June 23, 2020
AI4ESS

Recurrent networks in geosciences

CHAOPENG SHEN, DAPENG FENG, WEN-PING TSAI AND KUAI FANG

[1]CIVIL AND ENVIRONMENTAL ENGINEERING, **PENN STATE UNIVERSITY**

CSHEN@ENGR.PSU.EDU

Dapeng

https://github.com/mhpi

@ChaopengShen

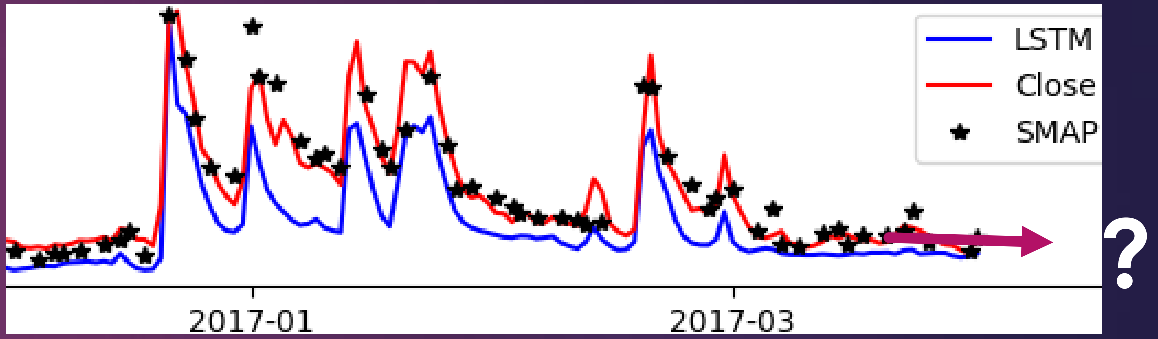
Kuai

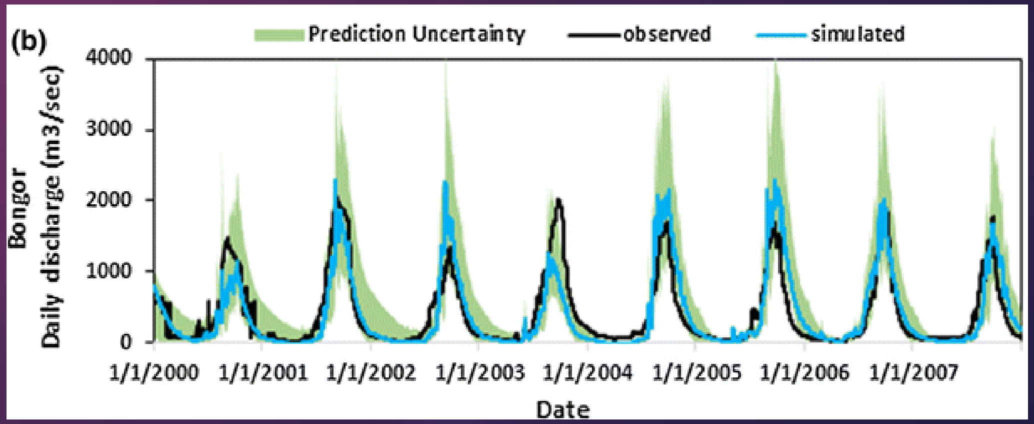
Overview

- Recurrent network applications in hydrology
- Code demo and walkthrough
- Hands-on!

What does the society ask of hydrologists?

- Future trends/risks in hydrologic responses under climate change
- Short-term forecast/states update
- Gauged vs ungauged locations
- Uncertainty quantification
- Water interactions w/ ecosystem/human systems (ET, moisture, GW, etc)

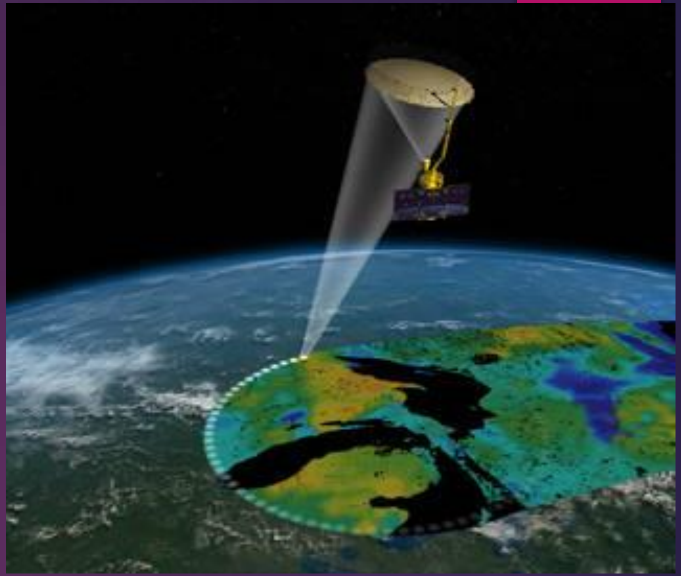




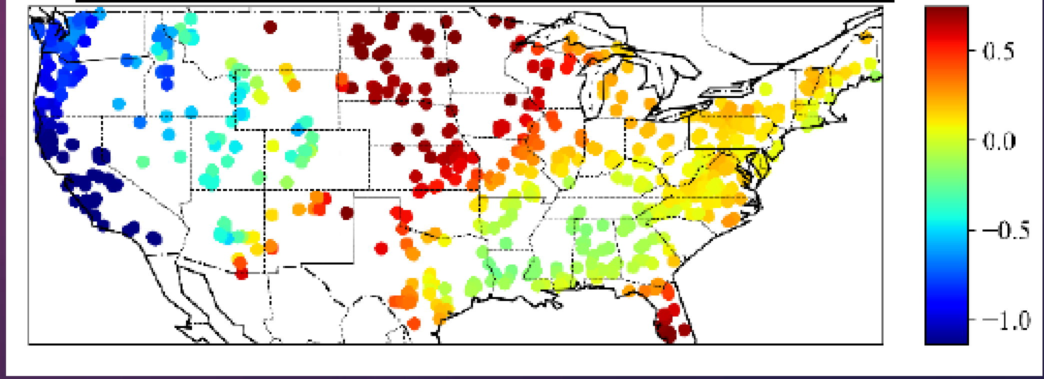
What relationship?

Two case studies

- Soil Moisture Active Passive (SMAP)
  - Launched recently (2015/04)
  - 2~3 days revisit time
  - Senses moisture-dependent top surface soil
- Streamflow modeling (beyond CAMELS →3000 basins)
  - Daily data
  - Accompanying attributes



Rainfall seasonality for USA basins



What is DL and why DL?

a rebranding of neural networks featuring

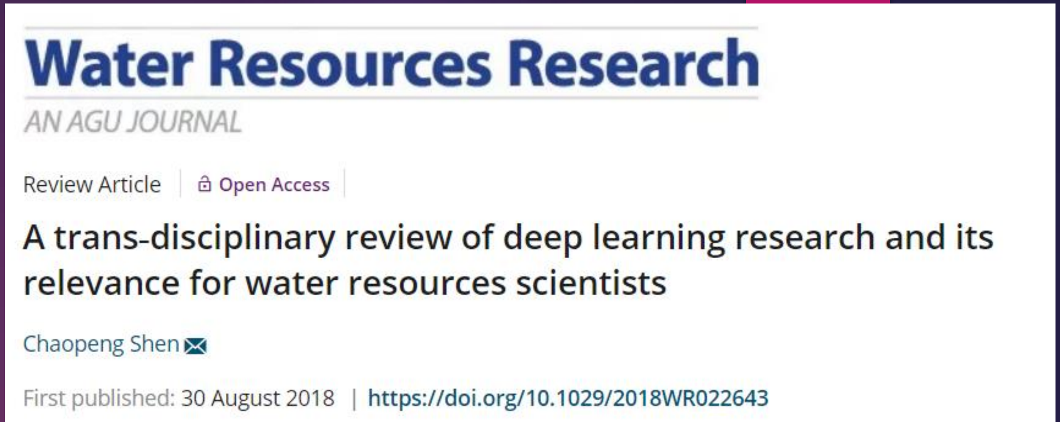
(i) Large capacity

(ii) Hidden layers that automatically extract features

(iii) Improved architecture/regularization

(iv) Working directly with data

a primary value proposition is the avoidance of expertise!



Water Resources Research

AN AGU JOURNAL

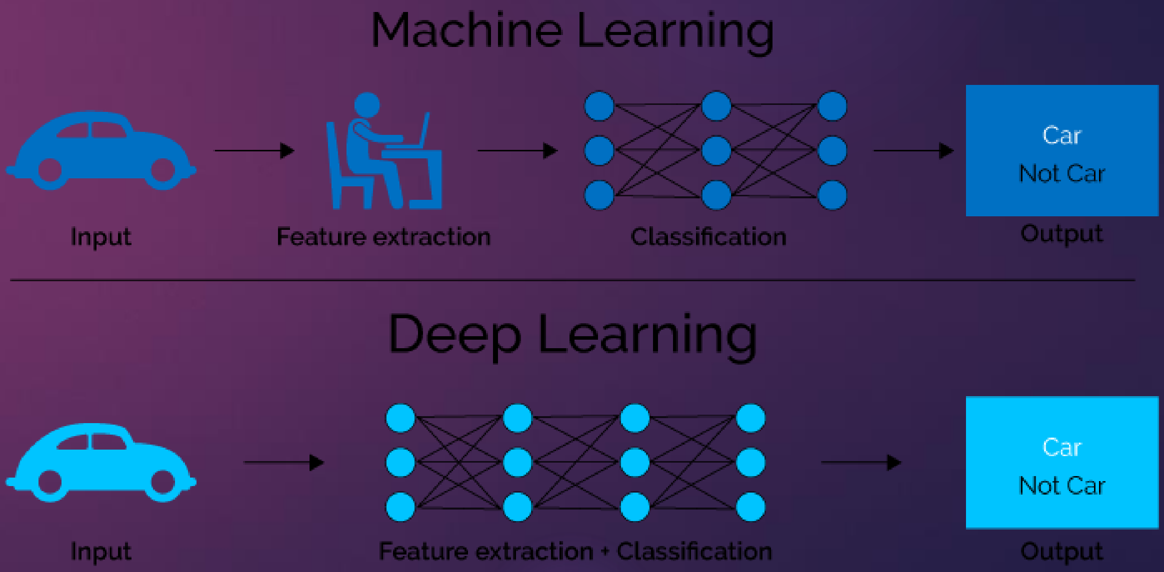
Review Article | Open Access

A trans-disciplinary review of deep learning research and its relevance for water resources scientists

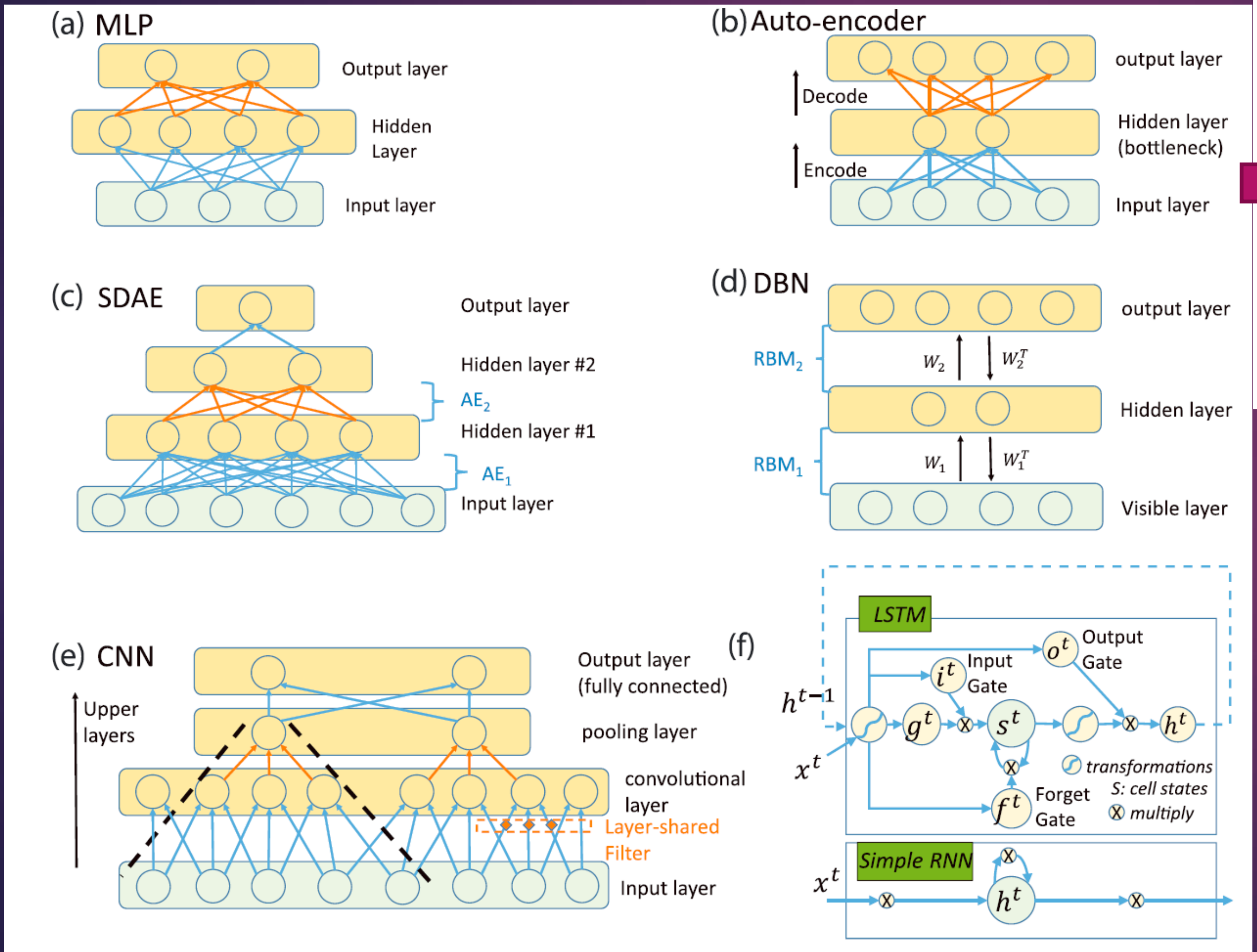
Chaopeng Shen

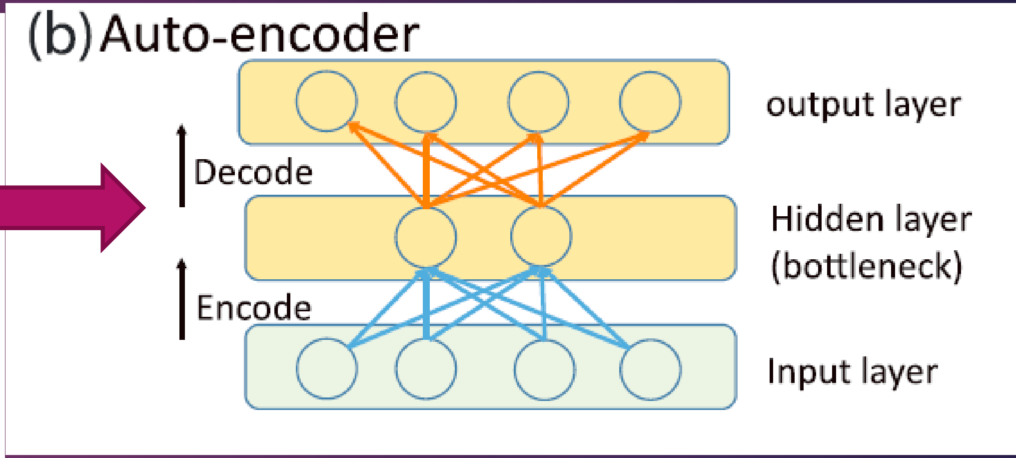
First published: 30 August 2018 | https://doi.org/10.1029/2018WR022643

$X \rightarrow Y$



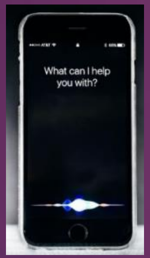
Some basic deep learning architectures





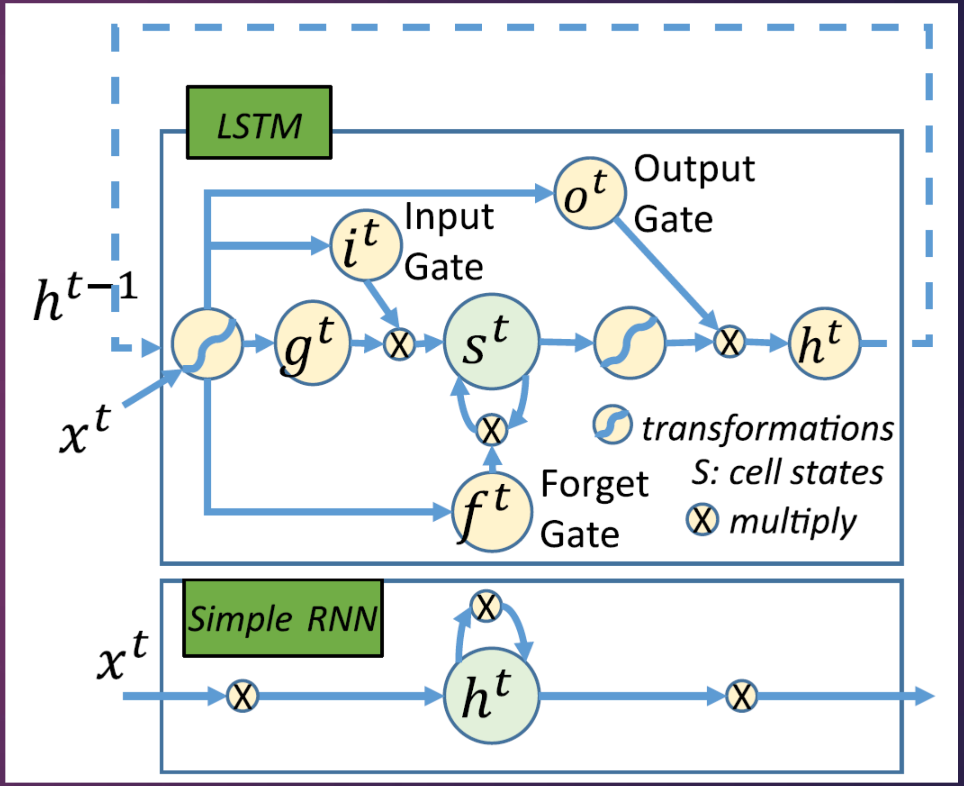
Long Short-Term Memory (LSTM)

- Time Series Deep Learning (DL)



→ Free from structural assumptions

Self-learned Memory system



(y)

A hydrologic model w/o structural assumptions…

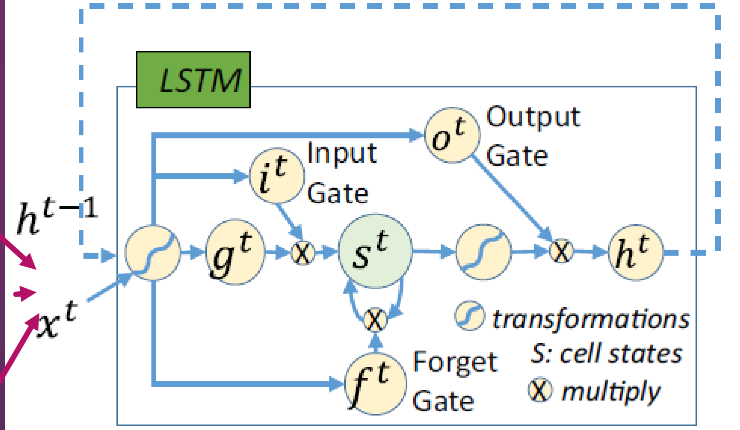
LSTM model

Atm. Forcings

(optional) Land Surface Model (Noah) solutions

Soil texture, slope, land cover, irrigation, depth to water table, etc

(x)



(y)

Target

train

Surface soil moisture

Long-term projections

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

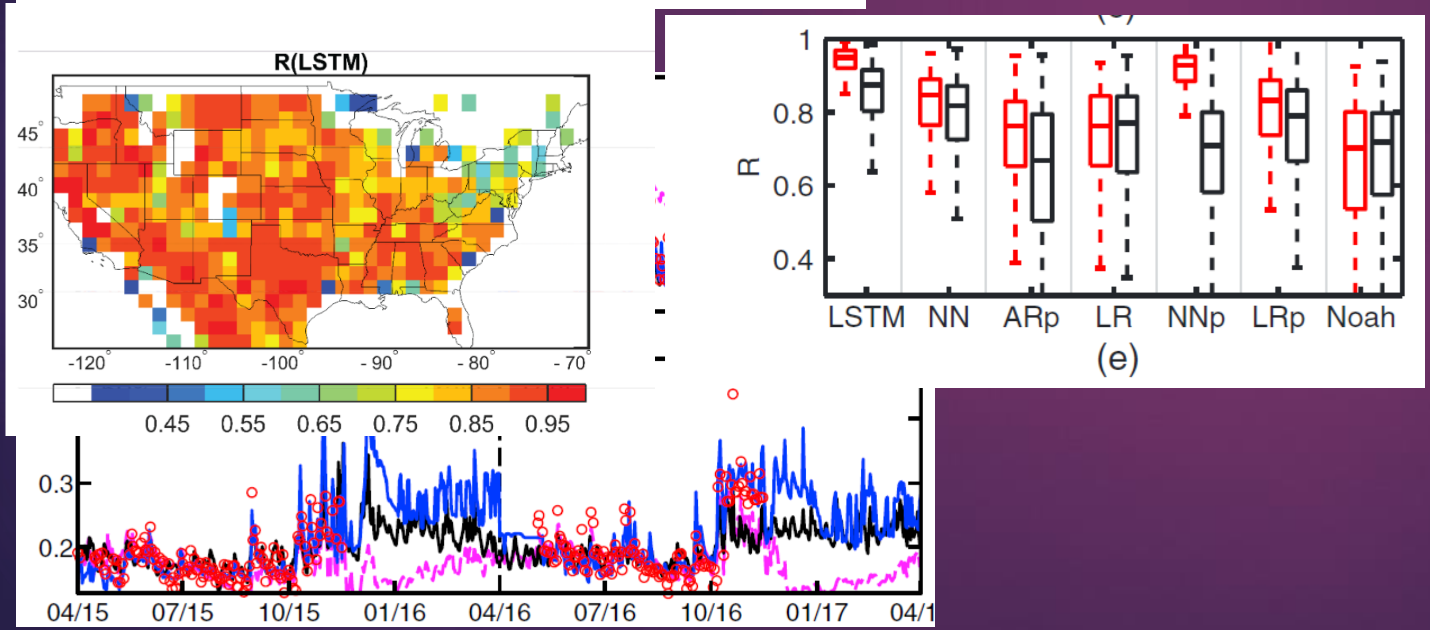
Geophysical Research Letters

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

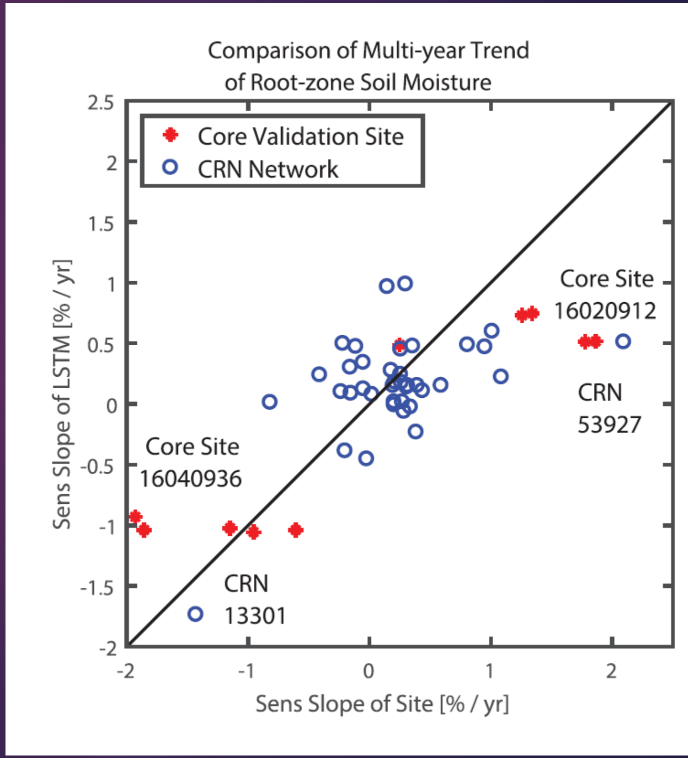


Fang et al., 2017

Journals & Magazines > IEEE Transactions on Geoscien... > Early Access

The Value of SMAP for Long-Term Soil Moisture Estimation With the Help of Deep Learning

3 Author(s) Kuai Fang ; Ming Pan ; Chaopeng Shen View All Authors



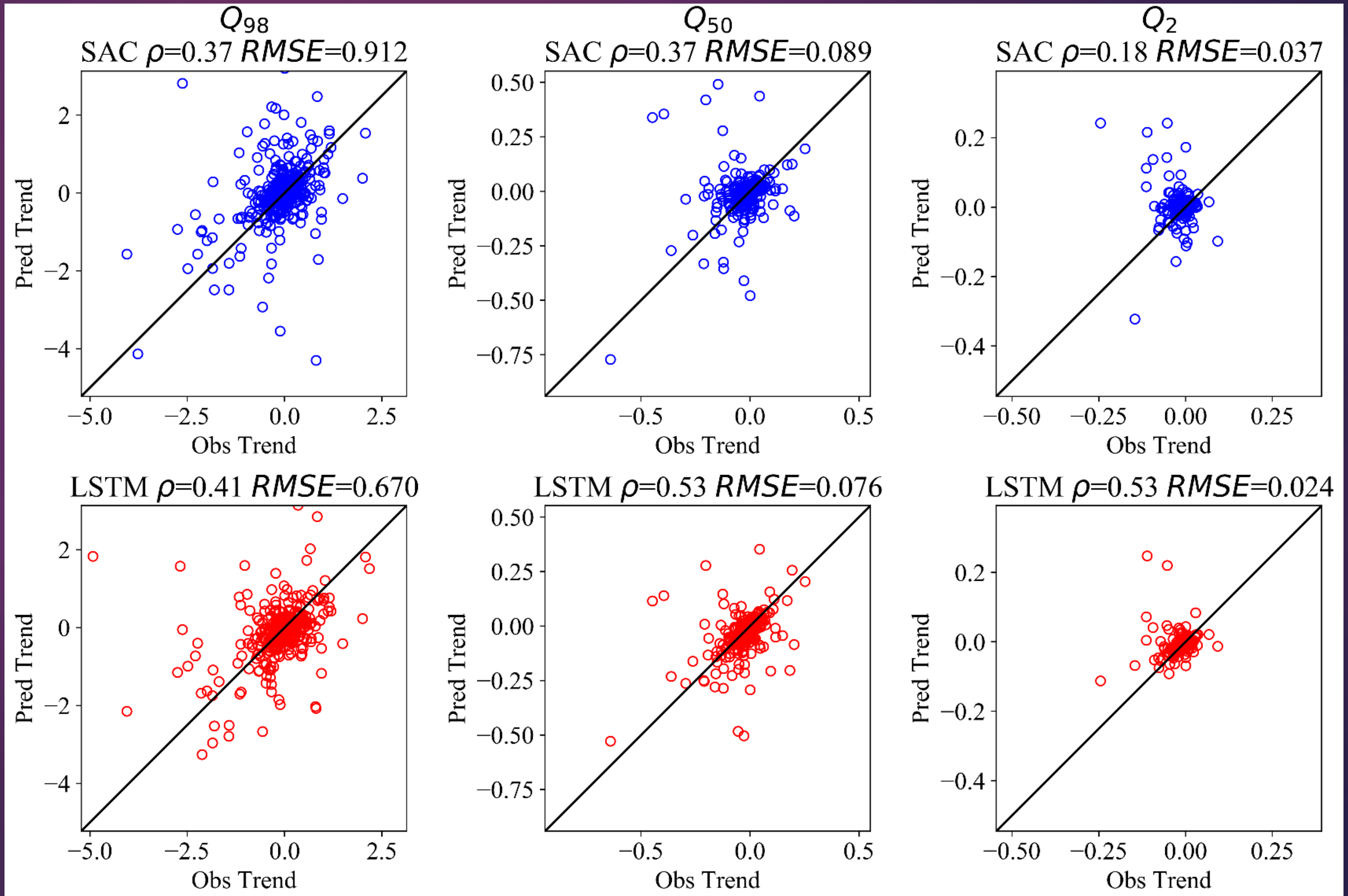
Fang et al., 2018

Long-term projections

- LSTM better than SAC-SMA at capturing the long-term trends

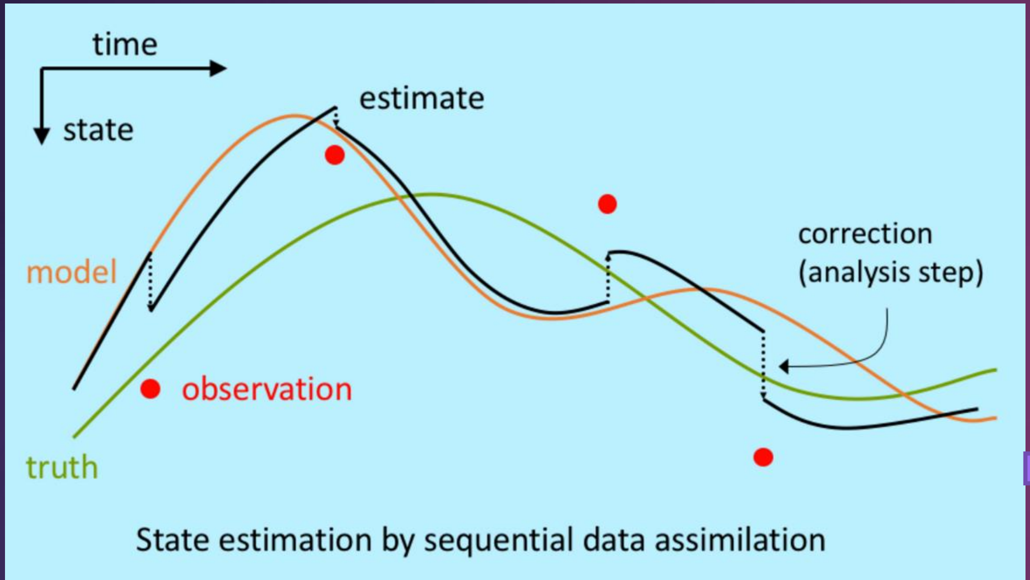
SAC

LSTM



Short-term forecast

- Traditional "data assimilation" scheme

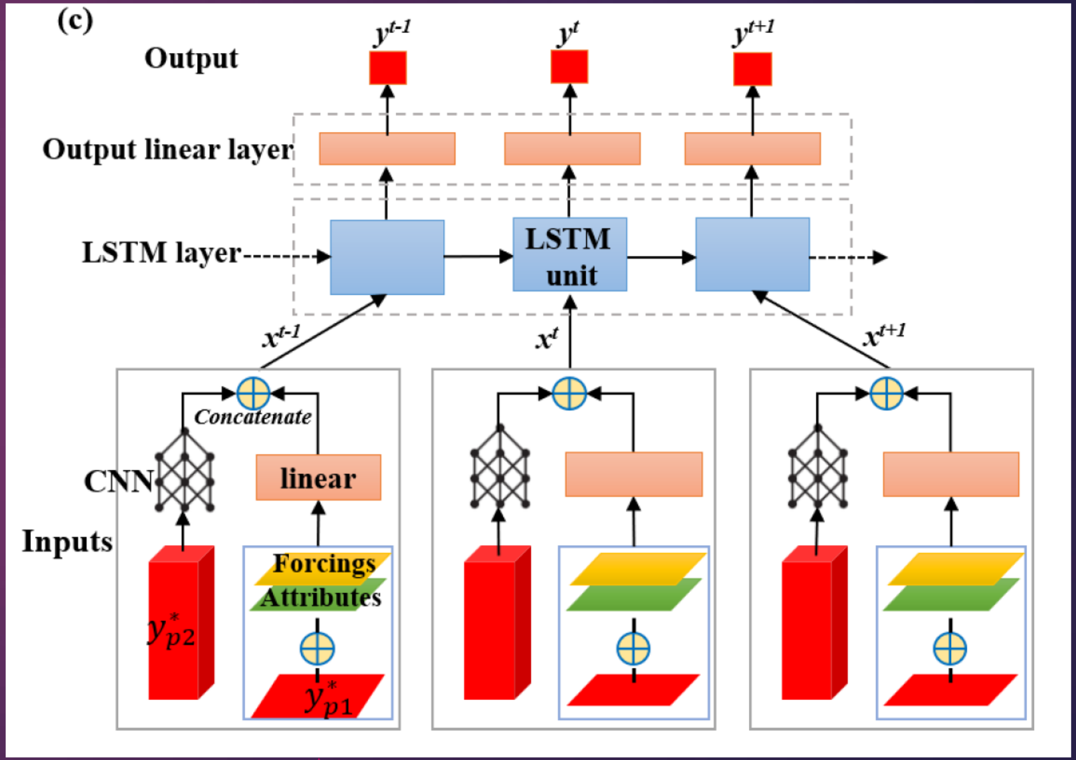


Simulation → Observation – (ENKF) → Correction

Choices: covariance matrix, what to include, how to solve, bias correction, etc.

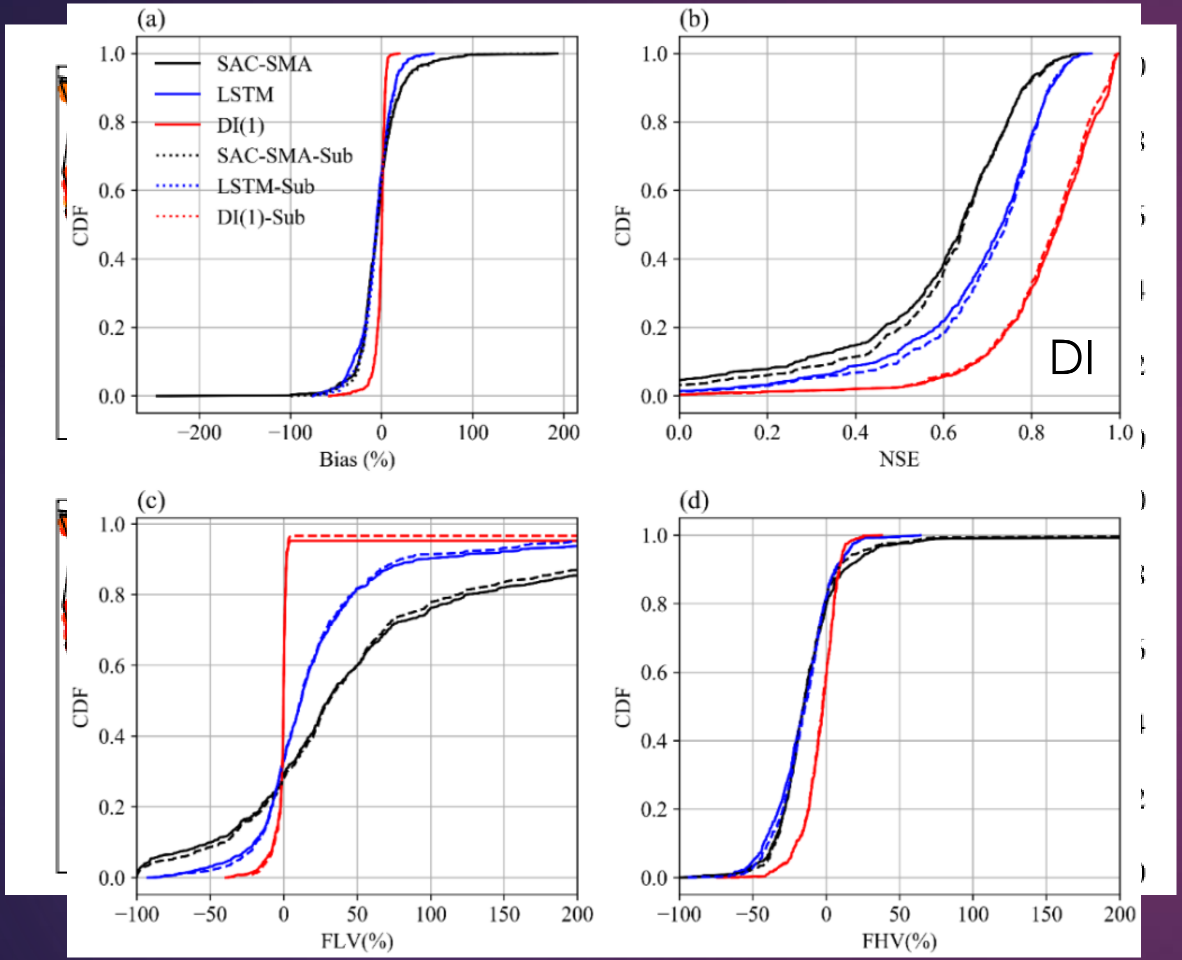
"Data Integration" (DI)

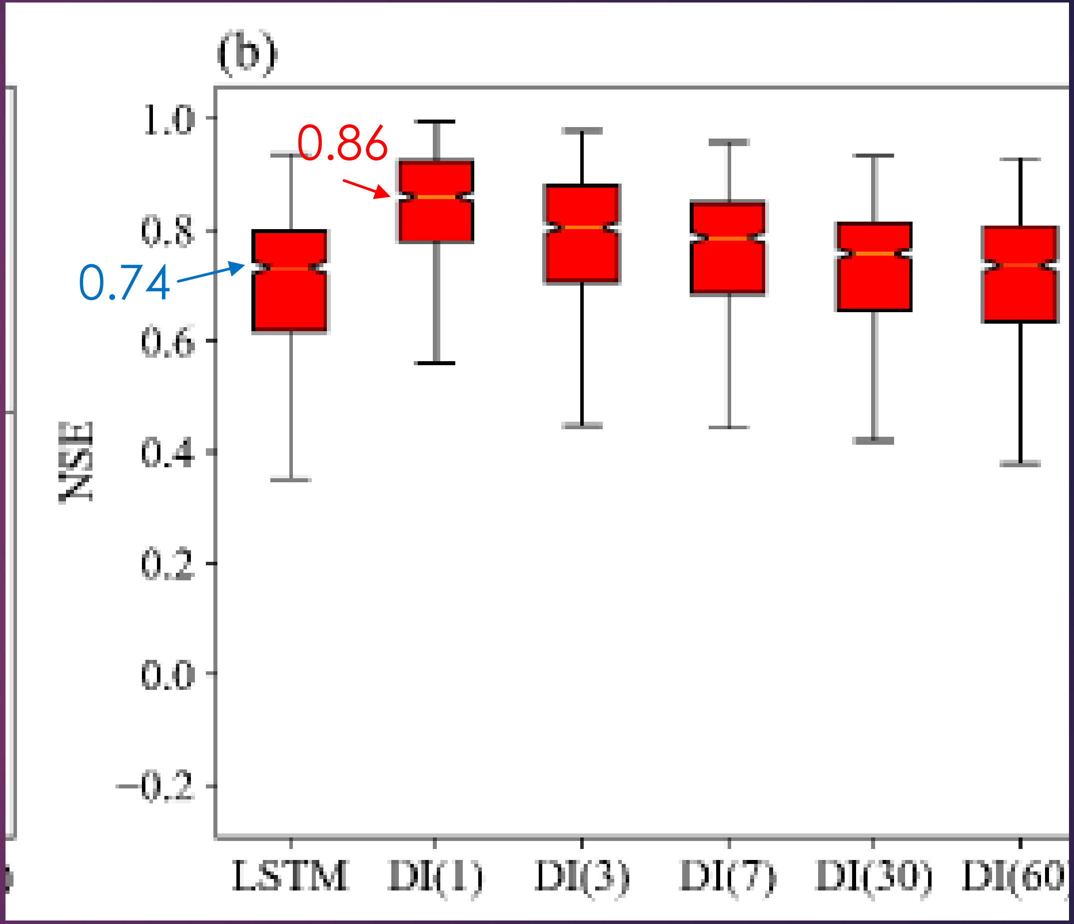
Observation → Corrected Simulation



Q^{t-1}, Q^{t-2}, etc

Forecast for streamflow

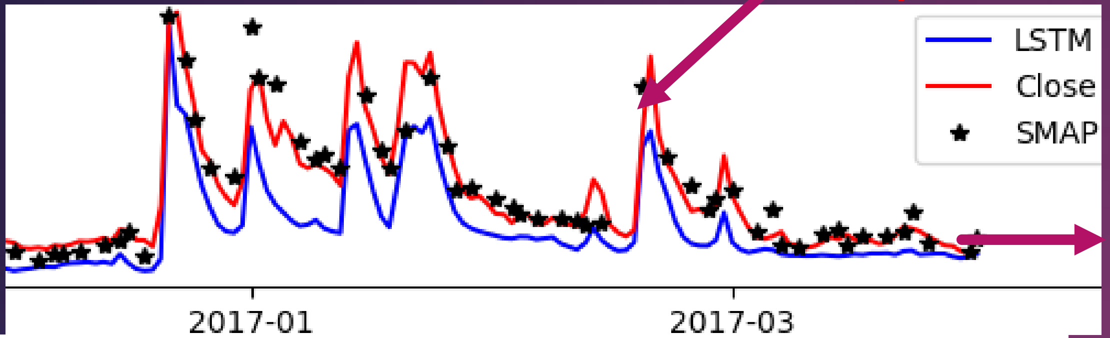


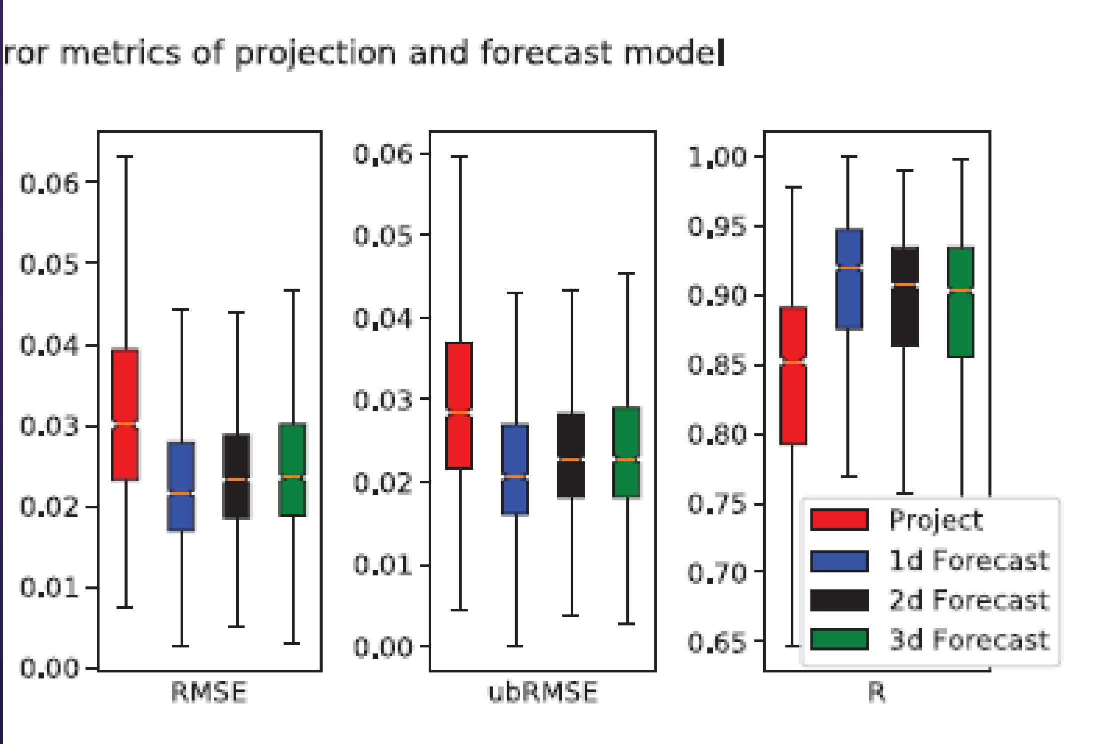


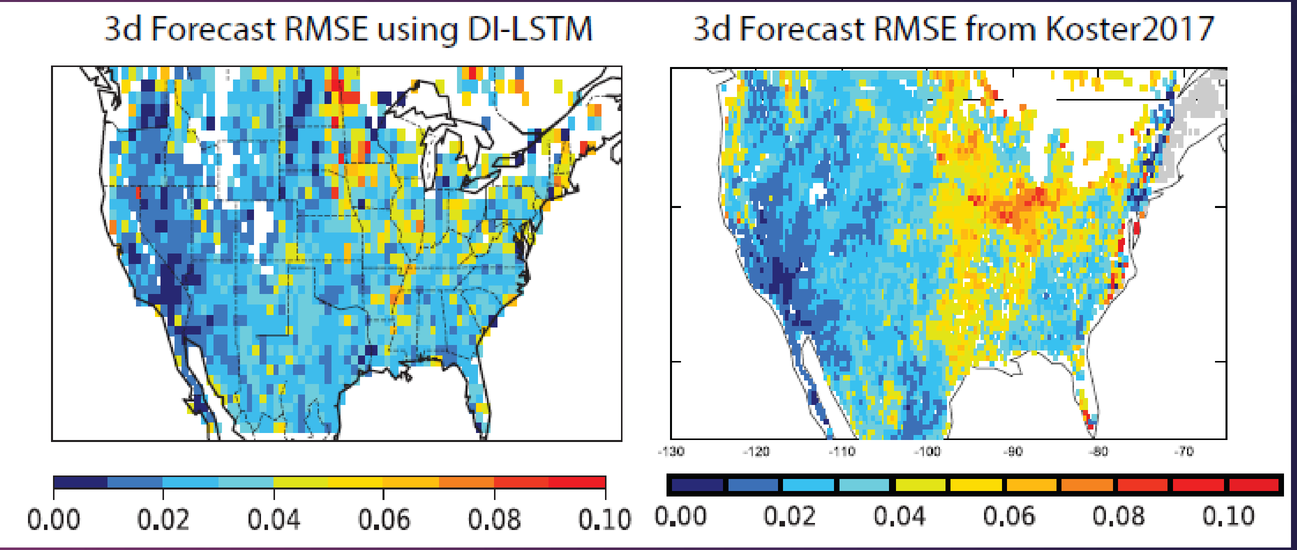
Feng, Shen et al., WRR, accepted, 2020
https://arxiv.org/abs/1912.08949

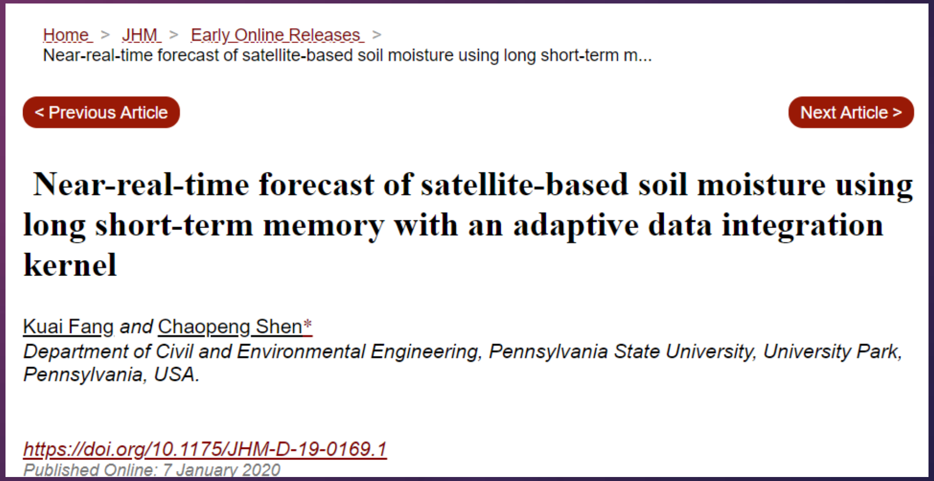
Forecast for (i) soil moisture (ii) streamflow

Gaps?









Home > JHM > Early Online Releases >
Near-real-time forecast of satellite-based soil moisture using long short-term m...

< Previous Article   Next Article >

Near-real-time forecast of satellite-based soil moisture using long short-term memory with an adaptive data integration kernel

Kuai Fang *and* Chaopeng Shen*
Department of Civil and Environmental Engineering, Pennsylvania State University, University Park, Pennsylvania, USA.

https://doi.org/10.1175/JHM-D-19-0169.1
Published Online: 7 January 2020

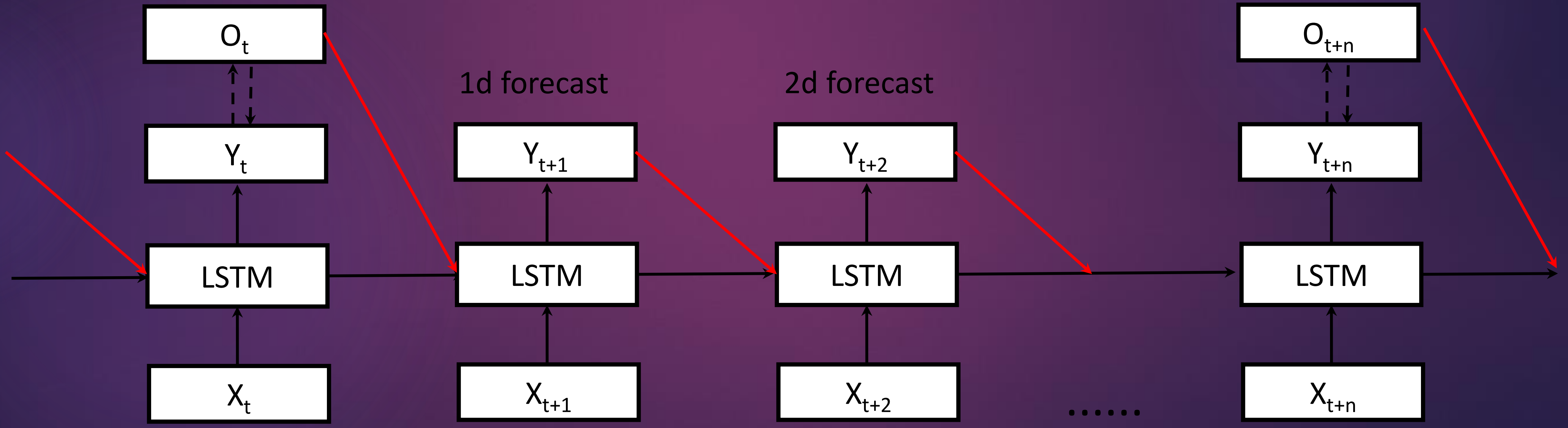
https://sites.google.com/view/mhpi/locust

What about missing data? Extensive w/ SMAP

X - model input
Y - model output
O - SMAP observation
dash lines - backward
solid lines - forward

-- use LSTM as a forward extrapolator!

-- a general solution for time-stepping problems where input data stream is also the output.



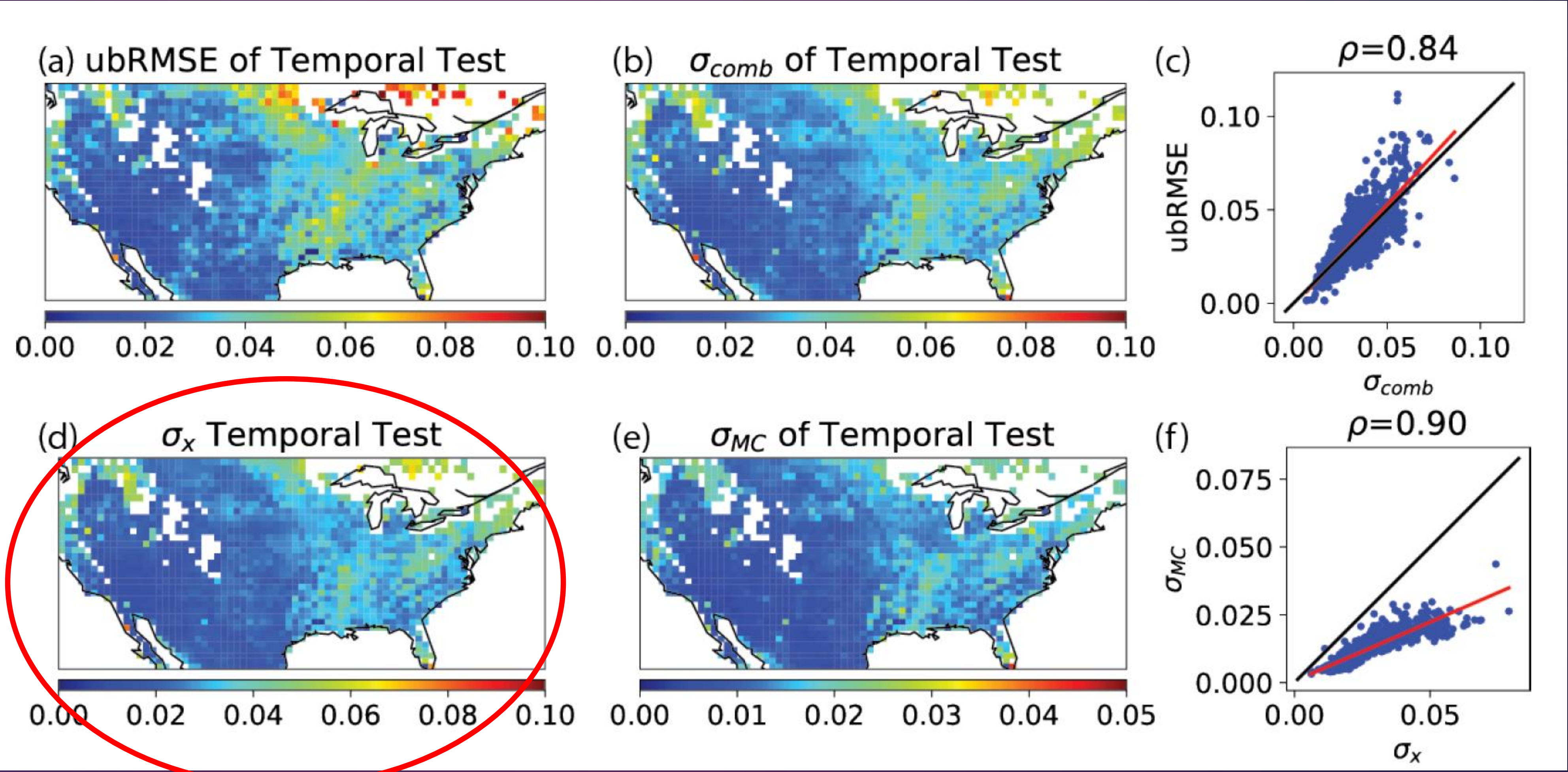
Gradual convergence of forward extrapolation and DI training. **H33A-04**

Uncertainty estimation

Training
- 2015/04 - 2016/04

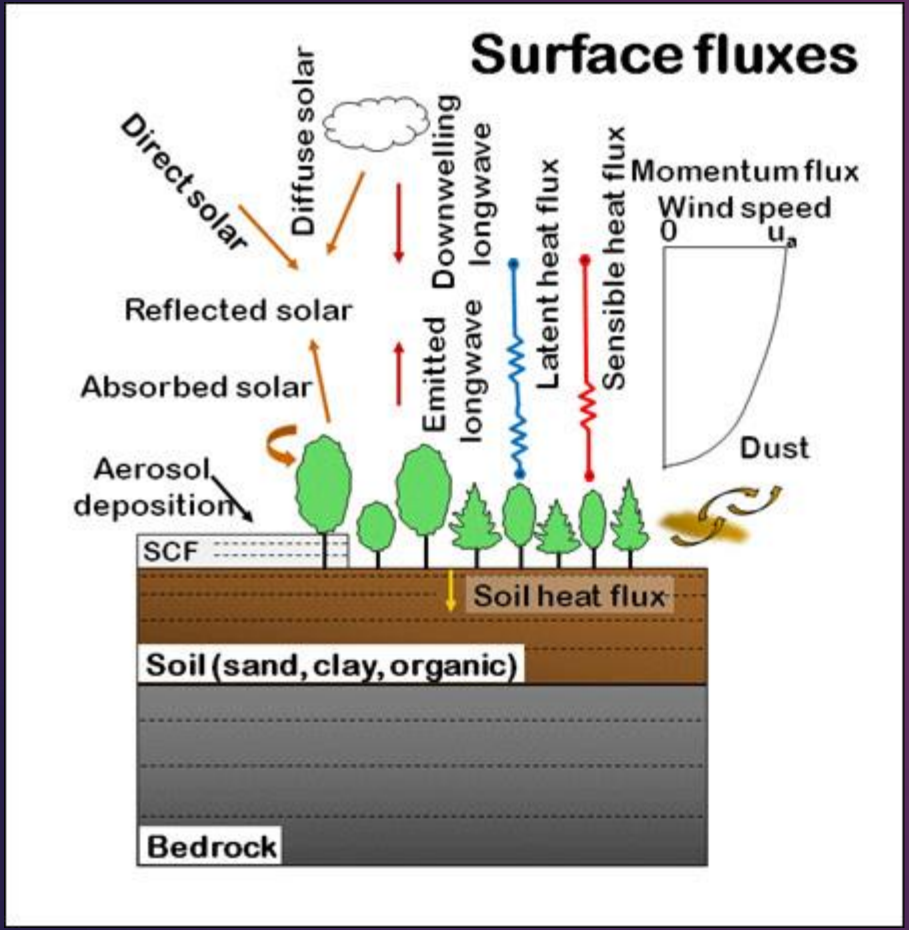
Validation
- 2016/04 - 2017/04
- hyper-parameter (σ_{mc} is a function of the dropout rate)

Temporal test
- 2017/04 - 2018/04
- same pixels as training set



Under review, https://www.essoar.org/doi/10.1002/essoar.10503330.1

Interactions w/ ecosystem/biogeochemistry



How about variables we cannot observe accurately on large scales?

ET, Groundwater, deeper soil moisture

From parameter calibration to →

parameter learning (fPL)

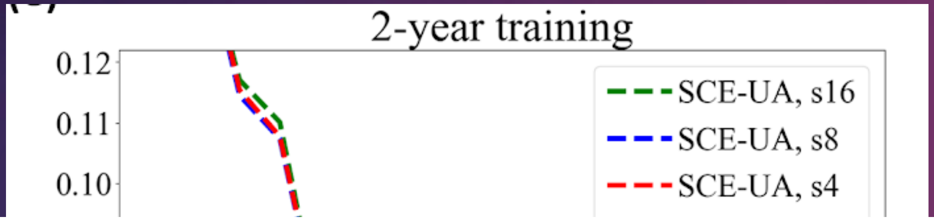
Forcings + y* → Parameters → VIC surrogate → y

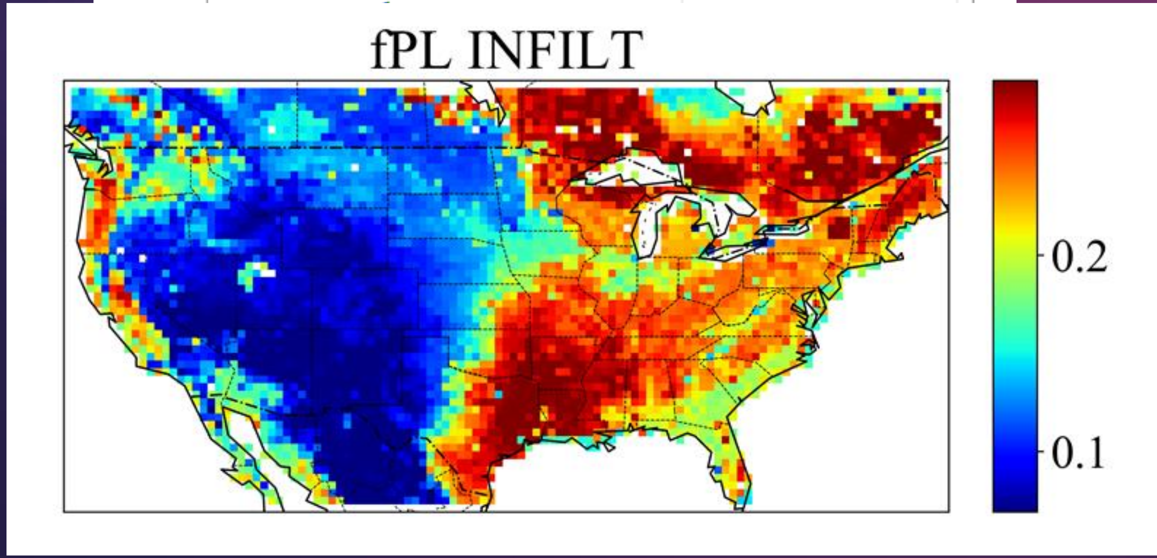
LSTM

We calibrated VIC model using SMAP data and the LSTM-based fPL scheme

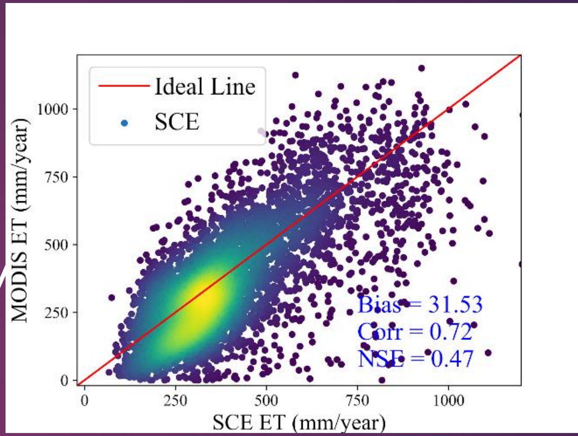
Parameter learning (*fPL*) -- results

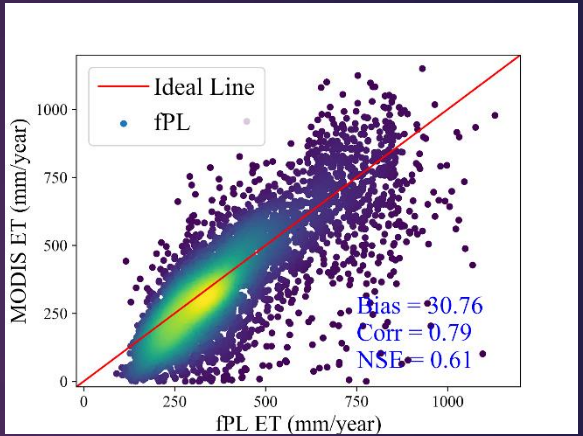
- Stronger than SCE-UA!
- Saves $O(10^4)$ computation!
- Now capable of modeling other variables such as ET and/or streamflow



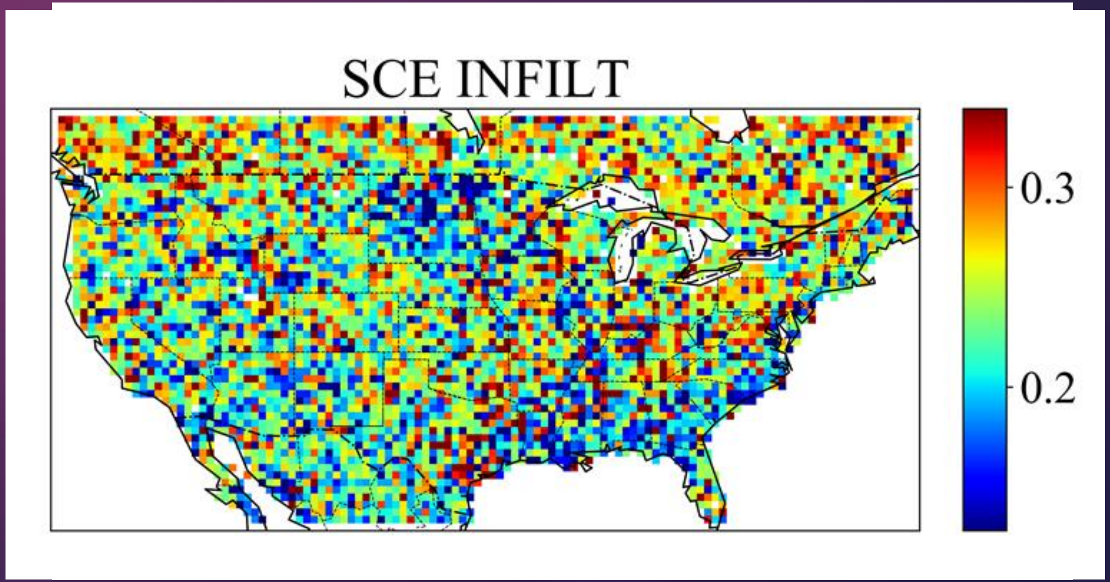


Uncalibrated variable: ET



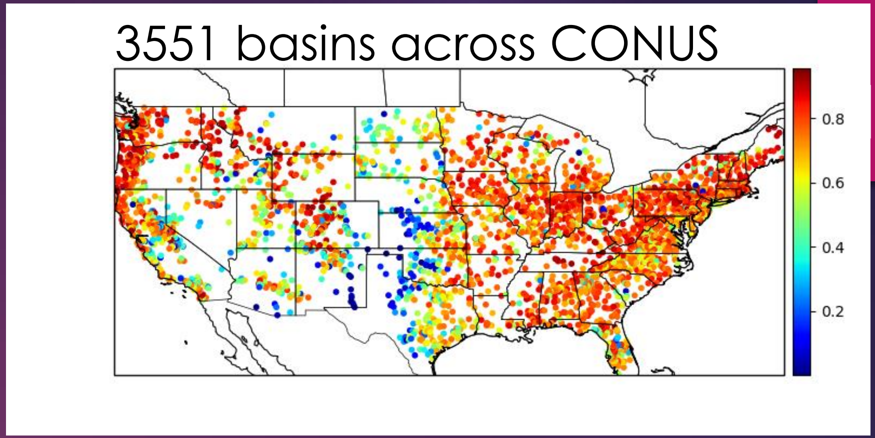


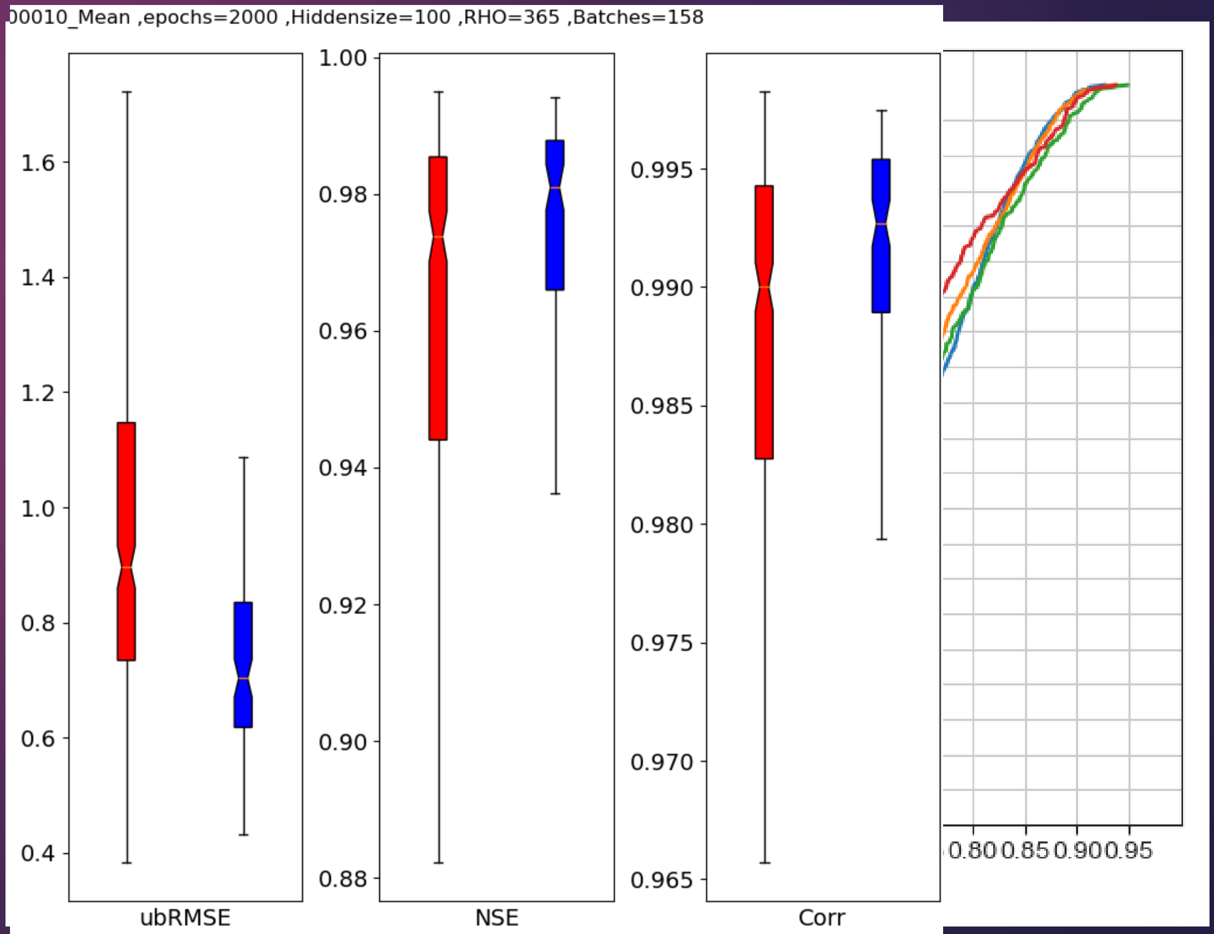
Spatial extrapolation

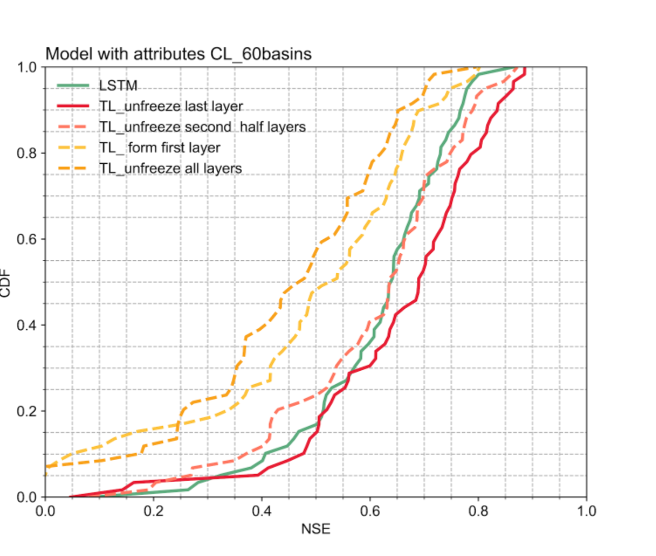


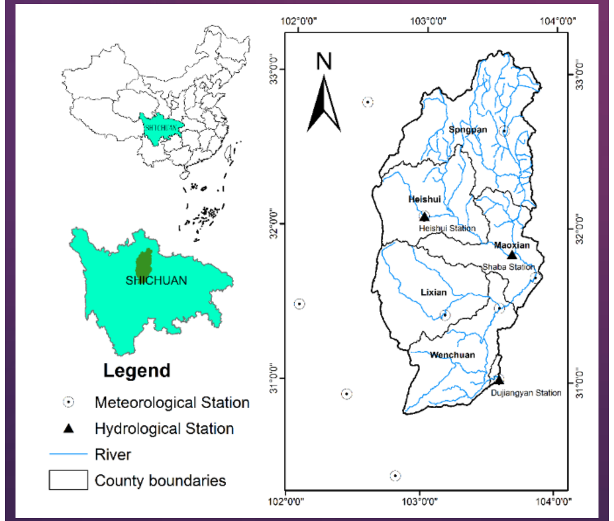
Ongoing LSTM work beyond CAMELS

- LSTM can also model (small) reservoirs
- We can migrate knowledge across continents and support modeling in sparsely gauged sites.
- Great results with stream water temperature as well!









Process-based modeling vs machine learning

===== **PBM** strength ====

- Built from the bottom-up to observe emergent patterns
- We know what we put in
- We can do experiments & identify causal relationships

===== Limitations ====

- Human biases
- Parameter calibration
- What we don't know?
- Errors compound?

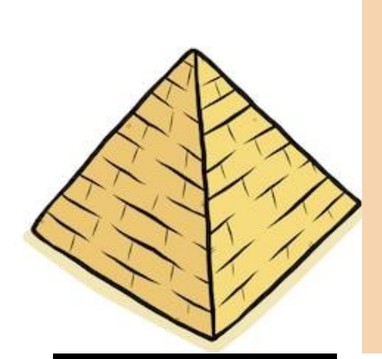
Synergy?

===== **BDML** strength =====

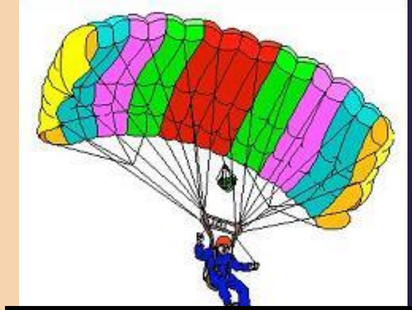
- Built from the top-down, directly from observations → accurate
- Less biased
- Identify things we don't know?
- Highly efficient in computation

==== Limitations ======

- Can't observe everything!
- May be difficult to interpret
- May not fully respect physical laws
- Does not understand causal relationships



Bottom up



Top down

Where to go from here?

- Hydrologic DL has launched a full-on assault to offer a full suite of hydrologic services with higher accuracy and lower cost.
- Time series DL will spread over to many geoscientific domains.
- DL will not replace PBM. On the contrary, there will be a class of **unified model** that links together PBM and DL.
- Powerful applications have emerged from hydrologic DL, while PINN near completes its proof of concept. In the future there PINN may see more growth.
- DL may be deeply ingrained into next generation models for science and practical operations

Perhaps one day DL will become an inalienable component of the hydrologic discipline itself

Citation: Shen, C. (2018), Deep learning: A next-generation big-data approach for hydrology, *Eos, 99*, https://doi.org/10.1029/2018EO095649. Published on 25 April 2018.

Thank you!

@ChaopengShen
cshen@engr.psu.edu

https://github.com/mhpi

http://water.engr.psu.edu/shen/hydroDL.html

CUAHSI cyberseminar series on BDML

WRR special issue on BDML

AGU Editor's review

Hydrol. Earth Syst. Sci., 22, 5639–5656, 2018
https://doi.org/10.5194/hess-22-5639-2018

Hydrology and Earth System Sciences
Open Access
EGU

HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community

Chaopeng Shen[1], Eric Laloy[2], Amin Elshorbagy[3], Adrian Albert[4], Jerad Bales[5], Fi-John Chang[6], Sangram Ganguly[7], Kuo-Lin Hsu[8], Daniel Kifer[9], Zheng Fang[10], Kuai Fang[1], Dongfeng Li[10], Xiaodong Li[11], and Wen-Ping Tsai[1]

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

sites.google.com/view/deepldb

deepLDB

deepLDB -- a machine-learning-based Landslide database and modeling system

Code walkthrough w/ an example

demo-LSTM-Tutorial.ipynb
(driver/gateway)

default.py:
(default definitions of options)
screen.py
(batch job support utility)

/master/master.py
(batch job interpreter)
loaddata()

train()

data

train.py;
(training job main driver,
prepare mini-batch)
trainmodel()

rnn.py;
(model computational core)
cudnnlstmModel() #LSTM+linear

cudnnlstm()
#implements cudnn version of LSTM

cpulstmmodel()
#when you don't have a GPU