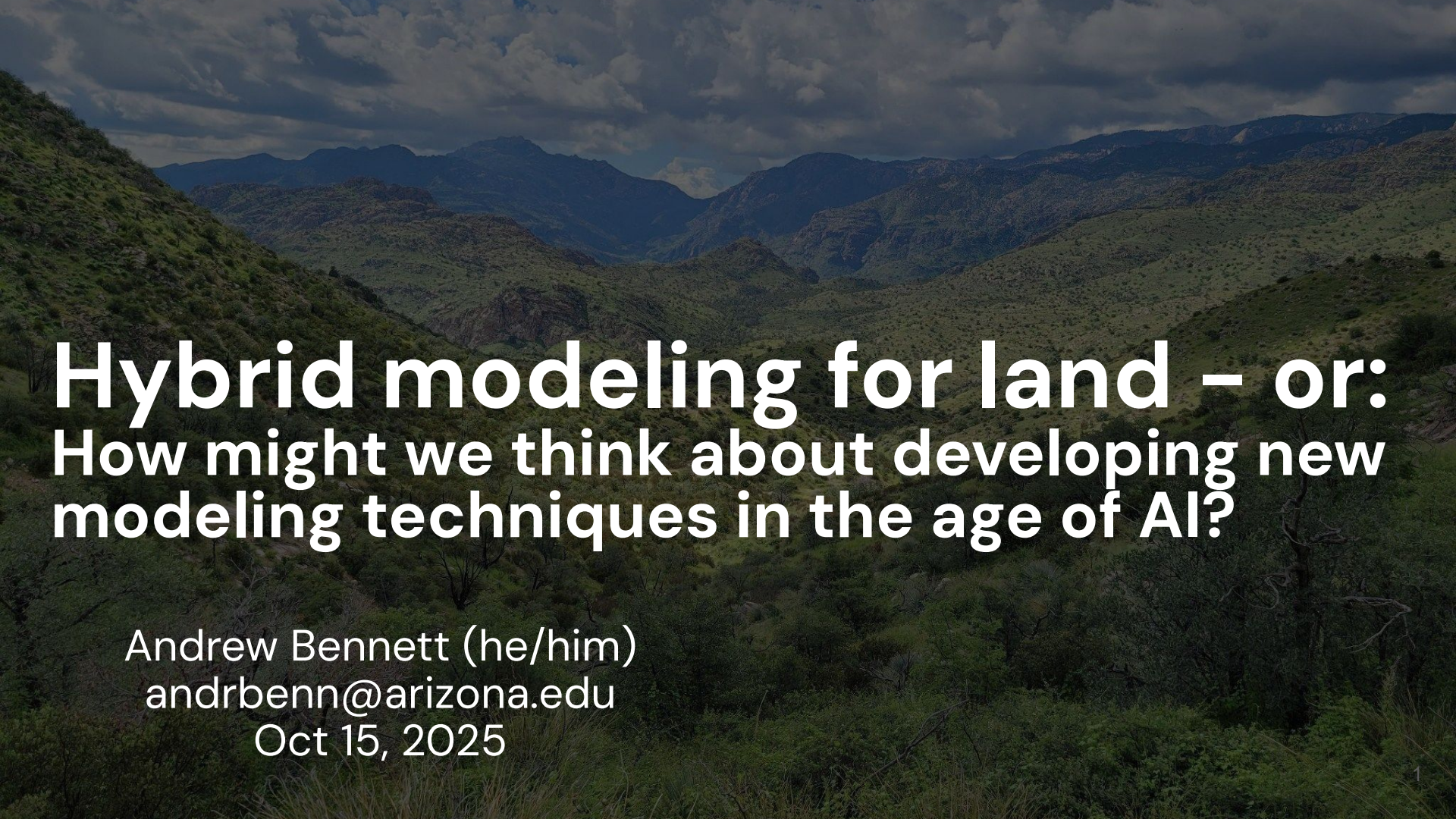




Hybrid modeling for land – or: How might we think about developing new modeling techniques in the age of AI?

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

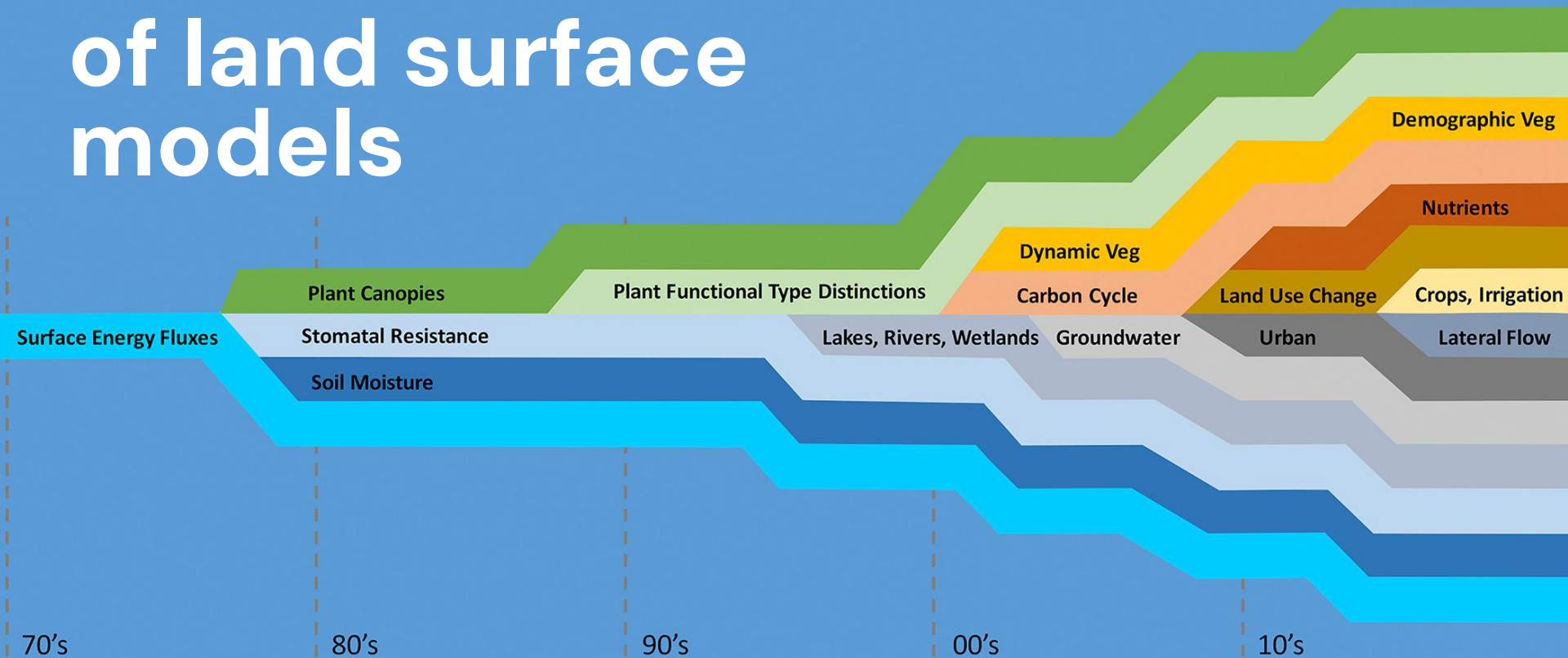
I've worked on a number of hydrologic modeling applications across scale and process

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





MEANWHILE...

Circa 2017-2019

The straw that broke the CAMELS back

Hydrol. Earth Syst. Sci., 22, 6005–6022, 2018
https://doi.org/10.5194/hess-22-6005-2018
© Author(s) 2018. This work is distributed under the Creative Commons Attribution 4.0 License.

Hydrology and
Earth System
Sciences
Open Access

Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks

Frederik Kratzert^{1,2}, Daniel Klotz¹, Claire Brenner¹, Karsten Schulz¹, and Mathew Herrnegger¹

Institute for
Resource
Management
and
Sustainable
Development
Correspondence to:
Frederik Kratzert
Received: 12 October 2017
Revised: 12 October 2017

Abstract
Long short-term
memory (LSTM)
networks offer
unprecedented
accuracy for
prediction in
ungauged basins.
We trained and
tested several
LSTMs on 531
basins from the
CAMELS data set
using k-fold
validation, so
that predictions
were made in
basins that
supplied no
training data. The
training and
test data set
included ~30
years of daily
rainfall–runoff
data from
catchments in
the United
States
ranging in size
from 0.22 to
5.20, and
including 12
of the 13
IGBP
vegetated
land cover
classifications.
This effectively
“ungauged”
model was
benchmarked
over a 15-year

Water Resources Research

TECHNICAL
REPORTS: METHODS
10.1029/2019WR026065

Special Section:

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

Key Points:

- Overall accuracy of LSTMs in ungauged basins is comparable to standard hydrology models in gauged basins
- There is sufficient information in catchment characteristics data to differentiate between catchment-specific rainfall–runoff behaviors

Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning

Frederik Kratzert¹, Daniel Klotz¹, Mathew Herrnegger², Alden K. Sampson³, Sepp Hochreiter¹, and Grey S. Nearing⁴

¹LIT AI Lab and Institute for Machine Learning, Johannes Kepler University, Linz, Austria, ²Institute for Hydrology and Water Management, University of Natural Resources and Life Sciences, Vienna, Austria, ³Upstream Tech, Natural Energy Inc., Alameda, CA, USA, ⁴Department of Geological Sciences, University of Alabama, Tuscaloosa, AL, USA

Abstract Long short-term memory (LSTM) networks offer unprecedented accuracy for prediction in ungauged basins. We trained and tested several LSTMs on 531 basins from the CAMELS data set using k-fold validation, so that predictions were made in basins that supplied no training data. The training and test data set included ~30 years of daily rainfall–runoff data from catchments in the United States ranging in size from 0.22 to 5.20 km² with aridity index from 0.22 to 5.20, and including 12 of the 13 IGBP vegetated land cover classifications. This effectively “ungauged” model was benchmarked over a 15-year

AGU100
ADVANCING
EARTH AND
SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
Regionalization in a Global Hydrological Model
From Physical Descriptors to Rainfall–Runoff

Yanbin Yin^{1,2}, Zhen-Hua Fan^{1,2}, and Xuebin Chen^{1,2}

¹State Key Laboratory of Hydrology and Water Resources, Hohai University, Nanjing, China, ²State Key Laboratory of Water Resources and Environment, Hohai University, Nanjing, China

Key Points:
• A global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance.
• The regionalization method is based on a machine learning algorithm.
• High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

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

Correspondence to:
Yanbin Yin, yin@hohai.edu.cn

Received: 12 October 2017
Accepted: 12 October 2017
Published: 12 October 2017

Editor: David M. Legler, Editor, Water Resources Research

Copyright: © 2017 Yin et al. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

How to Cite: Yin, Y., Fan, Z., & Chen, X. (2017). Regionalization in a Global Hydrological Model From Physical Descriptors to Rainfall–Runoff. *Water Resources Research*, 53, 6005–6022. doi:10.1029/2017WR021000

Abstract The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

Plain Language Summary The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

Introduction The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

1. Introduction The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

2. Data and Model Description The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

3. Results and Discussion The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

4. Conclusions The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

AGU100
ADVANCING
EARTH AND
SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
Regionalization in a Global Hydrological Model
From Physical Descriptors to Rainfall–Runoff

Yanbin Yin^{1,2}, Zhen-Hua Fan^{1,2}, and Xuebin Chen^{1,2}

¹State Key Laboratory of Hydrology and Water Resources, Hohai University, Nanjing, China, ²State Key Laboratory of Water Resources and Environment, Hohai University, Nanjing, China

Key Points:
• A global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance.
• The regionalization method is based on a machine learning algorithm.
• High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

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

Correspondence to:
Yanbin Yin, yin@hohai.edu.cn

Received: 12 October 2017
Accepted: 12 October 2017
Published: 12 October 2017

Editor: David M. Legler, Editor, Water Resources Research

Copyright: © 2017 Yin et al. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

How to Cite: Yin, Y., Fan, Z., & Chen, X. (2017). Regionalization in a Global Hydrological Model From Physical Descriptors to Rainfall–Runoff. *Water Resources Research*, 53, 6005–6022. doi:10.1029/2017WR021000

Abstract The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

Plain Language Summary The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

Introduction The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

1. Introduction The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

2. Data and Model Description The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

3. Results and Discussion The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

4. Conclusions The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

AGU100
ADVANCING
EARTH AND
SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
Regionalization in a Global Hydrological Model
From Physical Descriptors to Rainfall–Runoff

Yanbin Yin^{1,2}, Zhen-Hua Fan^{1,2}, and Xuebin Chen^{1,2}

¹State Key Laboratory of Hydrology and Water Resources, Hohai University, Nanjing, China, ²State Key Laboratory of Water Resources and Environment, Hohai University, Nanjing, China

Key Points:
• A global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance.
• The regionalization method is based on a machine learning algorithm.
• High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

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

Correspondence to:
Yanbin Yin, yin@hohai.edu.cn

Received: 12 October 2017
Accepted: 12 October 2017
Published: 12 October 2017

Editor: David M. Legler, Editor, Water Resources Research

Copyright: © 2017 Yin et al. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

How to Cite: Yin, Y., Fan, Z., & Chen, X. (2017). Regionalization in a Global Hydrological Model From Physical Descriptors to Rainfall–Runoff. *Water Resources Research*, 53, 6005–6022. doi:10.1029/2017WR021000

Abstract The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

Plain Language Summary The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

Introduction The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

1. Introduction The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

2. Data and Model Description The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

3. Results and Discussion The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.

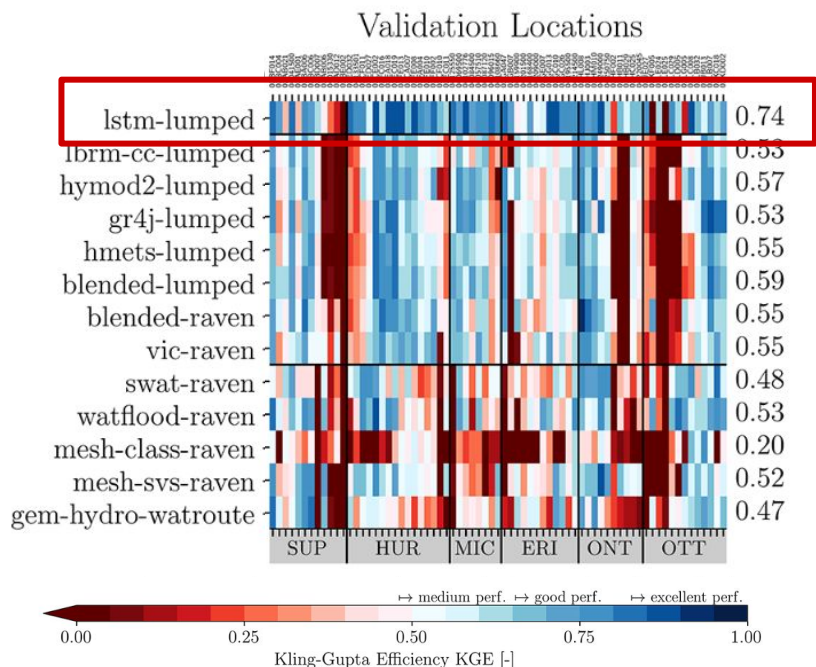
4. Conclusions The global hydrological model (GLIM-Hydro) is regionalized using physical descriptors to improve model performance. The regionalization method is based on a machine learning algorithm. High-dimensional descriptors (hydrological, meteorological, and land use) are used to regionalize the model.



All aboard the hype train!

On the ascendancy of ML methods
in hydrologic modeling

ML techniques surpass hydrologic models for streamflow prediction



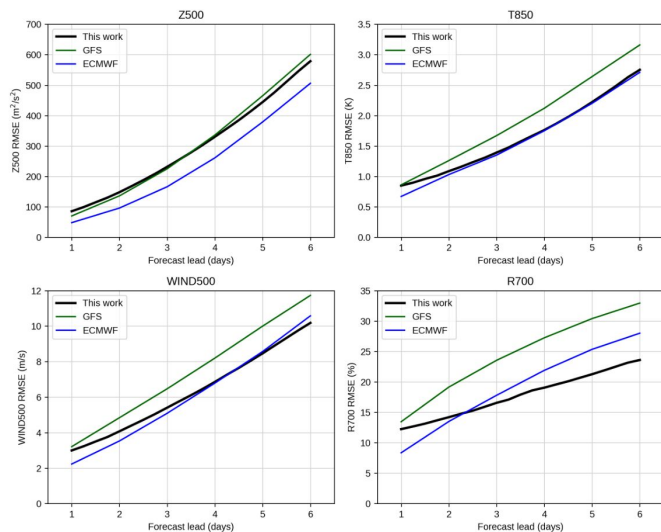
HYDROFORECAST		HydroForecast won out all five categories at the competition hosted by the United States Bureau of Reclamation, CEATI and hydropower utilities and verified by RTI International.		
		CATEGORY	DESCRIPTION	
1 ST	All-Around	ACROSS ALL metrics, horizons and flow ranges		
	Flood Forecaster	HIGHEST flow range		
	Eagle-Eye	LONGEST forecast horizon		
	Quick Draw	SHORTEST forecast horizon		
	Straight Shooter	LOWEST bias		

HydroForecast wins out at the Competition for Emerging Inflow Forecasting Technologies: “The first goal of the competition was to determine whether the HPC method utilized by the National Water Model, or the AI methods utilized by firms Upstream Tech and Sapere Consulting could beat the existing models used by the member agencies and River Forecast Centers in a fully operational setting. The second goal of the competition was to develop a record of forecast performance across products and regions that members could leverage to evaluate their own performance on their river systems”

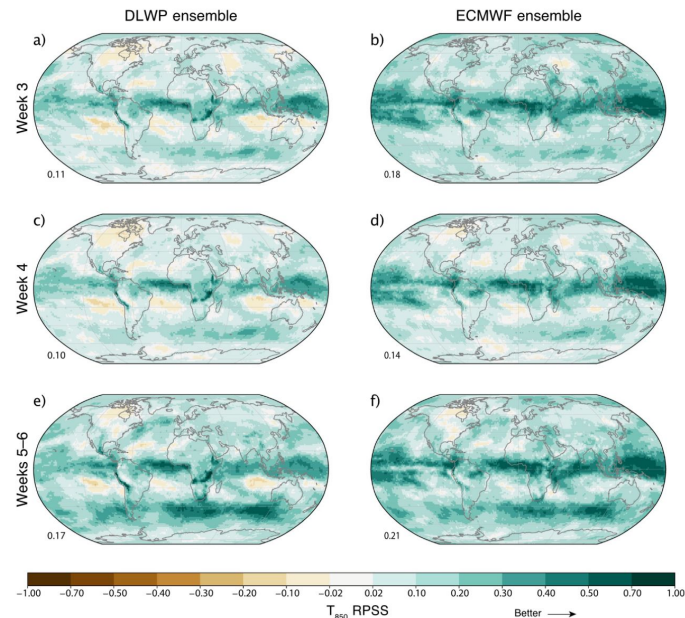
Mai, J., Shen, H., Tolson, B. A., Gaborit, É., Arsenault, R., Craig, J. R., et al. (2022). The Great Lakes Runoff Intercomparison Project Phase 4: the Great Lakes (GRIP-GL). *Hydrology and Earth System Sciences*, 26(13), 3537–3572. <https://doi.org/10.5194/hess-26-3537-2022>

<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!**

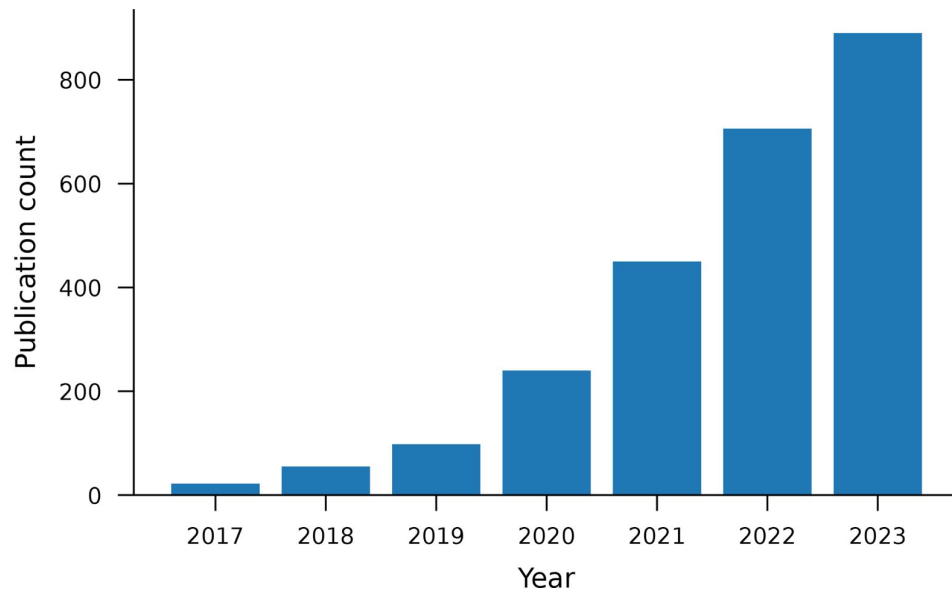


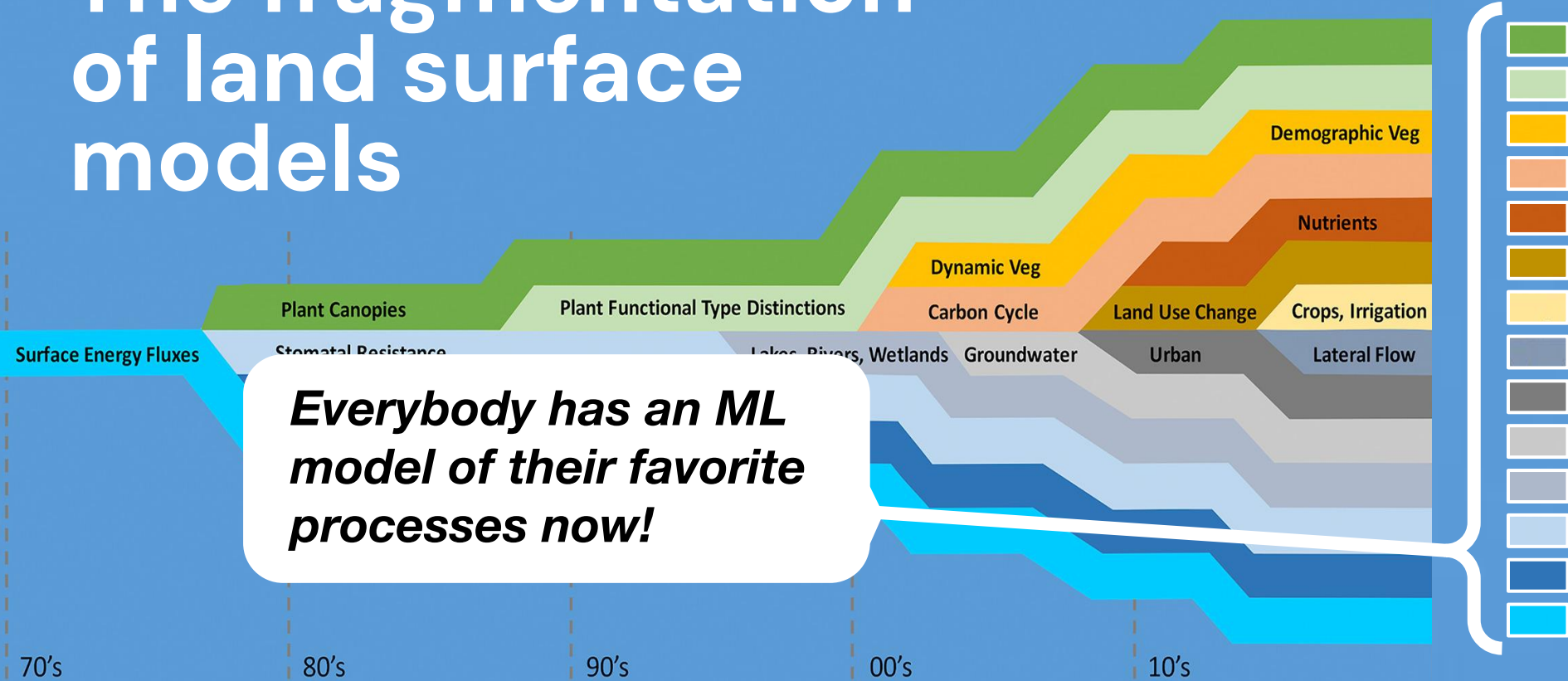
Figure 1. Number of hydrological publications related to rainfall–runoff modeling with LSTM networks over time based on data retrieved from Google Scholar in April 2024.



Oh no, it's not a train after all!

Model proliferation as a social problem

The fragmentation of land surface models



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



CLiMA
CLIMATE MODELING ALLIANCE

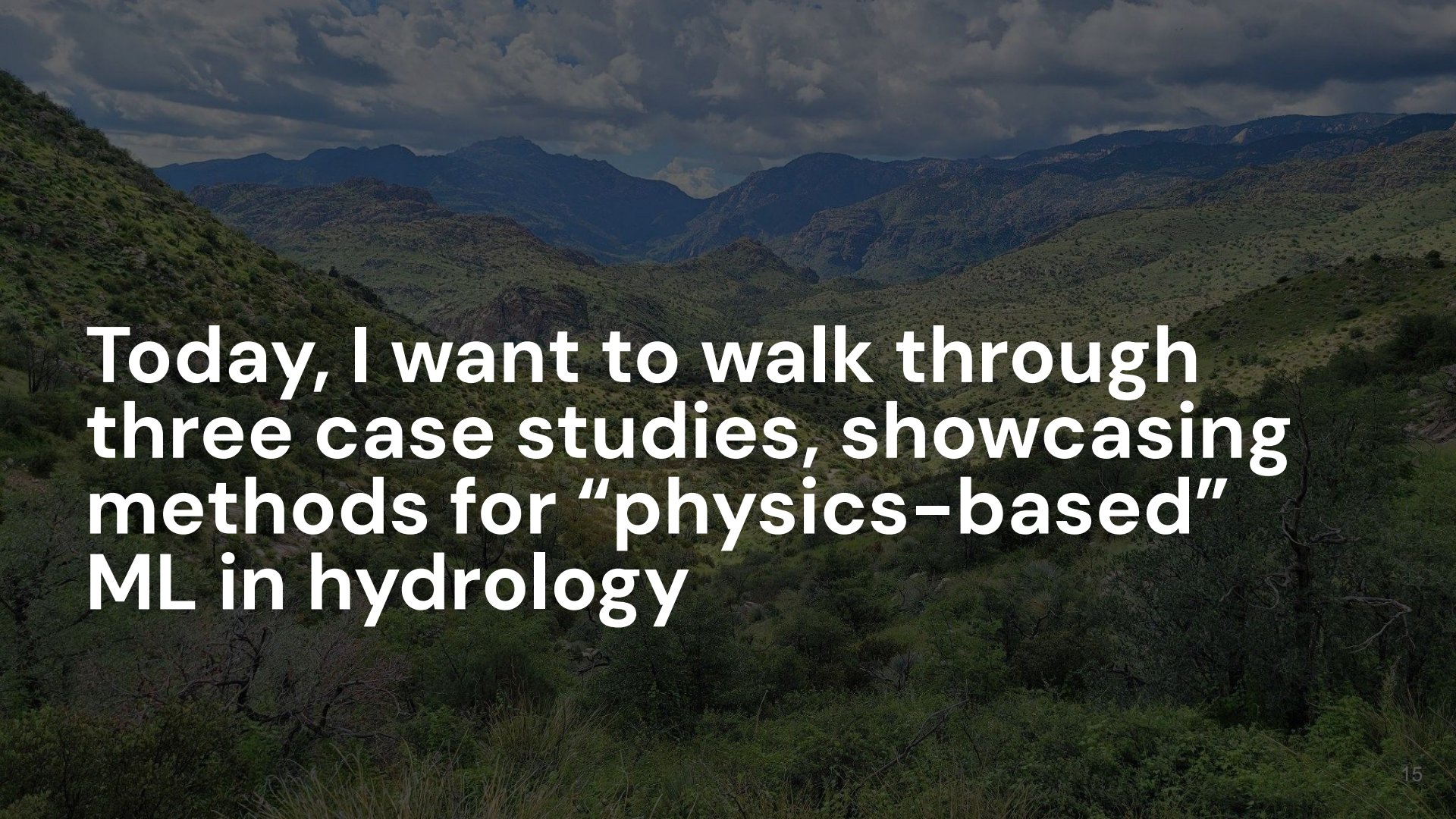


LEAP

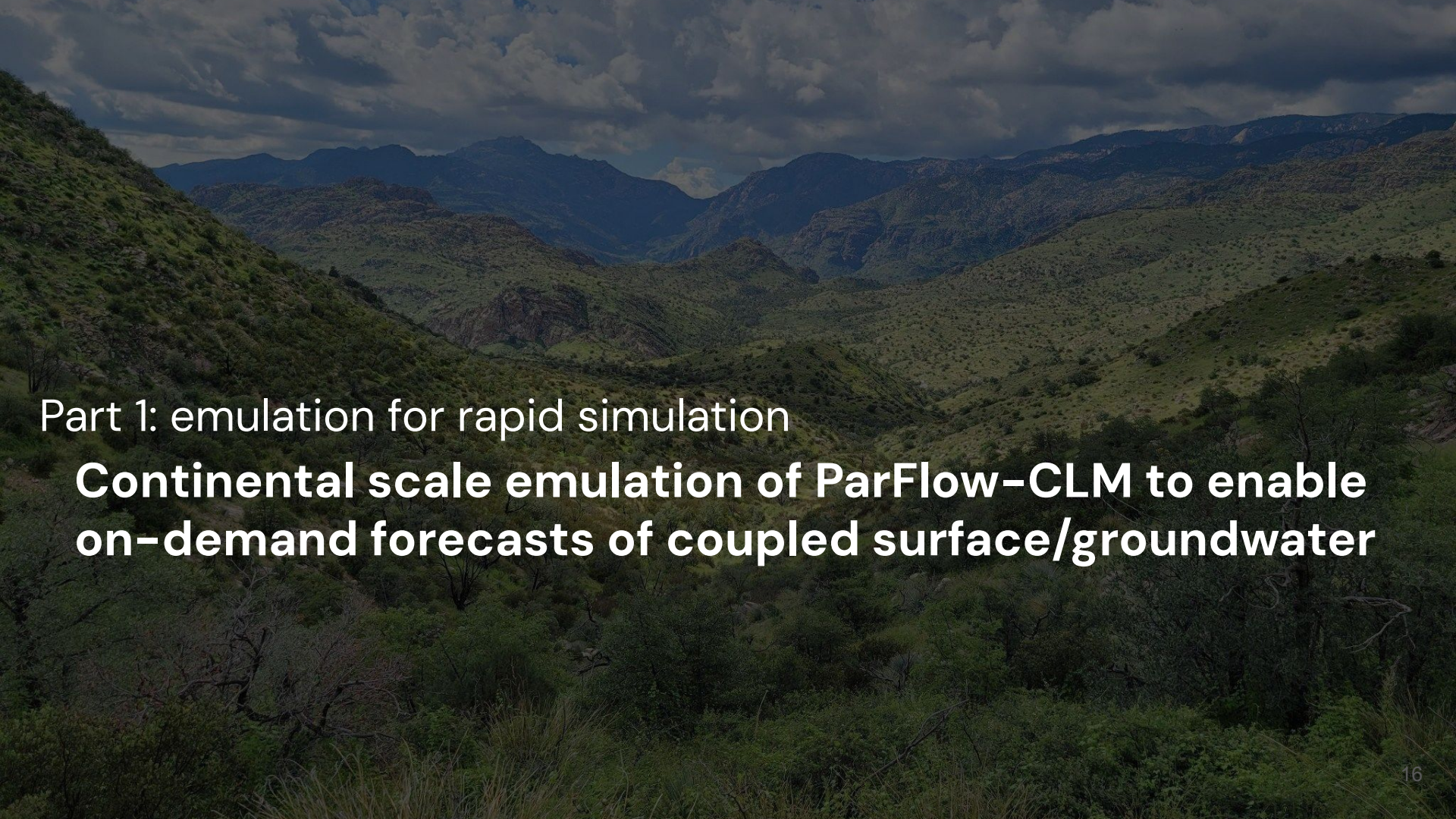


A landscape photograph of a mountain valley. The foreground and middle ground are filled with green, hilly terrain covered in sparse vegetation. In the background, there are several layers of mountain ranges, with the furthest ones appearing in shades of blue due to atmospheric perspective. The sky is filled with large, dark, dramatic clouds. The overall tone is somewhat somber due to the overcast sky.

**We need a vision for
integrating both
modeling approaches**

The background of the slide is a photograph of a vast, mountainous landscape. In the foreground, there are green, grassy hills with some sparse vegetation. The middle ground shows a series of rolling hills and valleys, with some rocky outcrops visible. In the far distance, more mountain ranges are visible under a sky filled with large, white and grey clouds. The overall tone of the image is natural and somewhat somber due to the overcast sky.

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

A landscape photograph of a mountain valley. The foreground and middle ground are filled with green, hilly terrain covered in sparse vegetation. In the background, there are several layers of mountain ranges, with the furthest ones appearing in shades of blue due to atmospheric perspective. The sky is filled with large, white and grey clouds.

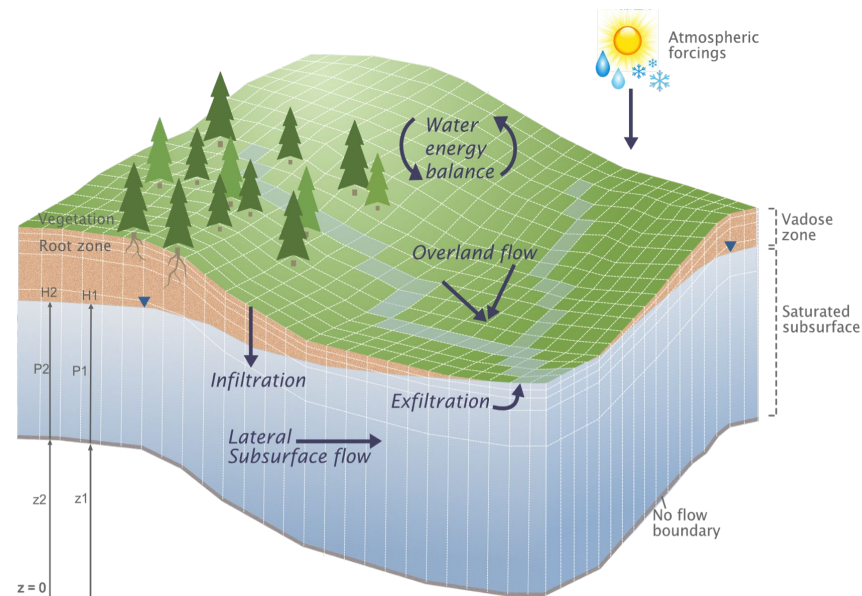
Part 1: emulation for rapid simulation

**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

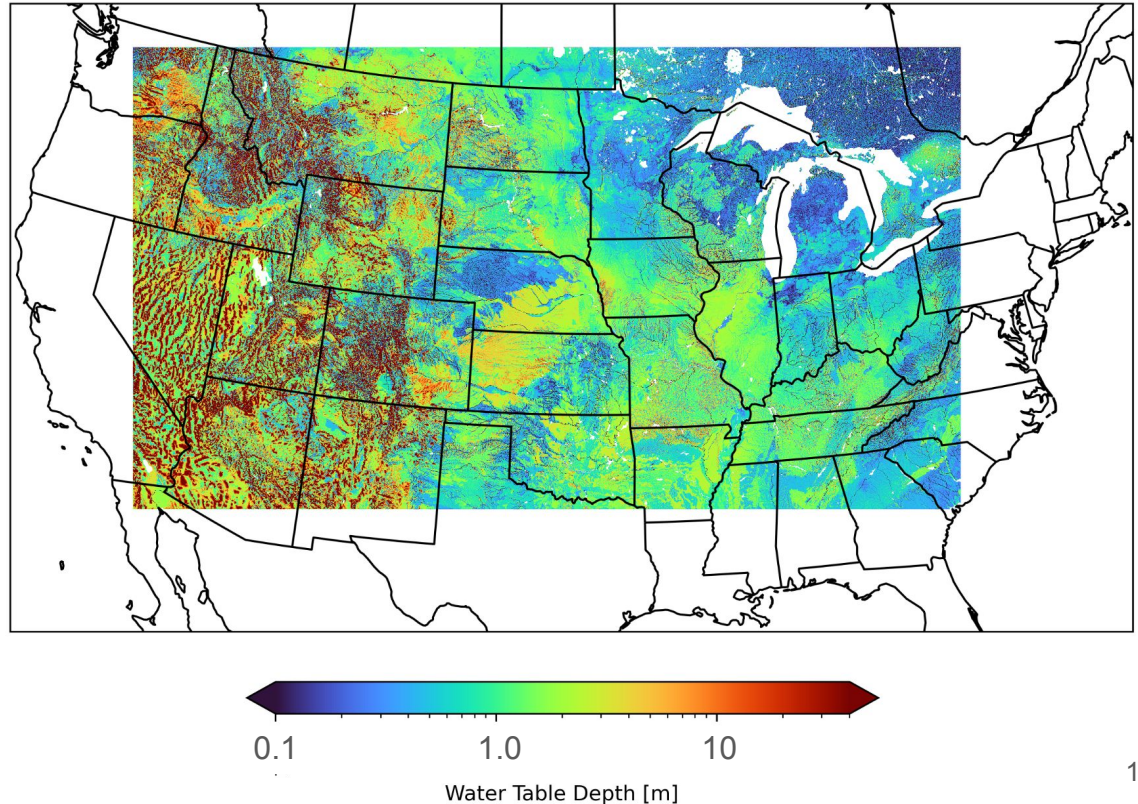


But, train emulators based on the models

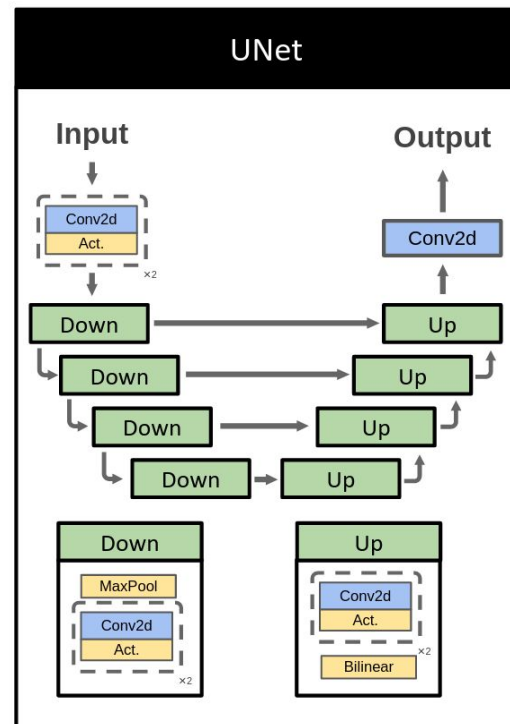
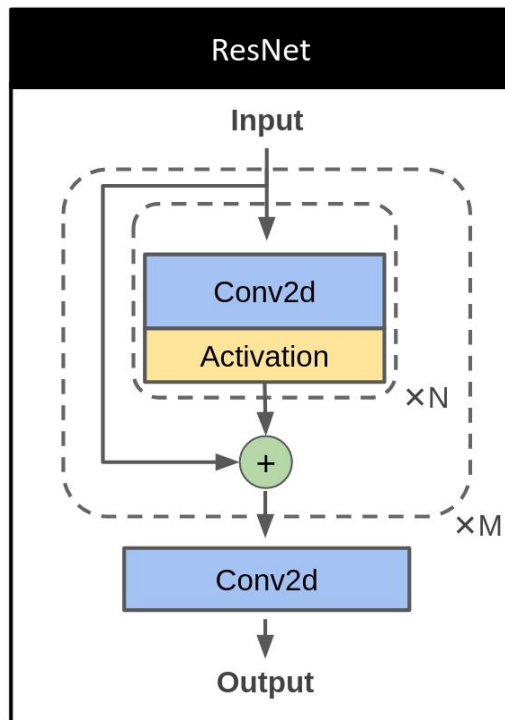
We are emulating
continental scale, high
resolution (1km)
simulations using
ParFlow-CLM

Resolving full
spatiotemporal dynamics
(3d + time)

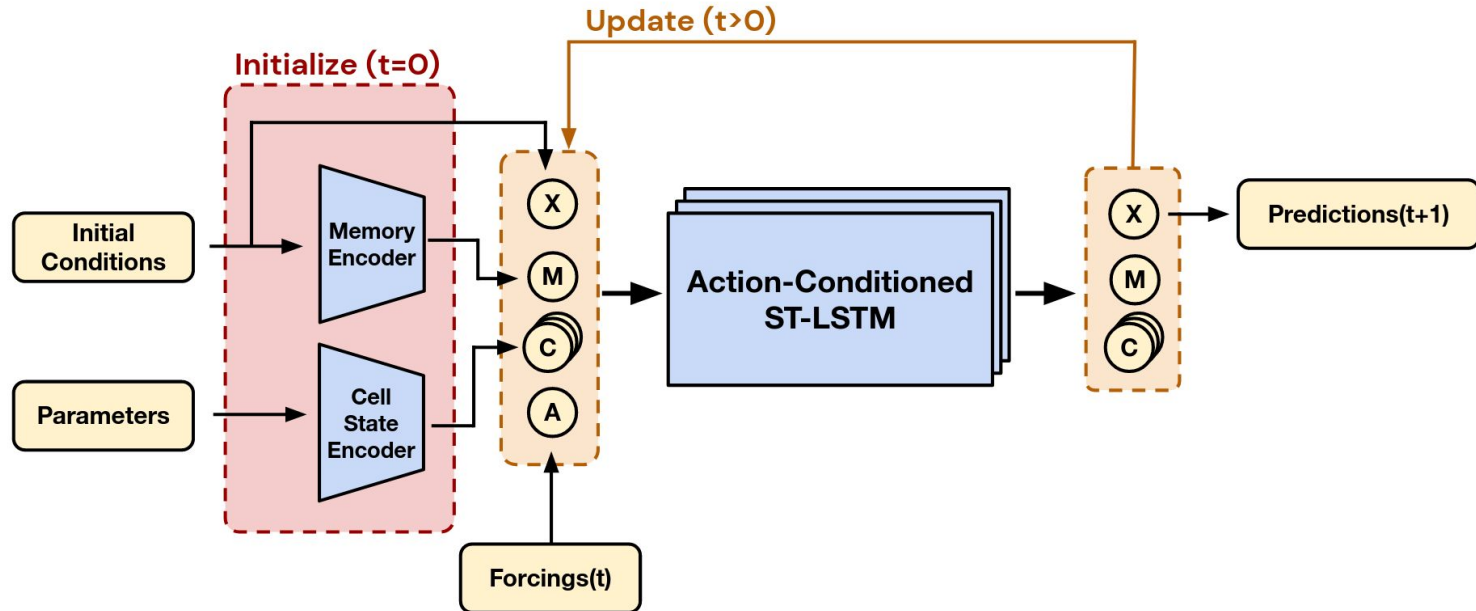
Processes vary across
orders of magnitude in
both space and time



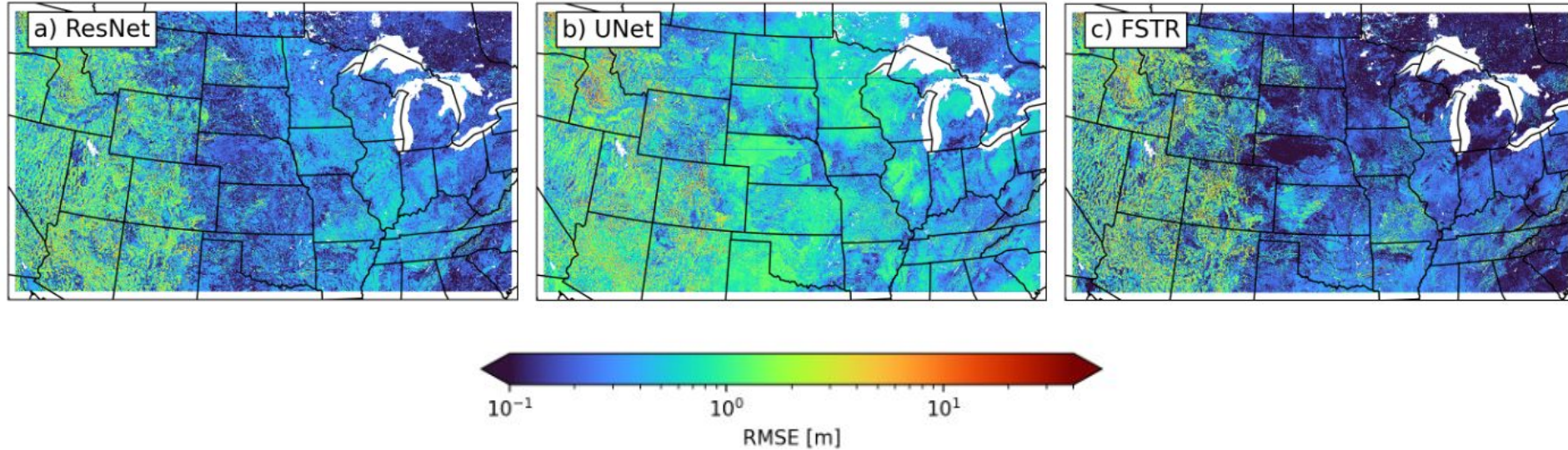
We used two baseline architectures for validation and intercomparison purposes



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

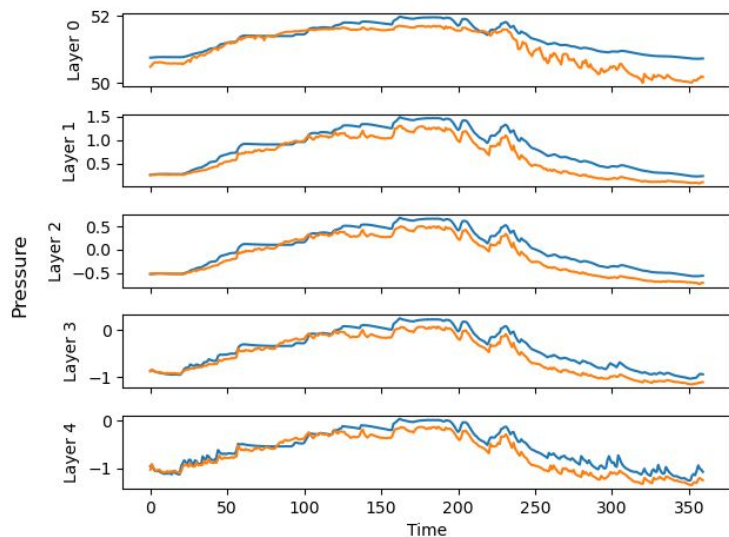


FSTR outperforms 2 standard neural network architectures on predictions of water table depth

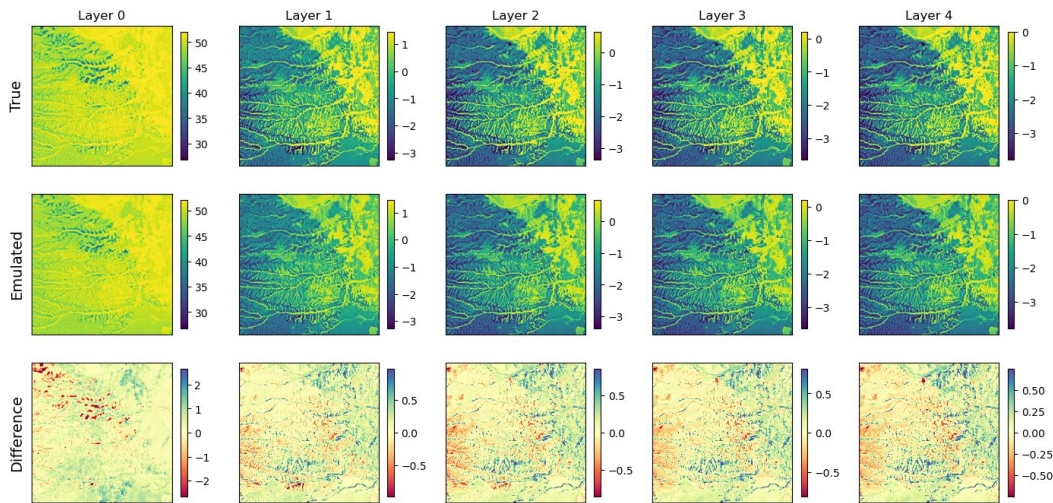


Overall, FSTR is extremely good at matching full 4-d spatiotemporal patterns

Capturing temporal variability



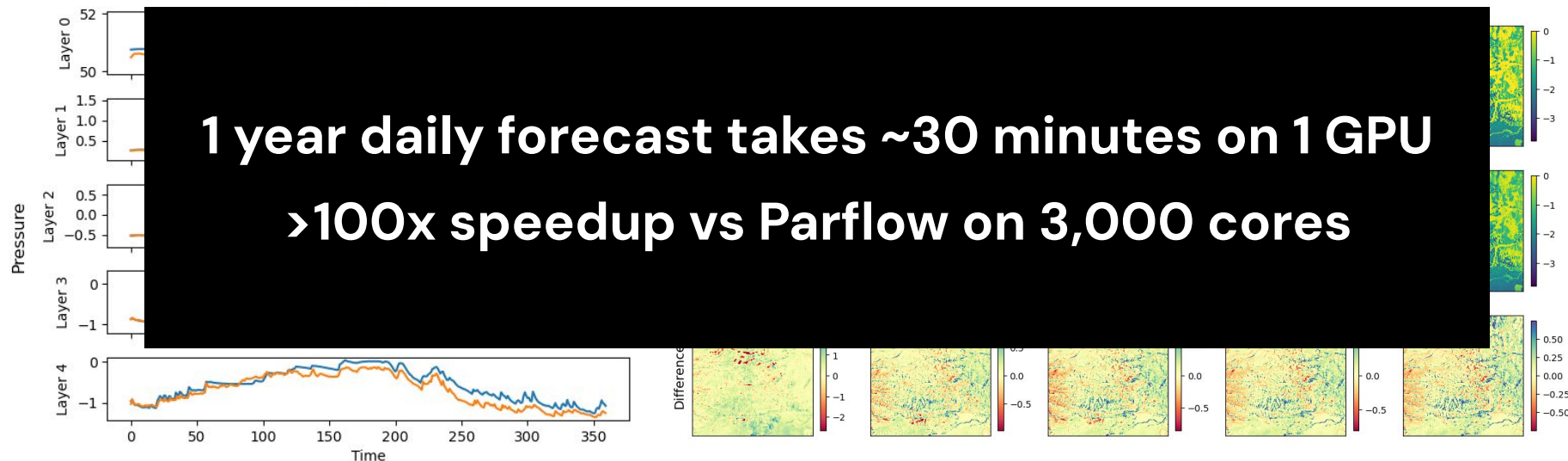
Capturing spatial variability



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

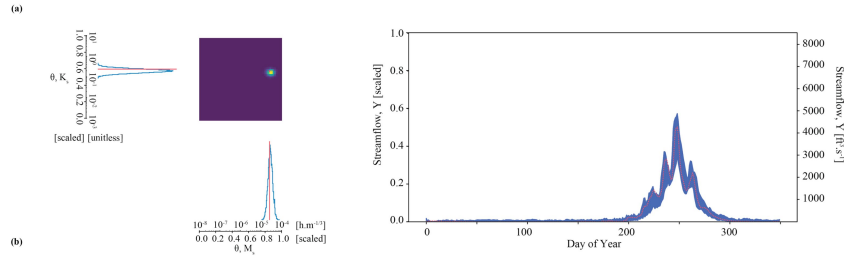
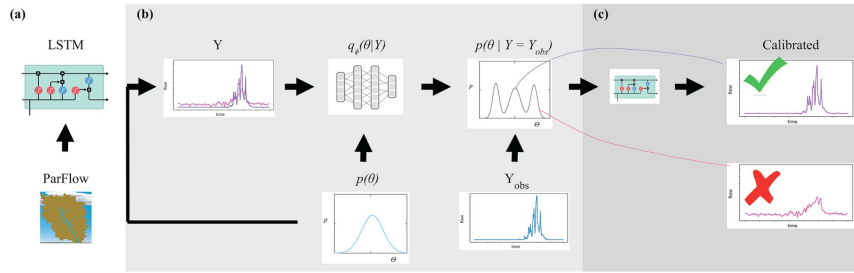
Capturing temporal variability

Capturing spatial variability



Bonus materials from this line of work:

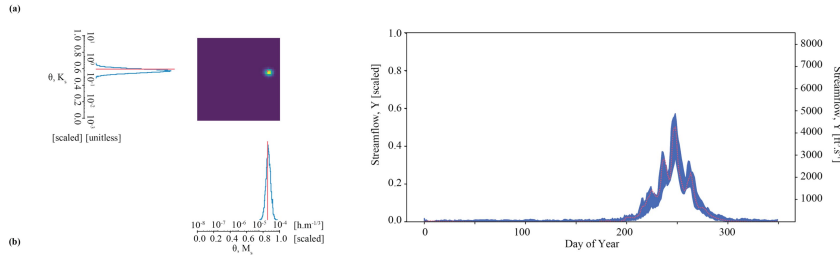
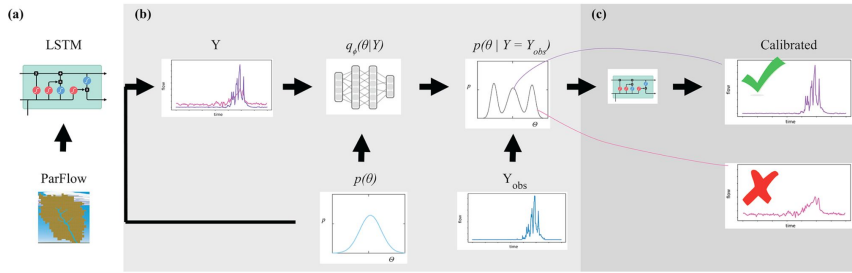
Simulation-based inference



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

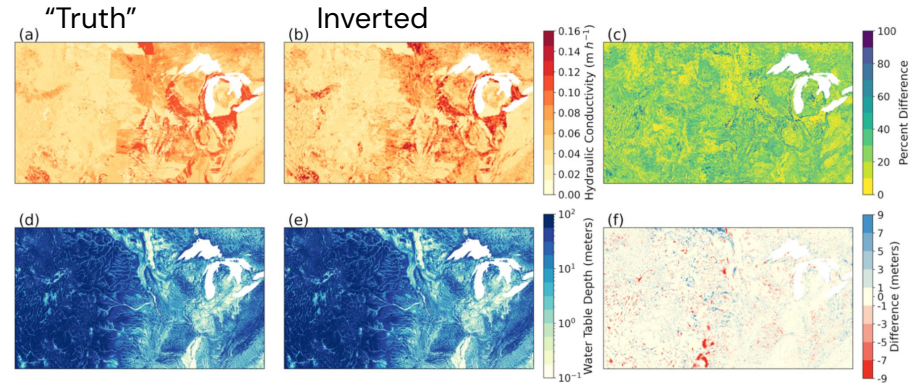
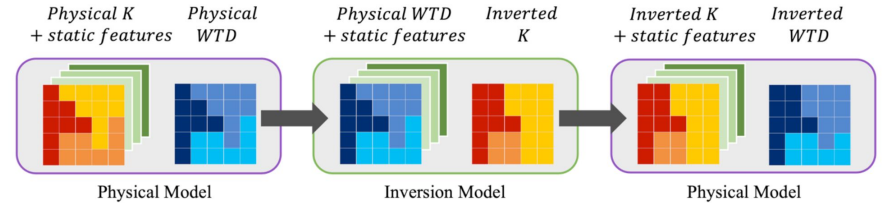
Bonus materials from this line of work:

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



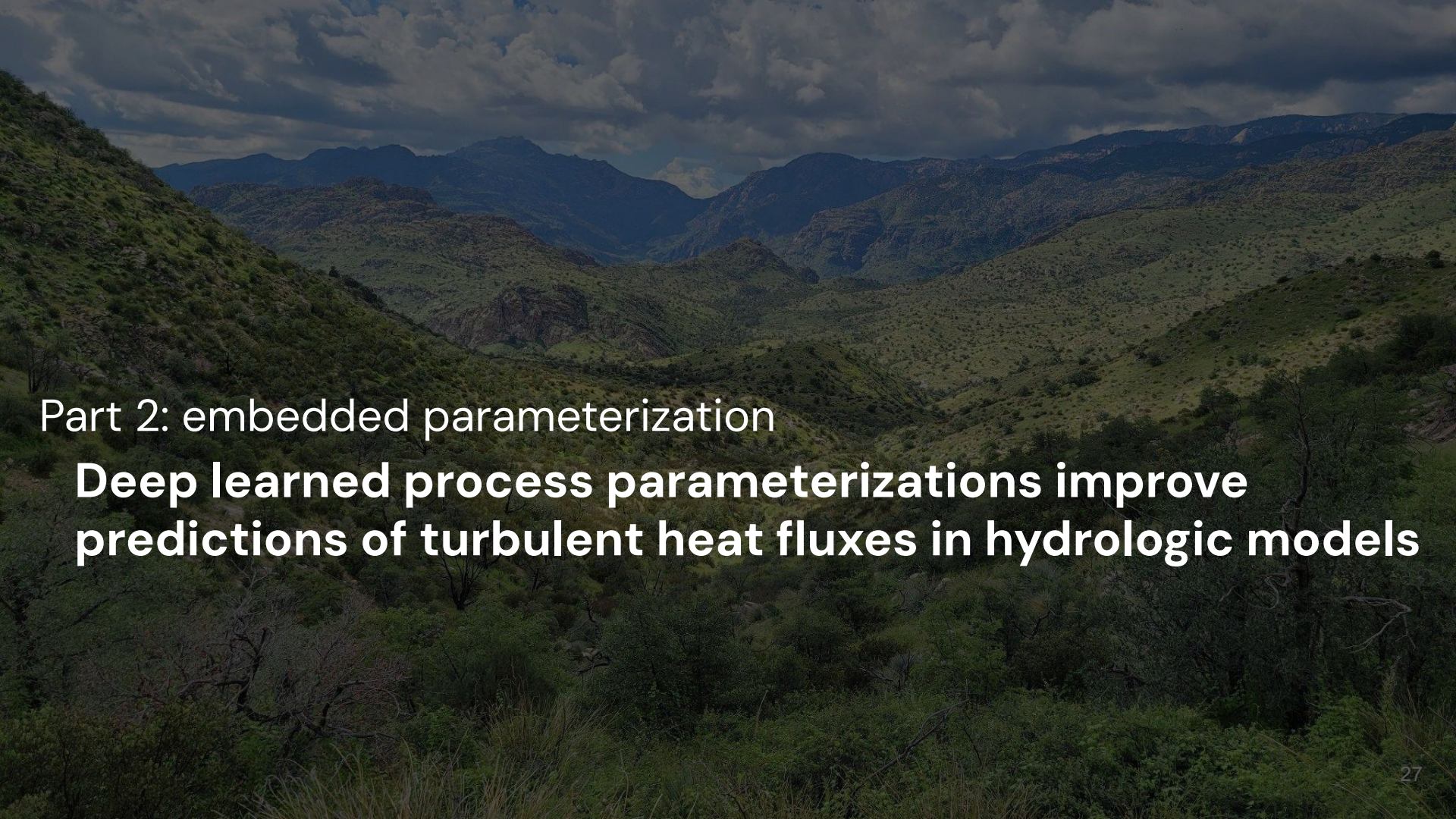
Amanda Triplett, Andrew Bennett, Laura Elizabeth Condon, et al. A Deep-Learning Based Parameter Inversion Framework for Large-Scale Groundwater Models. *Geophysical Research Letters*. In press.

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

A landscape photograph of a mountain valley. The foreground and middle ground are filled with green, hilly terrain covered in dense vegetation. In the background, there are several layers of mountain ranges, with the furthest ones appearing in shades of blue. The sky is filled with large, white and grey clouds. The overall lighting is somewhat dim, suggesting an overcast day.

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

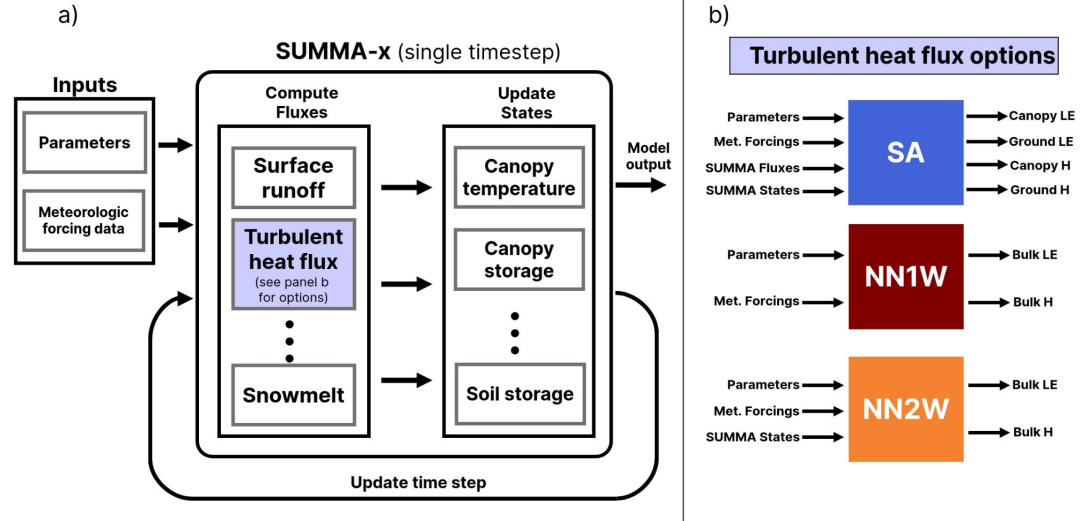
This work was in collaboration with Bart Nijssen (University of Washington)

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

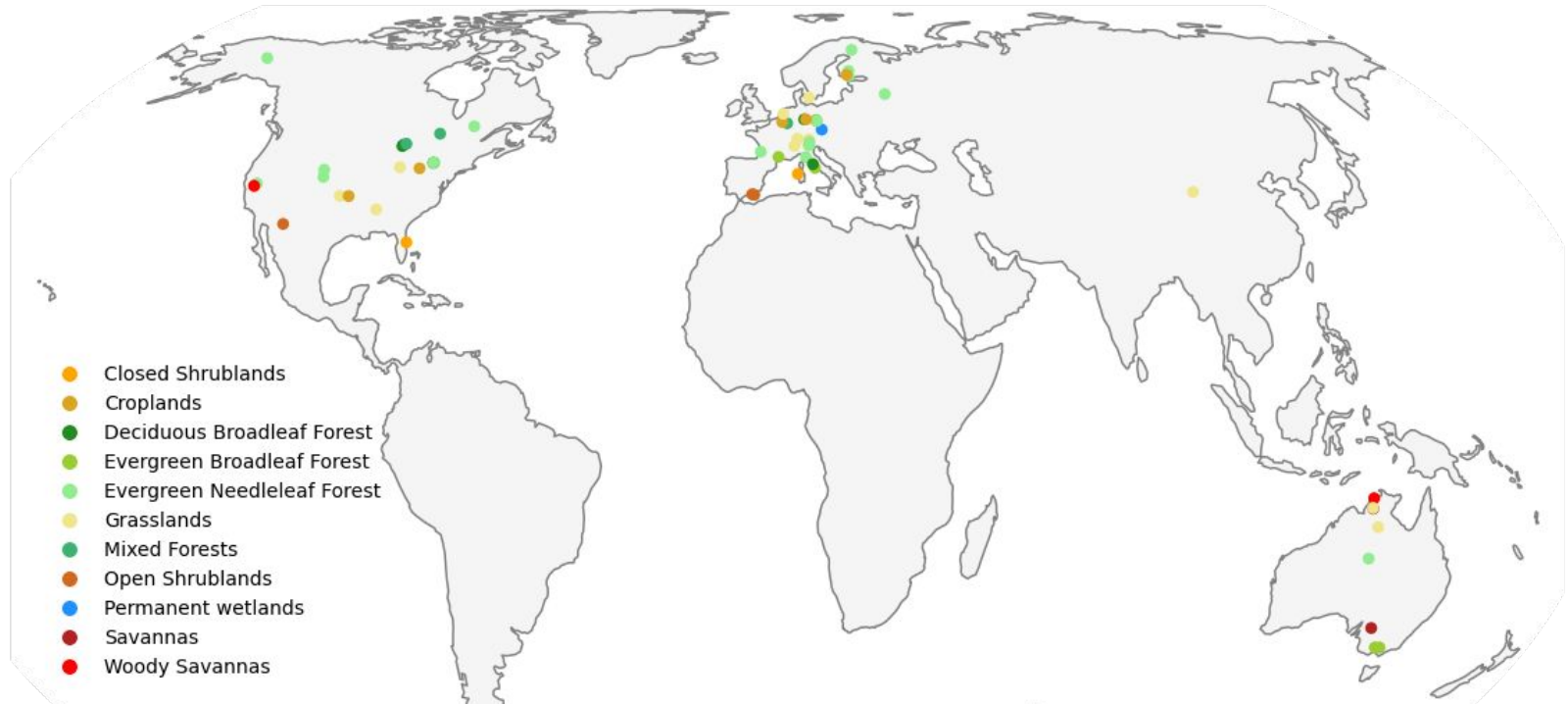
Our study focused on turbulent heat fluxes, which control ET and temperatures of the land surface

We explored three scenarios:

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

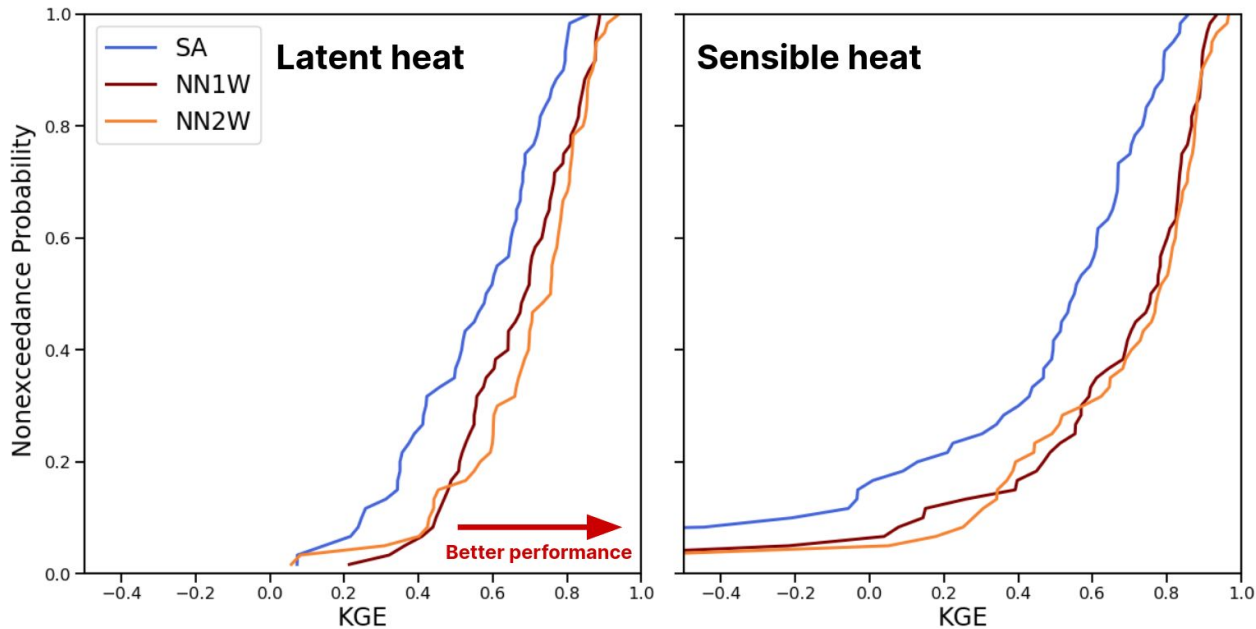


**We gathered data from 60 FluxNet sites,
totalling over 500 site-years of half-hourly data**

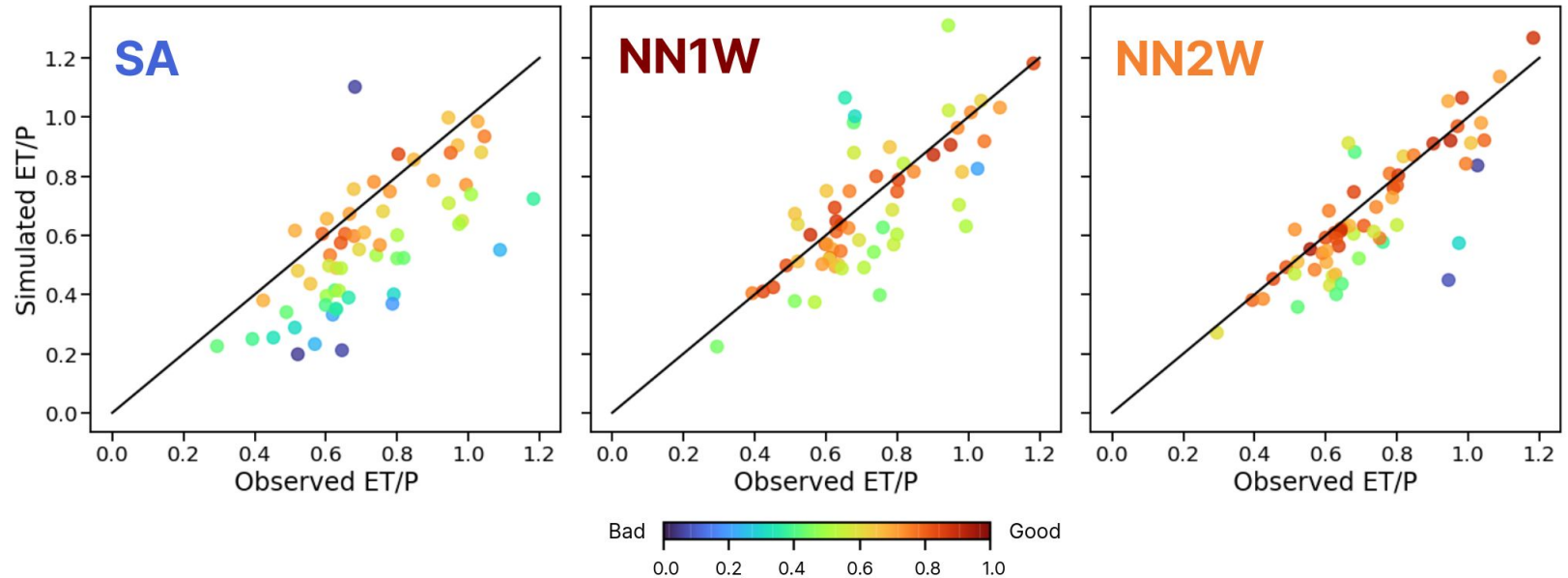


Both neural net methods show improved performance over process-based method

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

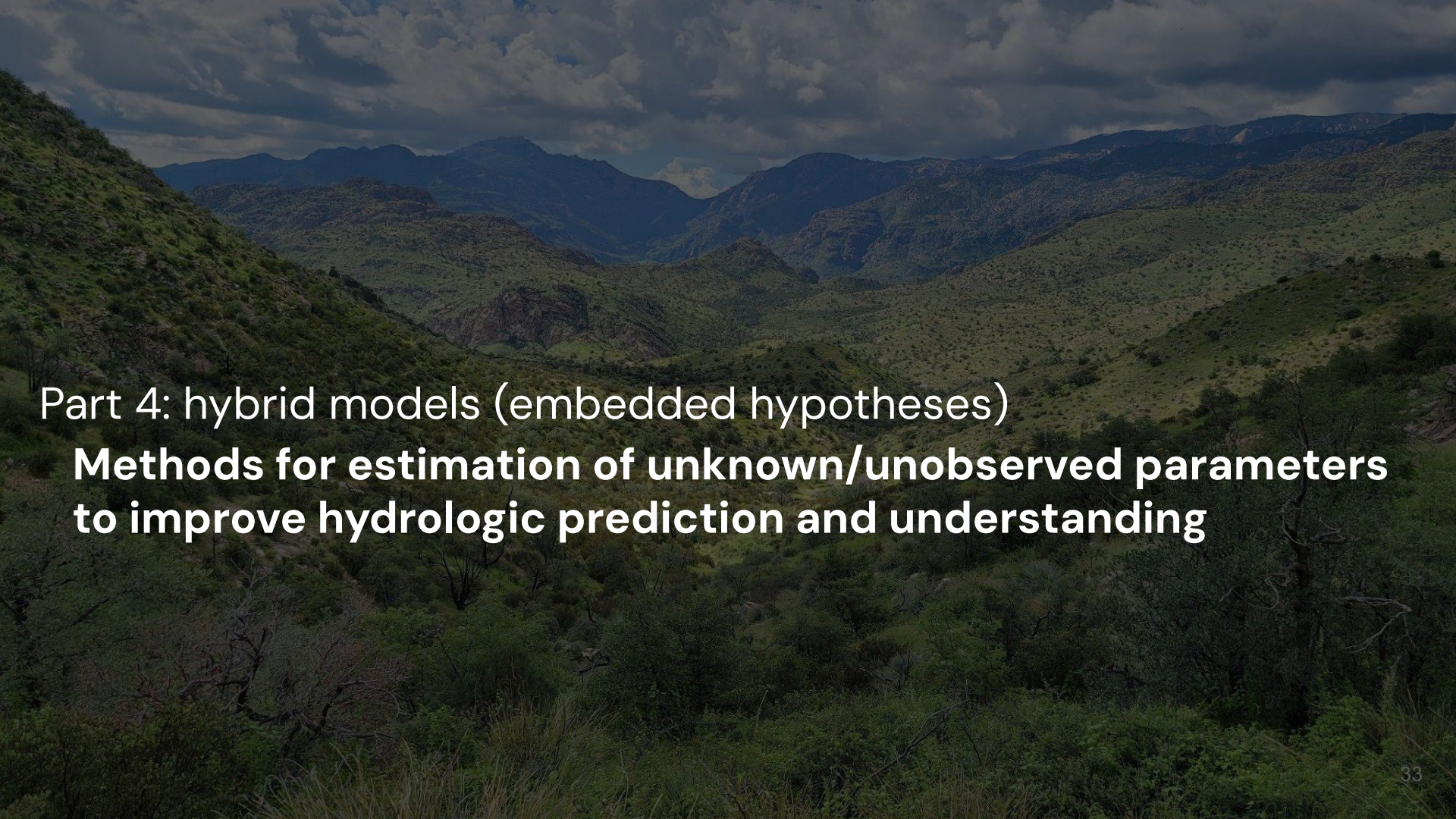


Key takeaways:

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

A landscape photograph of a mountain valley. The foreground and middle ground show rolling green hillsides covered in dense vegetation, including shrubs and small trees. In the background, a range of rugged mountains is visible under a sky filled with large, grey, overcast clouds. The lighting is somewhat dim, suggesting an overcast day.

Part 4: hybrid models (embedded hypotheses)

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

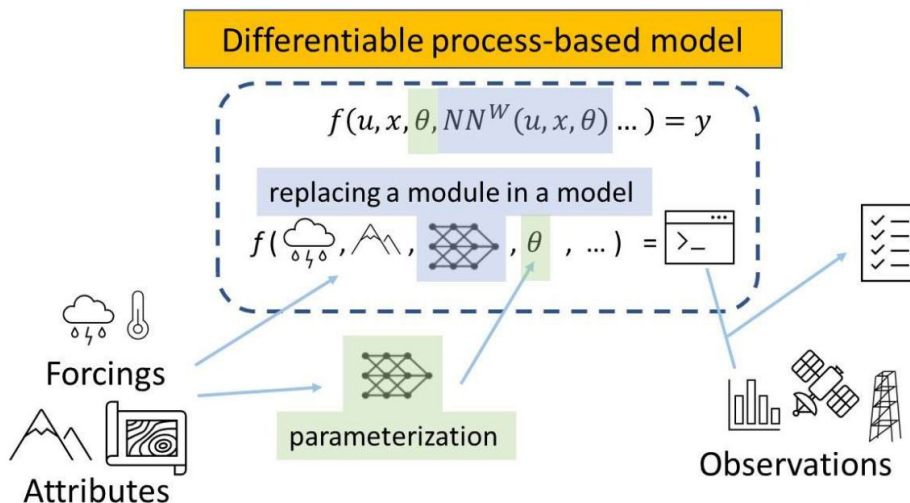
Methods for estimation of unknown/unobserved parameters to improve hydrologic prediction and understanding

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>

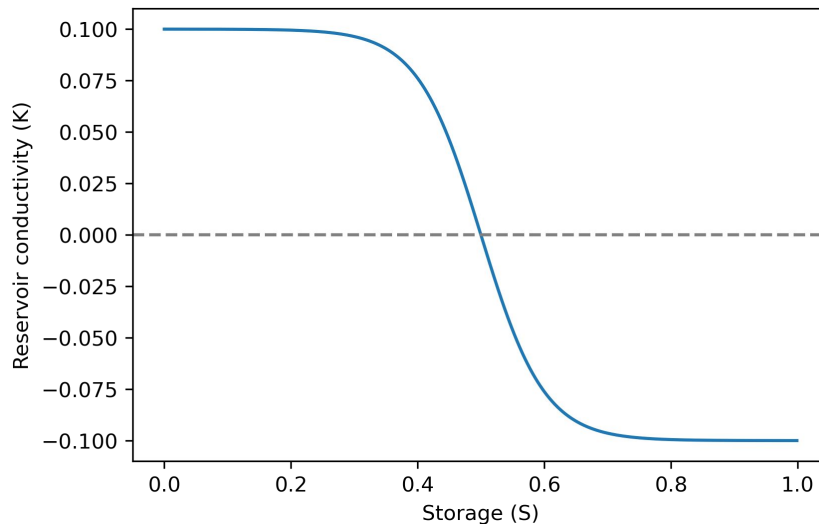
A primer on differentiable modeling

Consider a simple, nonlinear reservoir:

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

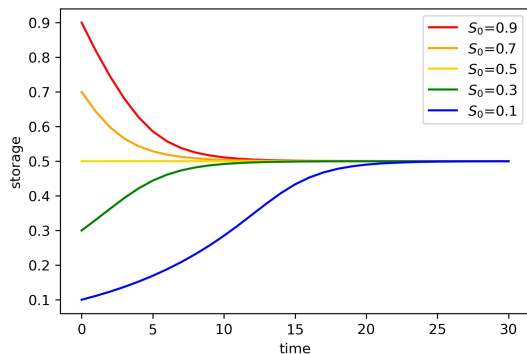
“unknown”

And 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”:

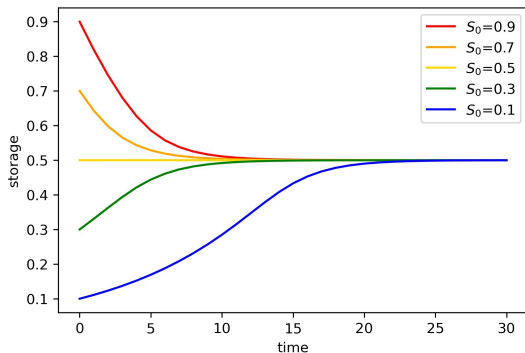


Fit a model where the parameterization is “learned” by a neural network:

$$\frac{dS}{dt} = \text{Neural Network} \cdot S$$

A primer on differentiable modeling

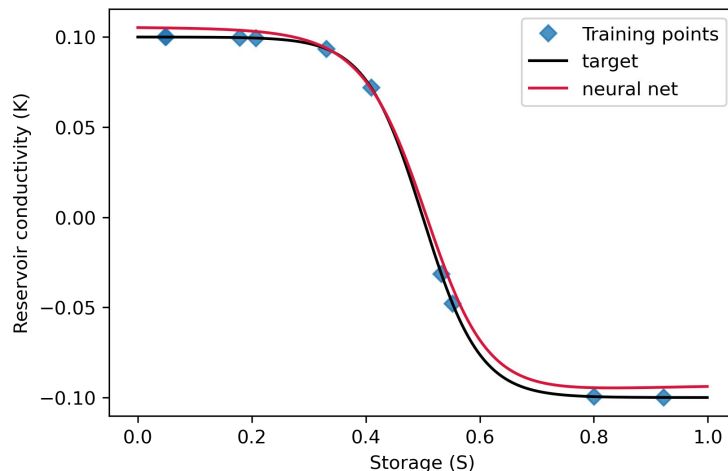
Given many samples derived from the system that we want to “learn”:



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

$$\frac{dS}{dt} = \text{[Neural Network Icon]} \cdot S$$

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



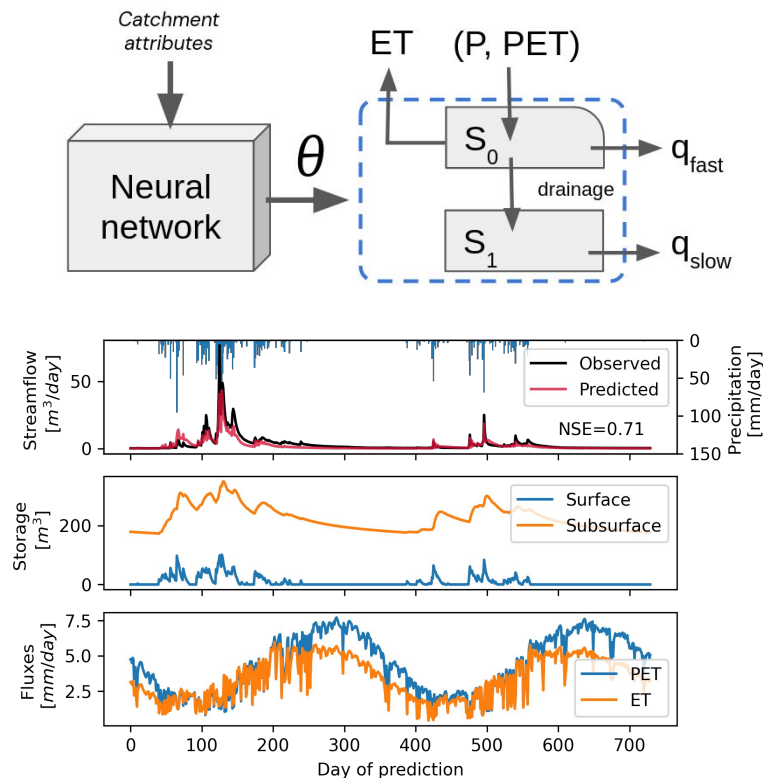
A primer on differentiable modeling

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:

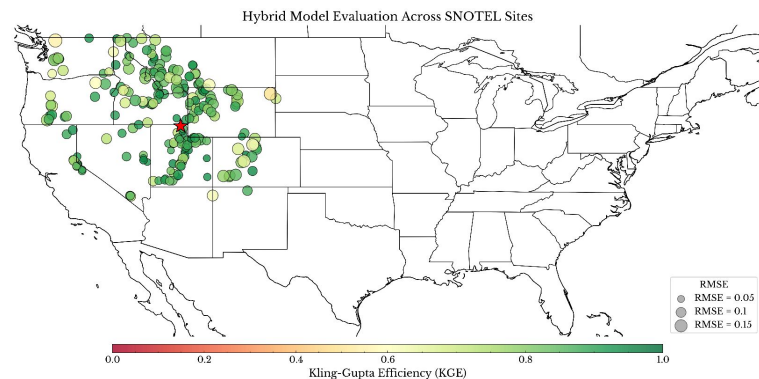
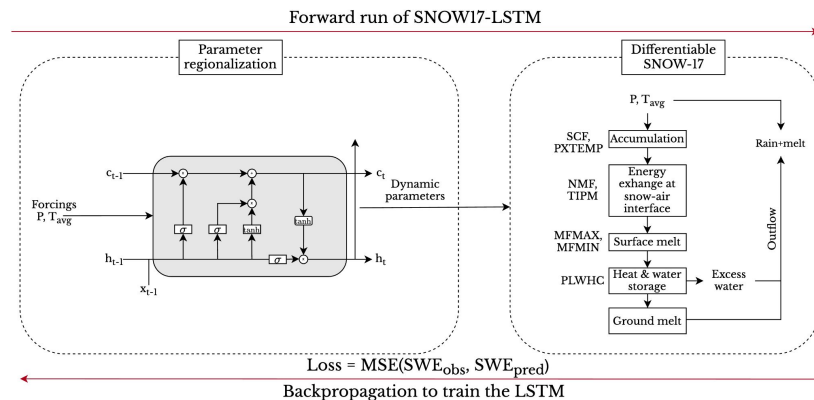
https://github.com/earth-artificial-intelligence/earth_ai_book_materials/blob/main/chapter_07/ai_for_physics_inspired_hydrology_modeling.ipynb



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 sites in the western US to predict snow water equivalent (SWE)



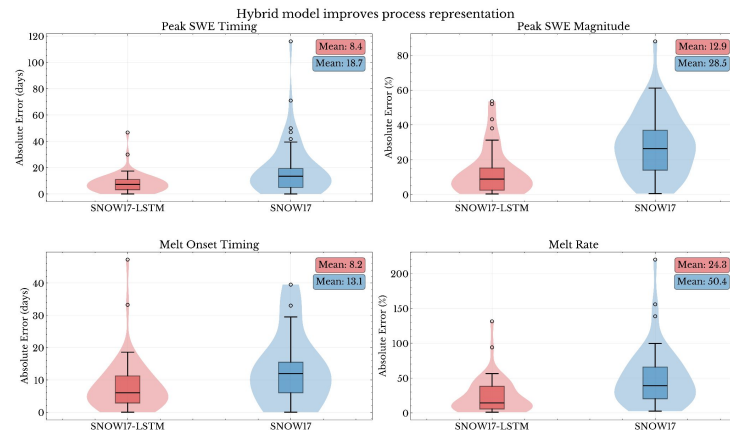
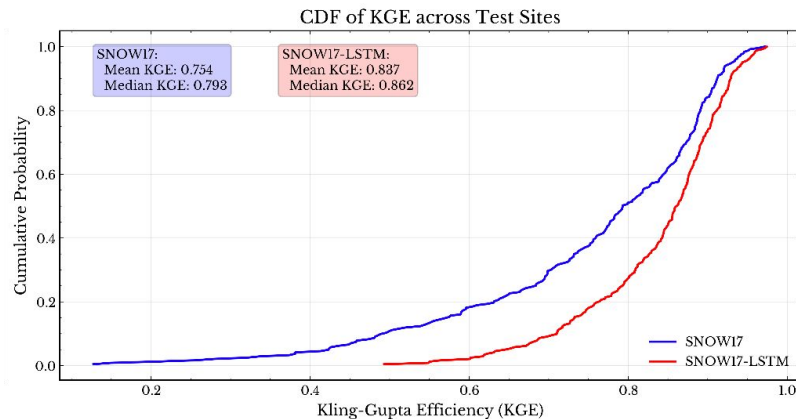
We are applying this method to multiple processes:

This work is led by Aamir Lamichhane

We parameterize the SNOW17 model with a LSTM, and train across Snotel sites 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

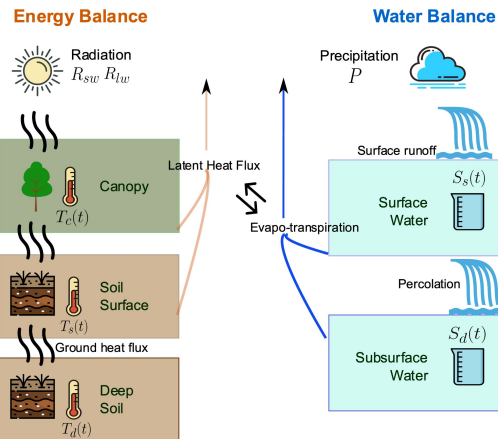


We are applying this method to multiple processes:

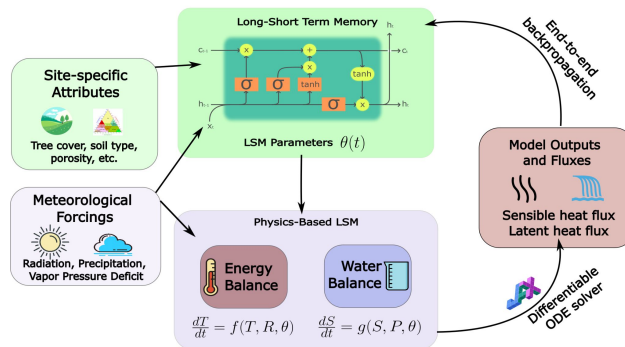
This work is led by Nabin Kalauni

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



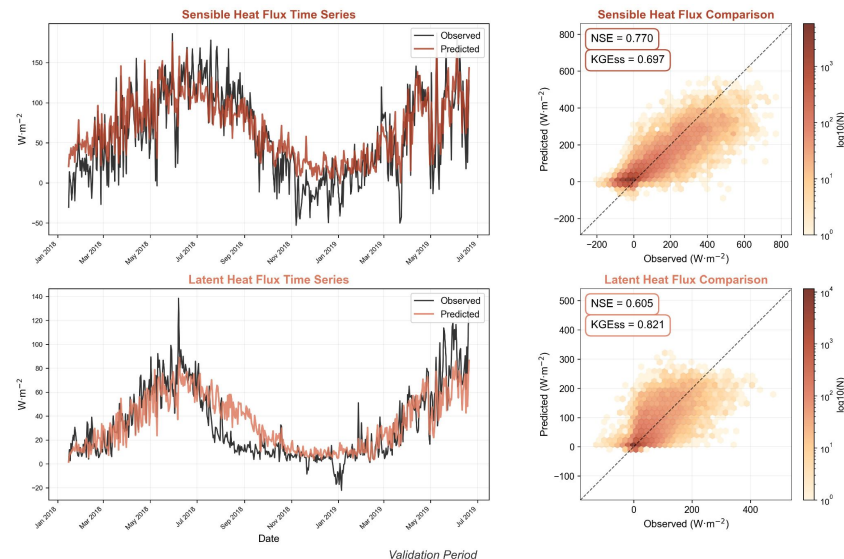
We are applying this method to multiple processes:

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!!

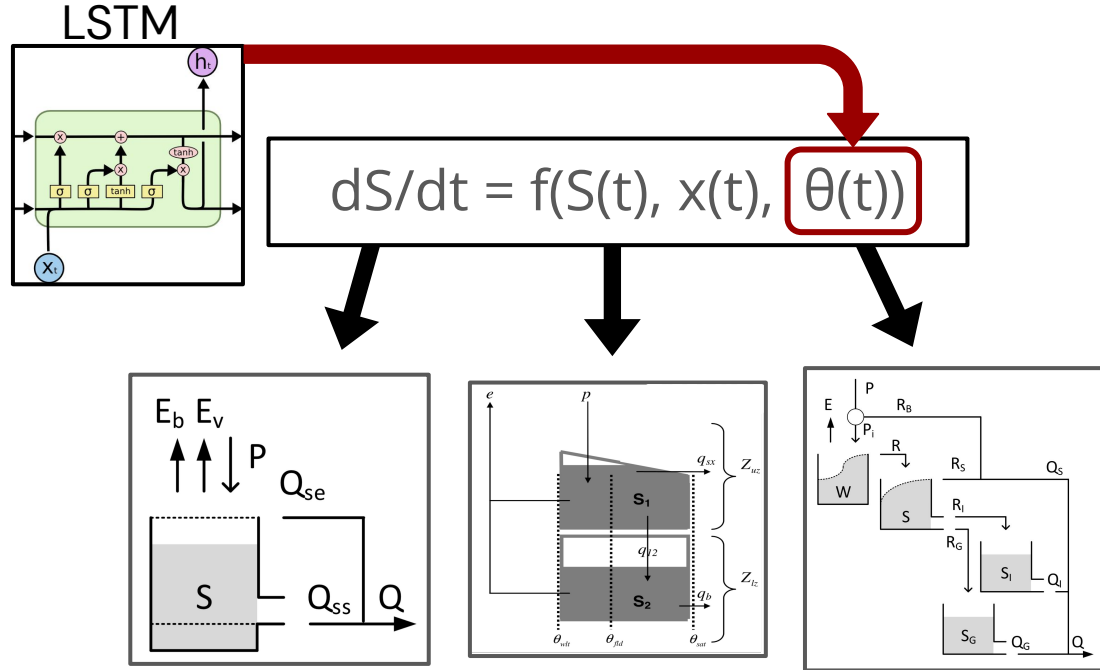


But, broader philosophical questions linger. We are tackling these with a multi-model approach

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

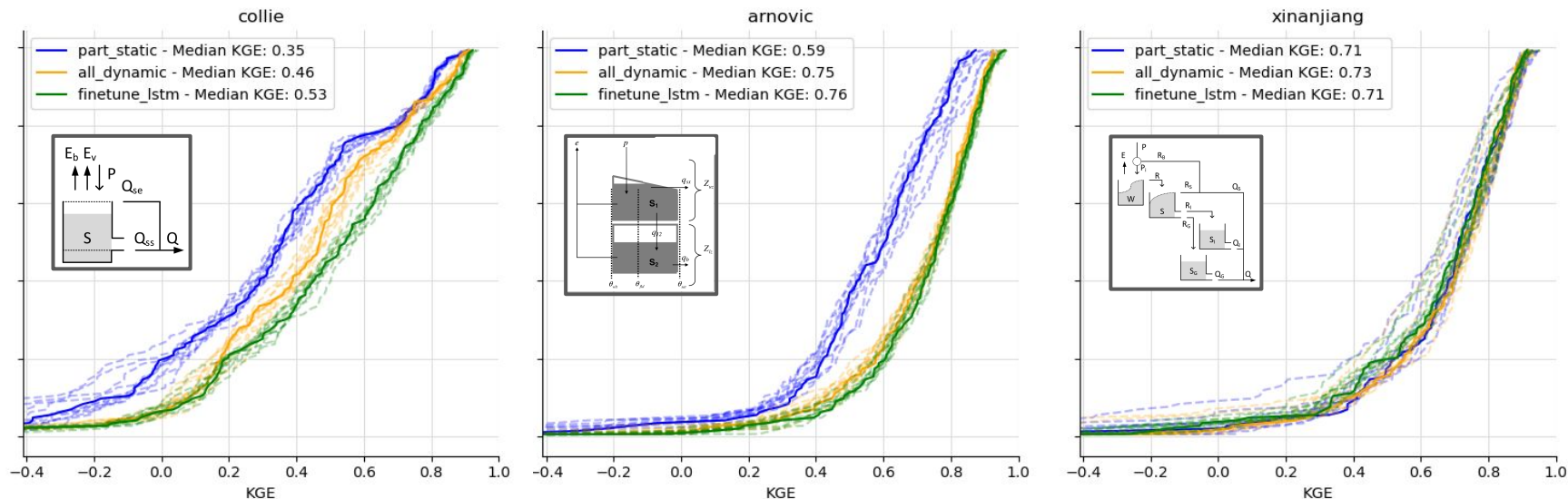
We do this for 140 catchments across the US



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

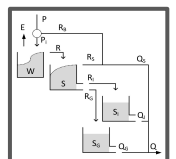
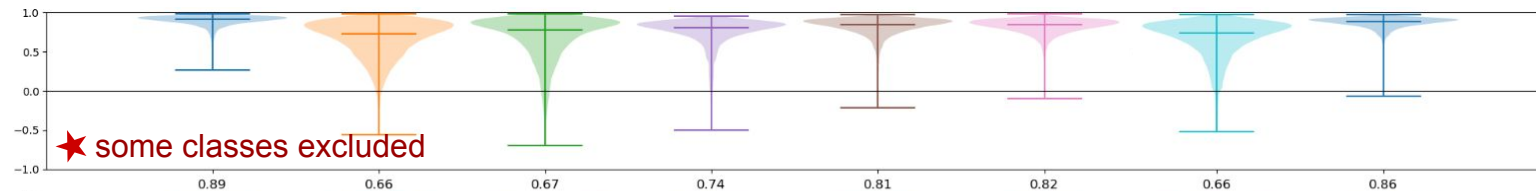
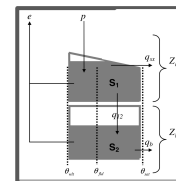
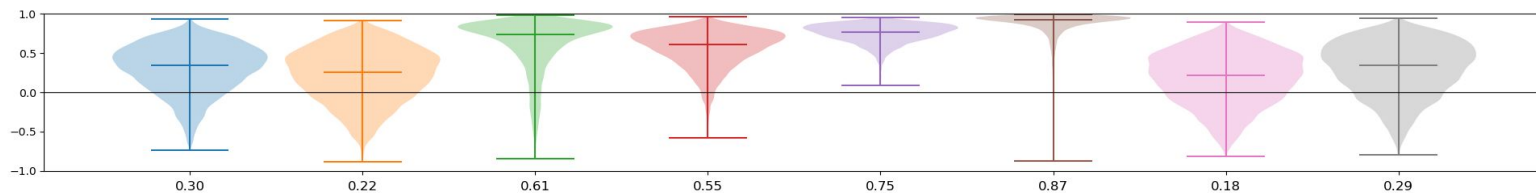
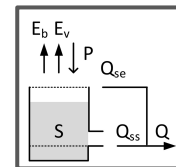
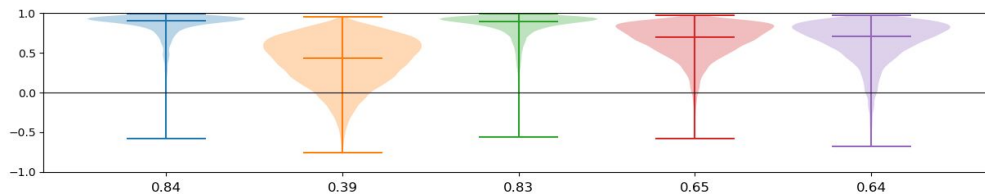
Figure shows CDF of performance across all 140 sites (SOTA: median KGE ~0.75)

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?**

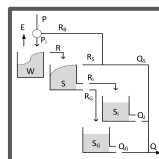
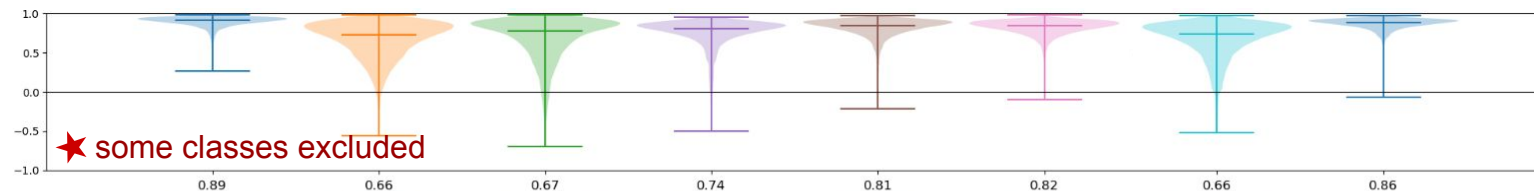
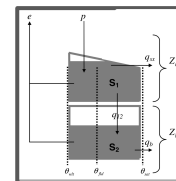
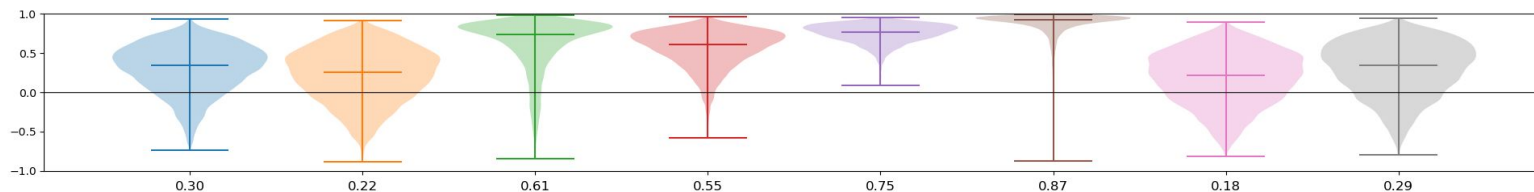
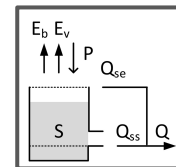
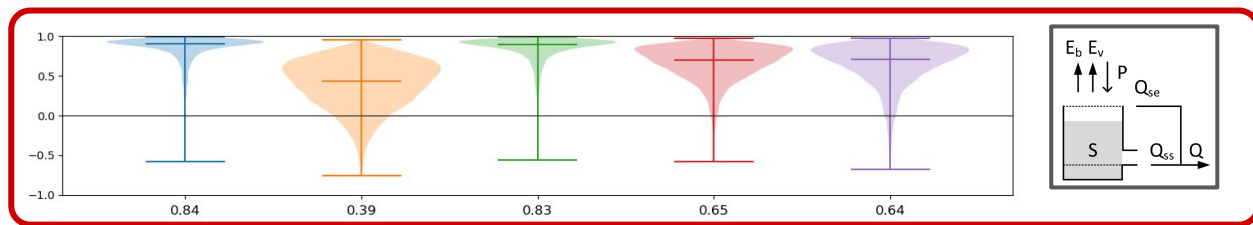
To find out we calculate the derived correlation between parameter values across all runs, and show the distributions as violin plots



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

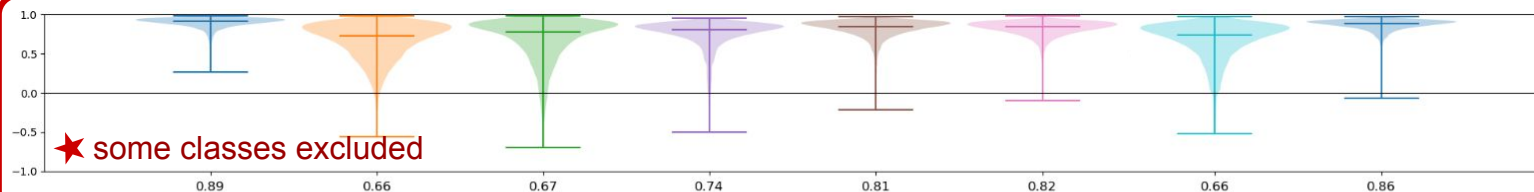
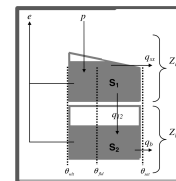
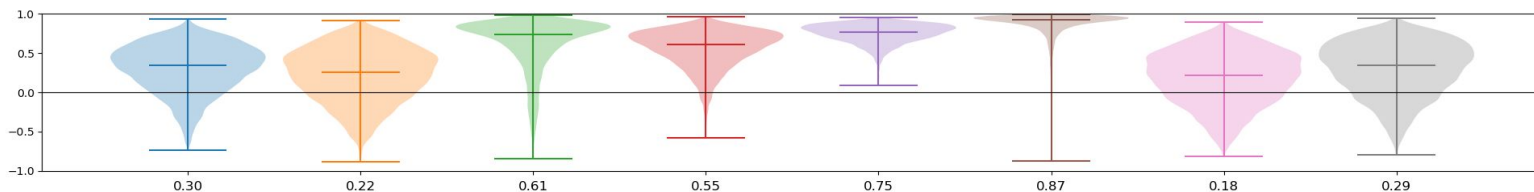
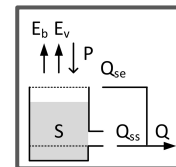
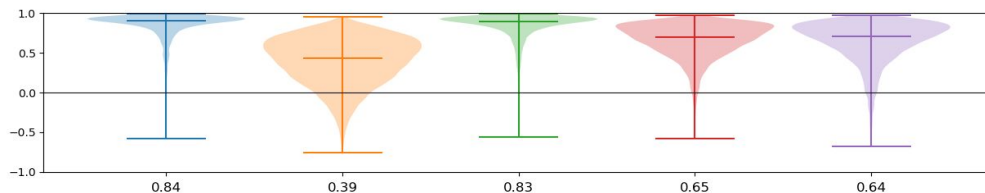


★ some classes excluded

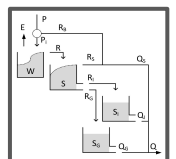
**For each of the model configurations this was done 10 times.
So, do the end models have similar hydrologic behaviors?**

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

**Good performance, and
high agreement in
parameter values**



★ some classes excluded



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

A scenic landscape of rolling green hills and mountains under a cloudy sky. The foreground shows dense green vegetation, while the background features layered mountain ranges under a dramatic, overcast sky.

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

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

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

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

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

Memory state update

$$g'_t = \tanh(W'_{xi} * X_t + W_{mg} * M_t^{l-1})$$

$$i'_t = \sigma(W'_{xi} * X_t + W_{mi} * M_t^{l-1})$$

$$f'_t = \sigma(W'_{xf} * X_t + W_{mf} * M_t^{l-1})$$

$$M_t^l = f'_t \odot M_t^{l-1} + i'_t \odot g'_t$$

Output, hidden, and action updates

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

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

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

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 Triplett¹, Andrew Bennett¹, Laura E. Condon¹, Peter Melchior^{2,3}, and Reed M. Maxwell^{4,5,6}

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

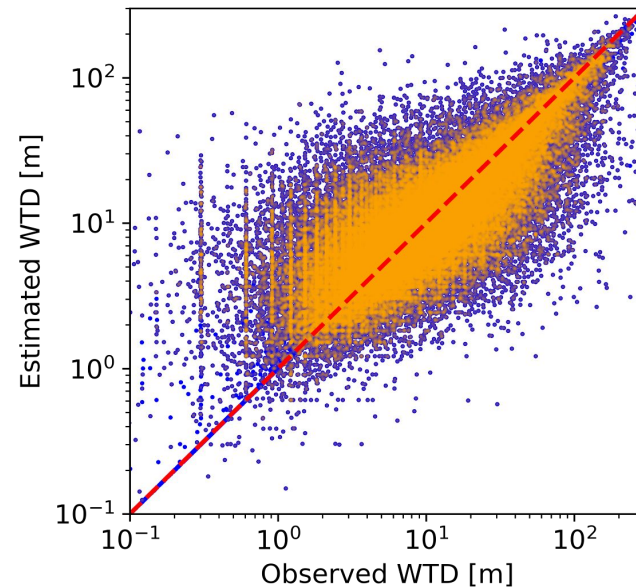
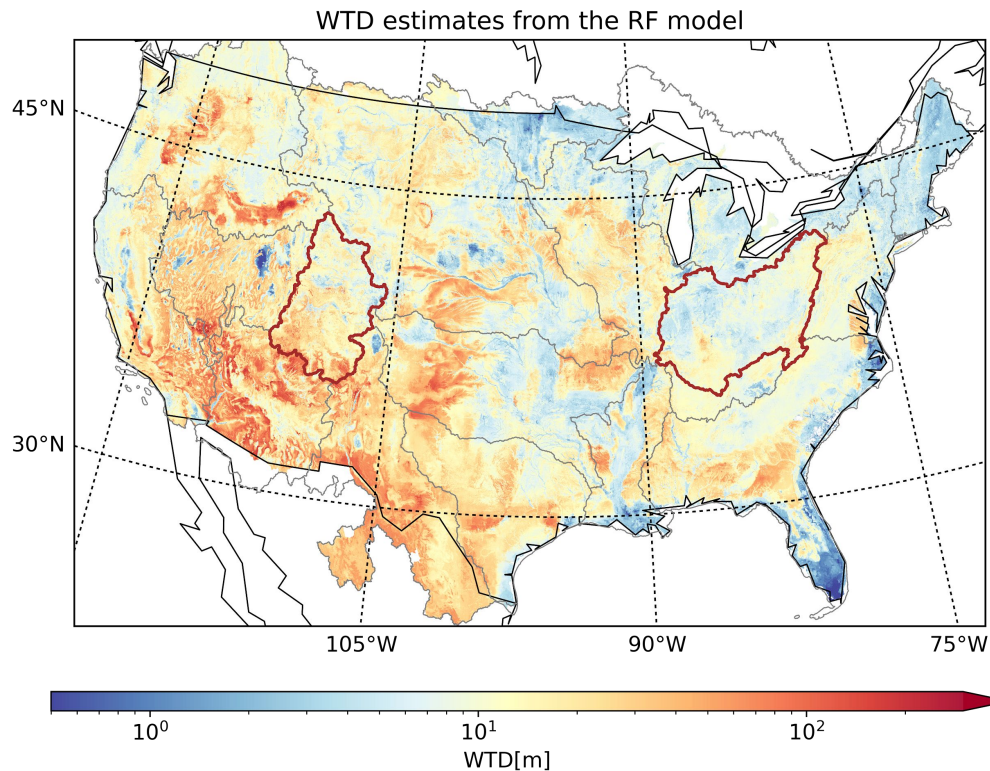
Abstract Hydrogeologic models generally require gridded subsurface properties, however these inputs are often difficult to obtain and highly uncertain. Parametrizing computationally expensive models where extensive calibration is computationally infeasible is a long standing challenge in hydrogeology. 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 physical models is a long standing challenge in hydrogeology.

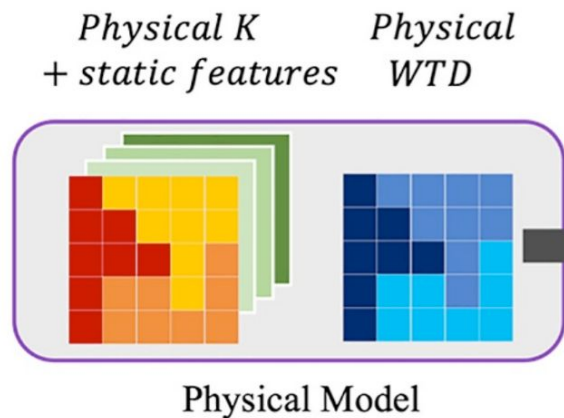
Originally we were motivated by work from Yueling Ma, who developed a ML estimate of WTD



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

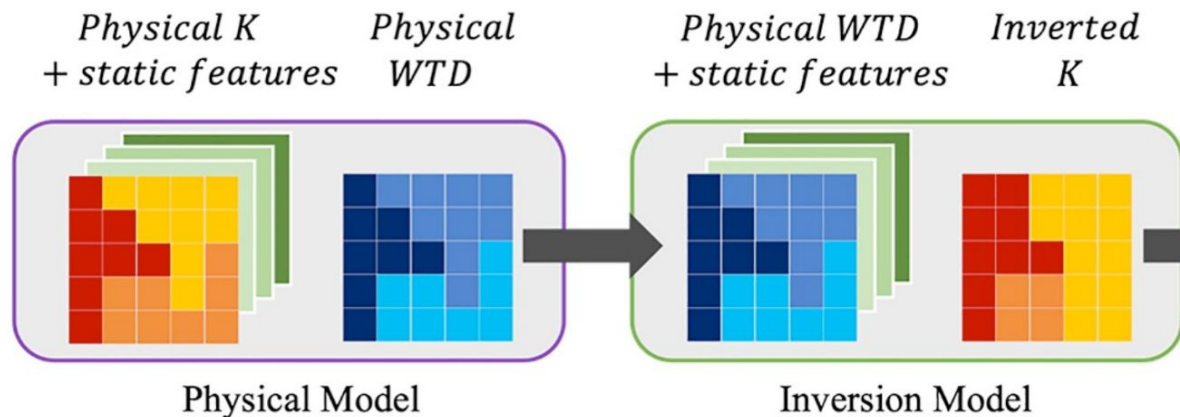
Develop the training data



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

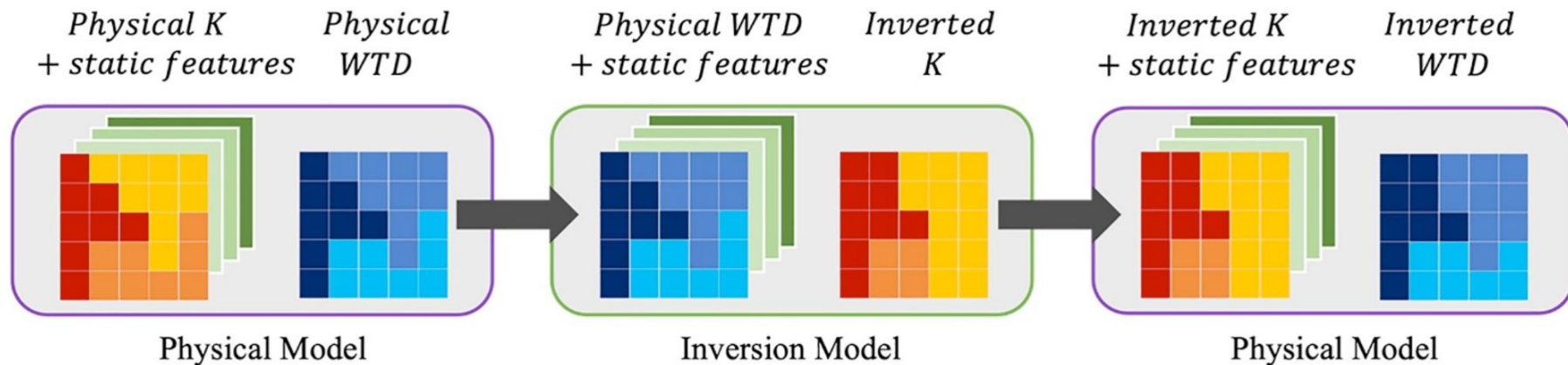


But first we must demonstrate that the overall method works, so Amanda Triplett led the effort to develop our inversion framework

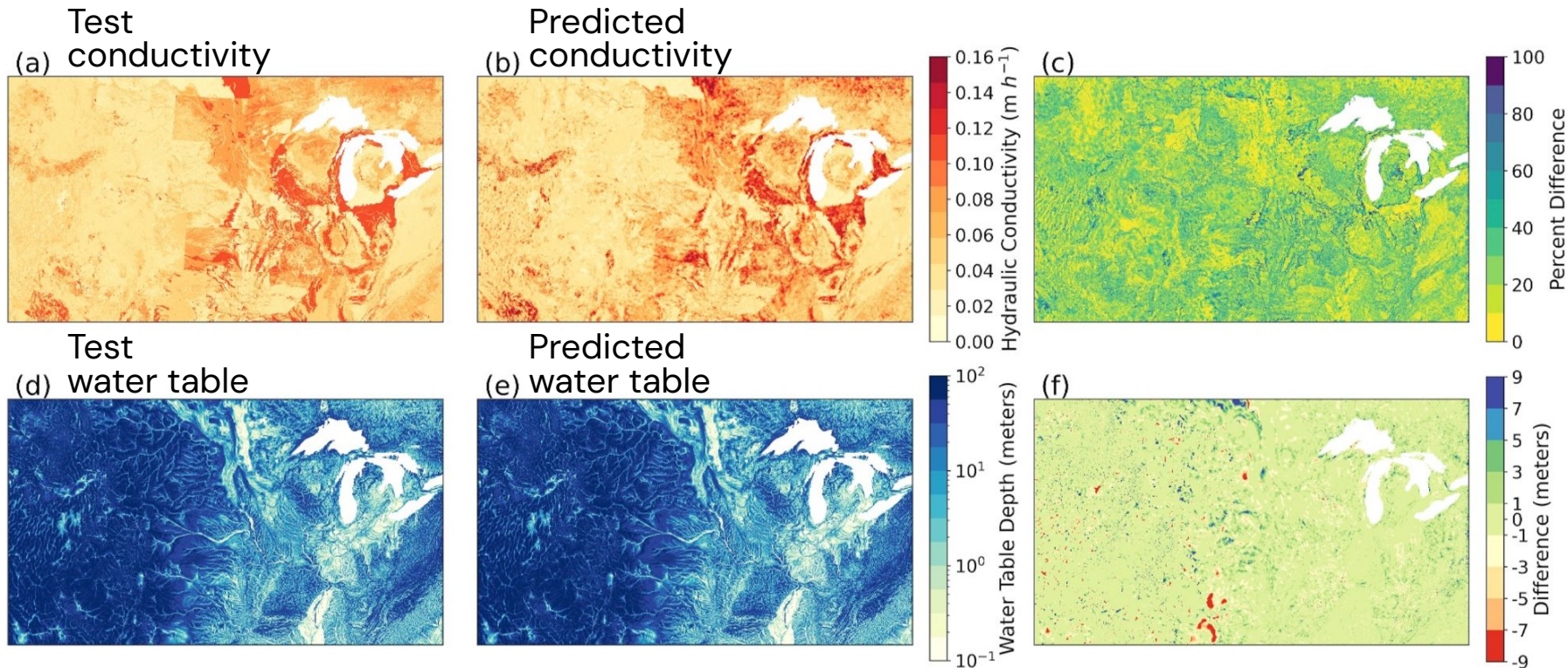
Develop the
training data

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)

