Water temperature

Rahmani et al., 2021, ERL



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



PNAS RESEARCH ARTICLE | ENVIRONMENTAL SCIENCES

Increasing phosphorus loss despite widespread concentration decline in US rivers

Wei Zhi[a,b,1], Hubert Baniecki[c,d], Jiangtao Liu[b], Elizabeth Boyer[e,f], Chaopeng Shen[b], Gary Shenk[g], Xiaofeng Liu[b,f], and Li Li[b,1]

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

Journal of Hydrology
Volume 639, August 2024, 131573

ELSEVIER

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

ELSEVIER

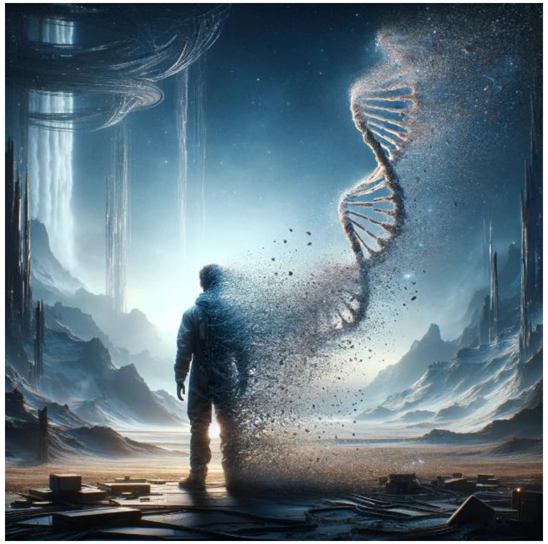
A deep learning-based novel approach to generate continuous daily stream nitrate concentration for nitrate data-sparse watersheds

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

[a] Department of Agricultural and Biological Engineering, The Pennsylvania State University, United States of America
[b] Department of Civil and Environmental Engineering, The Pennsylvania State University, United States of America

6

The "Good Genes of AI"



Genetic absorption of AI into our domains!

AI Gene	**enables…**
Large-depth NNs	Highly-complex functions
Minibatches and **GPU concurrency**	High data throughput
Differentiable programming	"End-to-end" training of large NNs
Knowledge management, etc……	

Learn from big data

Significant limitations

- **Not interpretable** --- No physical concepts
- **Right results for the wrong reason** (sensitivity)?
- **Unseen cases**? **Data-scarce** regions?
- **Scenarios?**
- Cannot answer **specific questions**

What if we can seamlessly connect NNs with process equations and learn from big data?

II. Physics-informed DL -- Differentiable Parameter Learning

nature COMMUNICATIONS

ARTICLE

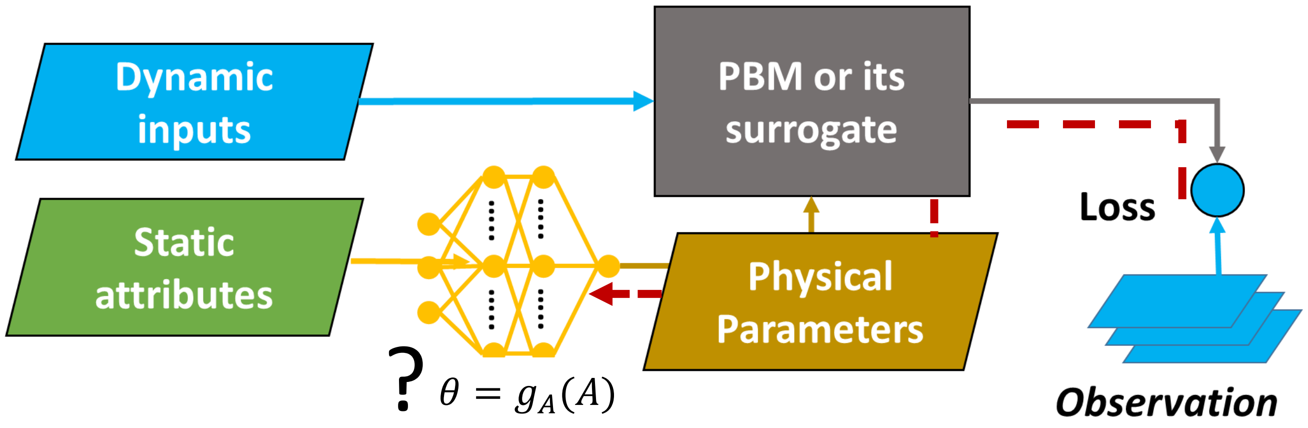
Check for updates

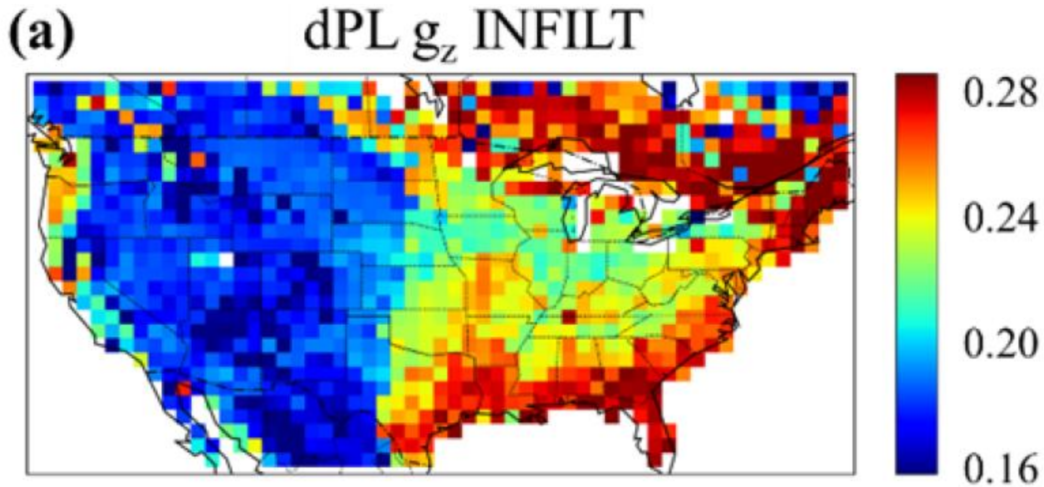
https://doi.org/10.1038/s41467-021-26107-z OPEN

From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling

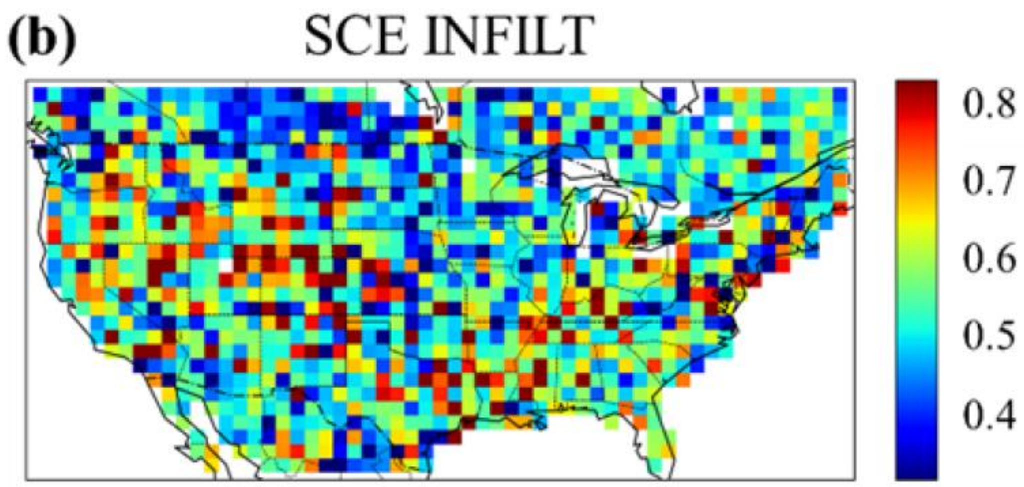
Wen-Ping Tsai[1], Dapeng Feng[1], Ming Pan[2,3], Hylke Beck[4], Kathryn Lawson[1,5], Yuan Yang[6,7], Jiangtao Liu[1] & Chaopeng Shen[1,5]✉

Address the "How" in Part III!





Regionalized parameterization
-- one technician learns to fix everyone's houses



Site-by-site calibration
-- we train a new technician for every house

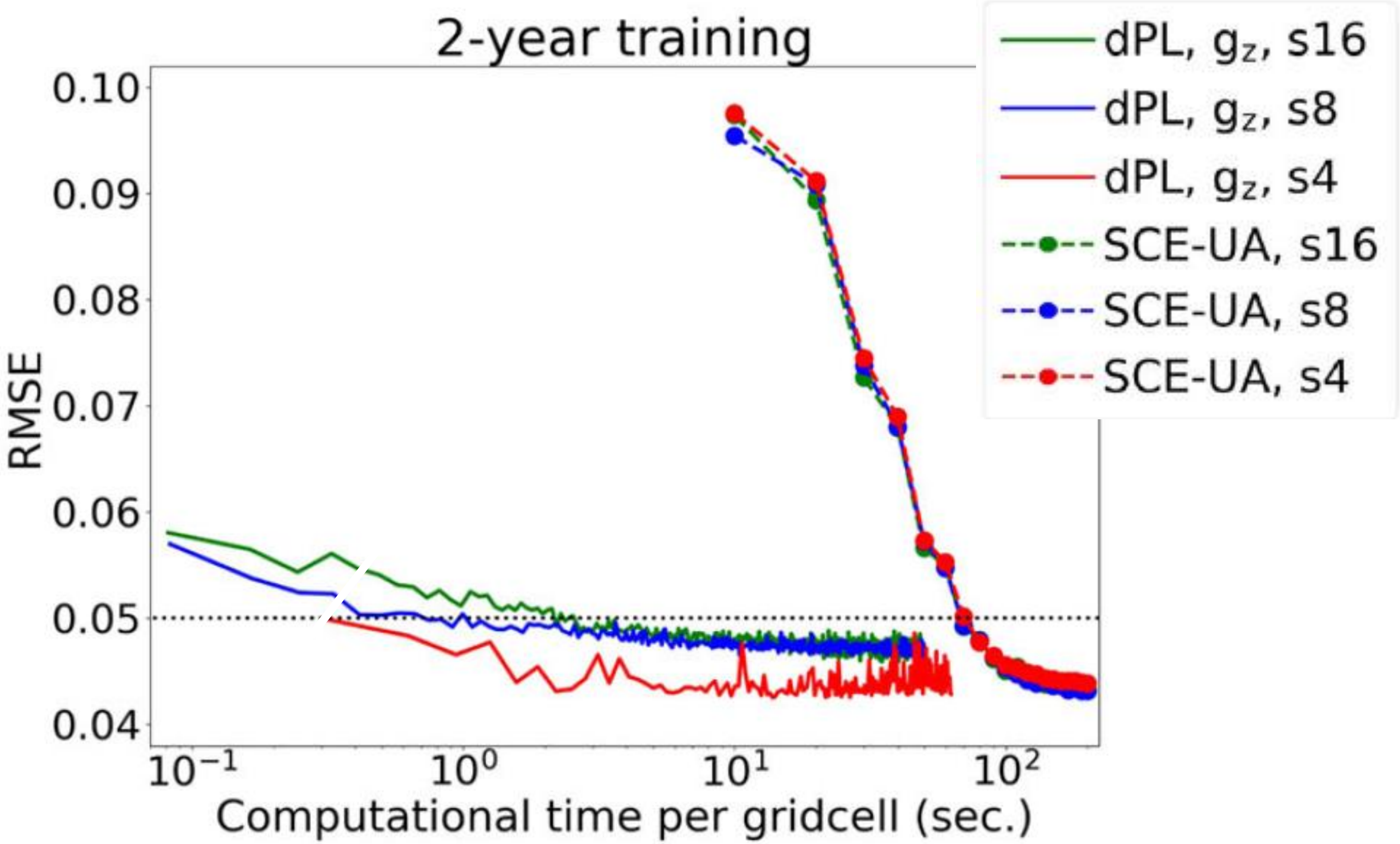
Data scaling relationships (network effect?)

1. dPL = SCEUA for lowest RMSE

2. dPL scales better with more data

3. Orders of magnitude more efficient: 100 proc 2-3 days vs. 1 GPU 2 hours

4. (not shown) better results for **untrained** variables and better **spatial generalization** than traditional approach!

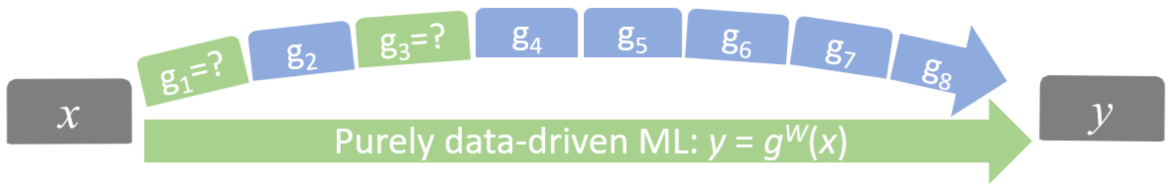


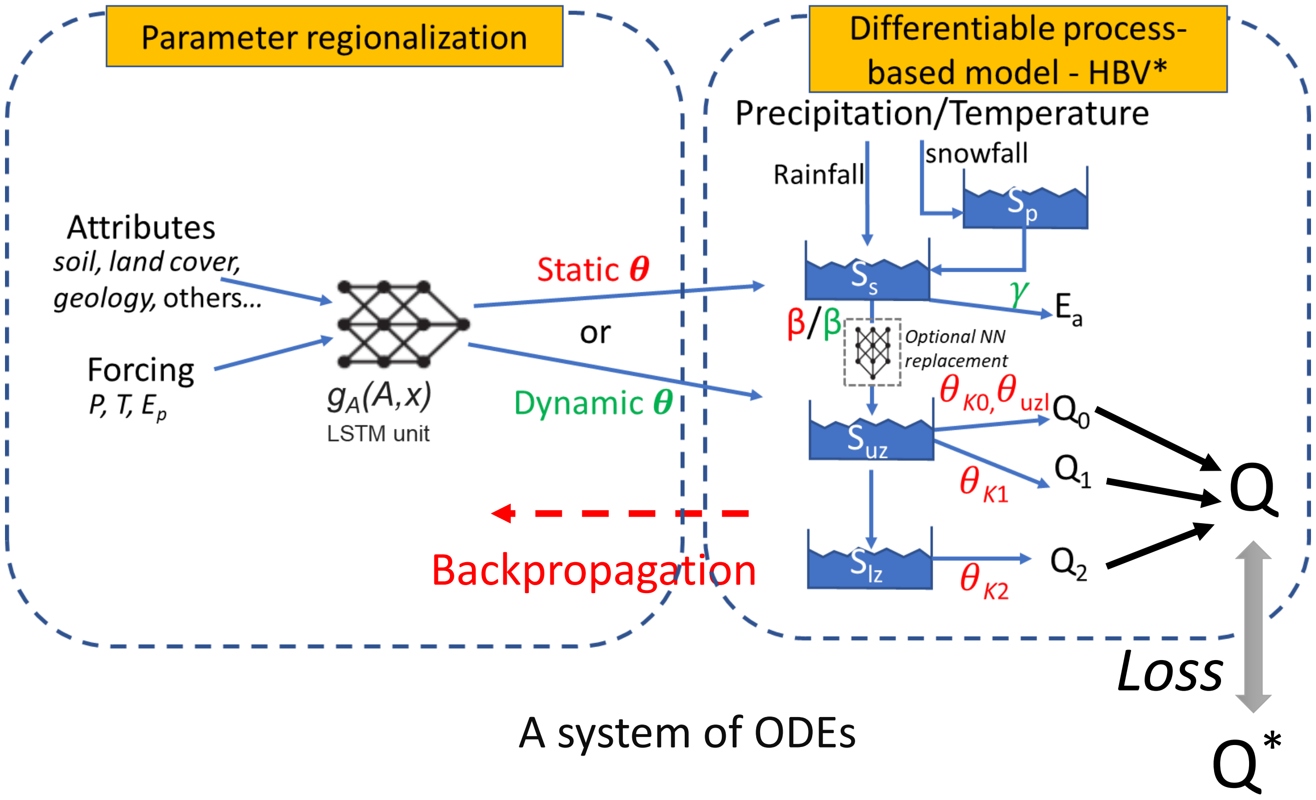
Tsai et al. 2021, Nature Communications

doi: 10.1038/s41467-021-26107-z

What is Differentiable Modeling?

DM

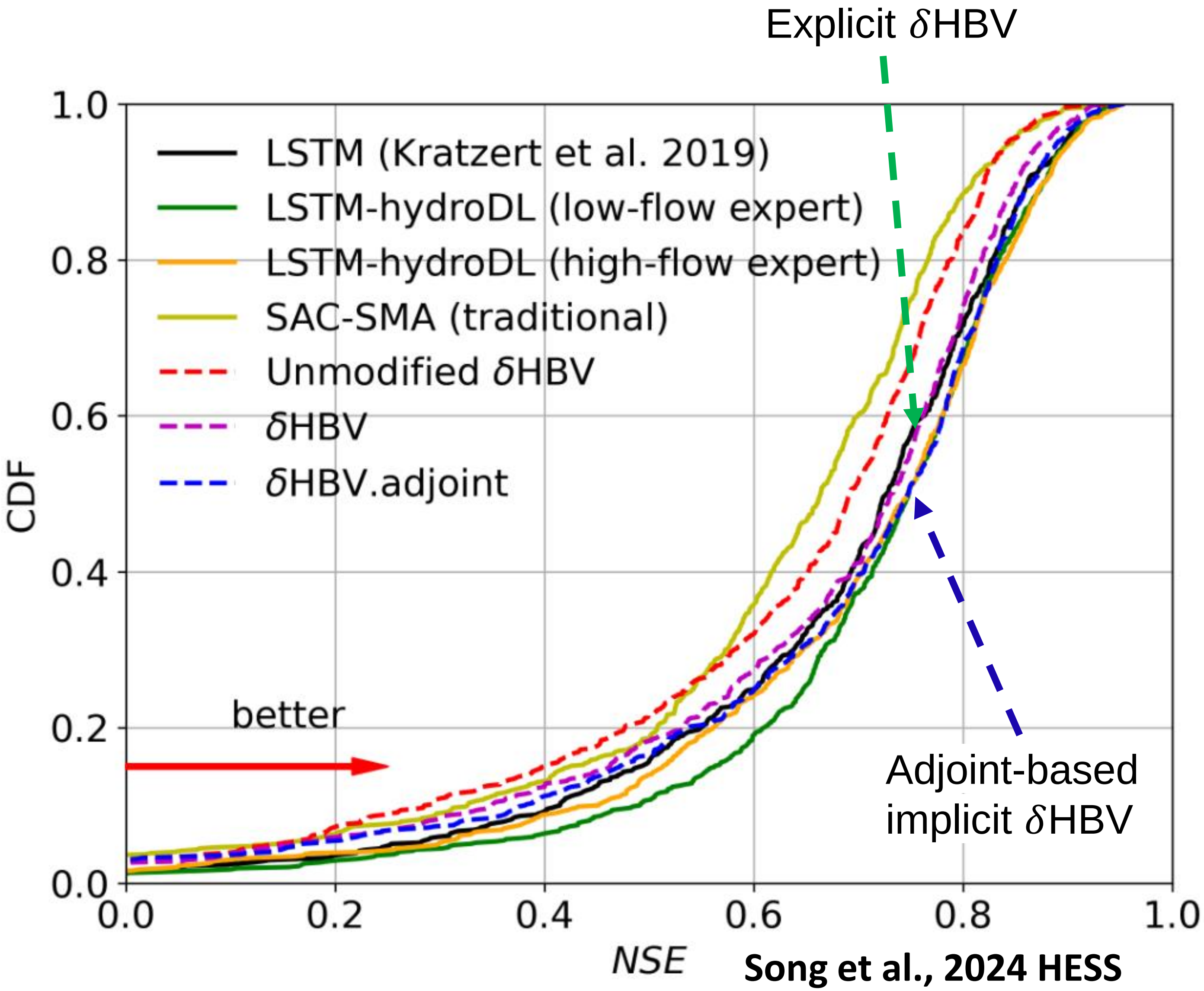




- NNs ("?") mixed w/ process-based equations (priors)
- Breaks a problem into parts, with some as priors
- "end-to-end" training on big data
- The priors constrain the learning to an interpretable scope.
- Can be used a forward simulator as well as

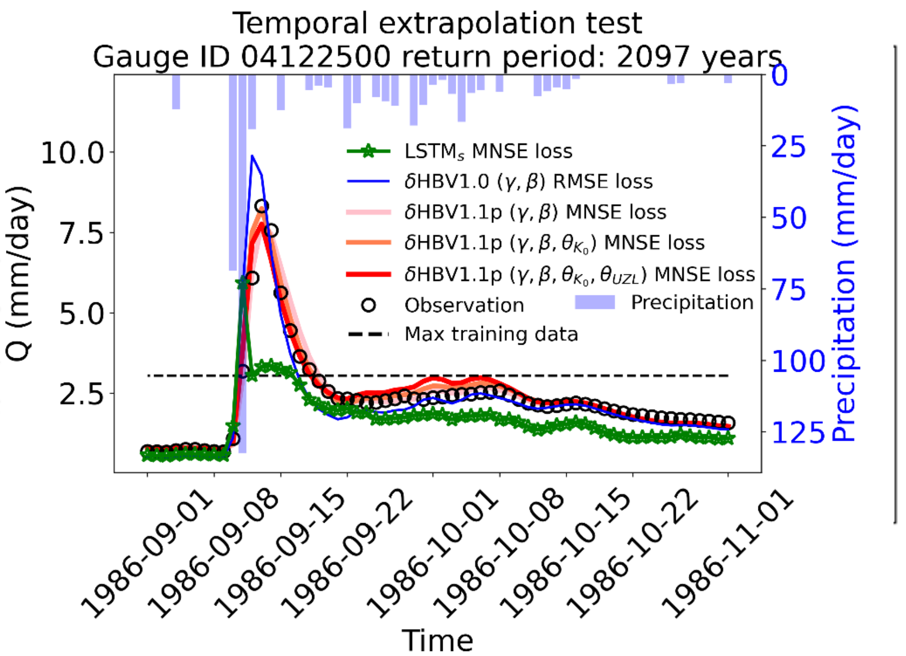
This is NOT physics-informed neural network (PINN)!

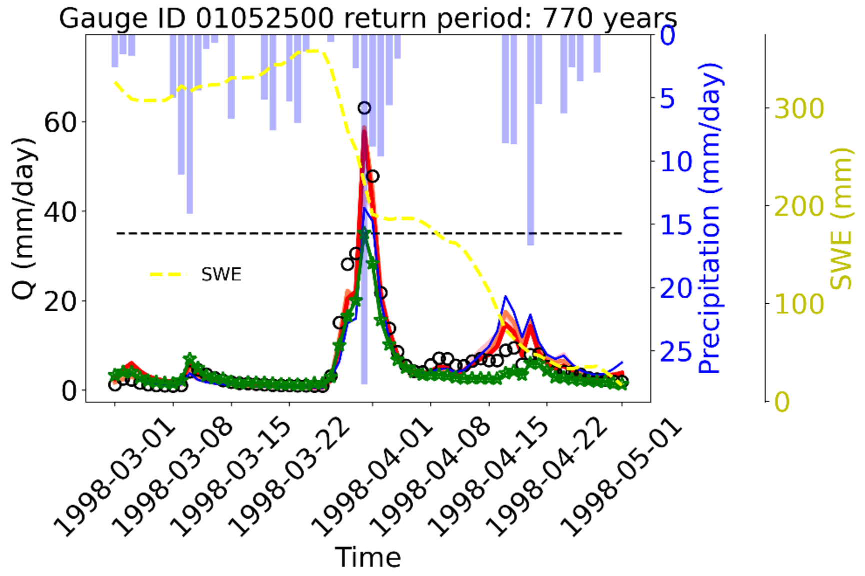
Performance fully equaling LSTM & better for unseen events



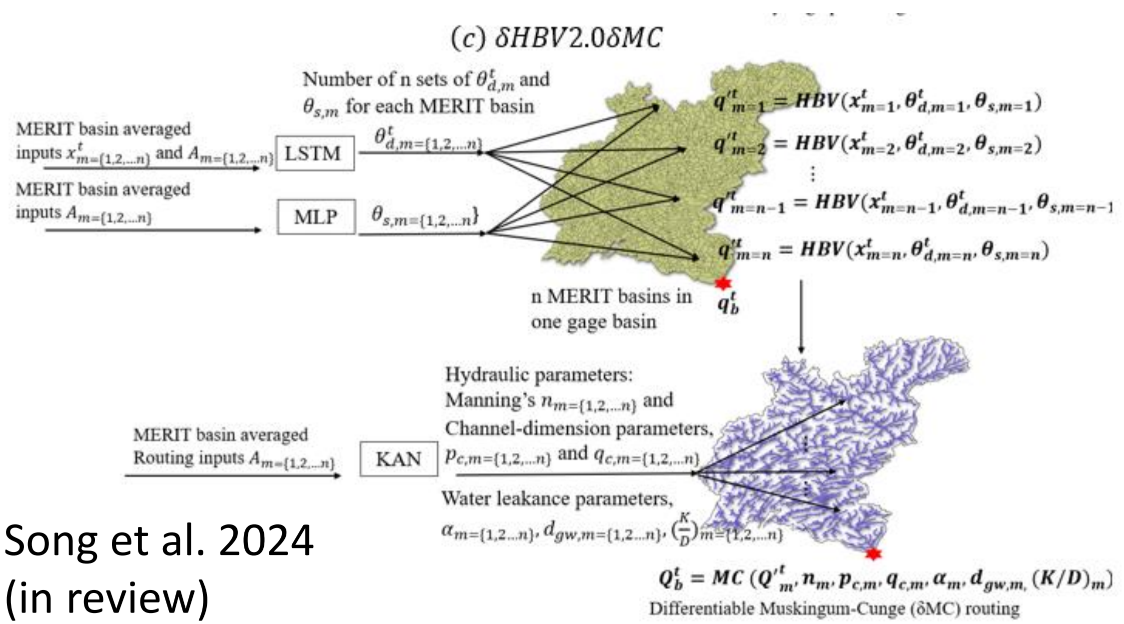
Song et al., 2024 HESS
doi: 10.5194/hess-2023-258

Evaluating on unseen extremes



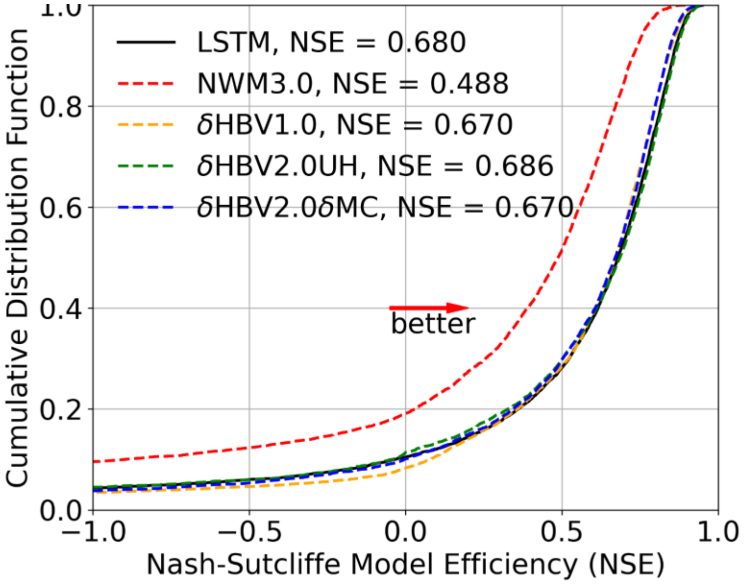


Large-scale, operational

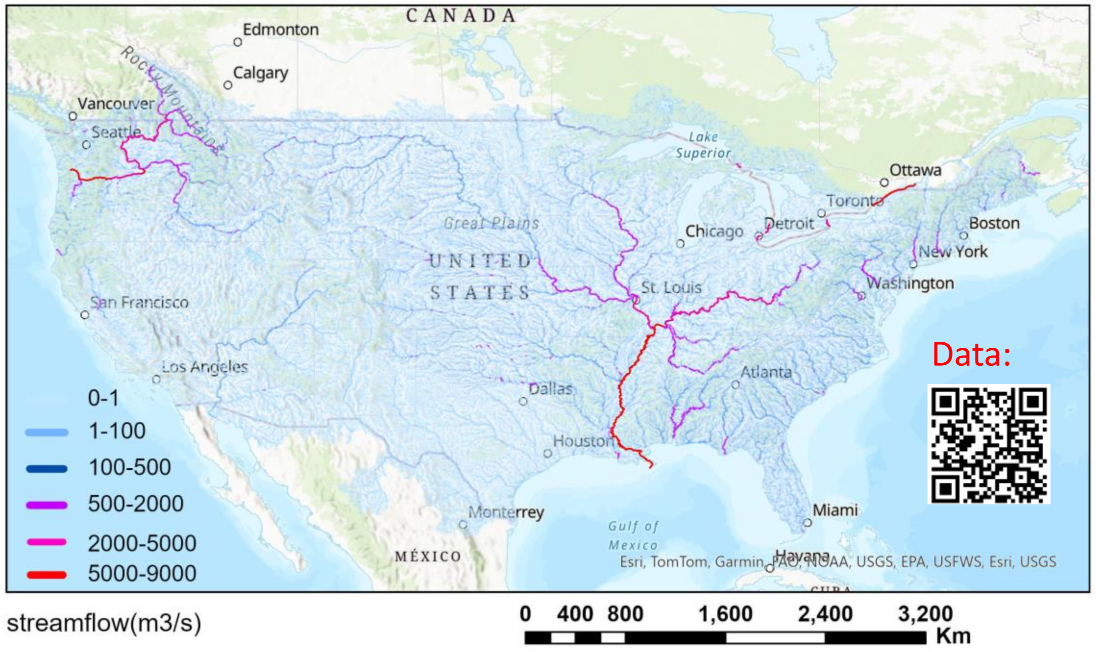


Song et al. 2024
(in review)

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

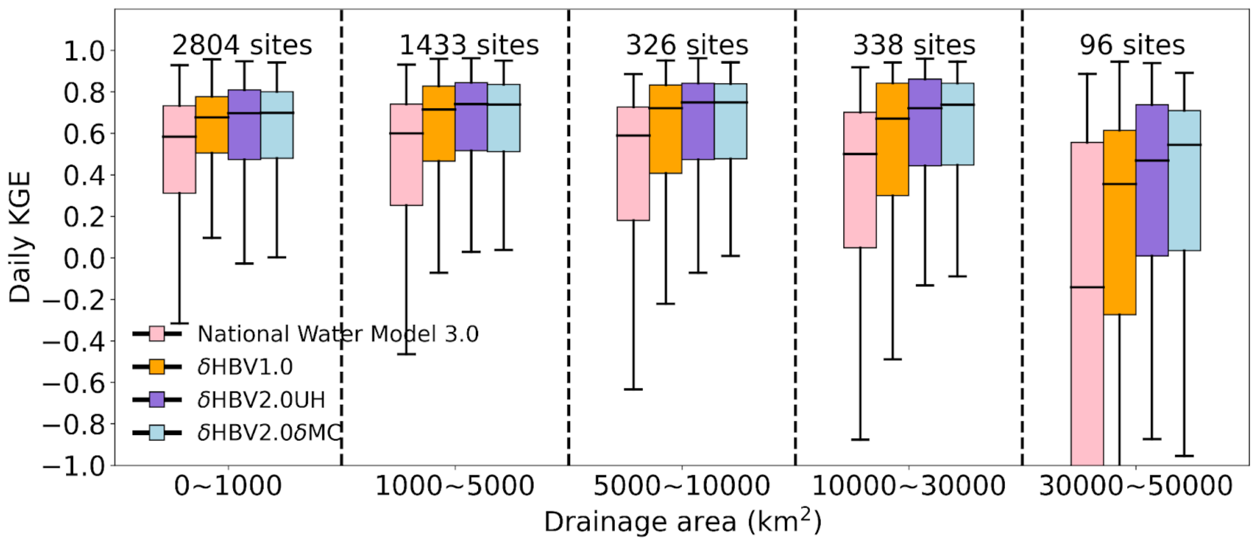


The mean seamless daily streamflow of 1995WY on the MERIT river network



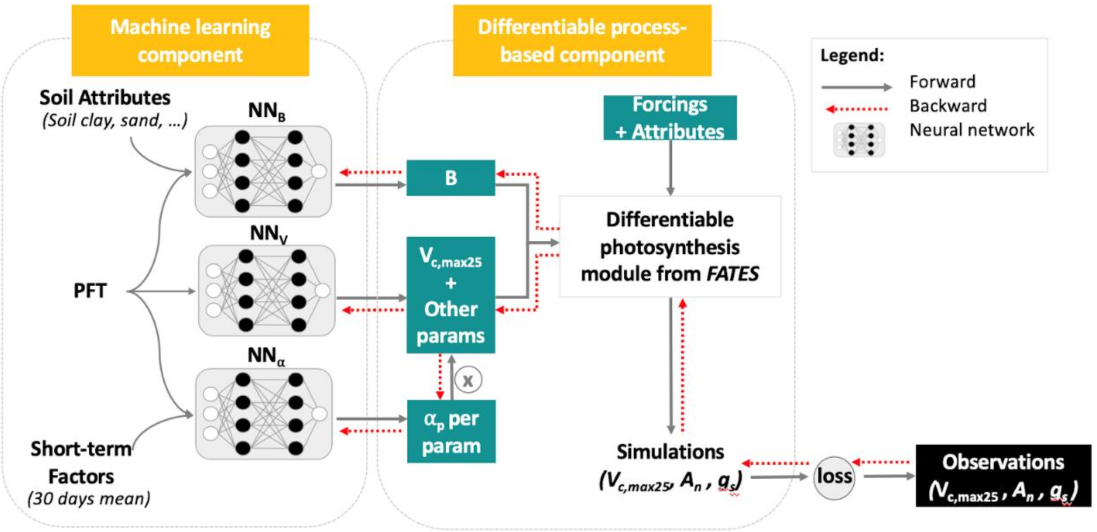
Data:

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



Why is DM transformative?

Widely & generically applicable

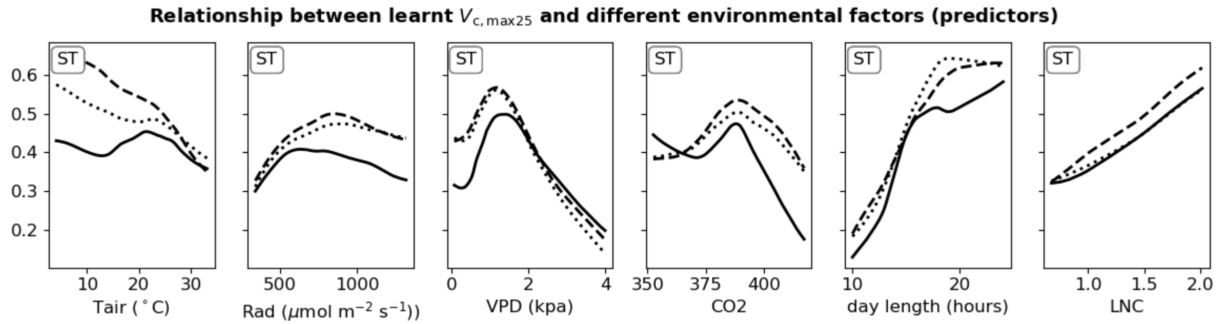


Building on our previous work:

Aboelyazeed et al., 2023 Biogeosciences

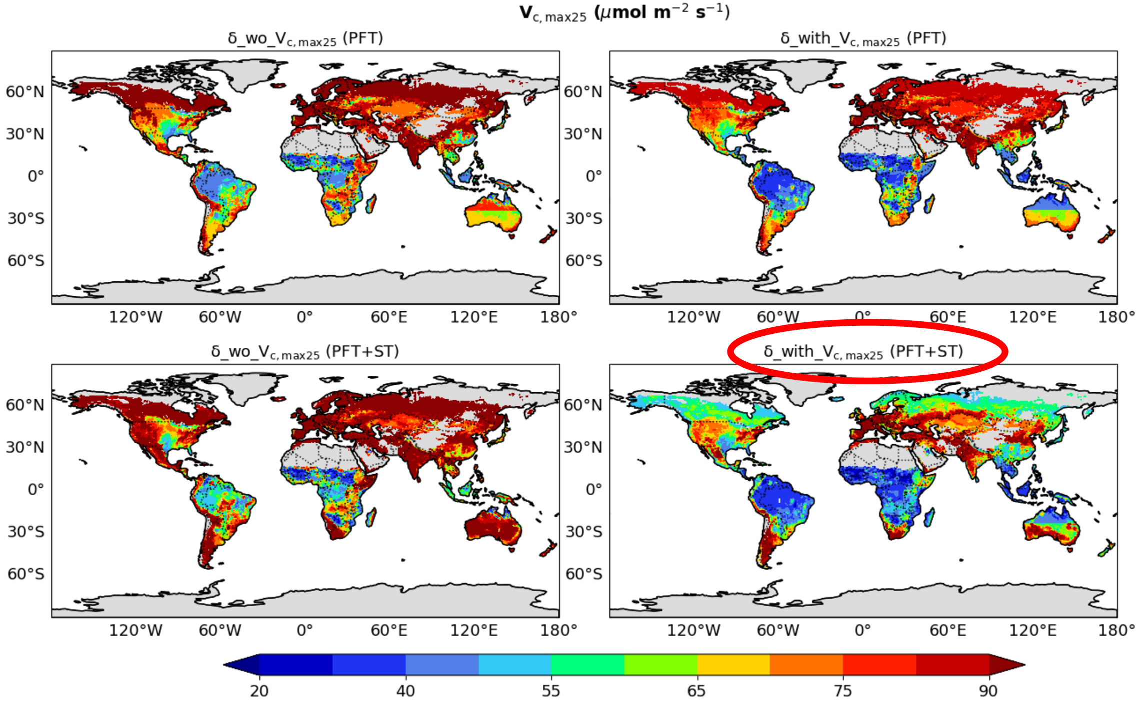
doi: 10.5194/bg-20-2671-2023

Learned acclimation functions:



Differentiable ecosystem modeling

New work:

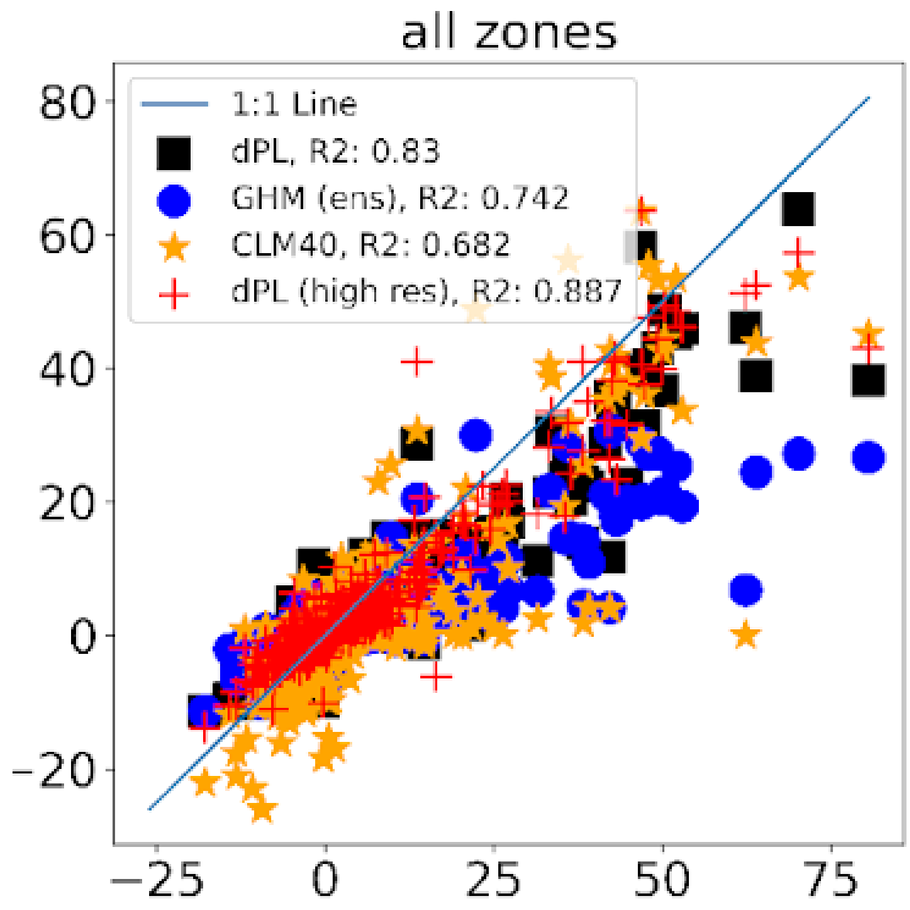


δ_{PFT} (V_{data})	$\delta_{PFT+Env}$	$\delta_{PFT+Env}$ (V_{data})
-18.47	3.14	-17.47

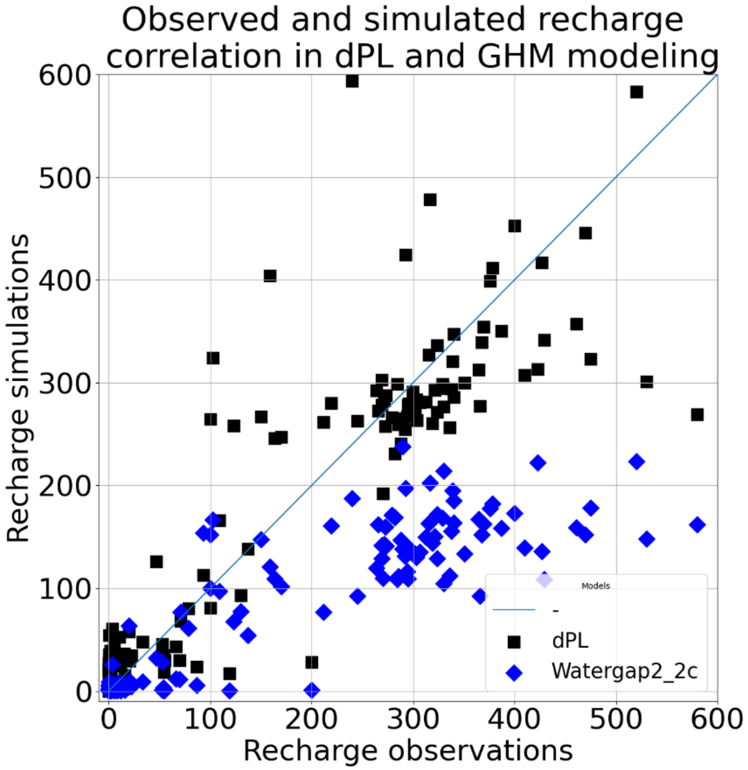
w/ Acclimation

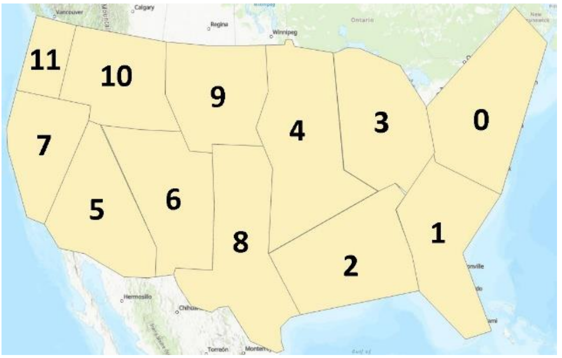
Training on both discharge & temperature

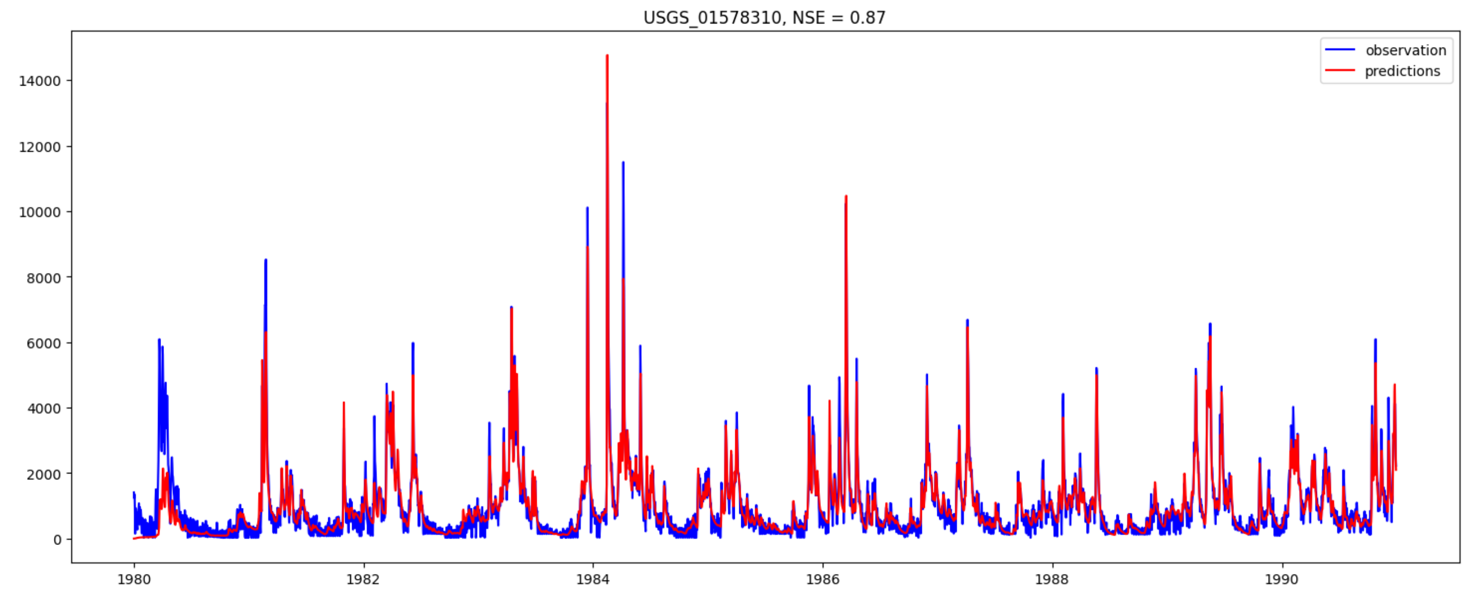
Baseflow trends

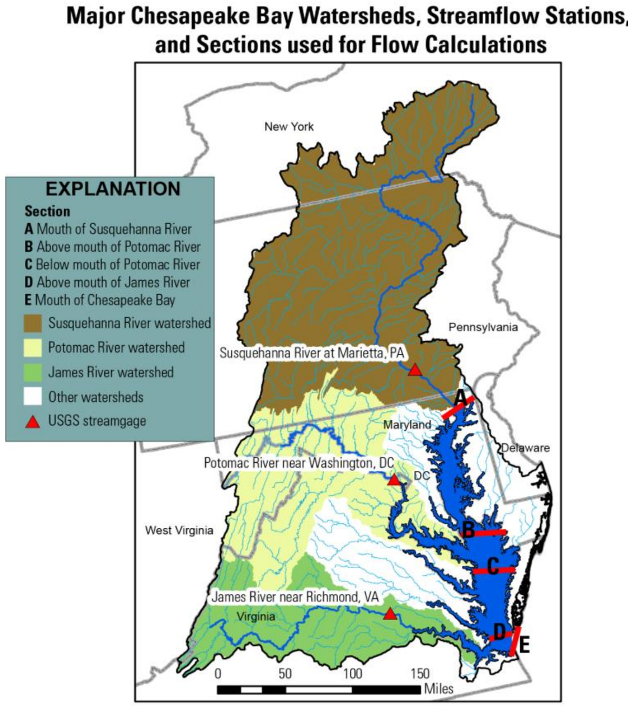


Recharge

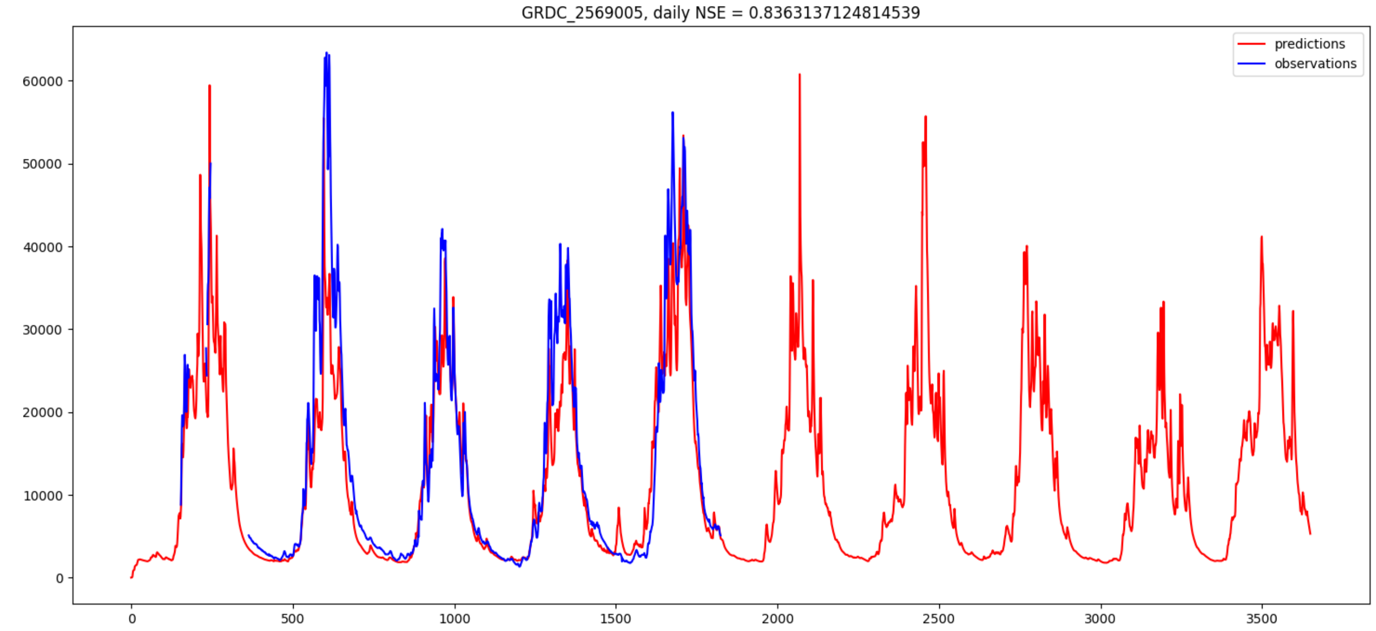








Chesapeake Bay

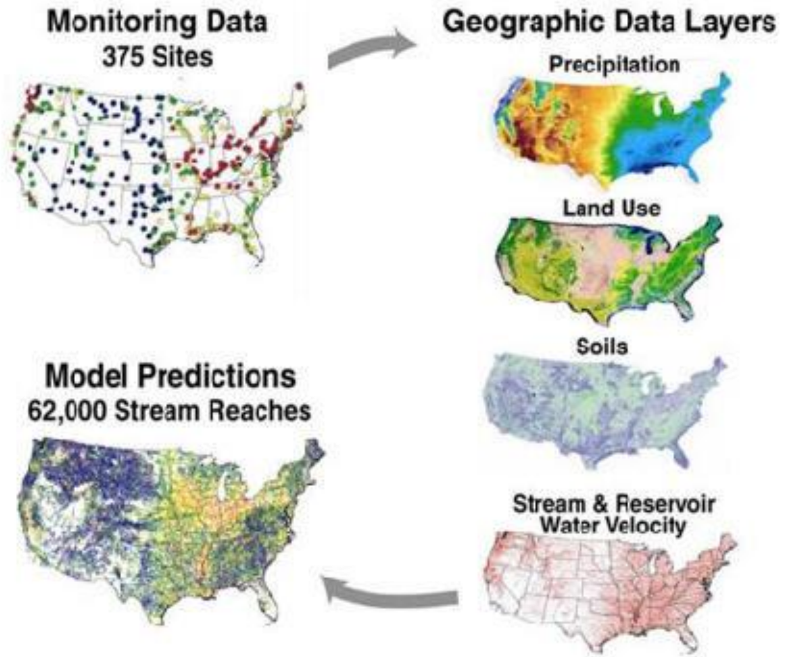


Mekong

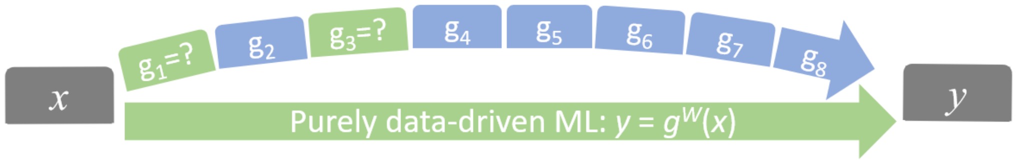
Outlooks for WQ

Our experiences from hydrology can directly translate into WQ modeling.

- Differentiable WQ modeling
- Scale-relevant predictions & diagnosis (landform, manage practices) w/ UQ
- Knowledge discovery (need priors!) learn from data what it can describe
- Foundation model for capturing coevolution
- Solving PDEs & Fluids



https://www.usgs.gov/mission-areas/water-resources/science/everything-you-need-know-about-sparrow



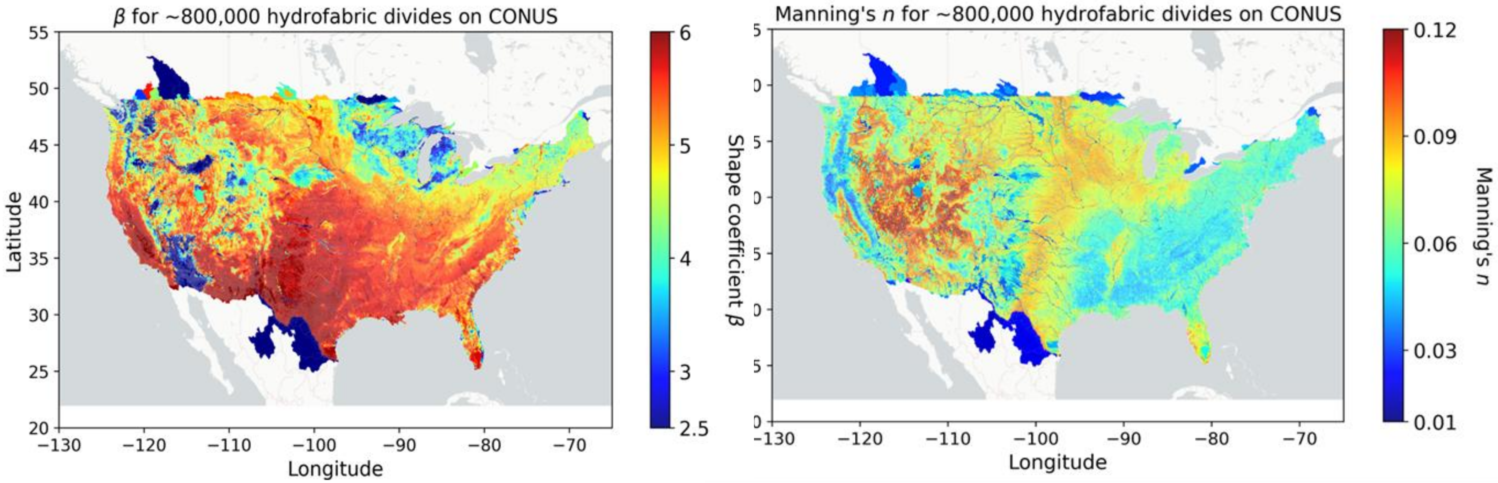
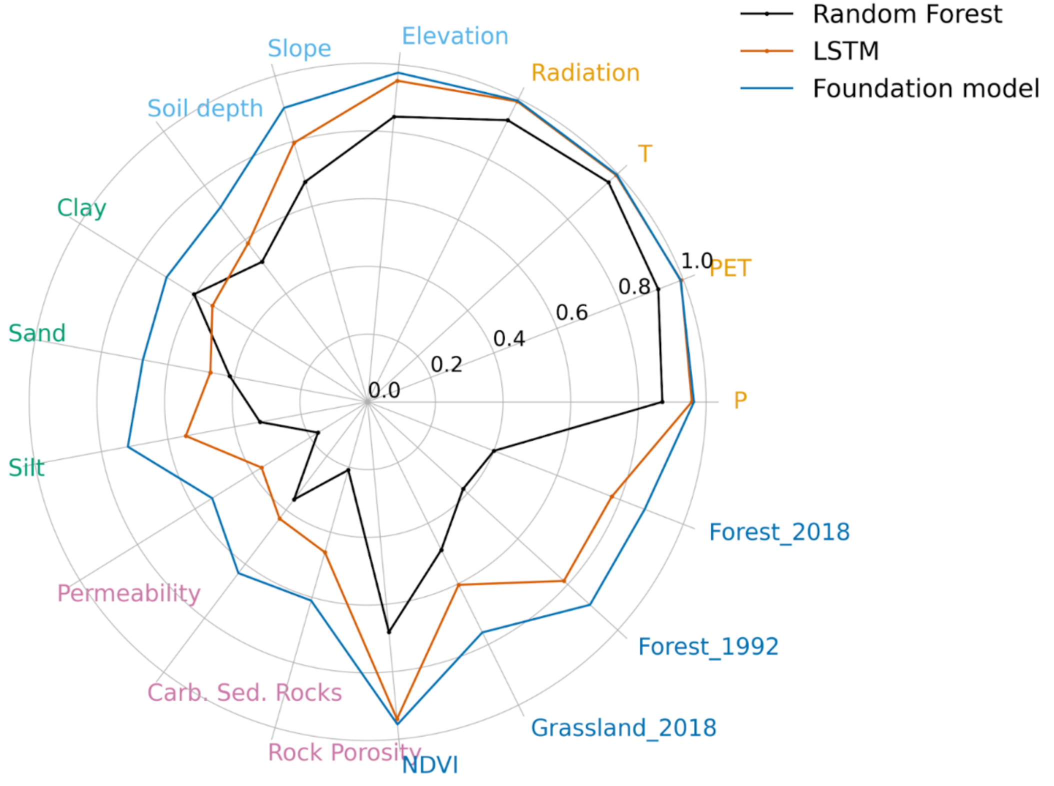
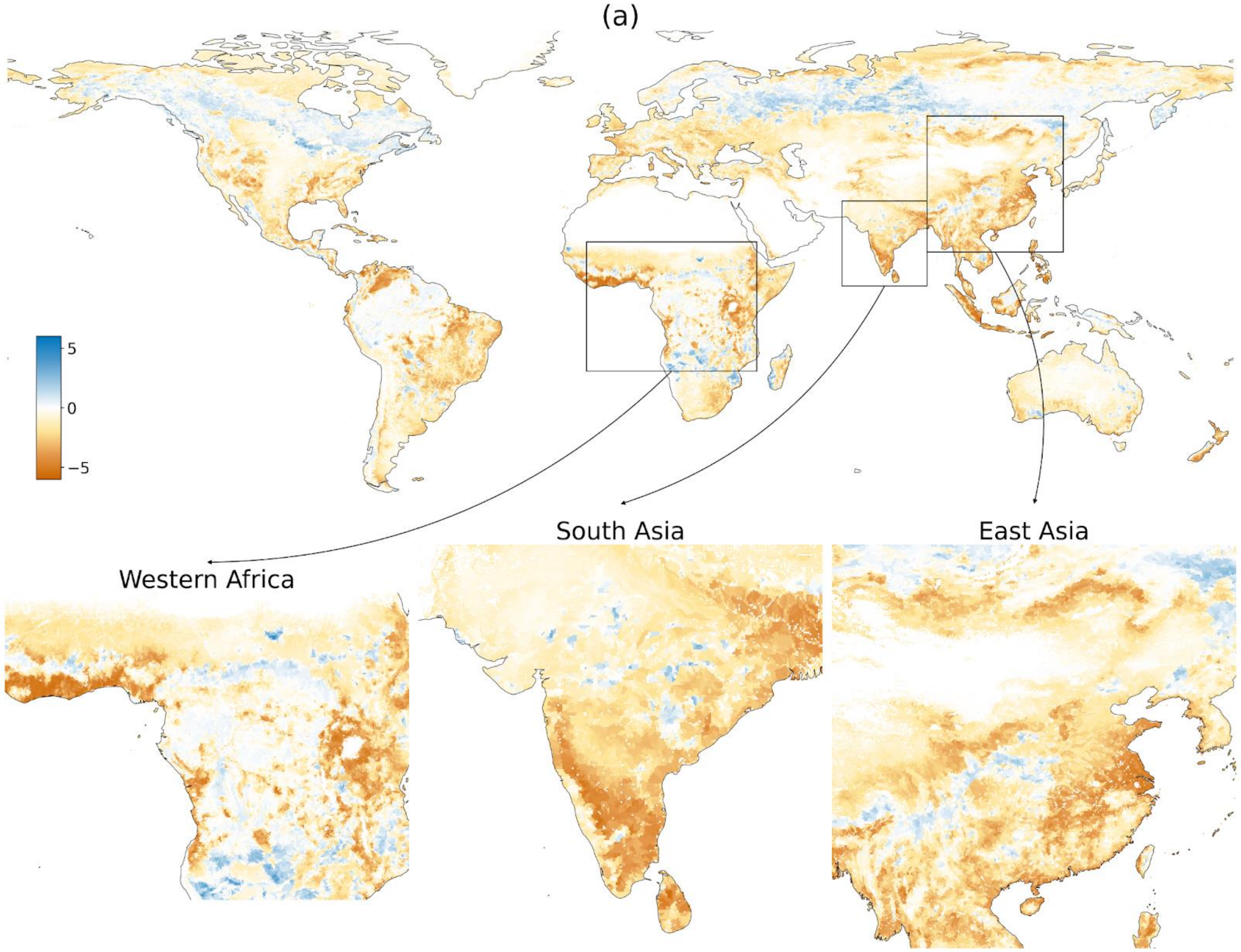


Figure 2

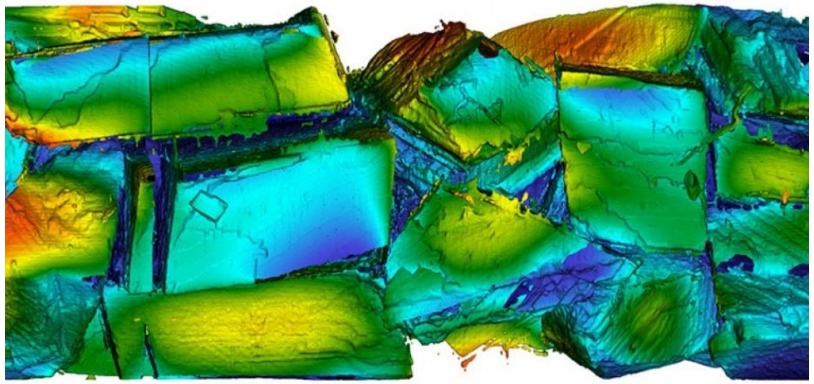
Foundation model: capturing the joint distribution of the landscape



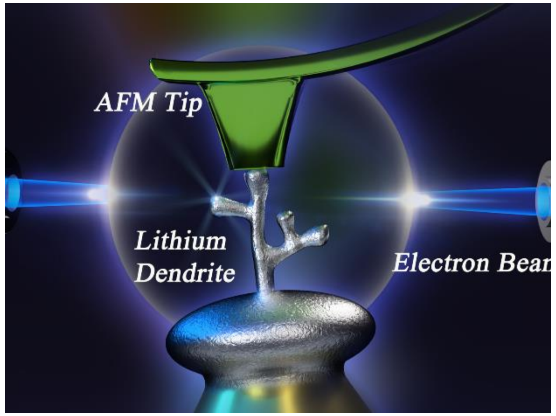


Outlooks --- what will be enabled in the future?

Reactive transport



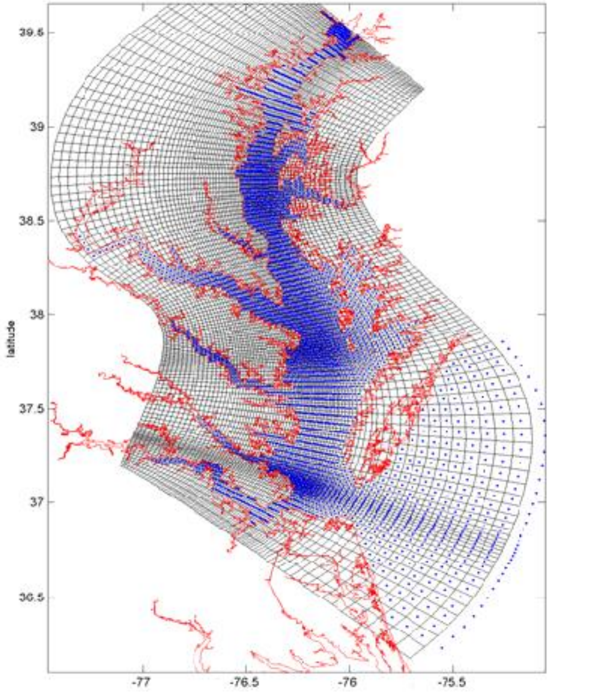
https://www.alcf.anl.gov/science/projects/chombo-crunch-modeling-pore-scale-reactive-transport-carbon-sequestration



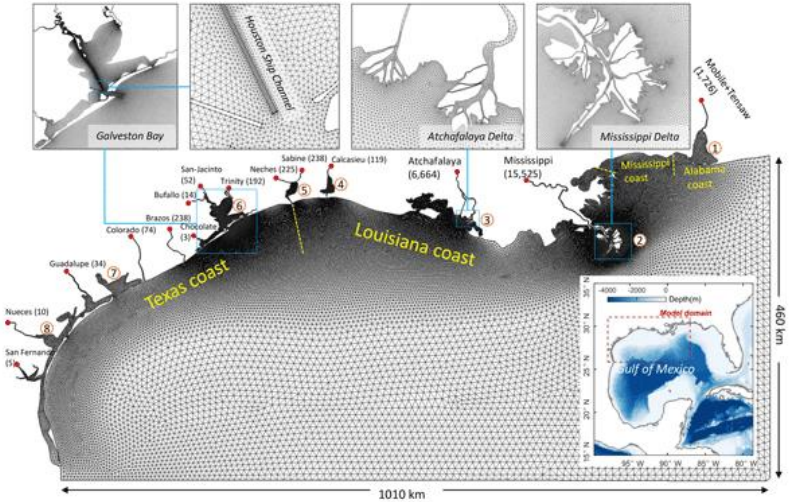
https://www.psu.edu/news/research/story/new-method-study-lithium-dendrites-could-lead-better-safer-batteries/

Battery dendrites

Lake circulation

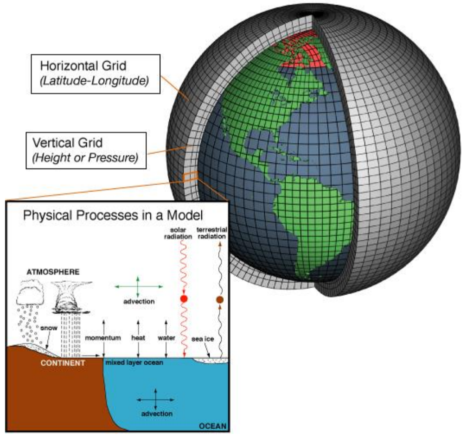


https://www.ices.dk/sites/pub/CM%20Doccuments/2006/O/O1106.pd



https://os.copernicus.org/articles/15/951/2019/

Coastal flooding

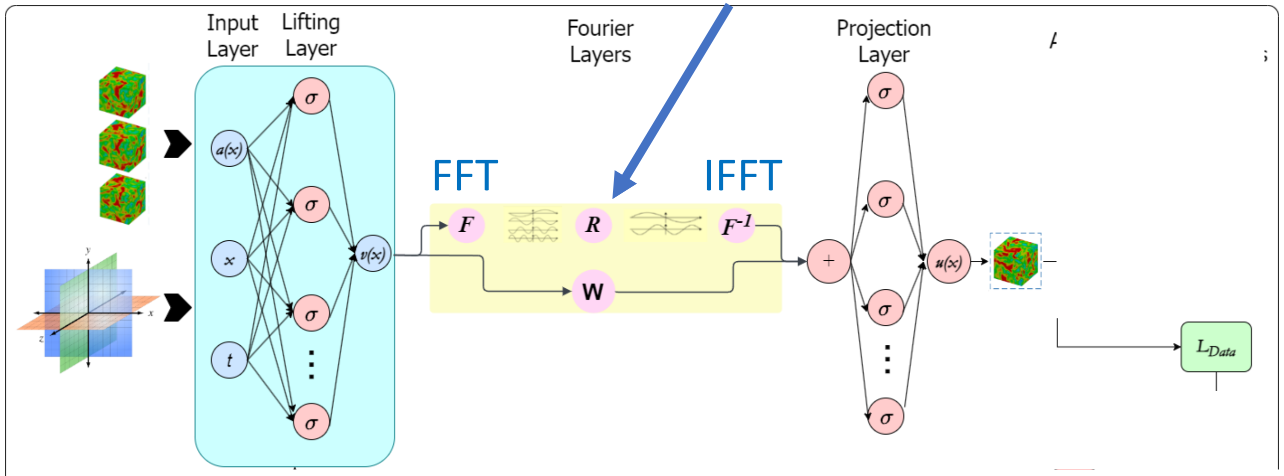


Numerical weather models

How can we do it faster & cheaper?

AI-fused Neural operators!

FNO learns operators in Fourier space
--- differential operators are multiplications in Freq space

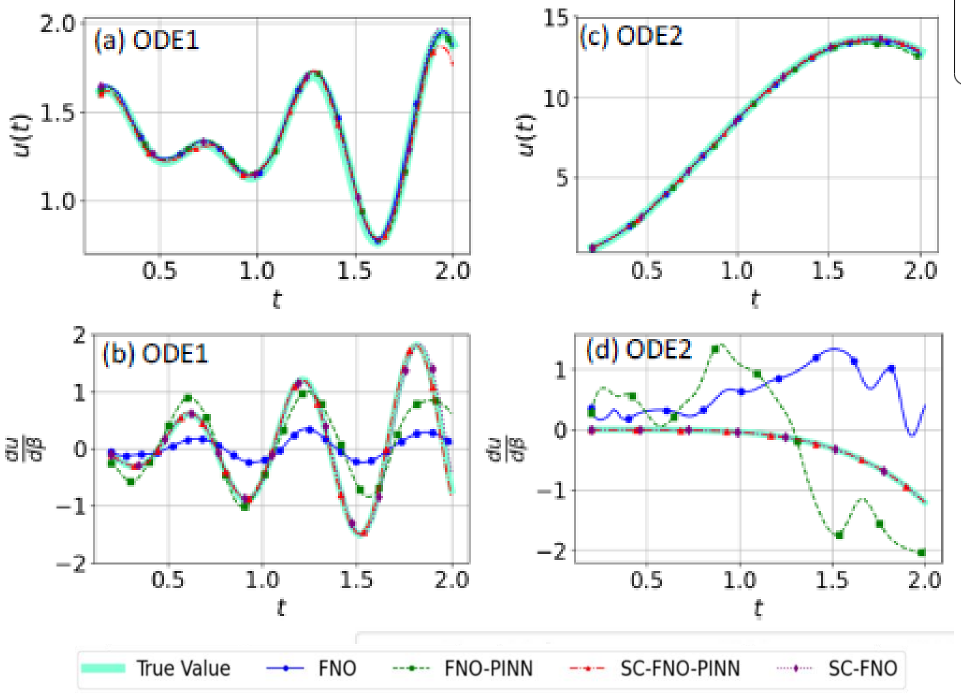


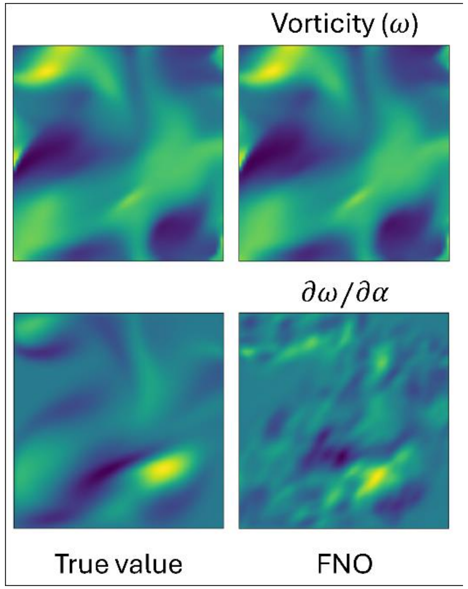
Zongyi Li… Anandkumar https://arxiv.org/abs/2010.08895

- du/dt = f(u, t, x, p)
- Fourier Neural Operator (FNO) solves PDEs: >O(10^4) faster!
- No time stepping is required!
- …. But but but, sensitivity (du/dp) often wrong
- Training via experience…. Very data intensive, what about training via instruction?

Solution u

Jacobian du/dp





2D Navier-Stokes

Behroozi et al., 2025 International Conference on Learning Representations (ICLR 2025)

Adding Jacobian into training: sensitivity much improved

- du/dt = f(u, t, **x**, p) (w/ a differentiable solver)
- Novel idea: train on not only the solution but also the Jacobian **du/dp,** leveraging our differentiable solver!

Adding the Jacobian loss (to the main solution loss):

$$L_s = \frac{1}{M}\sum_{j=1}^{M}\left\|\frac{\partial \hat{\mathbf{u}}(\mathbf{x}_j, t_j; \mathbf{p})}{\partial \mathbf{p}} - \frac{\partial \mathbf{u}(\mathbf{x}_j, t_j; \mathbf{p})}{\partial \mathbf{p}}\right\|^2$$

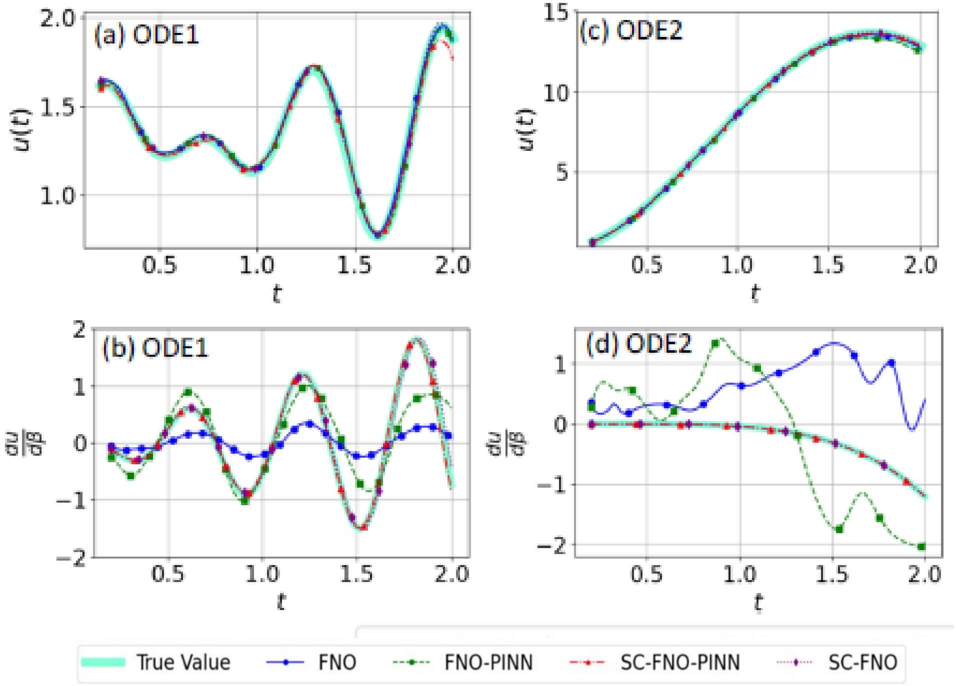
Saved from differentiable solver (w/ FD, AD or Adjoint)

Through the FNO

Within training input range

Solution u

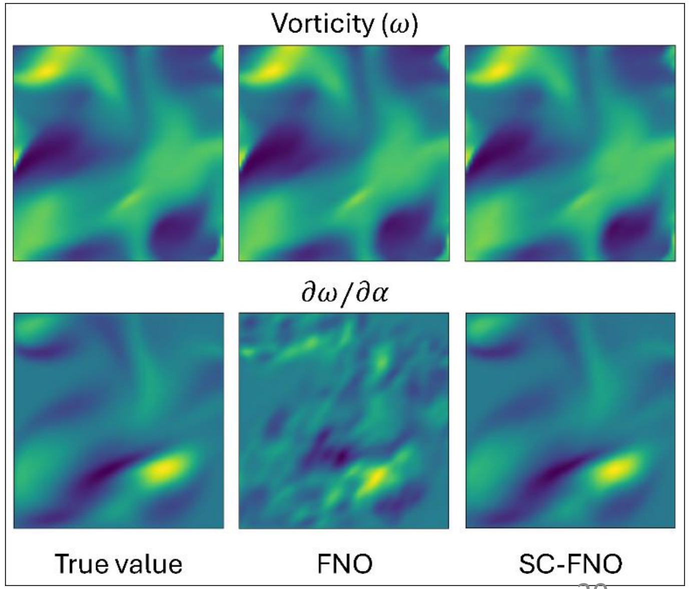
Jacobian du/dp



Within training input range

Solution u

Jacobian du/dp



2D Navier-Stokes

Behroozi et al., 2025 International Conference on Learning Representations (ICLR 2025)

Robust for perturbation which allows successful inversion (SC-FNO is NOT PINN)

Perturbation out of training range:

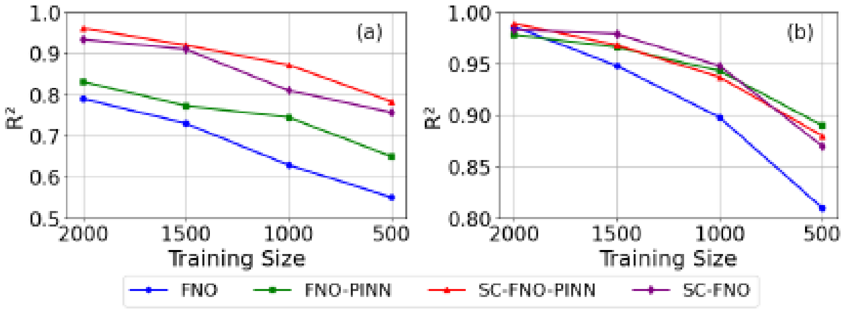


Figure 4: Models' performance on PDE1 across training sample sizes, (a) $\frac{\partial u}{\partial \omega}$ (b) $u(t)$.

Parameter Inversion:

1 parameter

5 parameters

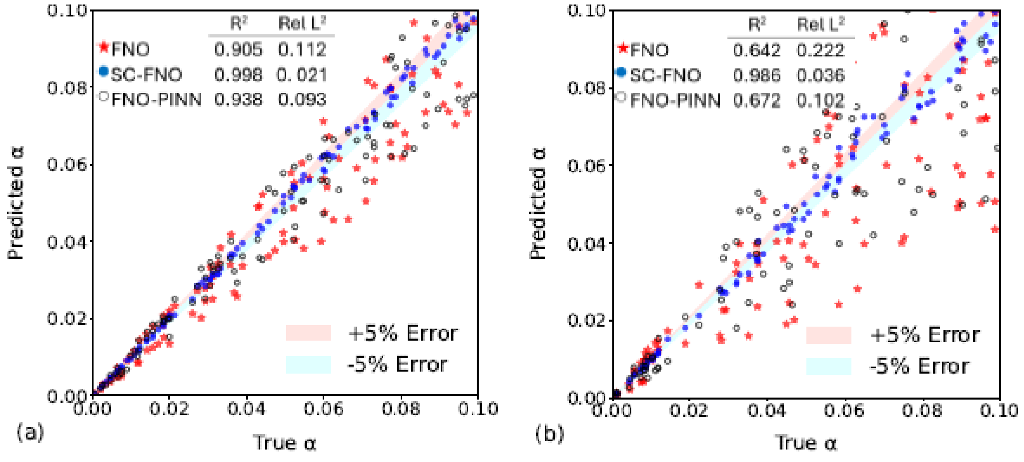


Figure 1: Inversion of the parameter α in PDE1 using FNO and SC-FNO models (a) single parameter inversion, (b) simultaneous multi-parameter inversion.

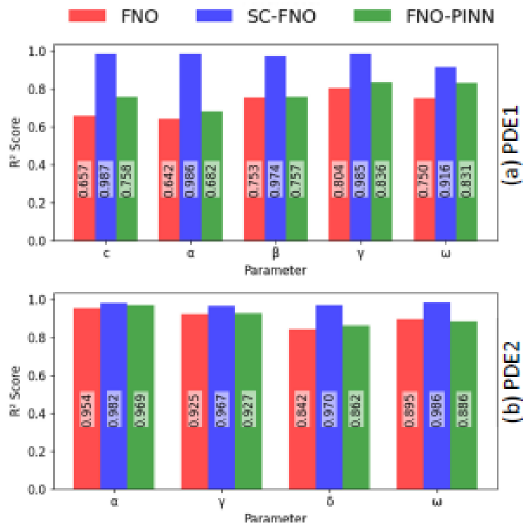


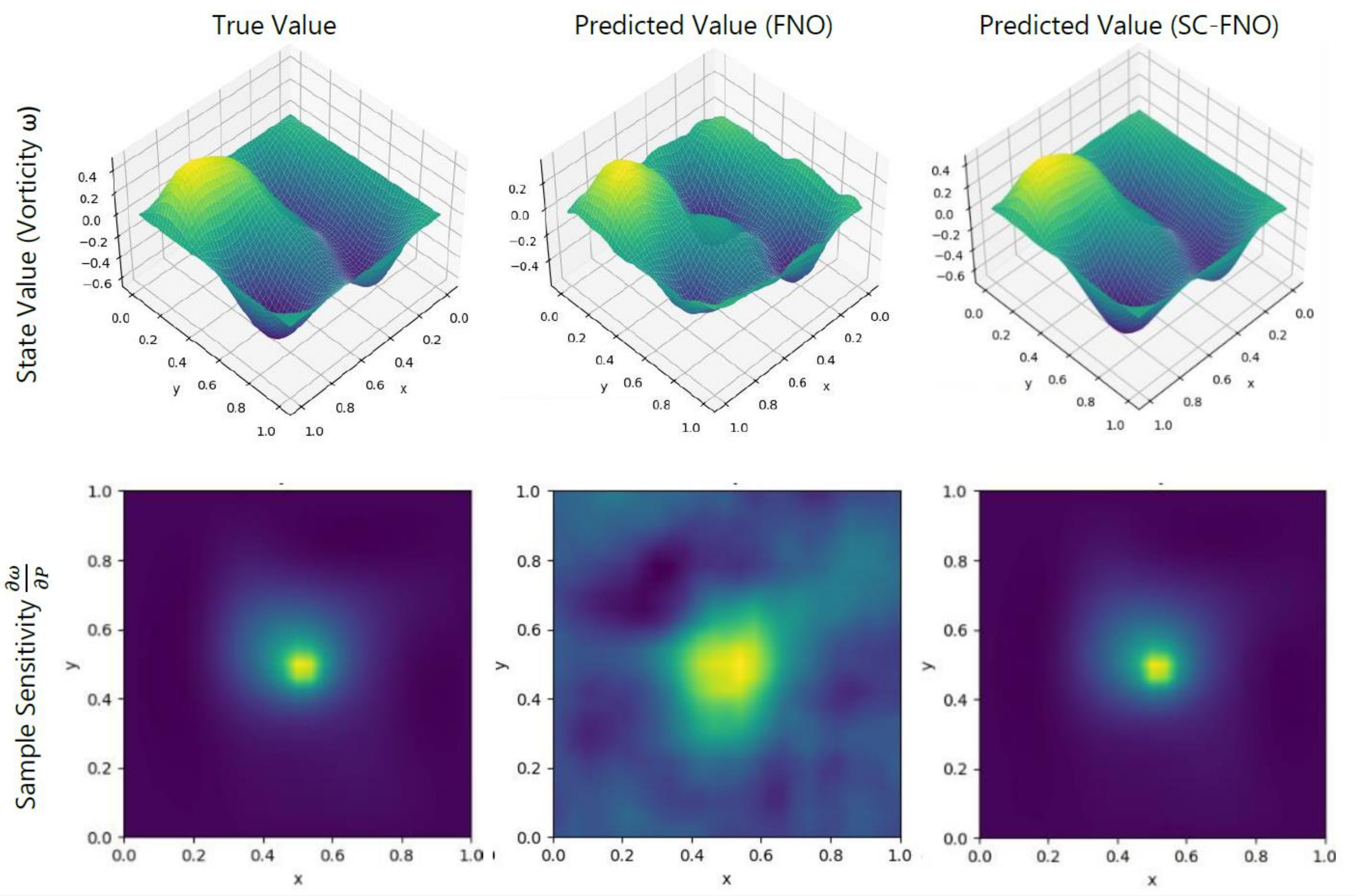
Figure 2: Simultaneous multi-parameter inversion accuracy for PDEs 1 and 2 using FNO and SC-FNO.

SC-FNO is NOT PINN.

PINN loss is only slightly useful: du/dp not in the equation!

The approach applies to distributed parameters (unpublished) --- forced Reynolds-Averaged Navier Stokes (RANS with spatially-varying forcing)

- 10^4 inputs+parameters!
- Default FNO --- not for inversion
- SC-FNO --- highly accurate solution & gradients



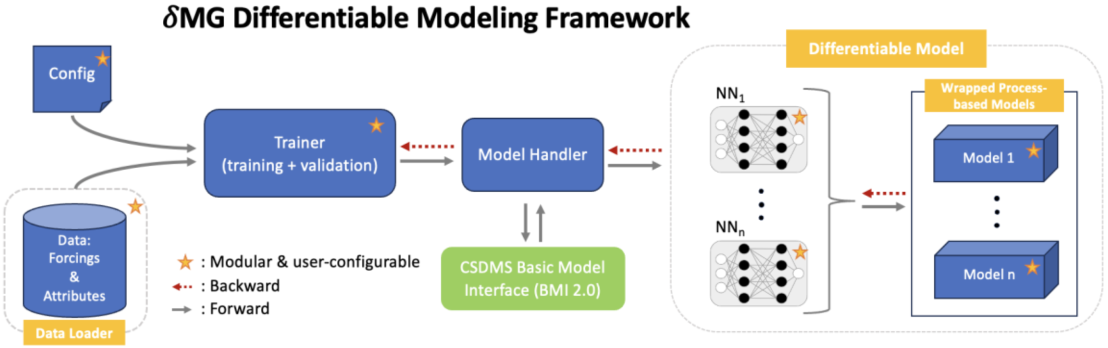
Fully-grown differentiable model

PBM

PBM+dPL

Summary

- Differentiable model is a game changer for continental-scale flood prediction or global scale climate change impact assessment for the water sector.
- SC-neural operator will serve very efficient and robust surrogate models.
- We now provide a **unified differentiable modeling framework, δMG**:



Our Code Collection: https://mhpi.github.io/
δMG: https://mhpi.github.io/codes/frameworks/

Poster (Lonzarich):
IN33B-2071

Water Resources Podcast: youtu.be/rzVU03OAIdQ

Apple Finch Pudding Science Podcast: https://youtu.be/nl30u65XZro

AGU TV: https://youtu.be/BIBBIM0BWaU

Flood Forecasting Workshop: https://bit.ly/3wvLV6N

Water Resources Research
REVIEW ARTICLE
10.1029/2018WR022643
Special Section: Big Data & Machine Learning in Water Sciences: Recent Progress and Their Use in Advancing Science

A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists

Chaopeng Shen[1]
[1]Civil and Environmental Engineering, Pennsylvania State University, University Park, PA, USA

nature COMMUNICATIONS
ARTICLE
https://doi.org/10.1038/s41467-021-26107-z OPEN

From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling

Wen-Ping Tsai[1], Dapeng Feng[1], Ming Pan[2,3], Hylke Beck[4], Kathryn Lawson[1,5], Yuan Yang[6,7], Jiangtao Liu[1] & Chaopeng Shen[1,5]

nature reviews earth & environment
https://doi.org/10.1038/s43017-023-00450-9
Perspective

Differentiable modelling to unify machine learning and physical models for geosciences

Chaopeng Shen[1], Alison P. Appling[2], Pierre Gentine[3], Toshiyuki Bandai[4], Hoshin Gupta[5], Alexandre Tartakovsky[6], Marco Baity-Jesi[7], Fabrizio Fenicia[7], Daniel Kifer[8], Li Li[1], Xiaofeng Liu[1], Wei Ren[9], Yi Zheng[10], Ciaran J. Harman[11], Martyn Clark[12], Matthew Farthing[13], Dapeng Feng[1], Praveen Kumar[6,14], Doaa Aboelyazeed[1], Farshid Rahmani[1], Yalan Song[1], Hylke E. Beck[15], Tadd Bindas[1], Dipankar Dwivedi[16], Kuai Fang[17], Marvin Höge[7], Chris Rackauckas[18], Binayak Mohanty[19], Tirthankar Roy[20], Chonggang Xu[21] & Kathryn Lawson[1]

PNAS
RESEARCH ARTICLE | ENVIRONMENTAL SCIENCES

Increasing phosphorus loss despite widespread concentration decline in US rivers

Wei Zhi[a,b,1], Hubert Baniecki[c,d], Jiangtao Liu[b], Elizabeth Boyer[e,f], Chaopeng Shen[b], Gary Shenk[g], Xiaofeng Liu[b,f], and Li Li[b,1]
Affiliations are included on p. 8.

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