Hybrid modeling for land - or:
How might we think about developing new modeling techniques in the age of AI?

Andrew Bennett (he/him)
andrbenn@arizona.edu
Oct 15, 2025

👋 Hi, I'm Andrew Bennett

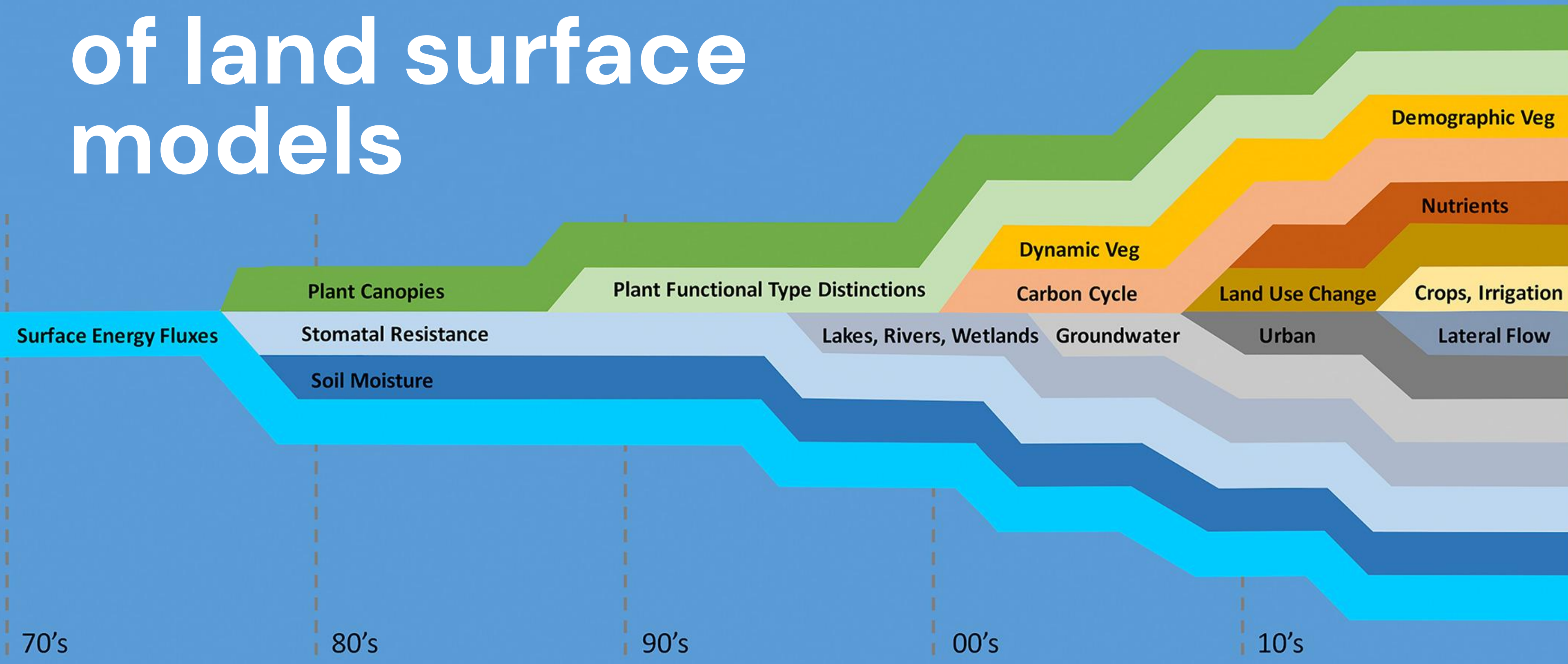
I am an assistant professor from the Dept. of Hydrology and Atmospheric Sciences and Statistics & Data Science Program at the University of Arizona

I've worked on a number of hydrologic modeling applications across scale and process

My main focus is using and developing deep learning models for hydrology

I am also interested in open source technologies for both research and teaching of Earth systems science

The evolution of land surface models



Fisher, R. A., and C. D. Koven (2020), Perspectives on the Future of Land Surface Models and the Challenges of Representing Complex Terrestrial Systems, *Journal of Advances in Modeling Earth Systems*, *12*(4), doi:10.1029/2018MS001453.

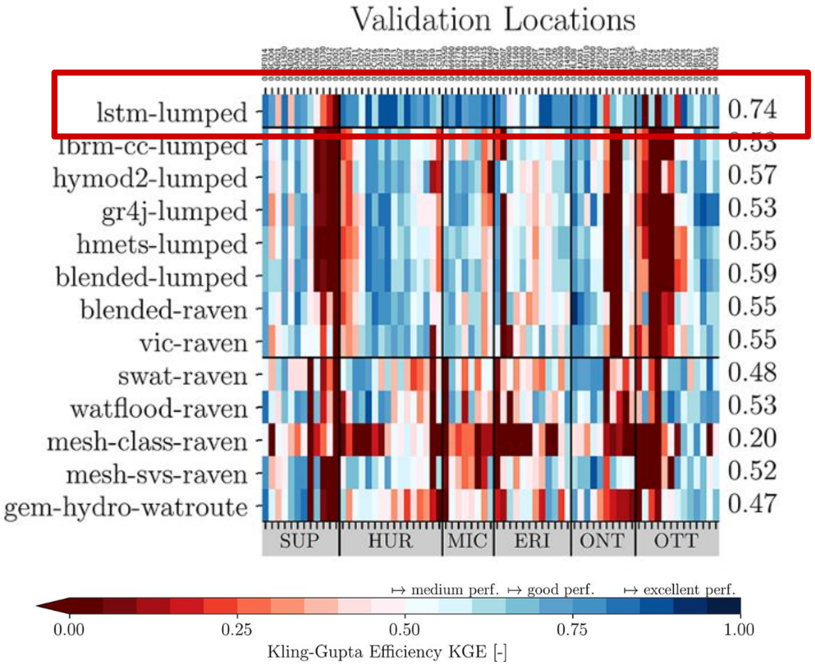
Circa 2017-2019

The straw that broke the CAMELS back

All aboard the hype train!

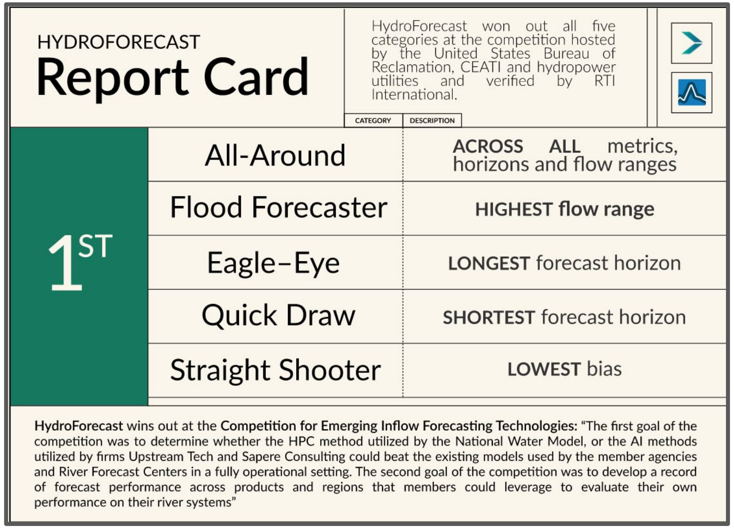
On the ascendency of ML methods in hydrologic modeling

ML techniques surpass hydrologic models for streamflow prediction



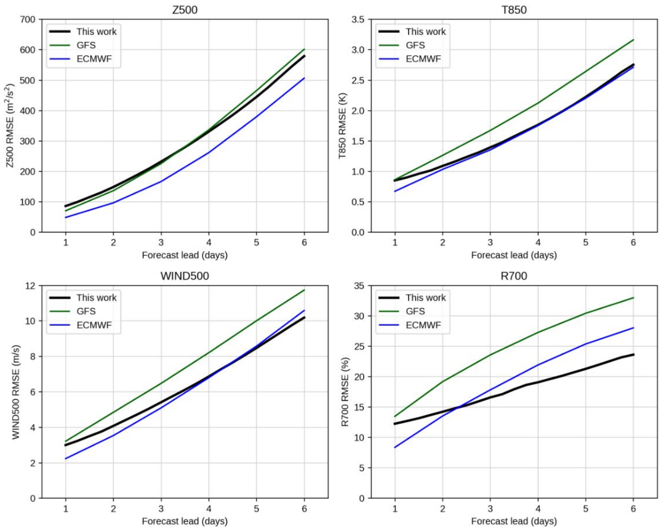
Mai, J., Shen, H., Tolson, B. A., Gaborit, É., Arsenault, R., Craig, J. R., et al. (2022). The Great Lakes Runoff Intercomparison Project Phase 4: the Great Lakes (GRIP-GL). *Hydrology and Earth System Sciences*, *26*(13), 3537–3572. https://doi.org/10.5194/hess-26-3537-2022

UPSTREAM TECH

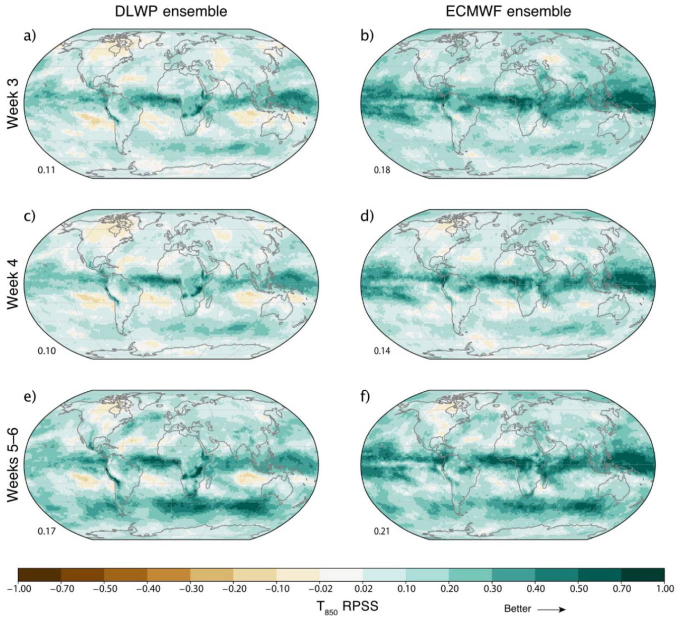


https://www.upstream.tech/hydroforecast

Deep learning uptake in atmospheric science has been rapid too



Keisler, R. (2022, February 15). Forecasting Global Weather with Graph Neural Networks. arXiv. Retrieved from http://arxiv.org/abs/2202.07575



Weyn, J. A., Durran, D. R., Caruana, R., & Cresswell-Clay, N. (2021). Sub-Seasonal Forecasting With a Large Ensemble of Deep-Learning Weather Prediction Models. *Journal of Advances in Modeling Earth Systems*, *13*(7), e2021MS002502. https://doi.org/10.1029/2021MS002502

Wait a second, this train is ***FAST*!**

There has been an explosion of ML papers applied to Earth system science, with strong performance

For instance, we're at almost 1000 LSTM papers published in 2023 in hydrology!

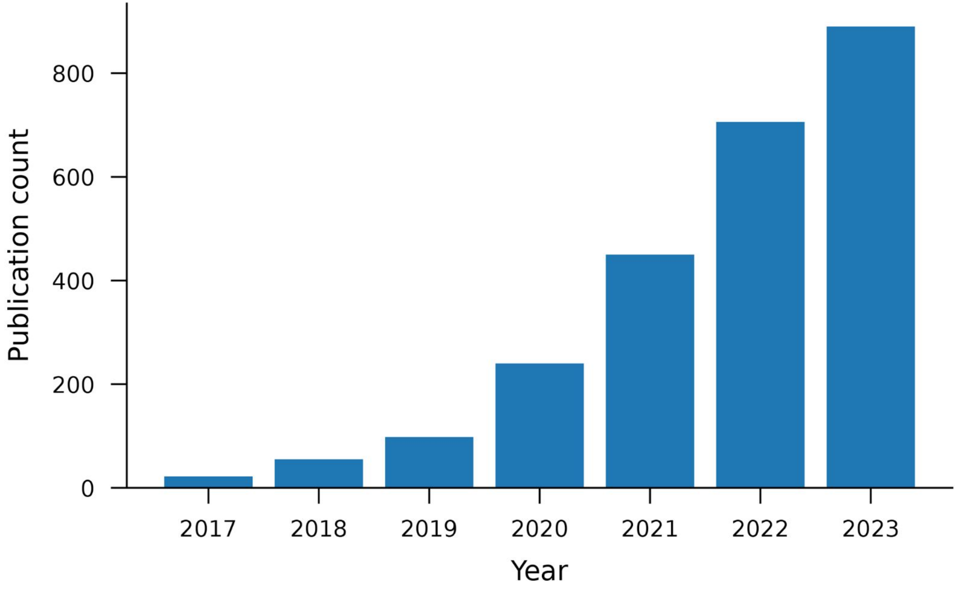
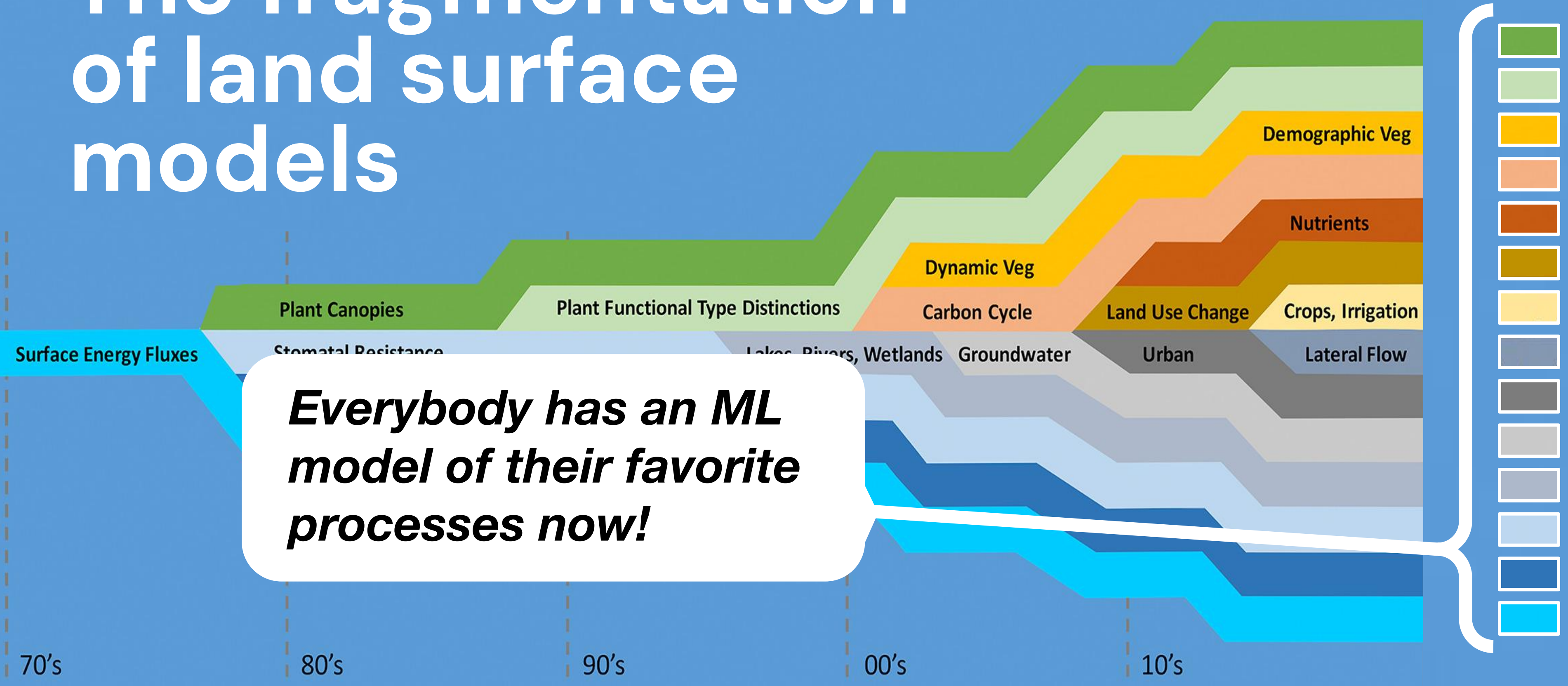


Figure 1. Number of hydrological publications related to rainfall–runoff modeling with LSTM networks over time based on data retrieved from Google Scholar in April 2024.

Oh no, it’s not a train after all!

Model proliferation as a social problem

The fragmentation of land surface models



Fisher, R. A., and C. D. Koven (2020), Perspectives on the Future of Land Surface Models and the Challenges of Representing Complex Terrestrial Systems, *Journal of Advances in Modeling Earth Systems*, *12*(4), doi:10.1029/2018MS001453.

Big efforts are starting, but the main focus is on atmospheric circulation

CliMA
CLIMATE MODELING ALLIANCE

LEAP

Ai2

We need a vision for integrating both modeling approaches

Today, I want to walk through three case studies, showcasing methods for “physics-based” ML in hydrology

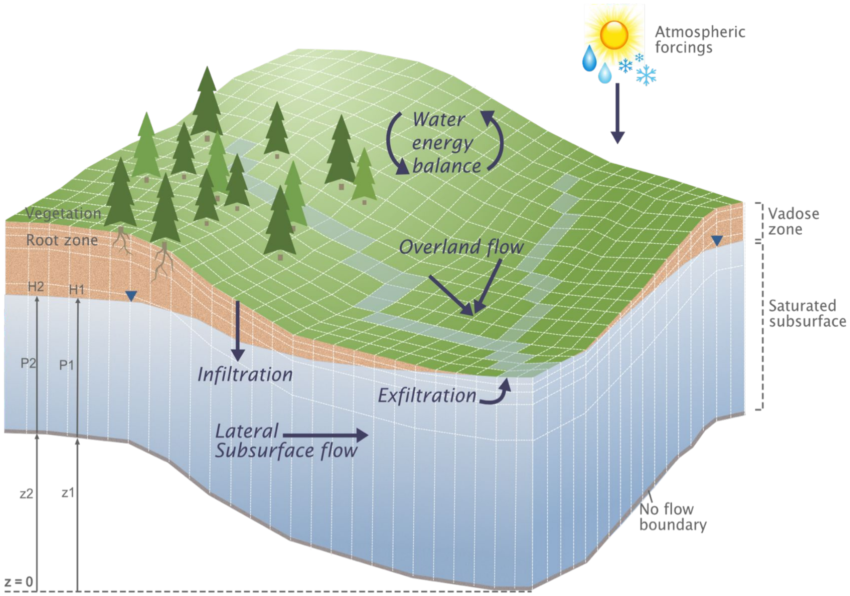
Part 1: emulation for rapid simulation

Continental scale emulation of ParFlow-CLM to enable on-demand forecasts of coupled surface/groundwater

Continental scale emulation of ParFlow-CLM to enable on-demand forecasts of coupled surface/groundwater

This work is in collaboration with a large team from Laura Condon's lab at University of Arizona and Reed Maxwell and Peter Melchior at Princeton University

Large, gridded models give spatiotemporally complete views of the state of a watershed, but are difficult to run and calibrate

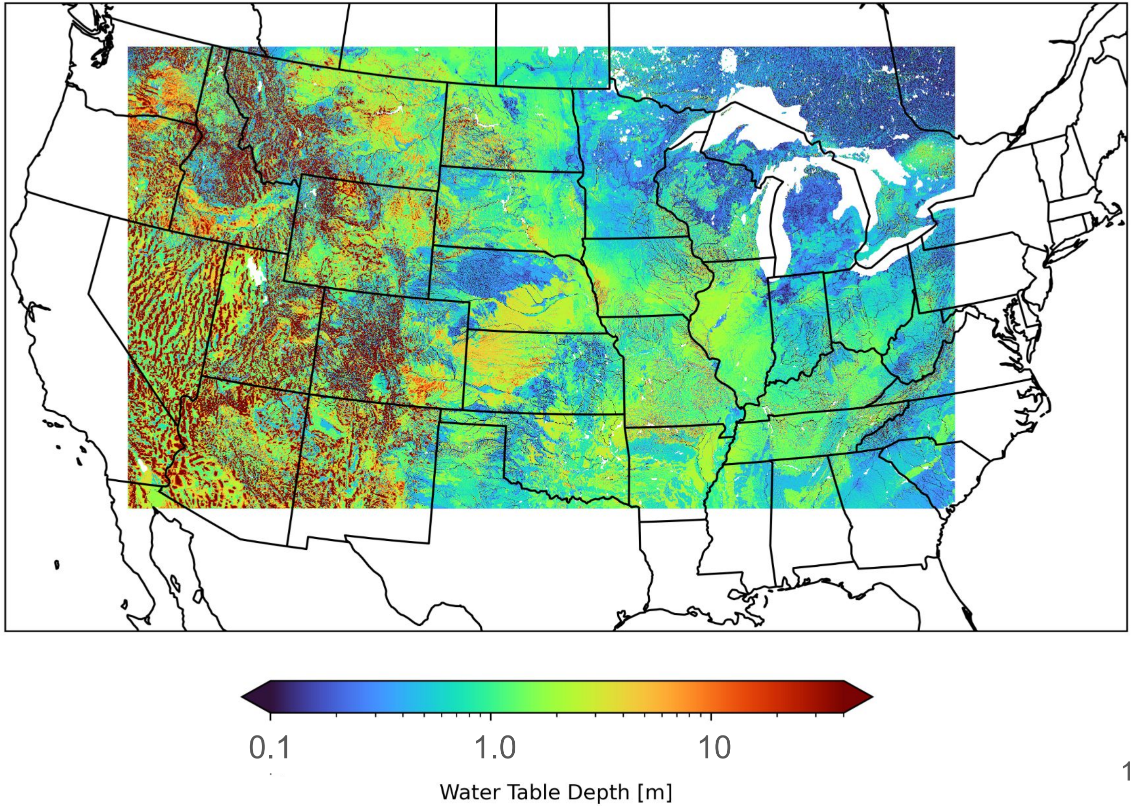


But, train emulators based on the models

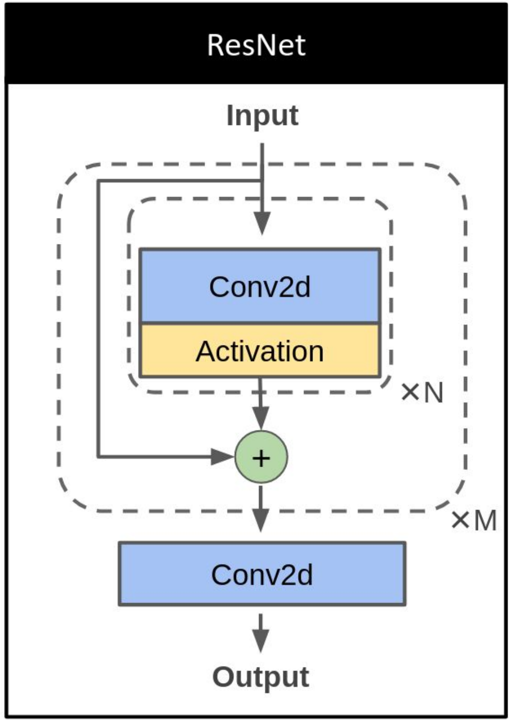
We are emulating continental scale, high resolution (1km) simulations using ParFlow-CLM

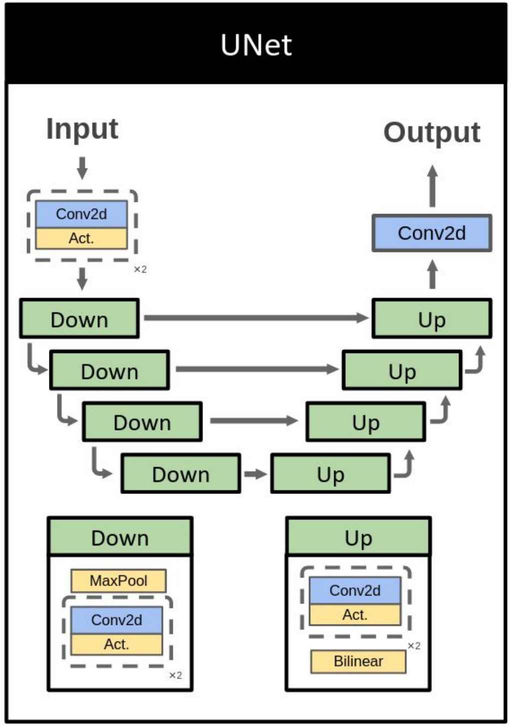
Resolving full spatiotemporal dynamics (3d + time)

Processes vary across orders of magnitude in both space and time

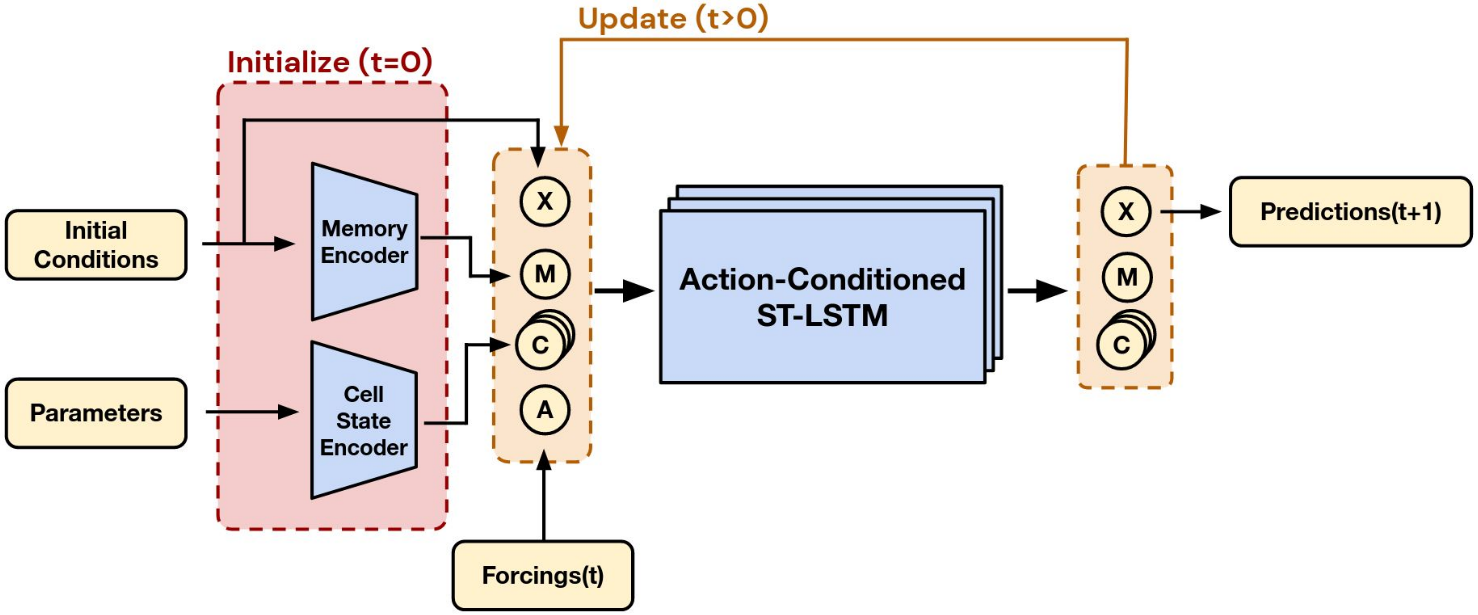


We used two baseline architectures for validation and intercomparison purposes

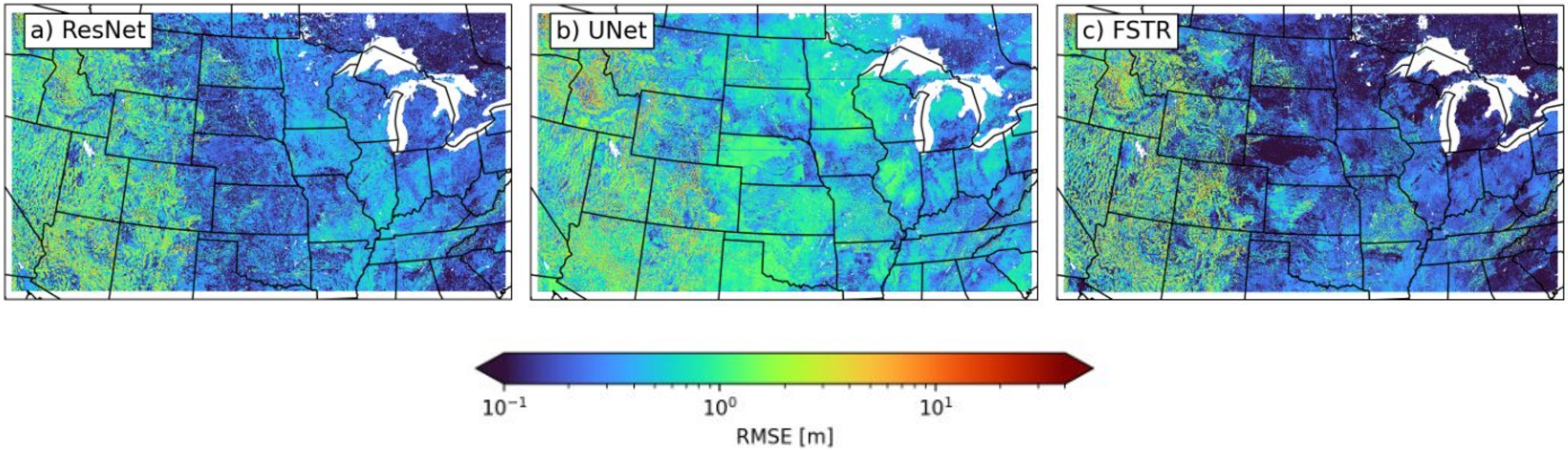




I developed a neural network architecture that works like a hydrologic model (FSTR)

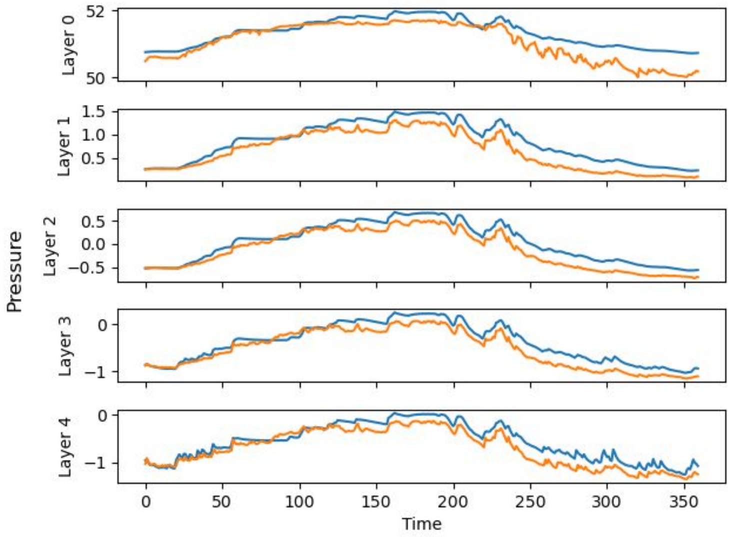


FSTR outperforms 2 standard neural network architectures on predictions of predictions of water table depth

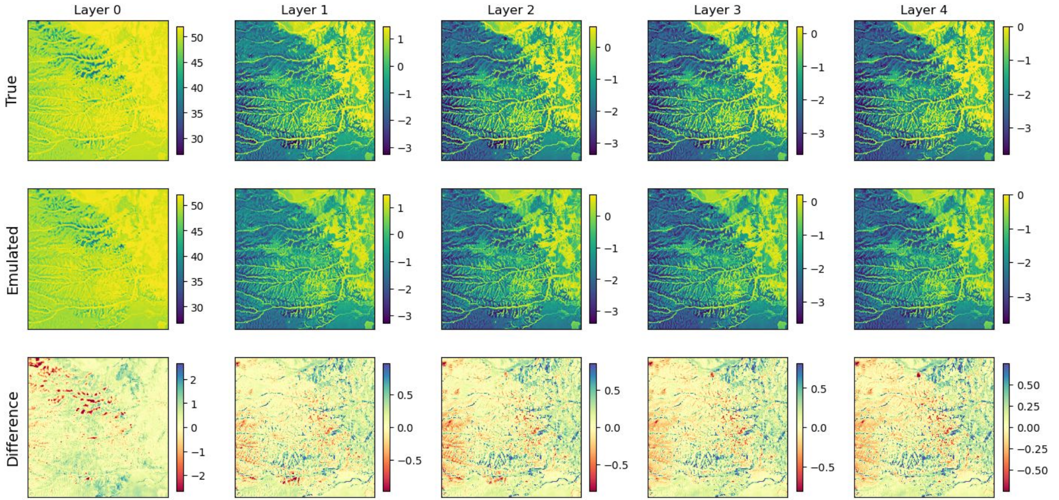


Overall, FSTR is extremely good at matching full 4-d spatiotemporal patterns

Capturing temporal variability



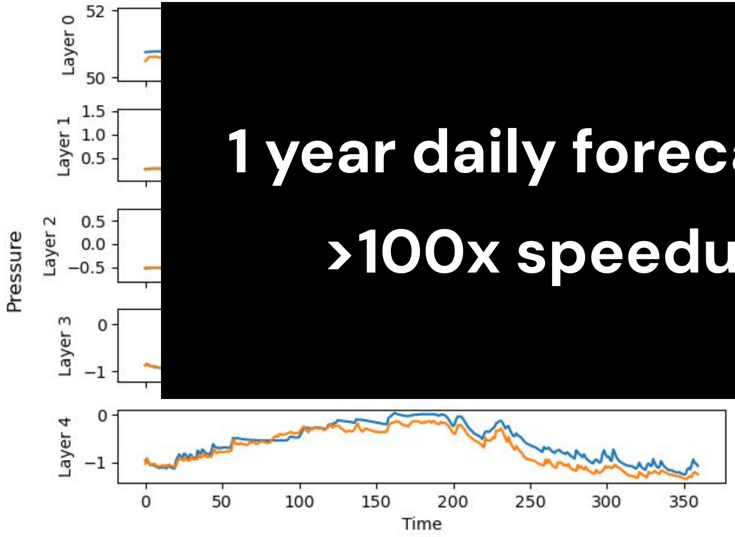
Capturing spatial variability



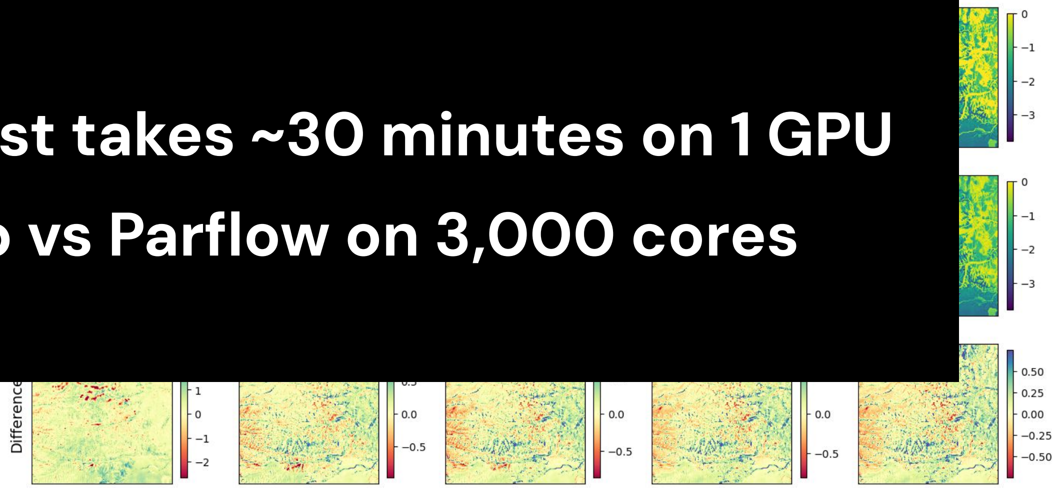
Bennett, A., Tran, H., De la Fuente, L., Triplett, A., Ma, Y., Melchior, P., et al. (2024). Spatio-temporal machine learning for regional to continental scale terrestrial hydrology. *Journal of Advances in Modeling Earth Systems*, 16, e2023MS004095. https://doi.org/10.1029/2023MS004095

Overall, FSTR is extremely good at matching full 4-d spatiotemporal patterns

Capturing temporal variability



Capturing spatial variability



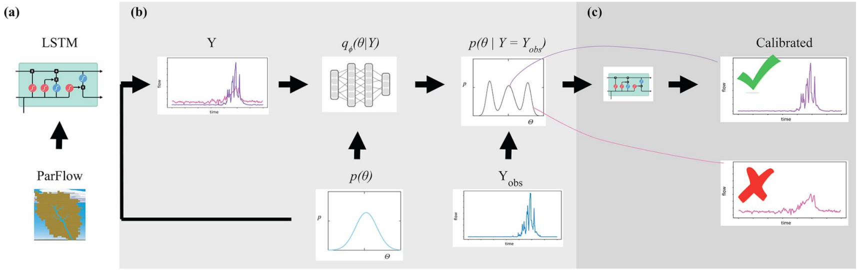
1 year daily forecast takes ~30 minutes on 1 GPU

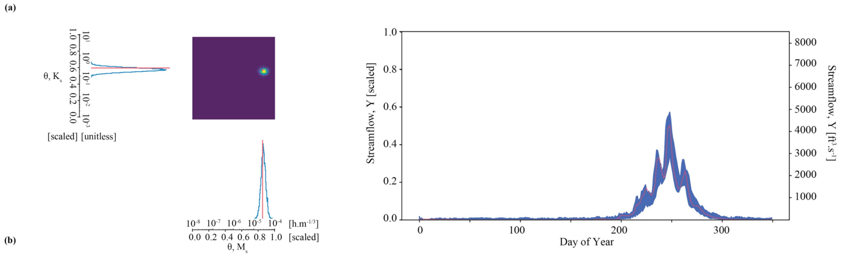
>100x speedup vs Parflow on 3,000 cores

Bennett, A., Tran, H., De la Fuente, L., Triplett, A., Ma, Y., Melchior, P., et al. (2024). Spatio-temporal machine learning for regional to continental scale terrestrial hydrology. *Journal of Advances in Modeling Earth Systems*, 16, e2023MS004095. https://doi.org/10.1029/2023MS004095

Bonus materials from this line of work:

Simulation-based inference

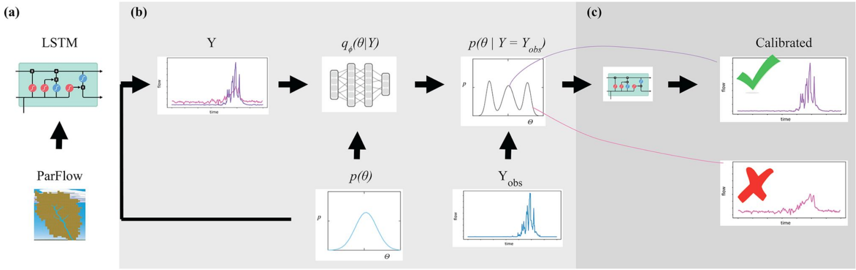


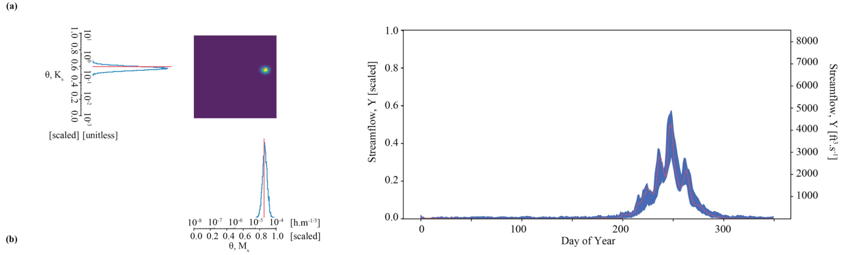


Hull, R., Leonarduzzi, E., De La Fuente, L., Viet Tran, H., Bennett, A., Melchior, P., Maxwell, R. M., and Condon, L. E.: Simulation-based inference for parameter estimation of complex watershed simulators, Hydrol. Earth Syst. Sci., 28, 4685–4713, https://doi.org/10.5194/hess-28-4685-2024, 2024

Bonus materials from this line of work:

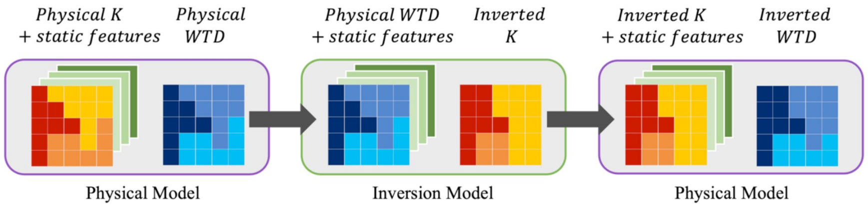
Simulation-based inference

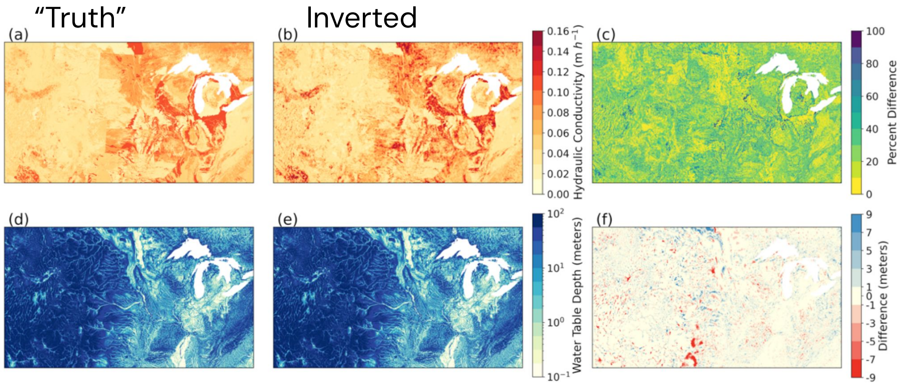




Hull, R., Leonarduzzi, E., De La Fuente, L., Viet Tran, H., Bennett, A., Melchior, P., Maxwell, R. M., and Condon, L. E.: Simulation-based inference for parameter estimation of complex watershed simulators, Hydrol. Earth Syst. Sci., 28, 4685–4713, https://doi.org/10.5194/hess-28-4685-2024, 2024

Large-scale parameter inversion

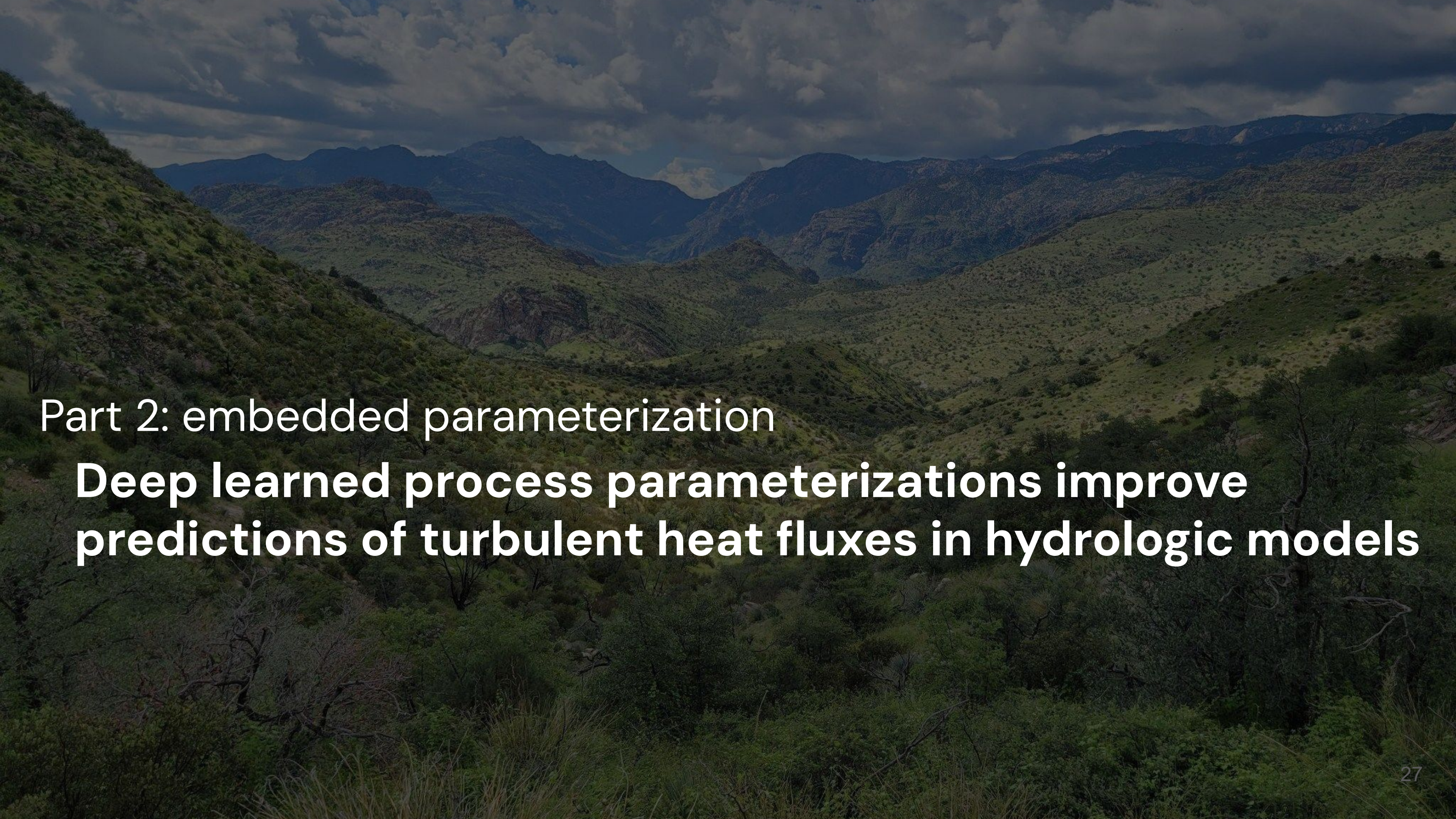




Amanda Triplett, Andrew Bennett, Laura Elizabeth Condon, et al. A Deep-Learning Based Parameter Inversion Framework for Large-Scale Groundwater Models. *Geophysical Research Letters* . In press.

Key takeaways:

- Model emulation can yield multiple orders of magnitude of computational speedup
- This enables ensemble methods, on-demand forecasting, parameter estimation/calibration and frugal deployment abilities
- Large scale models of the future will be expected to ship with emulated versions for rapid exploration, validation, and teaching



Part 2: embedded parameterization

Deep learned process parameterizations improve predictions of turbulent heat fluxes in hydrologic models

Deep learned process parameterizations improve predictions of turbulent heat fluxes in hydrologic models

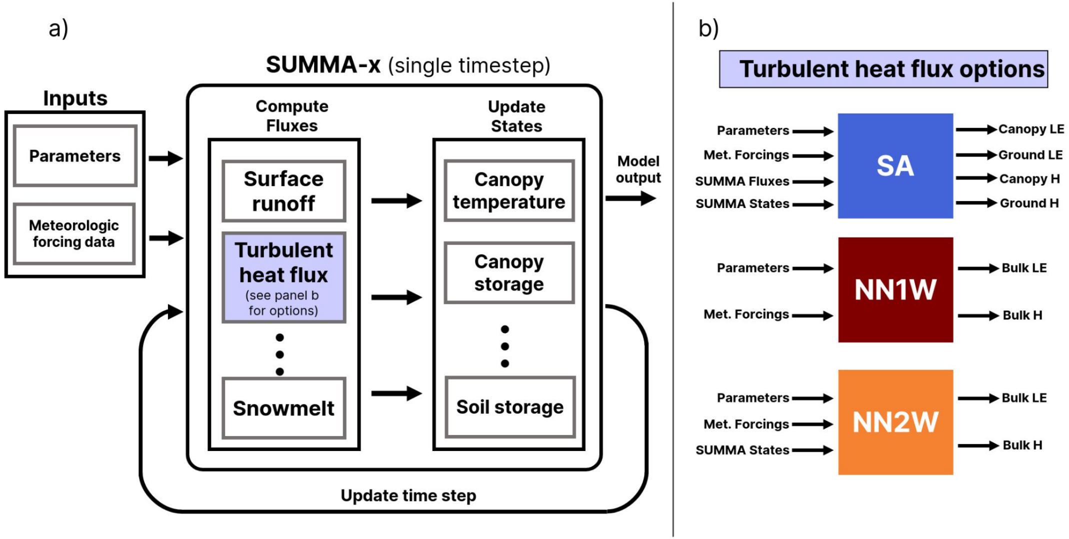
This work was in collaboration with Bart Nijssen (University of Washington)

Basic idea: Put a neural network into a process based model

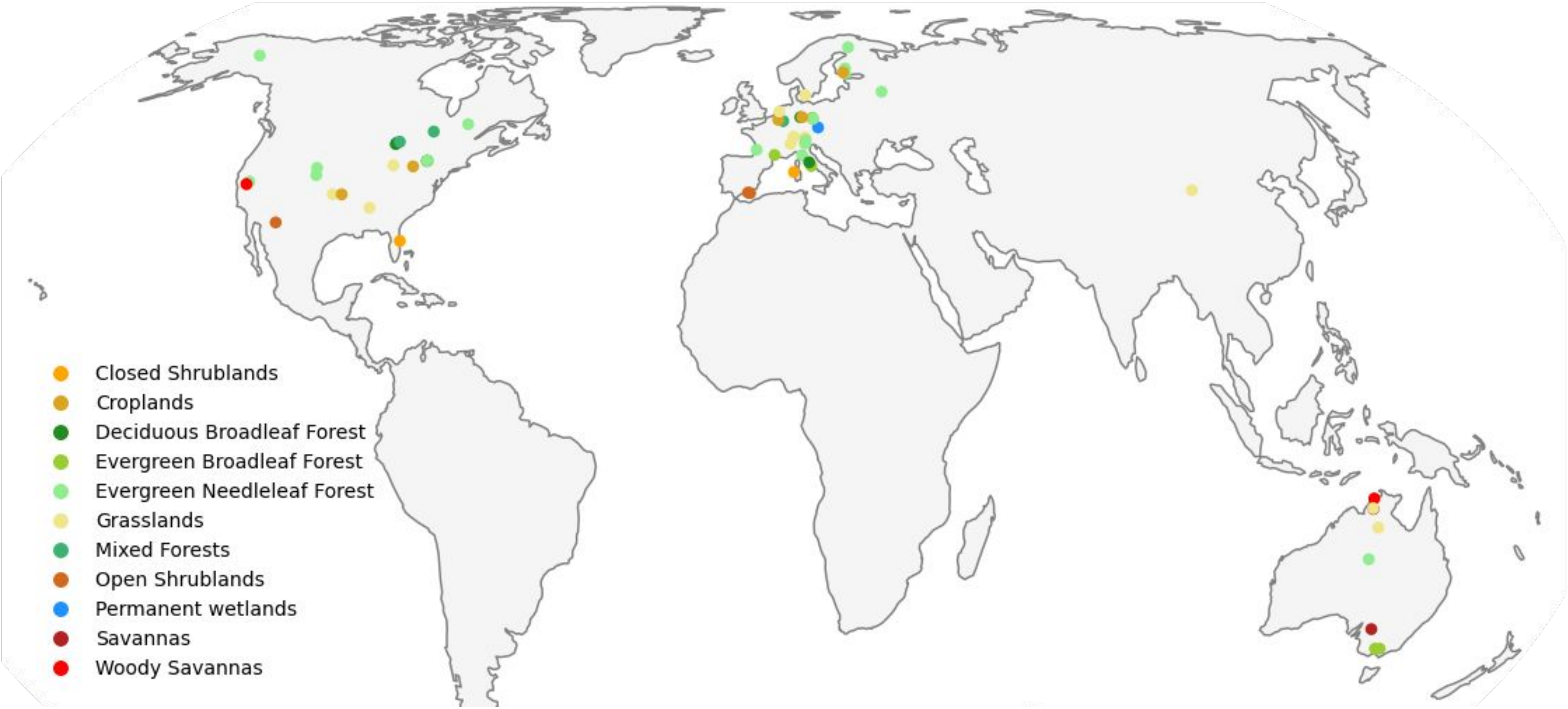
Our study focused on turbulent heat fluxes, which control ET and temperatures of the land surface

We explored three scenarios:

1. A calibrated PBHM
2. A one way coupled neural net
3. A two way coupled neural net

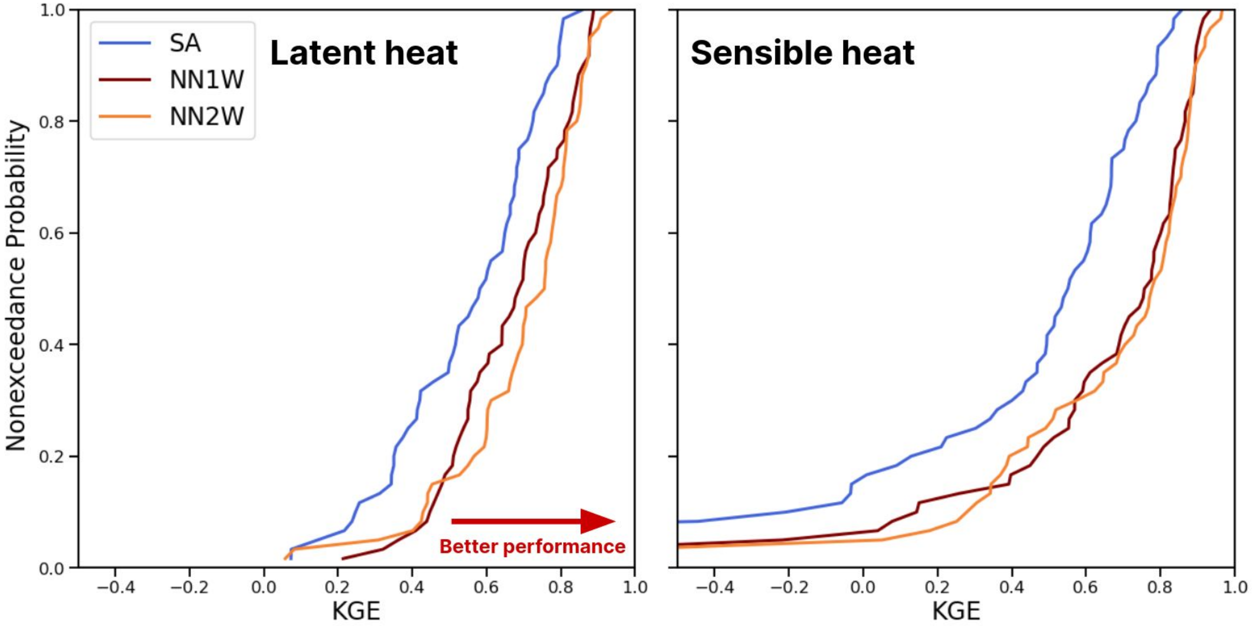


We gathered data from 60 FluxNet sites, totalling over 500 site-years of half-hourly data

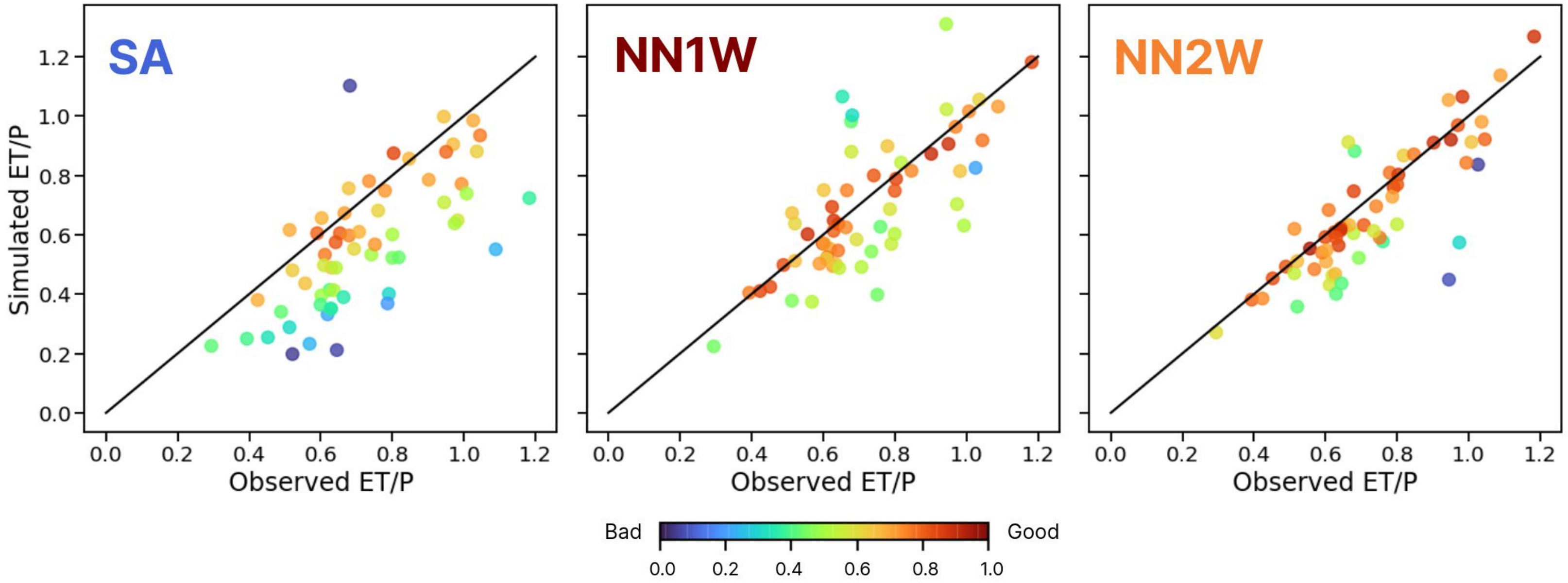


Both neural net methods show improved performance over process-based method

Figure shows CDF of performance across all 60 sites at 30 minute timescale



The introduction of the 2 way coupling also improved the long-term water balance compared to both SA and NN1W



Key takeaways:

Allowing model components to dynamically talk to each other we can have emergent improvements to our predictive capabilities

Better quantification of land-atmosphere interactions improves long-term water balance simulations

We currently do not have any scalable technologies to explore this space, and need to invest in such developments

Part 4: hybrid models (embedded hypotheses)

Methods for estimation of unknown/unobserved parameters to improve hydrologic prediction and understanding

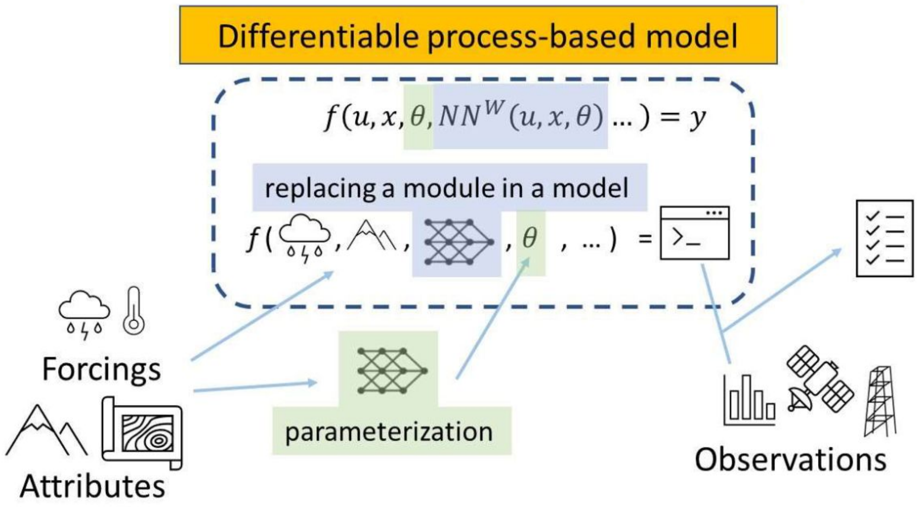
Methods for estimation of unknown/unobserved parameters to improve hydrologic prediction and understanding

First part is in mostly done by my students, Aamir Lamichhane & Nabin Kalauni

Last part of this was in collaboration with Frederik Kratzert (Google) and Wouter Knoben (University of Calgary)

Differentiable modeling is an emerging and hyped up technique, where you write the "physical" model in an ML framework

Then, you can use things like neural networks to replace processes or estimate parameter values for the internal equations



Shen, C., Appling, A.P., Gentine, P. *et al.* Differentiable modelling to unify machine learning and physical models for geosciences. *Nat Rev Earth Environ* **4**, 552–567 (2023). https://doi.org/10.1038/s43017-023-00450-9

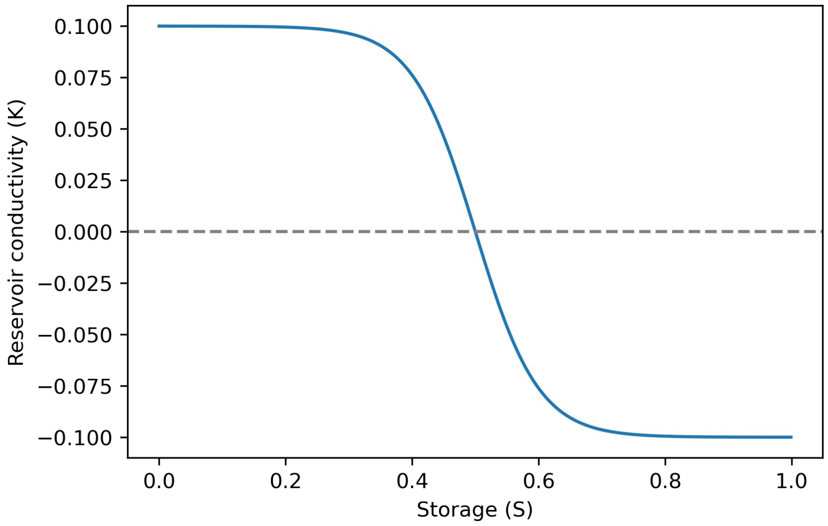
A primer on differentiable modeling

Consider a simple, nonlinear reservoir:

$$\frac{dS}{dt} = \boxed{K(S)} \cdot S$$

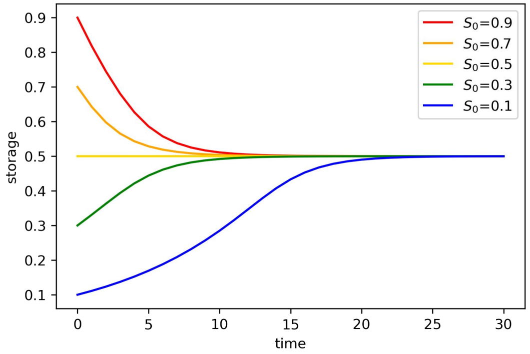
"unknown"

And a a K(S) function that we want to reproduce (or more accurately, find):

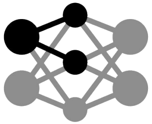


A primer on differentiable modeling

Given many samples derived from the system that we want to “learn”:

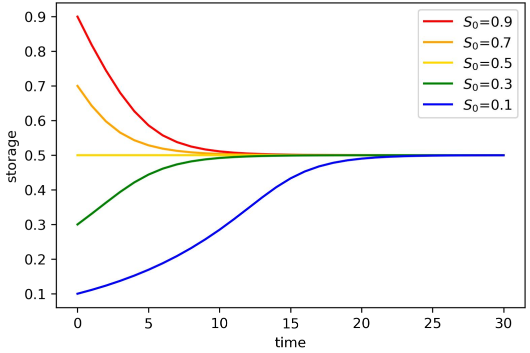


Fit a model where the parameterization is “learned” by a neural network:

$$\frac{dS}{dt} =$$  $$\cdot S$$

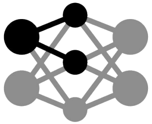
A primer on differentiable modeling

Given many samples derived from the system that we want to “learn”:



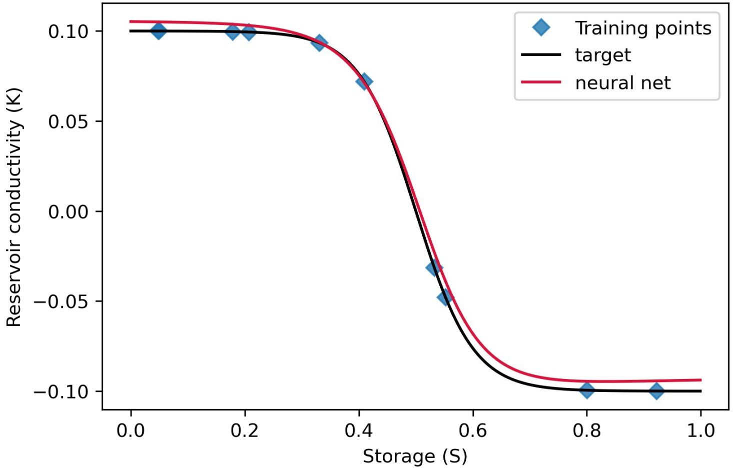
Fit a model where the parameterization is “learned” by a neural network:

$$\frac{dS}{dt} =$$



$$\cdot S$$

Inspecting the internals of the model we find a good fit:



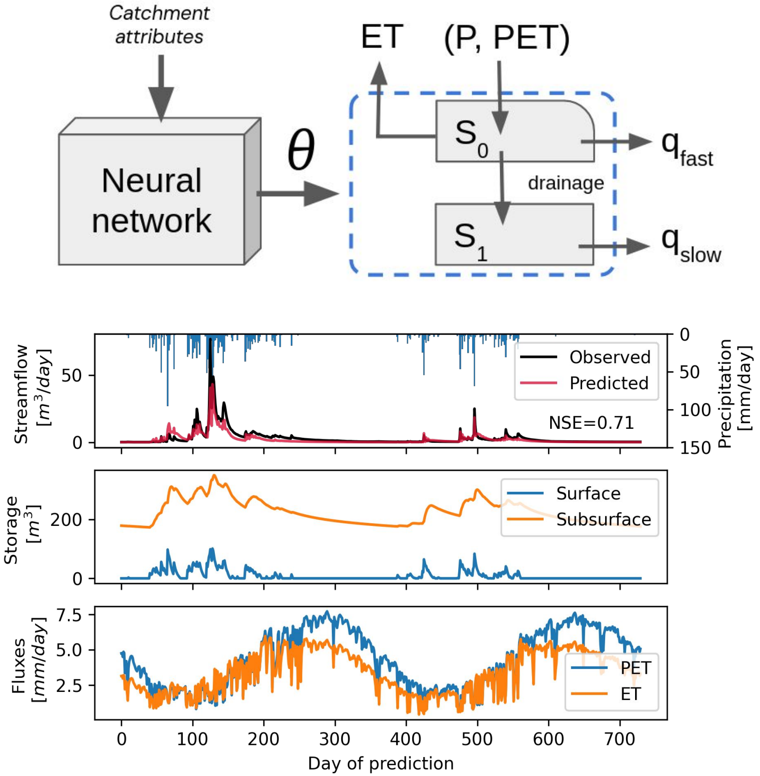
A primer on differentiable modeling

If you want to learn more I have a book chapter in EarthAI outlines the theory that drives "differentiable hydrology"

The chapter demonstrates how to build conceptual hydrologic models parameterized by neural networks

The code, data, and environment are all open source and publicly available:

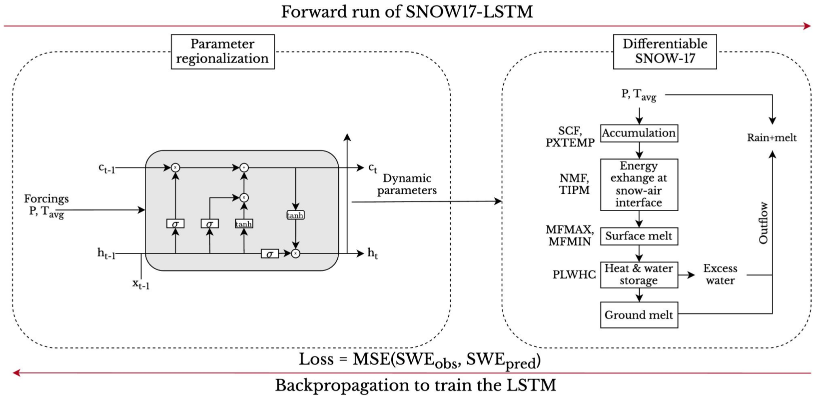
https://github.com/earth-artificial-intelligence/earth_ai_book_materials/blob/main/chapter_07/ai_for_physics_inspired_hydrology_modeling.ipynb

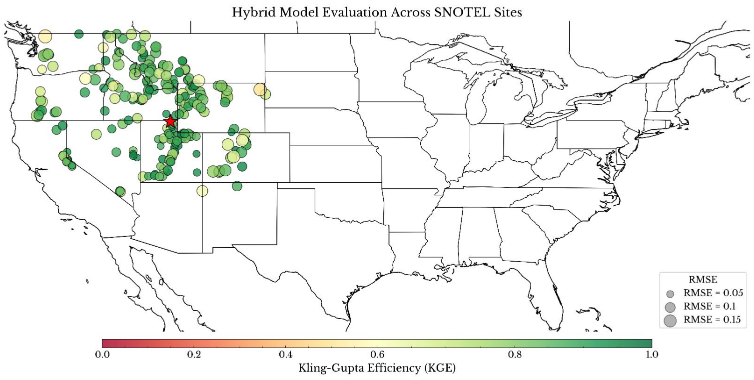


We are applying this method to multiple processes:

This work is led by Aamir Lamichhane

We parameterize the SNOW17 model with a LSTM, and train across Snotel sides in the western US to predict snow water equivalent (SWE)





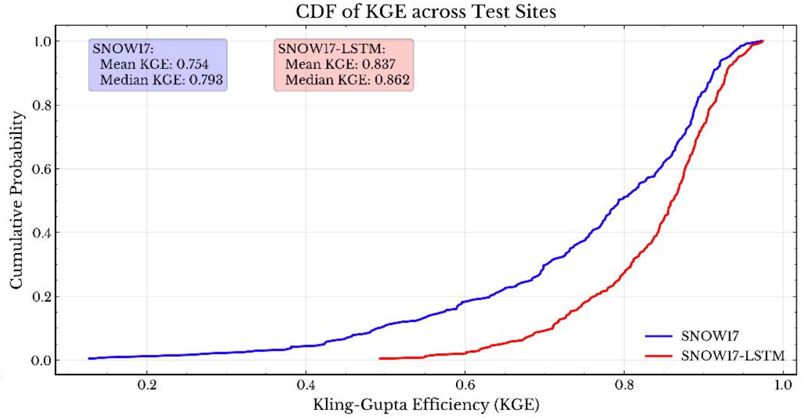
We are applying this method to multiple processes:

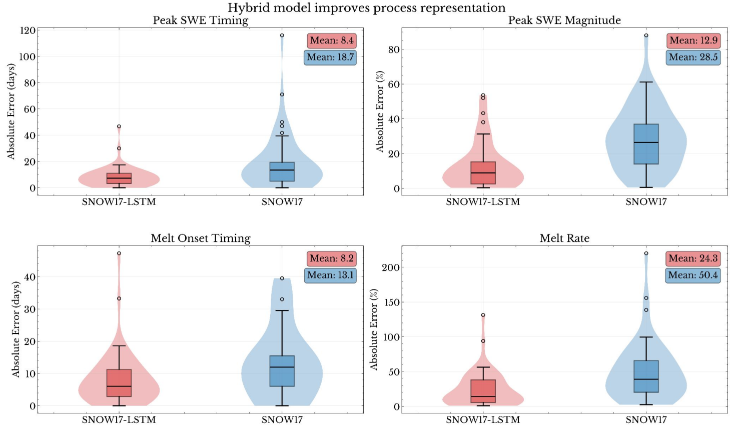
This work is led by Aamir Lamichhane

We parameterize the SNOW17 model with a LSTM, and train across Snotel sides in the western US to predict snow water equivalent (SWE)

Our model shows strong improvements in predictive capabilities broadly

We also find good improvement in key metrics such as peak SWE magnitude/timing, and melt periods



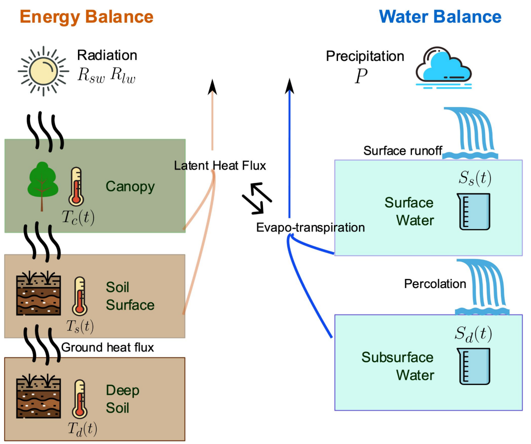


We are applying this method to multiple processes:

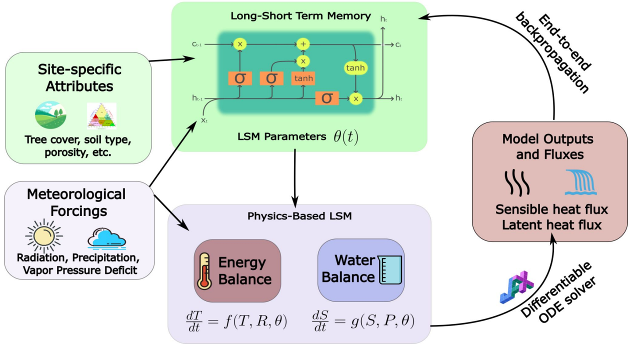
This work is led by Nabin Kalauni

We are developing a coupled mass/energy balance land surface model to simulate multiple processes, focusing currently on L-A interactions and evapotranspiration

Conceptual Coupled Land Surface Model



Hybrid Differentiable Land Surface Model



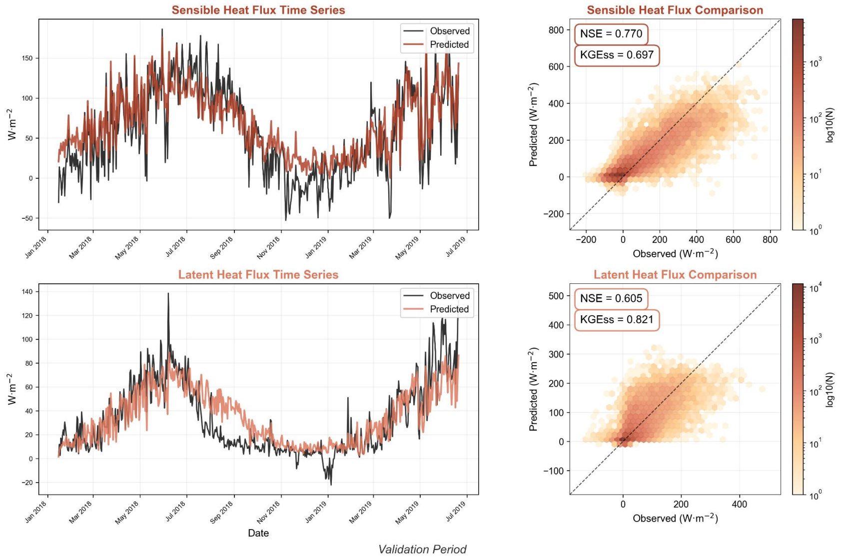
We are applying this method to multiple processes:

This work is led by Nabin Kalauni

We are developing a coupled mass/energy balance land surface model to simulate multiple processes, focusing currently on L-A interactions and evapotranspiration

Preliminary results show strong performance, though there are many process interactions and tradeoffs to explore

Big questions in training these types of models around initialization/spinup!!

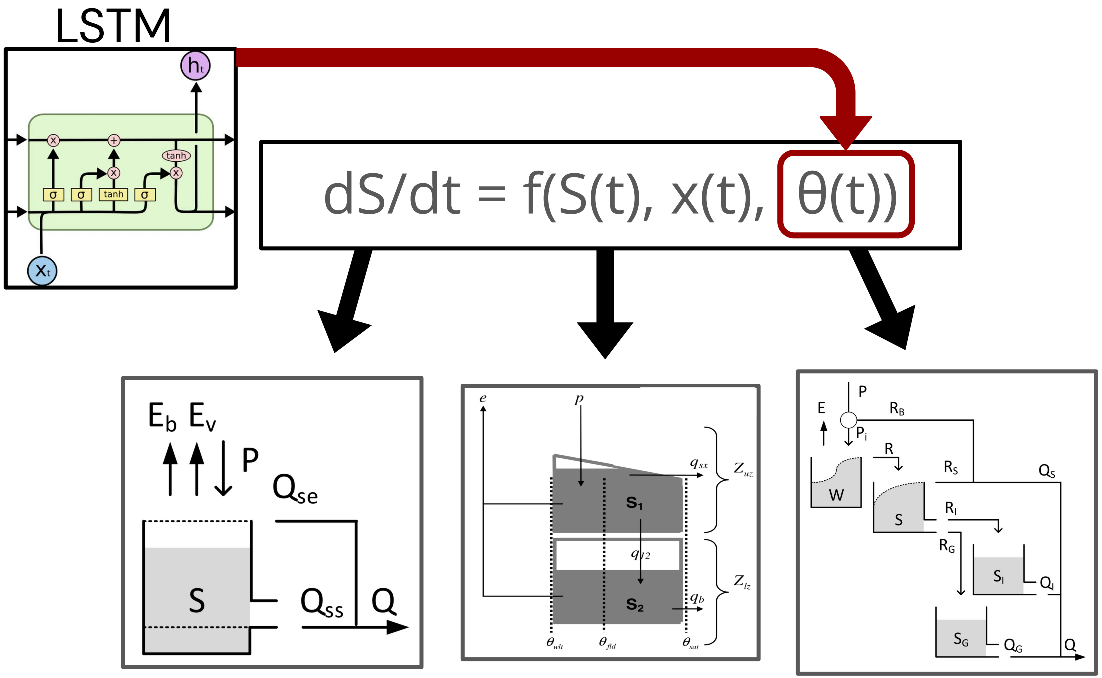


But, broader philosophical questions linger. We are tackling these with a multi-model approach

To do this, we used frozen LSTM models that were pre-trained from scratch

We then replace the last layer “regression head” with a layer that predicts parameter values for 3 different hydrologic model structures

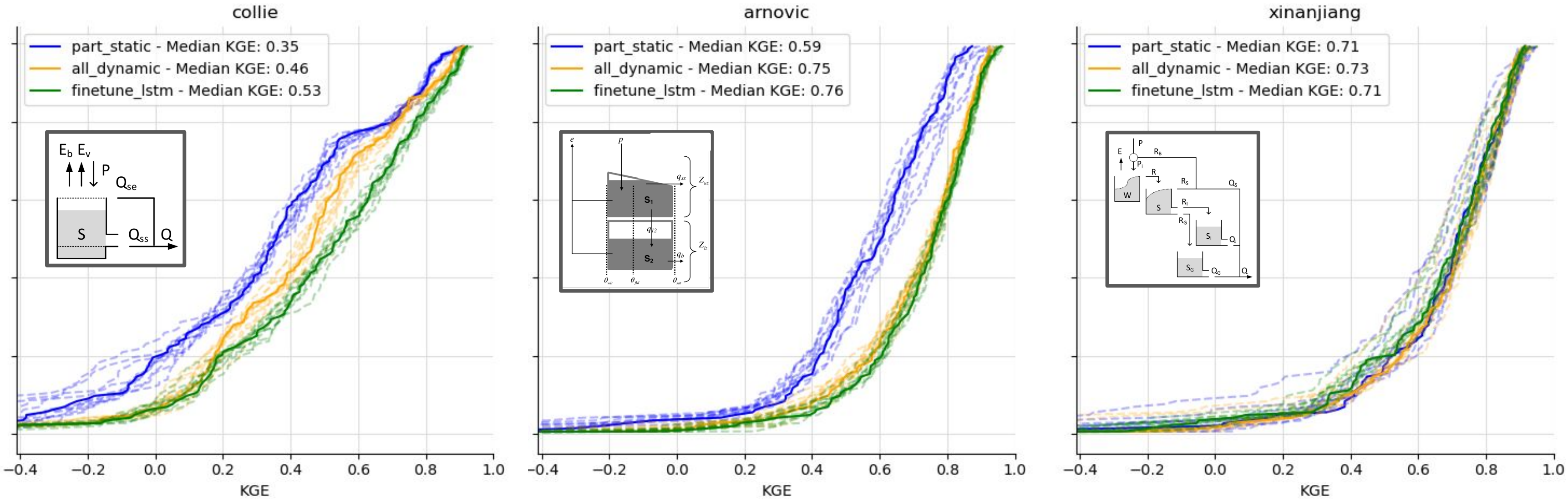
We do this for 140 catchments across the US



The medium and full complexity models show almost state of the art performance when configured flexibly

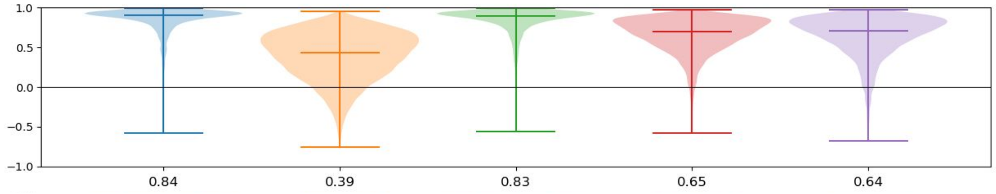
Figure shows CDF of performance across all 140 sites (SOTA: median KGE ~0.75)

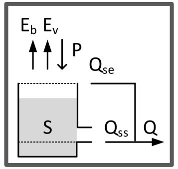
Note: The **exact same LSTM** is issued for all of these predictions, except green lines

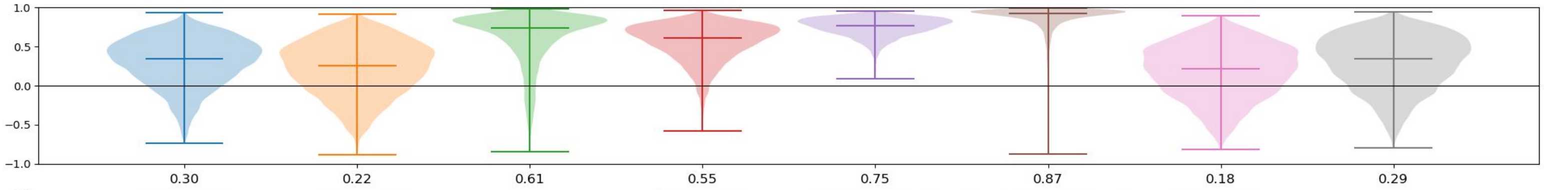


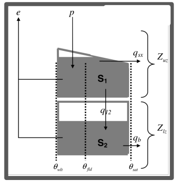
For each of the model configurations this was done 10 times. So, do the end models have similar hydrologic behaviors?

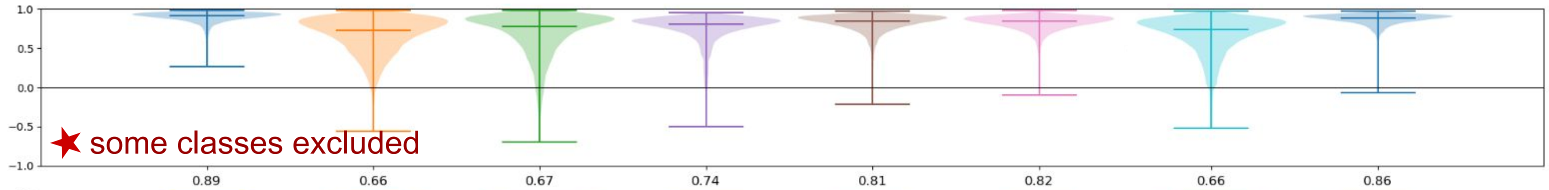
To find out we calculate the derived correlation between parameter values across all runs, and show the distributions as violin plots

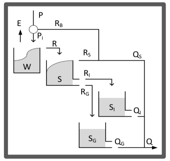








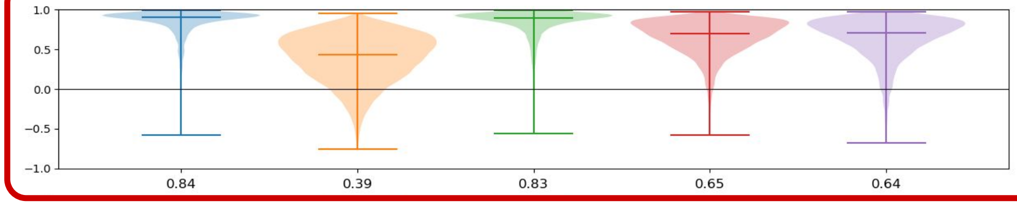


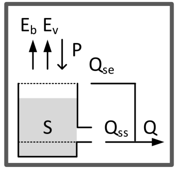


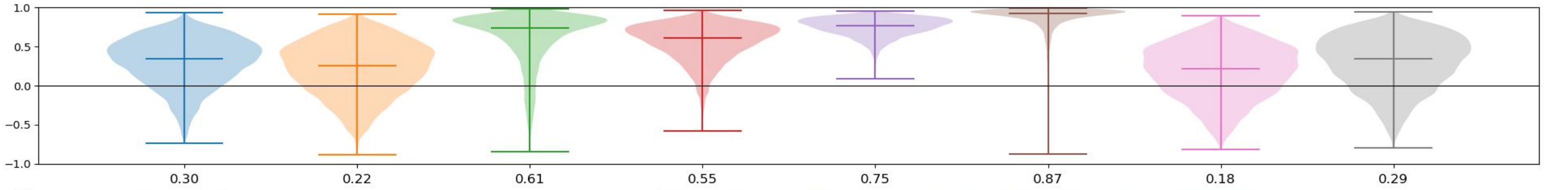
For each of the model configurations this was done 10 times. So, do the end models have similar hydrologic behaviors?

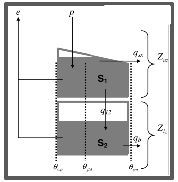
To find out we calculate the derived correlation between parameter values across all runs, and show the distributions as violin plots

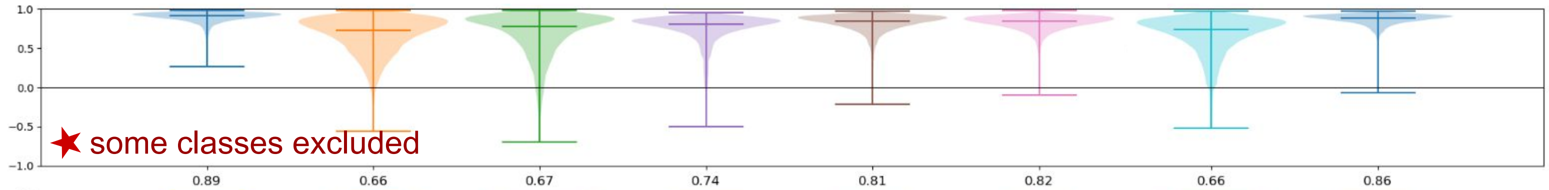
Shows decent parameter convergence, but poor performance



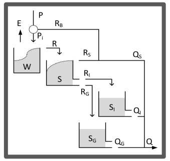








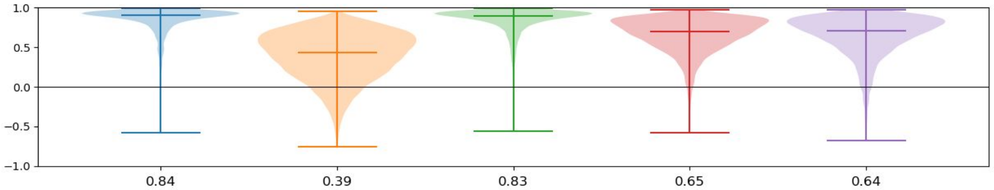
★ some classes excluded

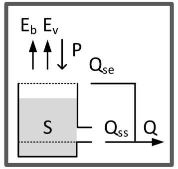


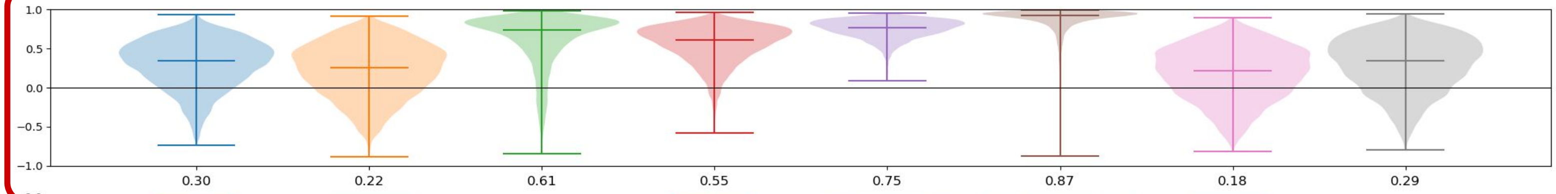
For each of the model configurations this was done 10 times. So, do the end models have similar hydrologic behaviors?

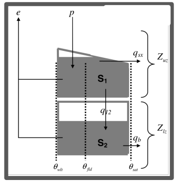
To find out we calculate the derived correlation between parameter values across all runs, and show the distributions as violin plots

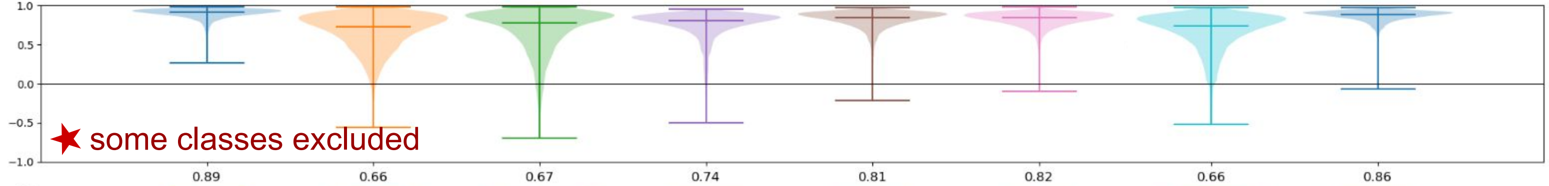
Had performance, but shows a large degree of disagreement in values



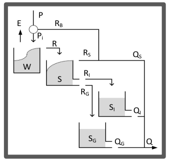








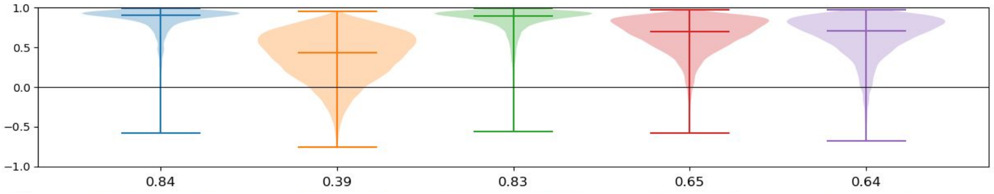
★ some classes excluded

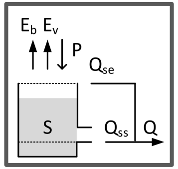


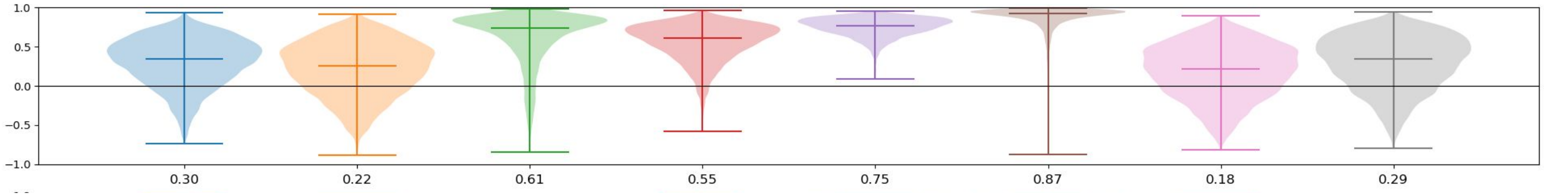
For each of the model configurations this was done 10 times. So, do the end models have similar hydrologic behaviors?

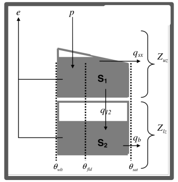
To find out we calculate the derived correlation between parameter values across all runs, and show the distributions as violin plots

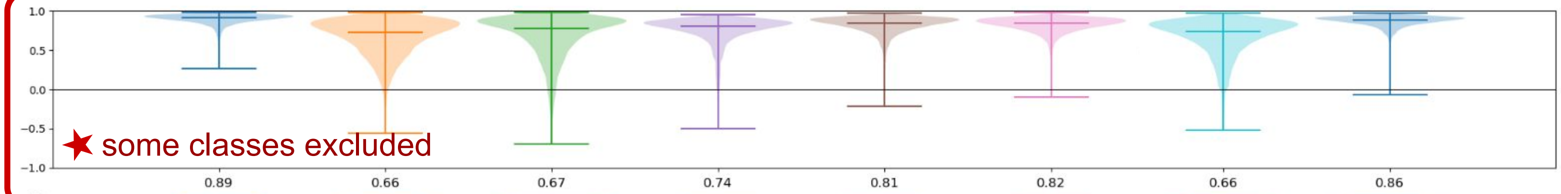
Good performance, and high agreement in parameter values

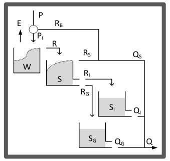












Key takeaways:

To me, at least, this seems to show that the LSTM is learning some “general” hydrologic concepts

Additionally, this is a whiff of evidence that the LSTM is probably driving the overall predictive performance

We (and others) have only implemented simple methods so far, there is still a ton to explore in this space

Thanks for listening! Looking forward to the discussion

Feel free to email any questions!

andrbenn@arizona.edu

Core AC-ST-LSTM equations

Cell state update

$$g_t = tanh\left(W_{xg} * X_t + W_{hg} * H_{t-1}^l\right)$$

$$i_t = \sigma\left(W_{xi} * X_t + W_{hi} * H_{t-1}^l\right)$$

$$f_t = \sigma\left(W_{xf} * X_t + W_{hf} * H_{t-1}^l\right)$$

$$C_t^l = f_t \odot C_{t-1}^l + i_t \odot g_t$$

Memory state update

$$g_t' = tanh\left(W_{xi}' * X_t + W_{mg} * M_t^{l-1}\right)$$

$$i_t' = \sigma\left(W_{xi}' * X_t + W_{mi} * M_t^{l-1}\right)$$

$$f_t' = \sigma\left(W_{xf}' * X_t + W_{mf} * M_t^{l-1}\right)$$

$$M_t^l = f_t' \odot M_t^{l-1} + i_t' \odot g_t'$$

Output, hidden, and action updates

$$o_t = \sigma\left(W_{xo} * X_t + W_{ho} * H_{t-1}^l + W_{co} * C_t^l + W_{mo} * M_t^l\right)$$

$$H_t^l = o_t \odot tanh\left(W_{1\times1} * \left[C_t^l, M_t^l\right]\right)$$

$$V_t^l = \left(W_{hv} * H_{t-1}^l\right) \odot \left(W_{av} * A_{t-1}\right)$$

Even with fast and robust emulators, we still want to be able to improve the predictions of our models by (hopefully) finding better parameter values

https://doi.org/10.1029/2024GL114285

Check for updates

AGU ADVANCING EARTH AND SPACE SCIENCES

Geophysical Research Letters

RESEARCH LETTER
10.1029/2024GL114285

Special Collection:
Advancing Interpretable AI/ML Methods for Deeper Insights and Mechanistic Understanding in Earth Sciences: Beyond Predictive Capabilities

Key Points:
- We train a machine learning model to invert hydraulic conductivity fields from water table depth for a ParFlow model of the US
- Our method provides good results while requiring many fewer forward simulations than traditional model calibration approaches
- Our method maintains physical relationships between variables including metamorphic relationships not directly provided in training

Supporting Information:
Supporting Information may be found in the online version of this article.

Correspondence to:
L. E. Condon,
lecondon@arizona.edu

Citation:
Triplett, A., Bennett, A., Condon, L. E., Melchior, P., & Maxwell, R. M. (2025). A deep-learning based parameter inversion framework for large-scale groundwater models. *Geophysical Research Letters*, *52*, e2024GL114285. https://doi.org/10.1029/2024GL114285

Received 20 DEC 2024
Accepted 11 APR 2025

A Deep-Learning Based Parameter Inversion Framework for Large-Scale Groundwater Models

Amanda Triplett[1], Andrew Bennett[1], Laura E. Condon[1], Peter Melchior[2,3], and Reed M. Maxwell[4,5,6]

[1]Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ, USA, [2]Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA, [3]Center for Statistics and Machine Learning, Princeton University, Princeton, NJ, USA, [4]Integrated Groundwater Modeling Center, Princeton University, Princeton, NJ, USA, [5]Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA, [6]High Meadows Environmental Institute, Princeton University, Princeton, NJ, USA

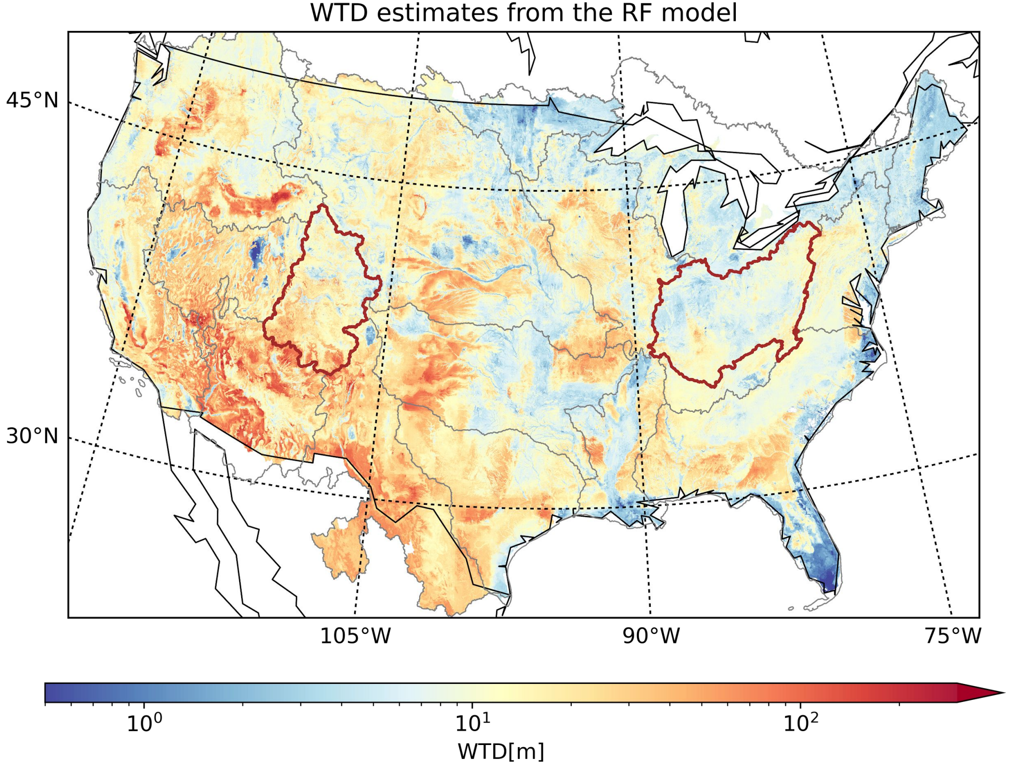
Abstract Hydrogeologic models generally require gridded subsurface properties, however these inputs are often difficult to obtain and highly uncertain. Parametrizing computationally expensive models where extensive calibration is computationally infeasible is a long standing challenge in hydrogeology. Here we present a machine learning framework to address this challenge. We train an inversion model to learn the relationship between water table depth and hydraulic conductivity using a small number of physical simulations. For a 31M grid cell model of the US we demonstrate that the inversion model can produce a reliable K field using only 30 simulations for training. Furthermore, we show that the inversion model captures physically realistic relationships between variables, even for relationships that were not directly trained on. While there are still limitations for out of sample parameters, the general framework presented here provides a promising approach for parametrizing expensive models.

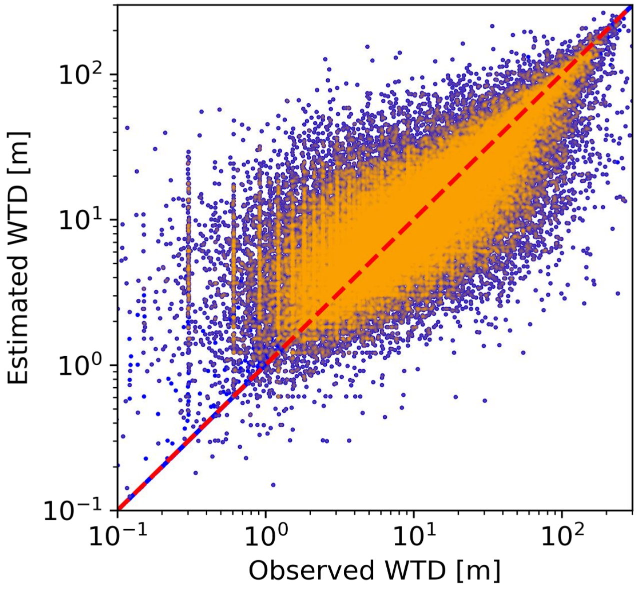
Plain Language Summary Numerical models that simulate groundwater flow often require gridded data about subsurface properties that are uncertain and difficult to obtain. Parameter adjustments (or calibration) is often needed to improve model performance. There are many existing approaches for parameter calibration; however a common limitation is that they require many model simulations, which can be very computationally expensive. Here we present a machine learning (ML) based inversion method for subsurface parameterization. We train a ML model to learn the relationship between subsurface parameters and simulated water table depth across the US. Our method produces accurate results and requires a minimal number of simulations. Furthermore, we demonstrate that this approach has learned physically accurate relationships.

1. Introduction

Hydrogeologic models rely heavily on estimates of subsurface parameters to accurately simulate groundwater flow. Large, gridded, hydrogeologic models can have millions of cells, each requiring parameterization. These parameters, such as hydraulic conductivity, are extremely difficult to measure at scale and uncertainty impacts the accuracy and utility of simulations. Determining values for unknown parameters in large, spatially distributed

Originally we were motivated by work from Yueling Ma, who developed a ML estimate of WTD

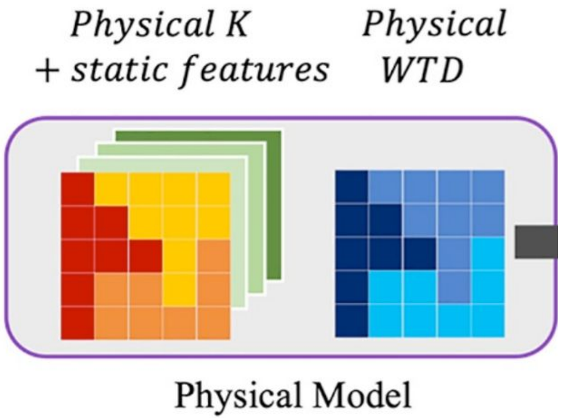




Ma, Y., Leonarduzzi, E., Defnet, A., Melchior, P., Condon, L.E. and Maxwell, R.M. (2024), Water Table Depth Estimates over the Contiguous United States Using a Random Forest Model. Groundwater, 62: 34-43. https://doi.org/10.1111/gwat.13362

But first we must demonstrate that the overall method works, so Amanda Triplett led the effort to develop our inversion framework

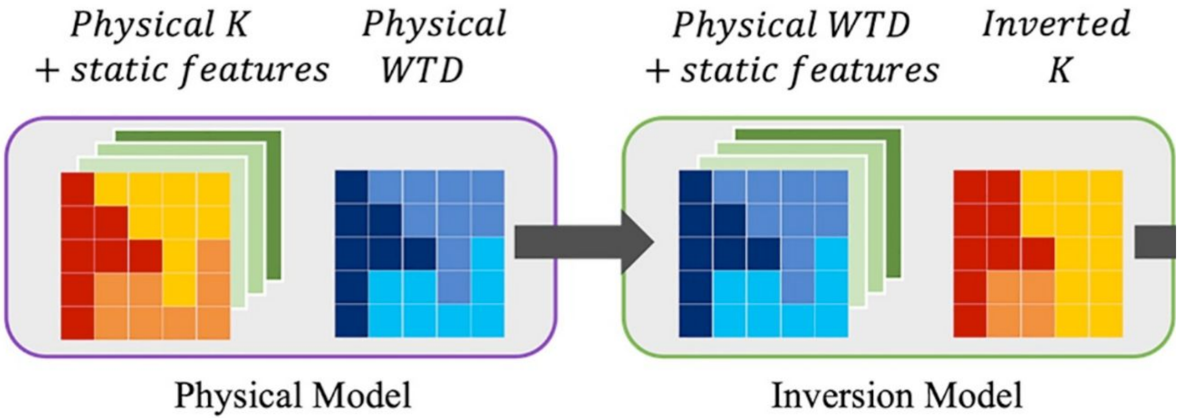
Develop the training data



But first we must demonstrate that the overall method works, so Amanda Triplett led the effort to develop our inversion framework

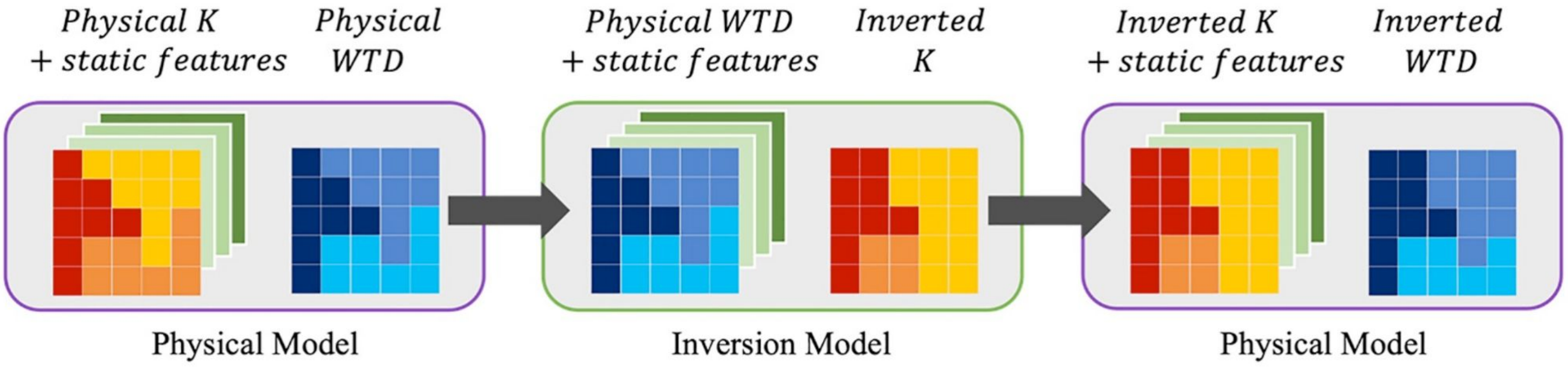
Develop the training data

Train the inversion model

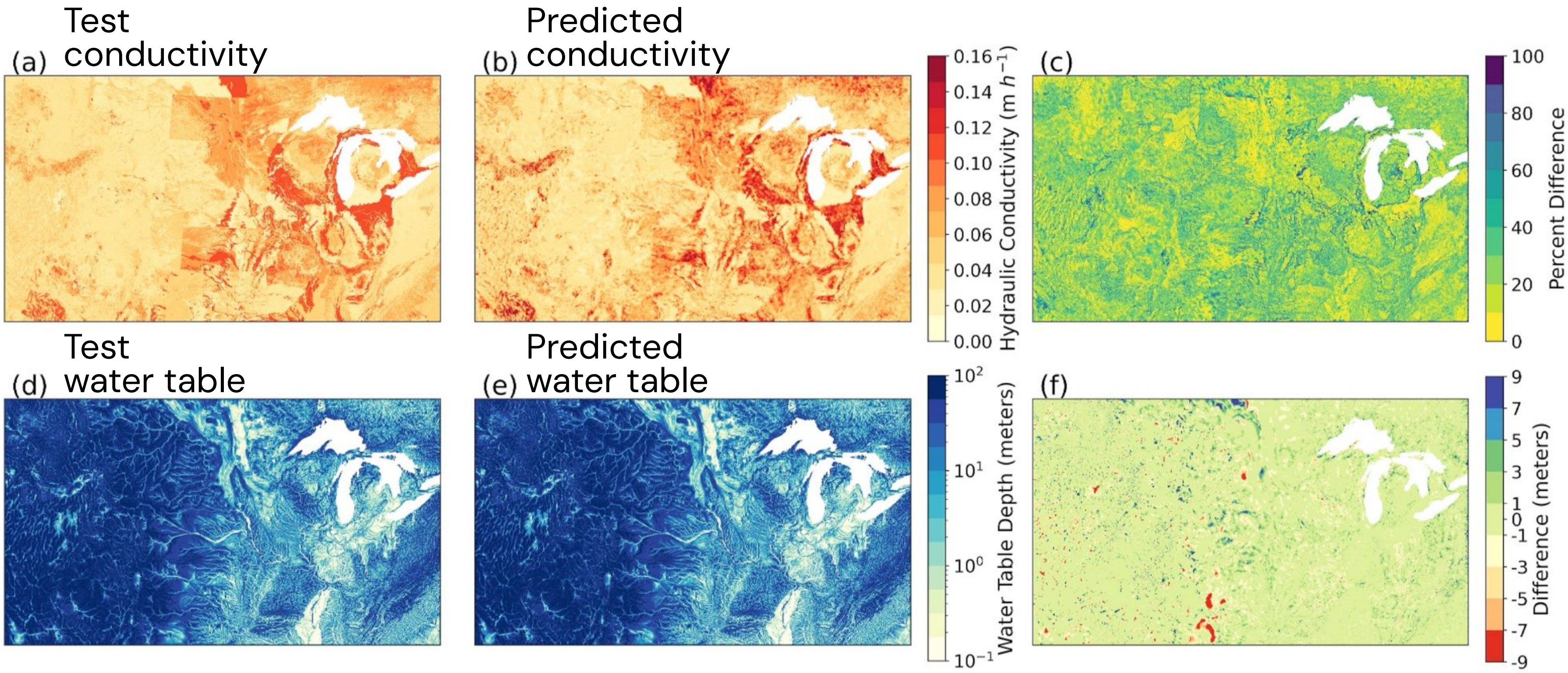


But first we must demonstrate that the overall method works, so Amanda Triplett led the effort to develop our inversion framework

Develop the training data	Train the inversion model	Ensure the parameters work



This works well! Showing an example from the test set



The model also learns physically plausible relationships

Of course, we used a particular set of recharge values for training, but they are uncertain

To test robustness we used a “metamorphic test” by applying perturbations to the recharge, holding WTD constant

Overall results follow hydrologic reasoning - to maintain WTD with more water, a higher conductivity is required (and vice versa)