Machine Learning for Ecological Forecasting

Near-term Ecological Forecasting Initiative Short Course 2022

Jake Zwart (he/him)

With lots of input from USGS Data Science co-workers

To put |

THE NEW YORKER

Adapted from *The Next Word. Where will predictive text take us? By John Seabrook*

To put it crudely, speaking or writing is a box whose input is a meaning plus a communicative intent, and whose output is a string of words; comprehension is a box with the opposite information flow. What is essentially wrong with this perspective is that it assumes that meaning and intent are inextricably linked. Their separation, the learning scientist Phil Zuckerman has argued, is an illusion that we have built into our brains, a false sense of coherence.

THE NEW YORKER

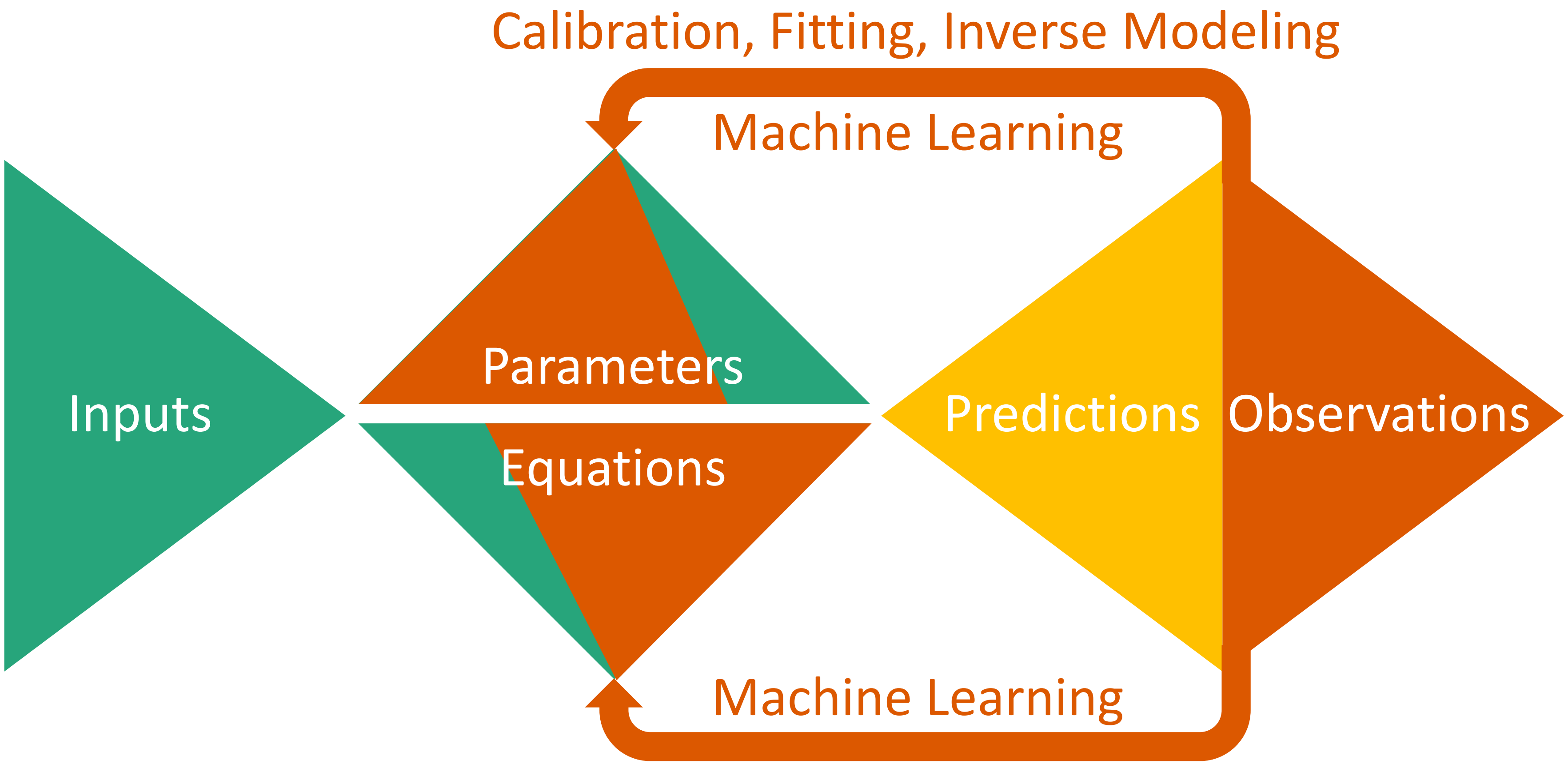
Adapted from *The Next Word. Where will predictive text take us? By John Seabrook*

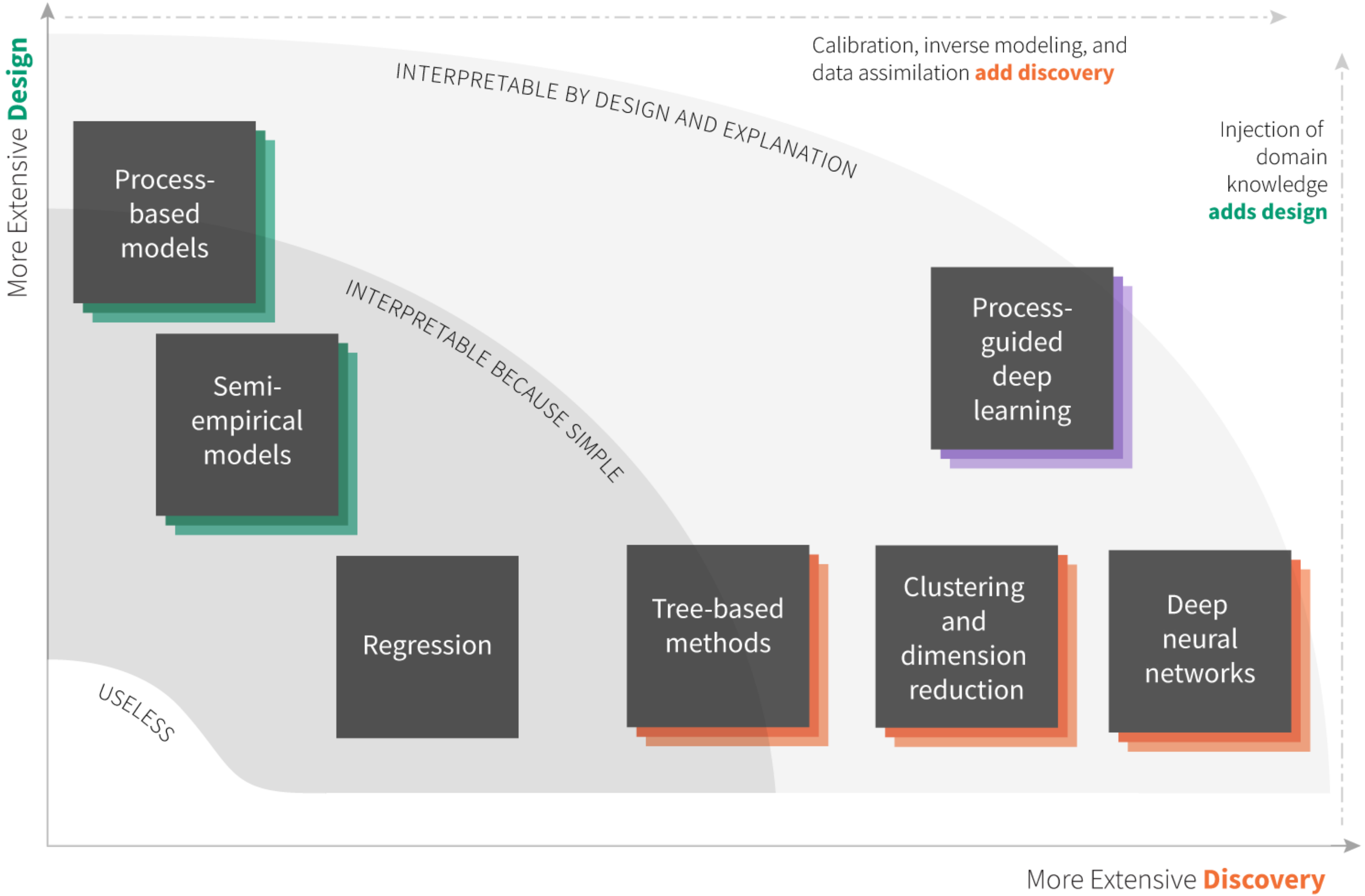
What do we need for ecological forecasting?

1. Produce accurate predictions
2. Characterize prediction uncertainty
3. Make use of recent observations
4. Improve ecological understanding

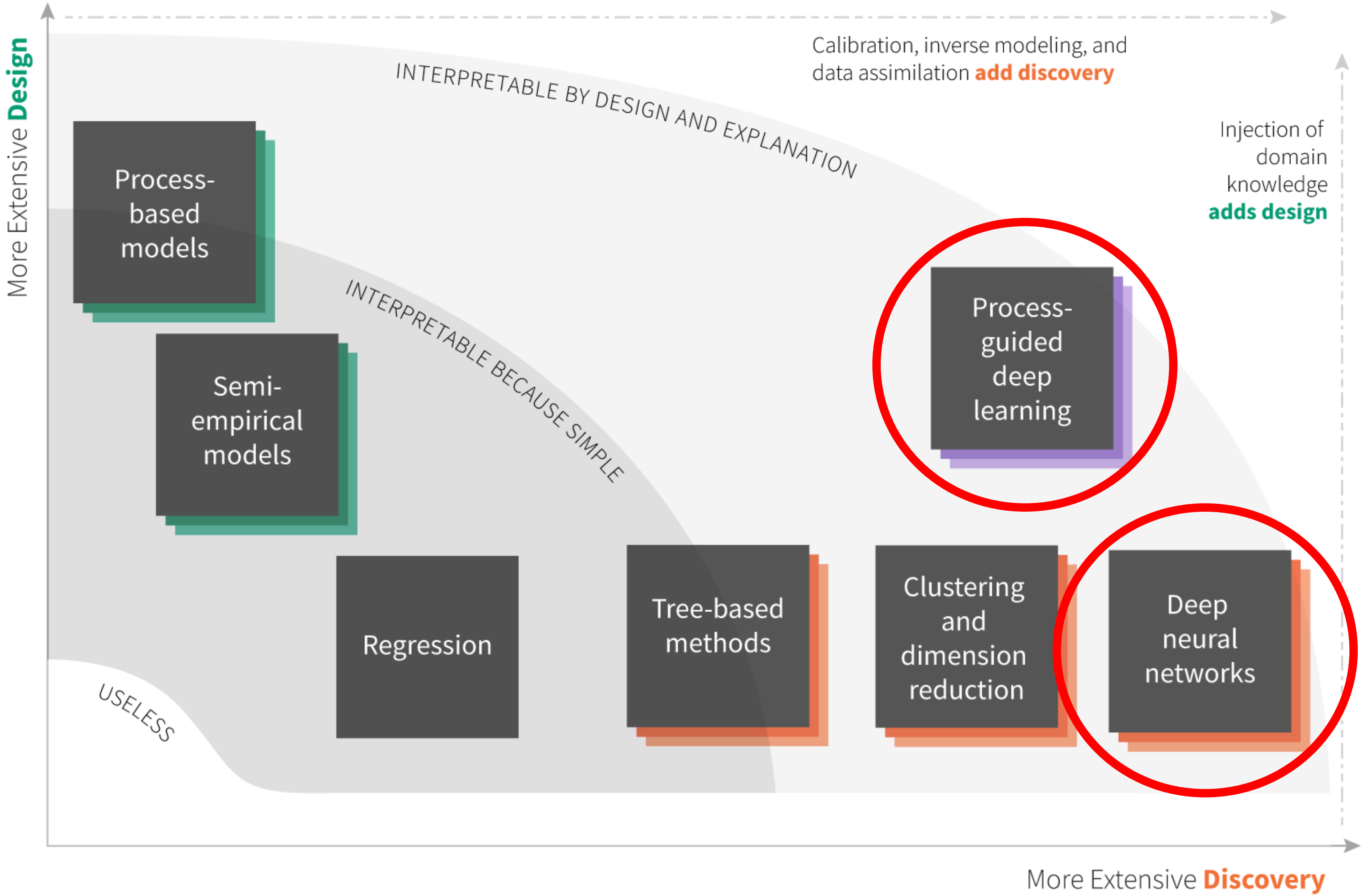
Machine learning models are another tool in our arsenal

Model components are designed...or discovered





Appling et al., *In press*



Appling et al., *In press*

Deep Learning

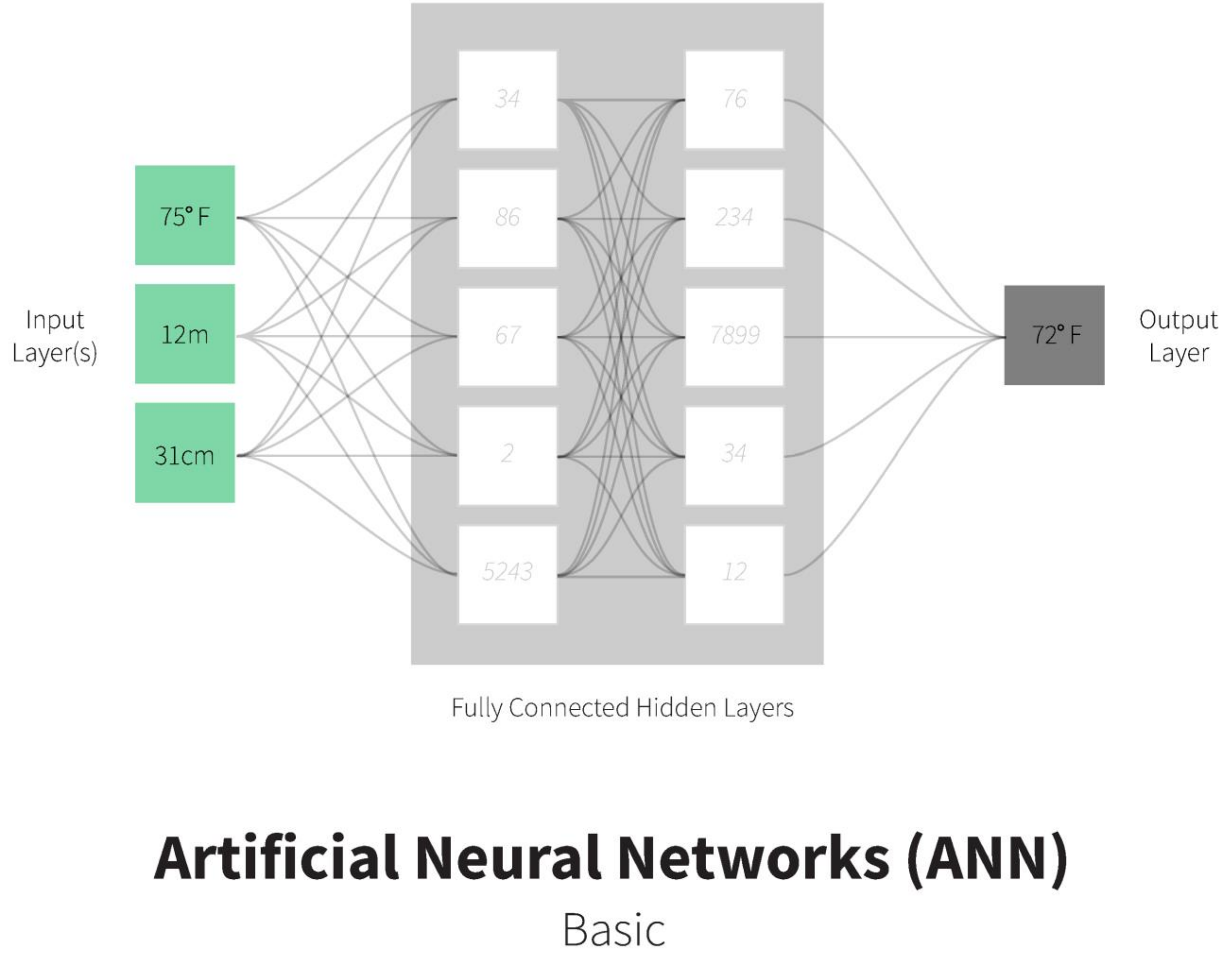
Overview: Artificial Neural Networks

Advantages:

- **Models dynamic, non-linear and noisy data**
- **Low computational cost of predicting**
- **Can be applied to many types of problems**

Disadvantage:

- **Can yield instable outputs**
- **Slow convergence speed**
- **Hyperparameter tuning can be difficult**



Appling et al. *In Press*; Cruz et al. 2021

Deep Learning

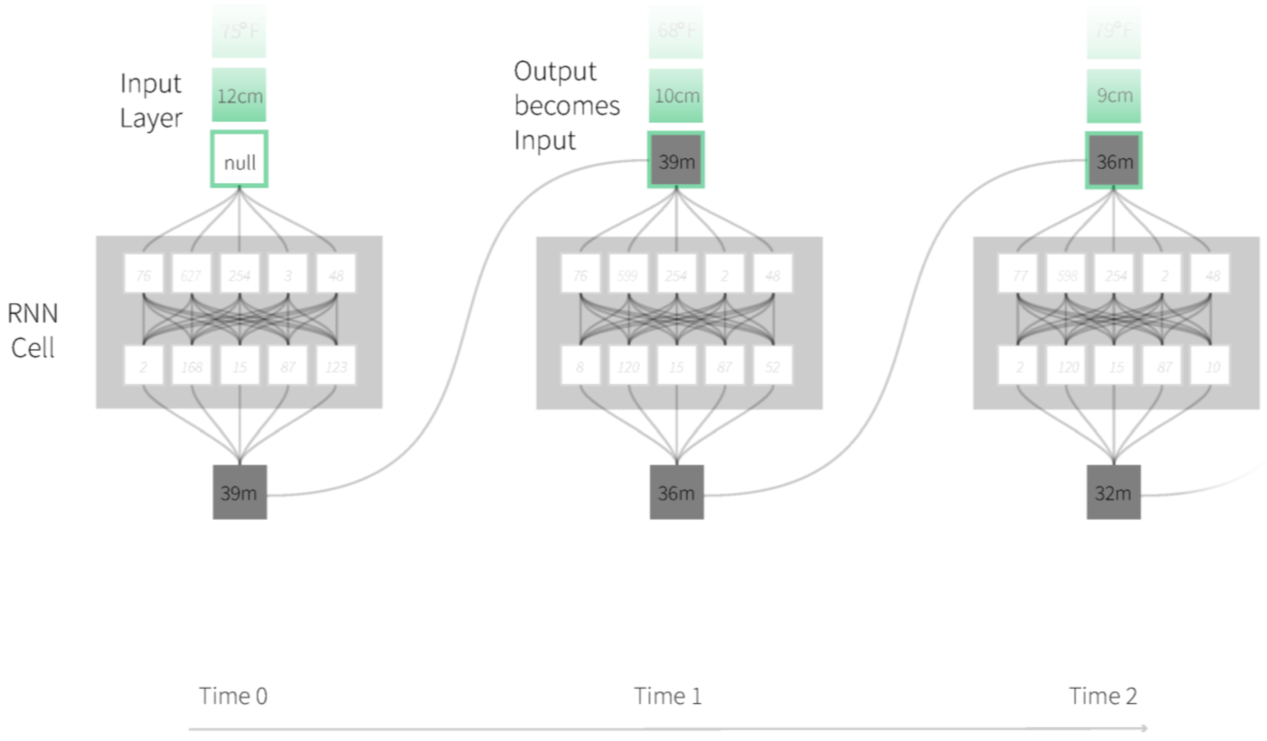
Overview: Recurrent Neural Networks

Advantages:

- **Captures temporal dependencies over variable time periods**

Disadvantage:

- **Higher complexity and computational cost**



Recurrent Neural Networks (RNN)

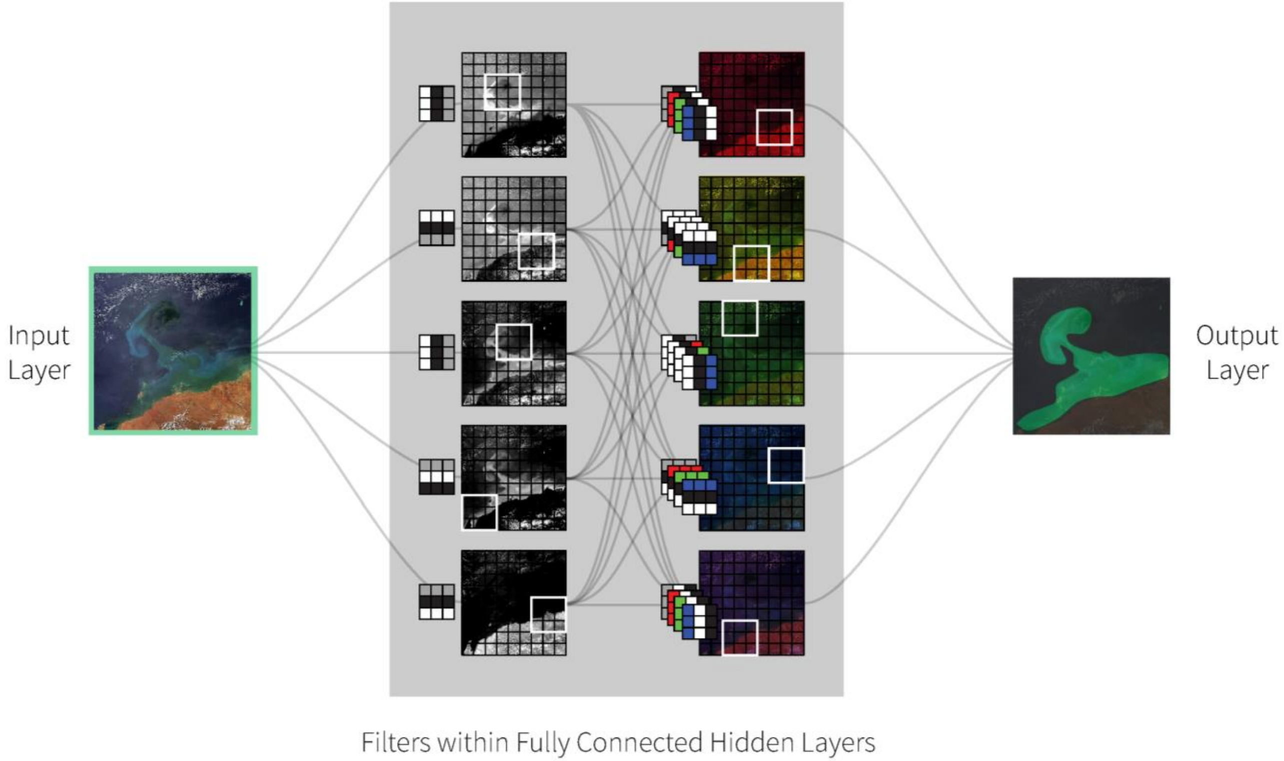
Time-Aware

Appling et al. *In Press*; Cruz et al. 2021

Deep Learning
Overview: Convolutional Neural Networks

Advantages:
- Captures spatial relationships
- Small number of trainable weights

Disadvantage:
- Difficult to capture long-term dependencies

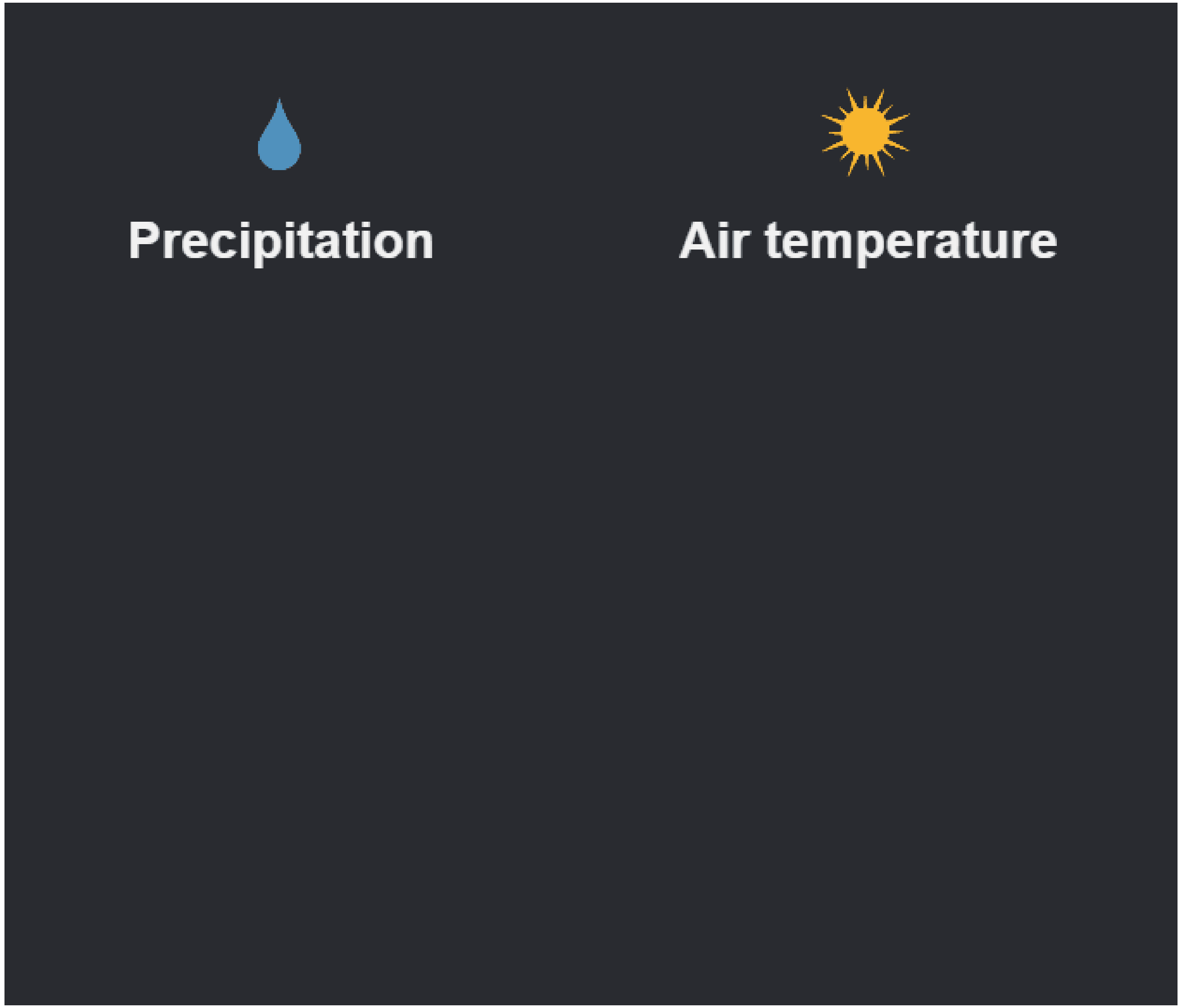


Convolutional Neural Networks (CNN)
Space-Aware

Appling et al. *In Review*; Cruz et al. 2021

Deep Learning

Example: Stream temperature prediction



https://labs-beta.waterdata.usgs.gov/visualizations/temperature-prediction/index.html#/modeling

What do we need for ecological forecasting?

1. Produce accurate predictions
2. Characterize prediction uncertainty
3. Make use of recent observations
4. Improve ecological understanding

Machine learning models are another tool in our arsenal

Water Resources Research

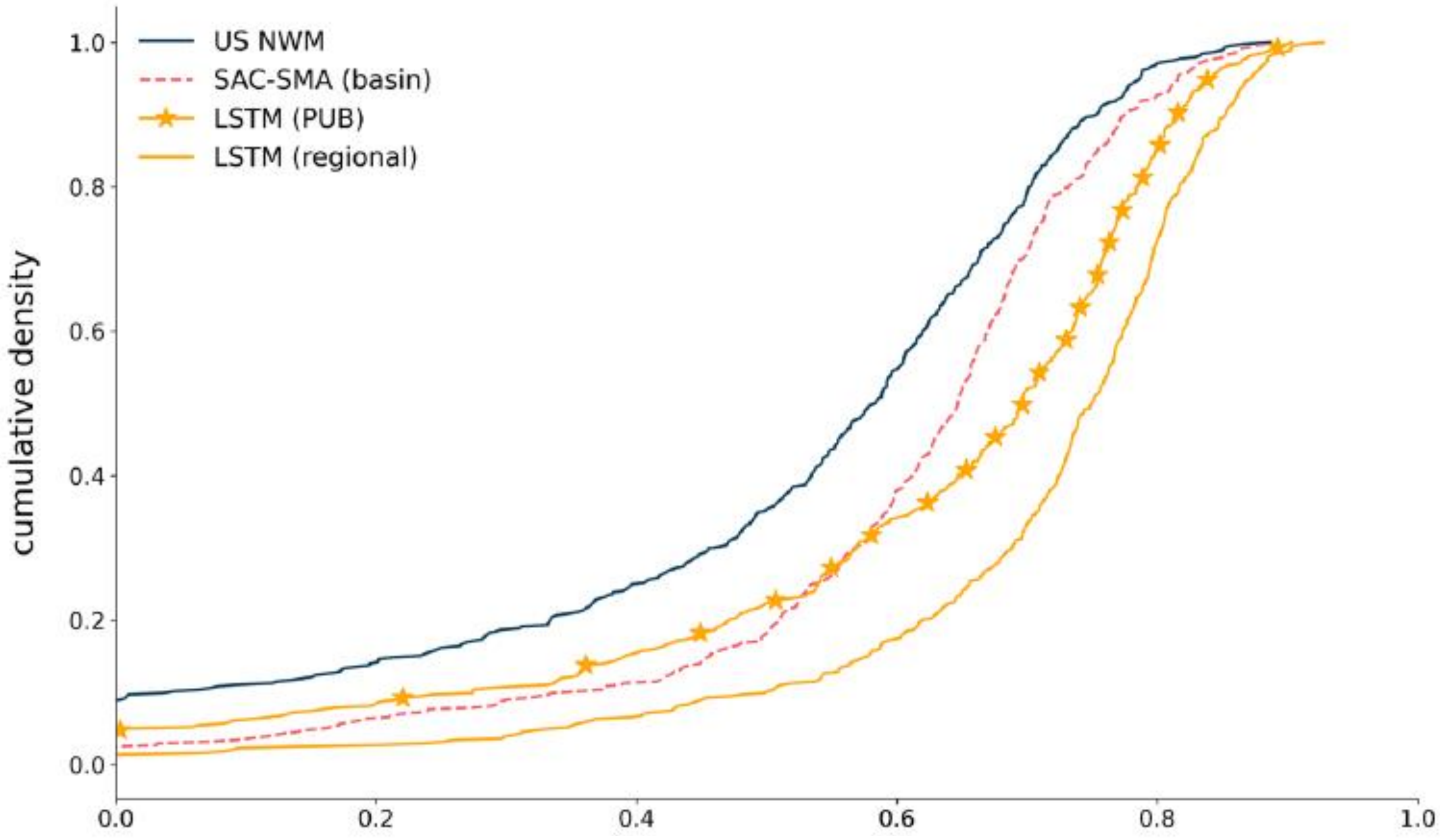
COMMENTARY

10.1029/2020WR028091

Special Section:
Big Data & Machine Learning in Water Sciences: Recent

What Role Does Hydrological Science Play in the Age of Machine Learning?

Grey S. Nearing[1], Frederik Kratzert[2], Alden Keefe Sampson[3], Craig S. Pelissier[4], Daniel Klotz[2], Jonathan M. Frame[1], Cristina Prieto[5], and Hoshin V. Gupta[6]



Predictive accuracy

Nearing et al. 2020

Water Resources Research

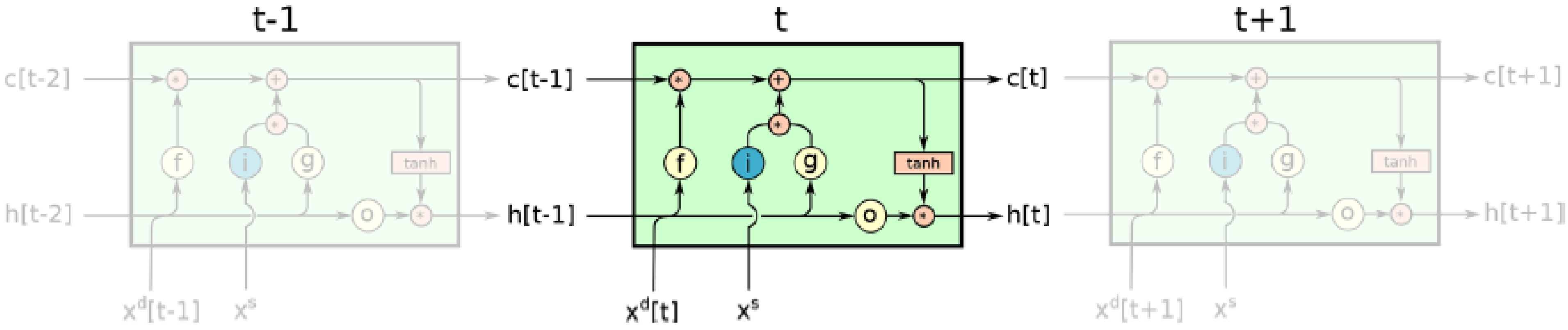
COMMENTARY
10.1029/2020WR028091

Special Section:
Big Data & Machine Learning
in Water Sciences: Recent

What Role Does Hydrological Science Play in the Age of Machine Learning?

Grey S. Nearing[1], Frederik Kratzert[2], Alden Keefe Sampson[3], Craig S. Pelissier[4], Daniel Klotz[2], Jonathan M. Frame[1], Cristina Prieto[5], and Hoshin V. Gupta[6]

(b) EA-LSTM



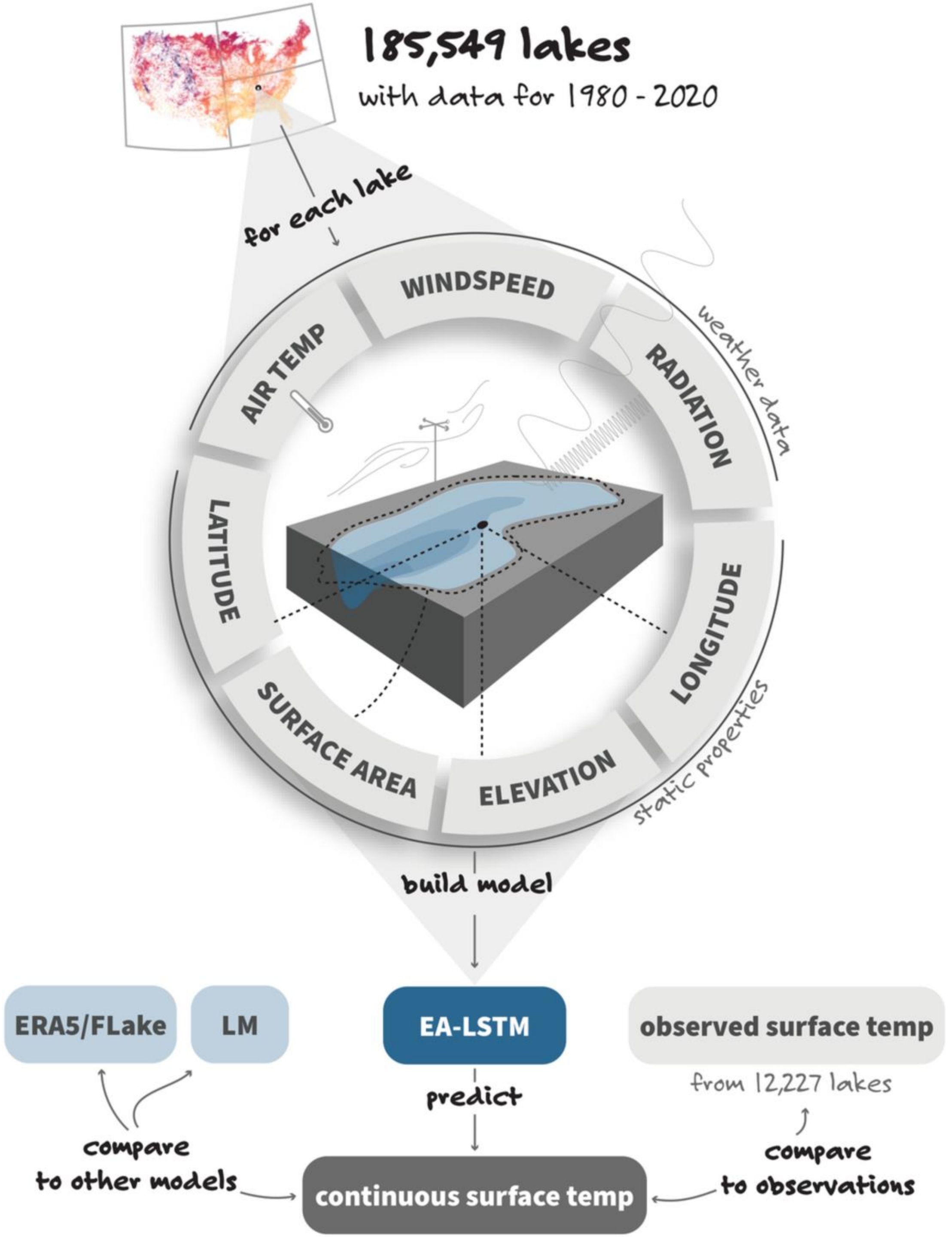
Kratzert et al. 2019

LIMNOLOGY AND OCEANOGRAPHY
LETTERS
ASLO
Open Access

Data Article | Open Access

Daily surface temperatures for 185,549 lakes in the conterminous United States estimated using deep learning (1980–2020)

Jared D. Willard, Jordan S. Read, Simon Topp, Gretchen J. A. Hansen, Vipin Kumar



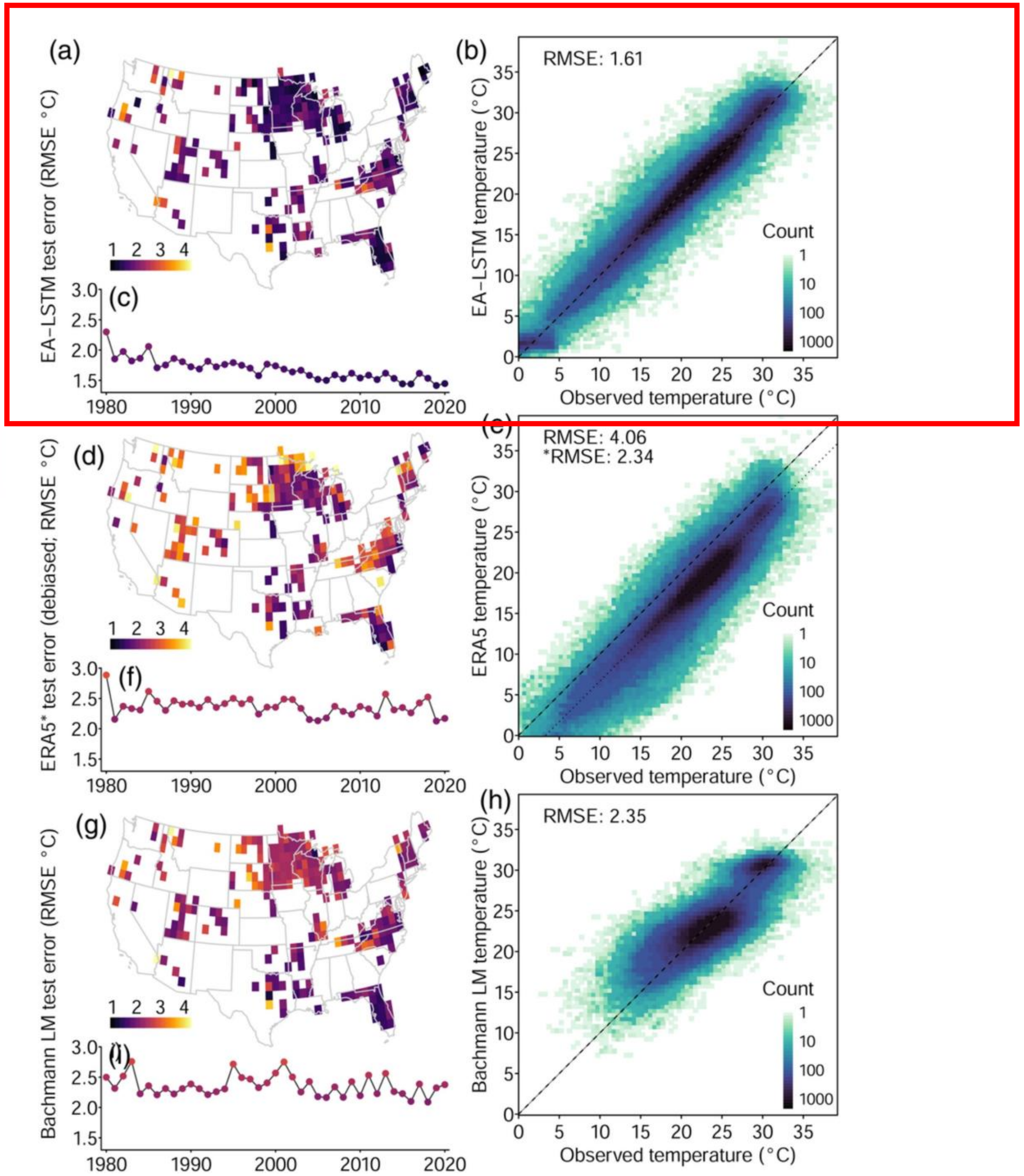
ML model most accurate for 185,000+ lakes

LIMNOLOGY AND OCEANOGRAPHY
LETTERS
ASLO
Open Access

Data Article | Open Access

Daily surface temperatures for 185,549 lakes in the conterminous United States estimated using deep learning (1980–2020)

Jared D. Willard, Jordan S. Read, Simon Topp, Gretchen J. A. Hansen, Vipin Kumar



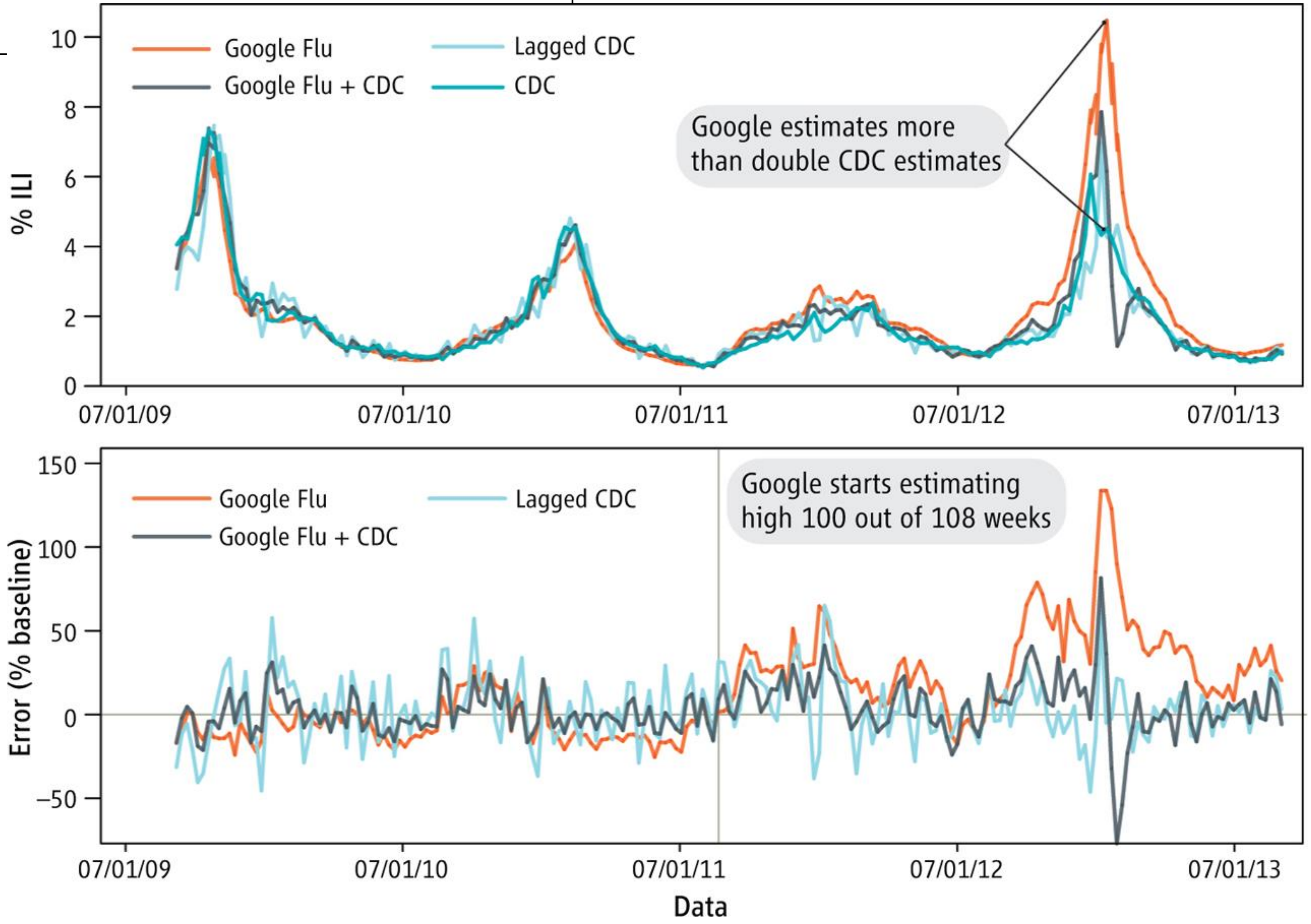
POLICYFORUM

BIG DATA

The Parable of Google Flu: Traps in Big Data Analysis

David Lazer,[1,2*] Ryan Kennedy,[1,3,4] Gary King,[3] Alessandro Vespignani[5,6,3]

Large errors in flu prediction were largely avoidable, which offers lessons for the use of big data.



- Overfit to few data points
- Not accounting for changes in search

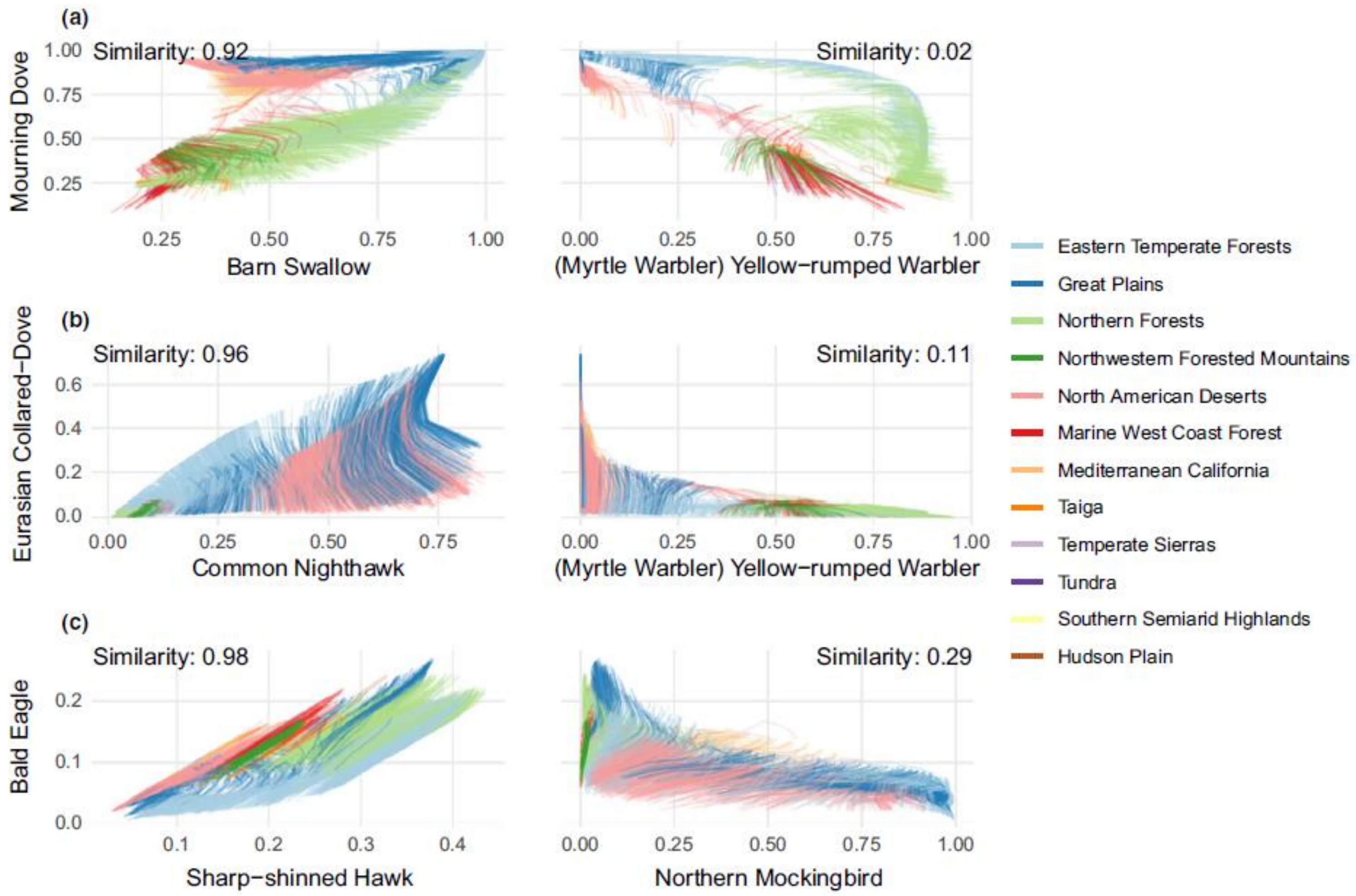
ECOLOGY LETTERS

Method | Full Access

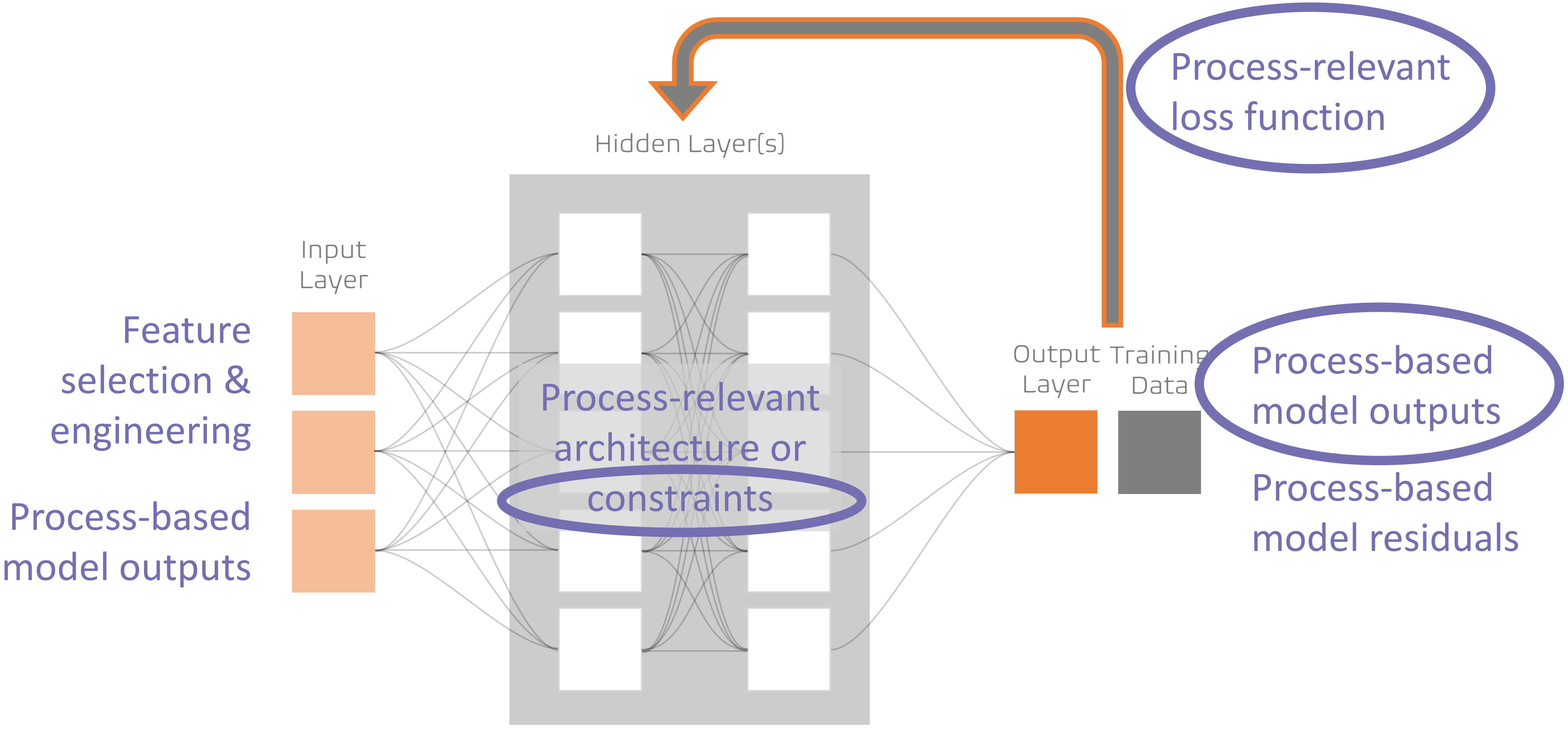
Neural hierarchical models of ecological populations

Maxwell B. Joseph

- Use neural networks to parameterize a species occupancy model
- Model structure is pre-defined



PGDL for lake and stream temperatures



Concepts expanded from Willard et al. preprint
Figure by Ellen Bechtel, modified from Appling et al. in Press

PGDL for lake and stream temperatures

- **Structural awareness of time**
  (LSTM; lakes and streams)

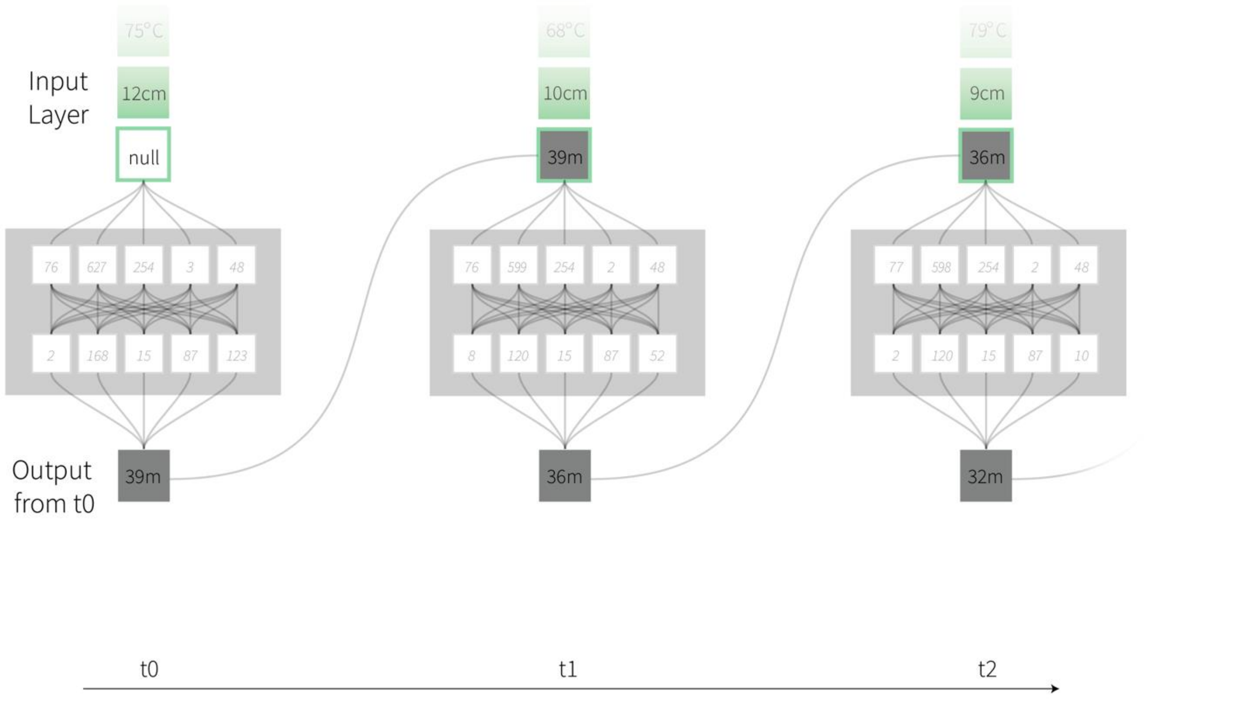
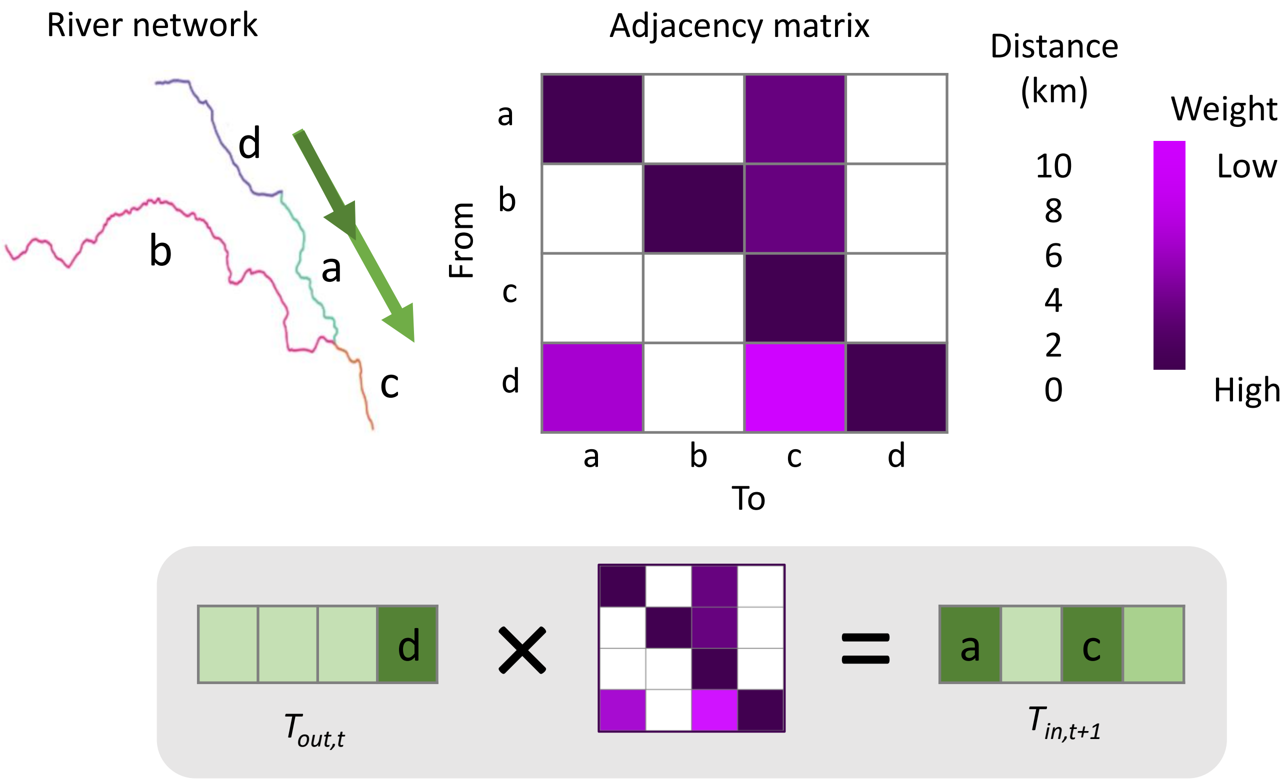


Figure by Ellen Bechtel, modified from Appling et al. in Press
Streams: Jia et al., in review & arXiv 2020; Sadler et al., in prep
Lakes: Karpatne et al. arXiv 2017; Jia et al., Proc. SIAM, 2019; Read et al. WRR, 2019

PGDL for lake and stream temperatures

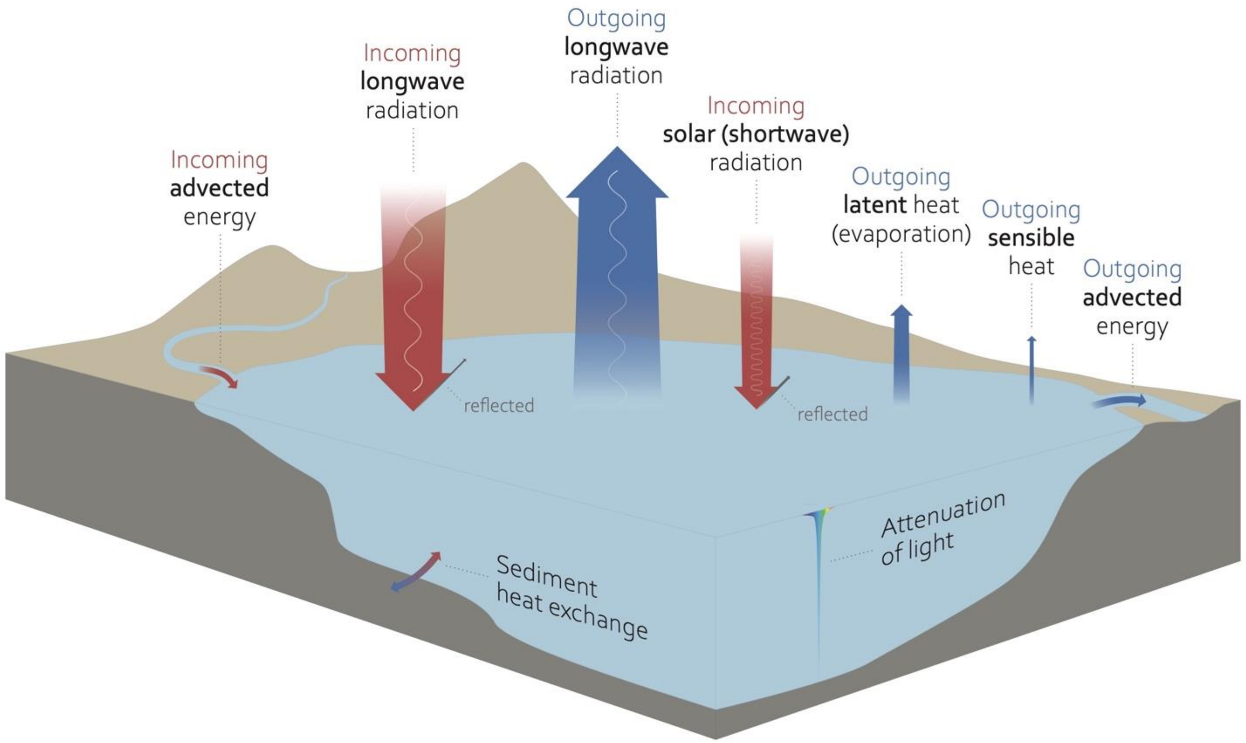
- **Structural awareness of time** (LSTM; both) **and space** (GCNN; streams)



Graph figures: inspired by Jeff Sadler's
Streams: Jia et al., in review & arXiv 2020; Sadler et al., in prep
Lakes: Karpatne et al. arXiv 2017; Jia et al., Proc. SIAM, 2019; Read et al. WRR, 2019

PGDL for lake and stream temperatures

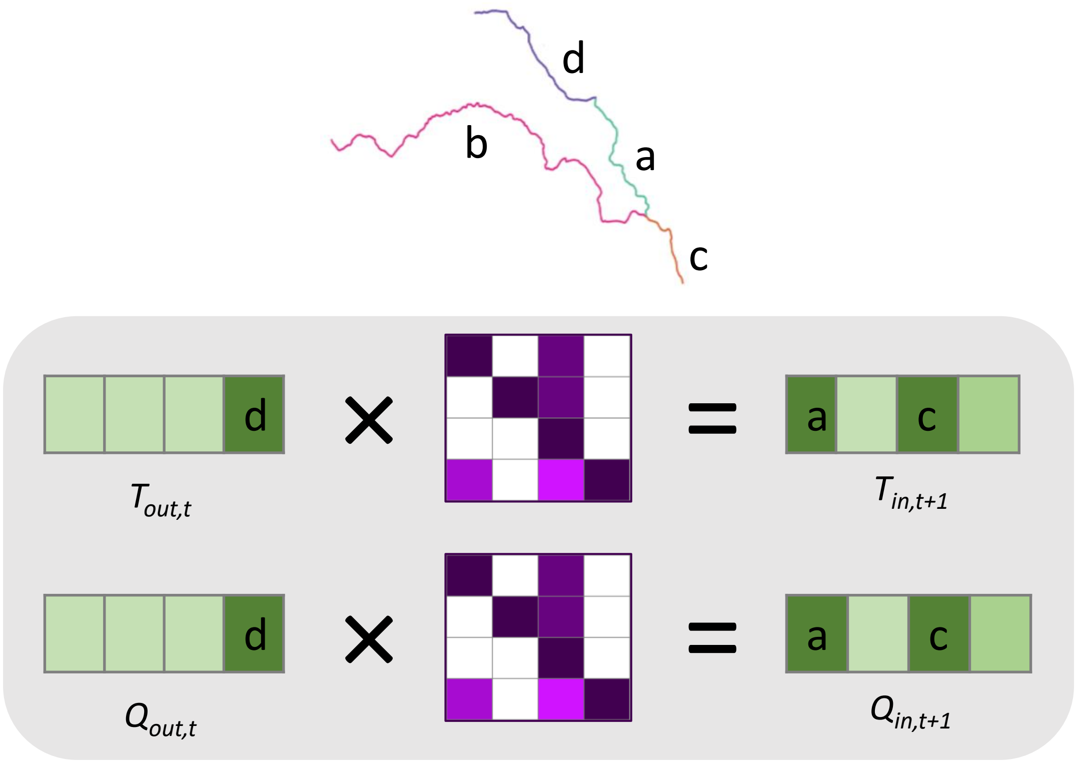
- **Structural awareness of time** (LSTM; both) **and space** (GCNN; streams)
- **Custom loss function:** energy balance (lakes)



Energy balance figure: Hayley Corson-Dosch
Streams: Jia et al., in review & arXiv 2020; Sadler et al., in prep
Lakes: Karpatne et al. arXiv 2017; Jia et al., Proc. SIAM, 2019; Read et al. WRR, 2019

PGDL for lake and stream temperatures

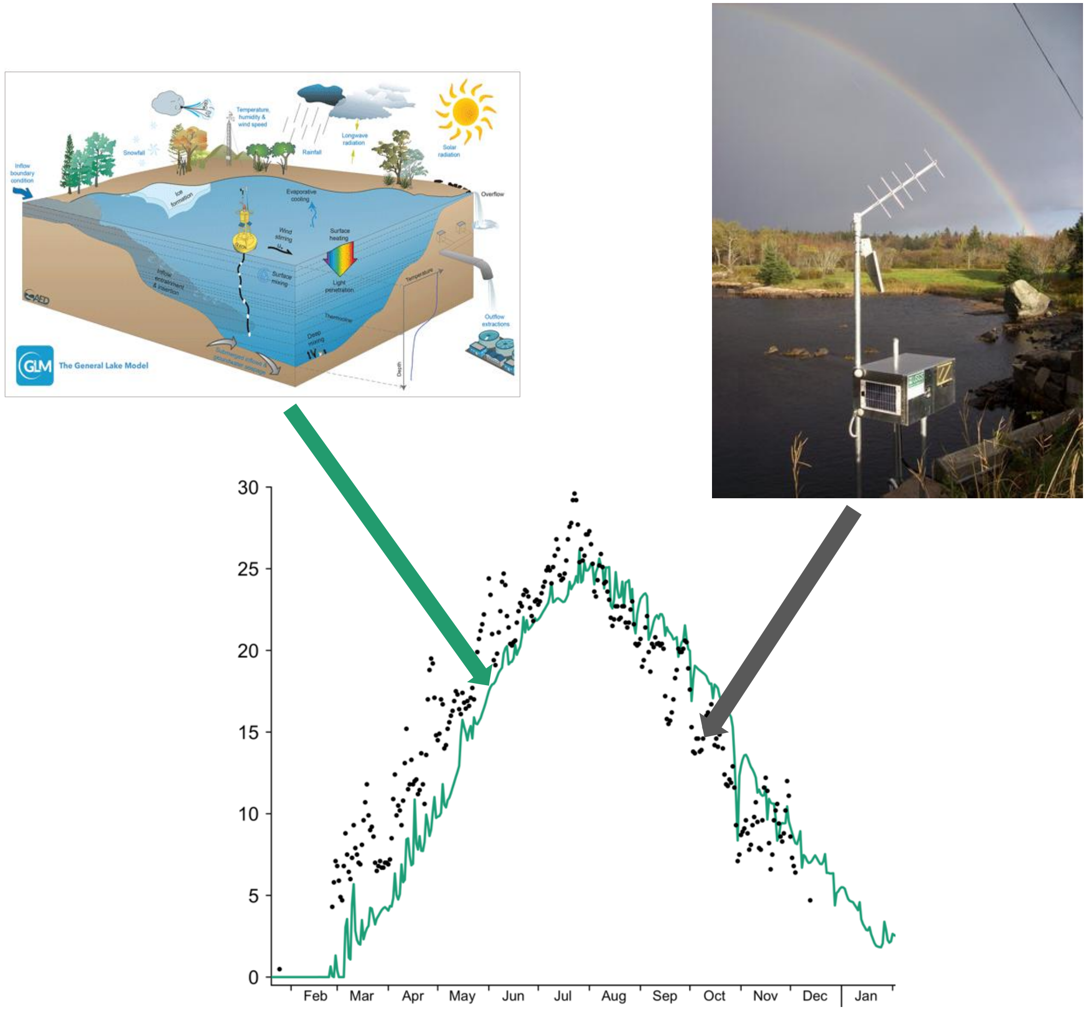
- **Structural awareness of time** (LSTM; both) **and space** (GCNN; streams)
- **Custom loss function:** energy balance (lakes), heat & flow info shared downstream (streams)



Energy balance figure: Hayley Corson-Dosch
Streams: Jia et al., in review & arXiv 2020; Sadler et al., in prep
Lakes: Karpatne et al. arXiv 2017; Jia et al., Proc. SIAM, 2019; Read et al. WRR, 2019

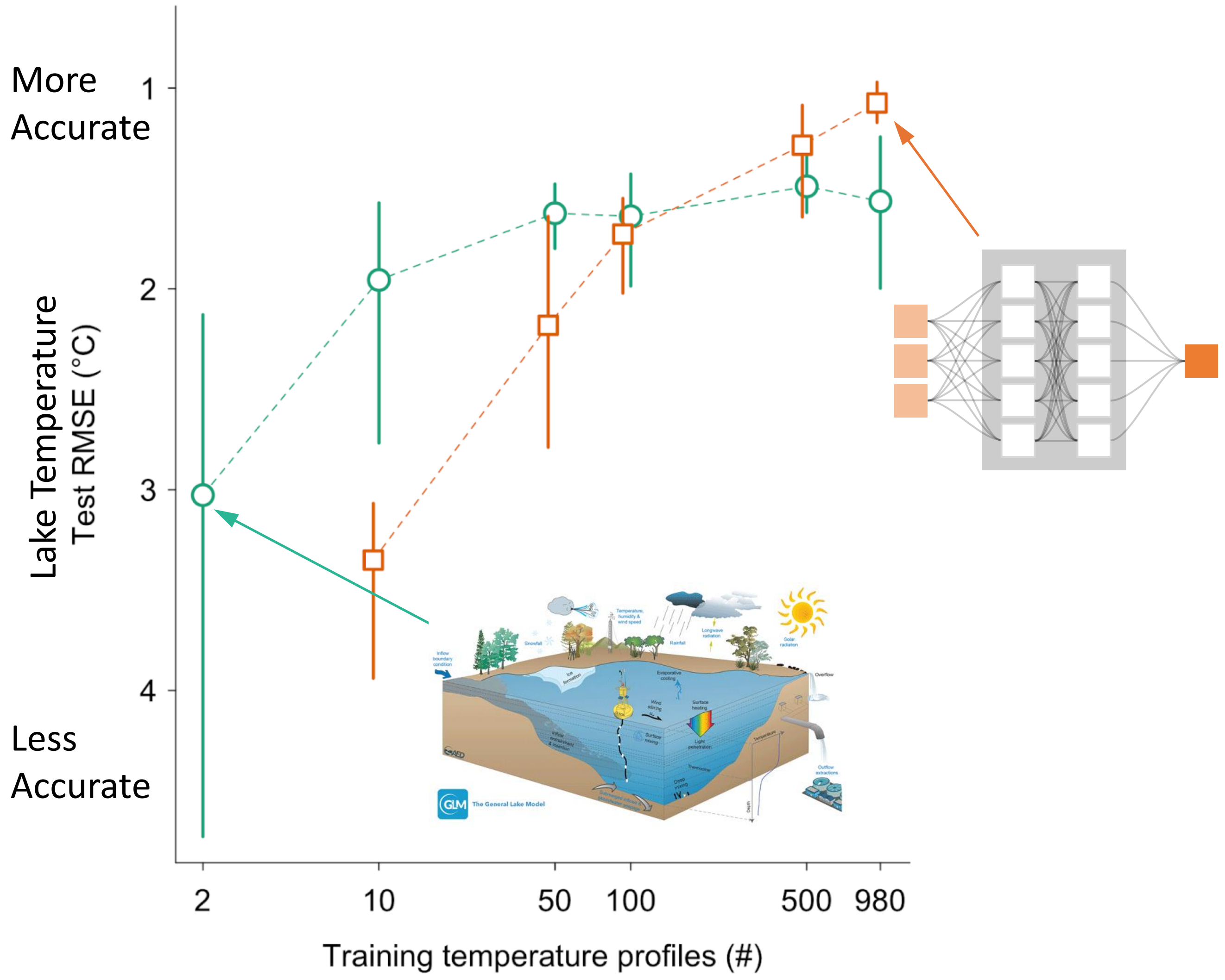
PGDL for lake and stream temperatures

- **Structural awareness of time** (LSTM; both) **and space** (GCNN; streams)
- **Custom loss function:** energy balance (lakes), heat & flow info shared downstream (streams)
- **Pretraining** on process model outputs (both)



Streams: Jia et al., in review & arXiv 2020; Sadler et al., in prep
Lakes: Karpatne et al. arXiv 2017; Jia et al., Proc. SIAM, 2019; Read et al. WRR, 2019

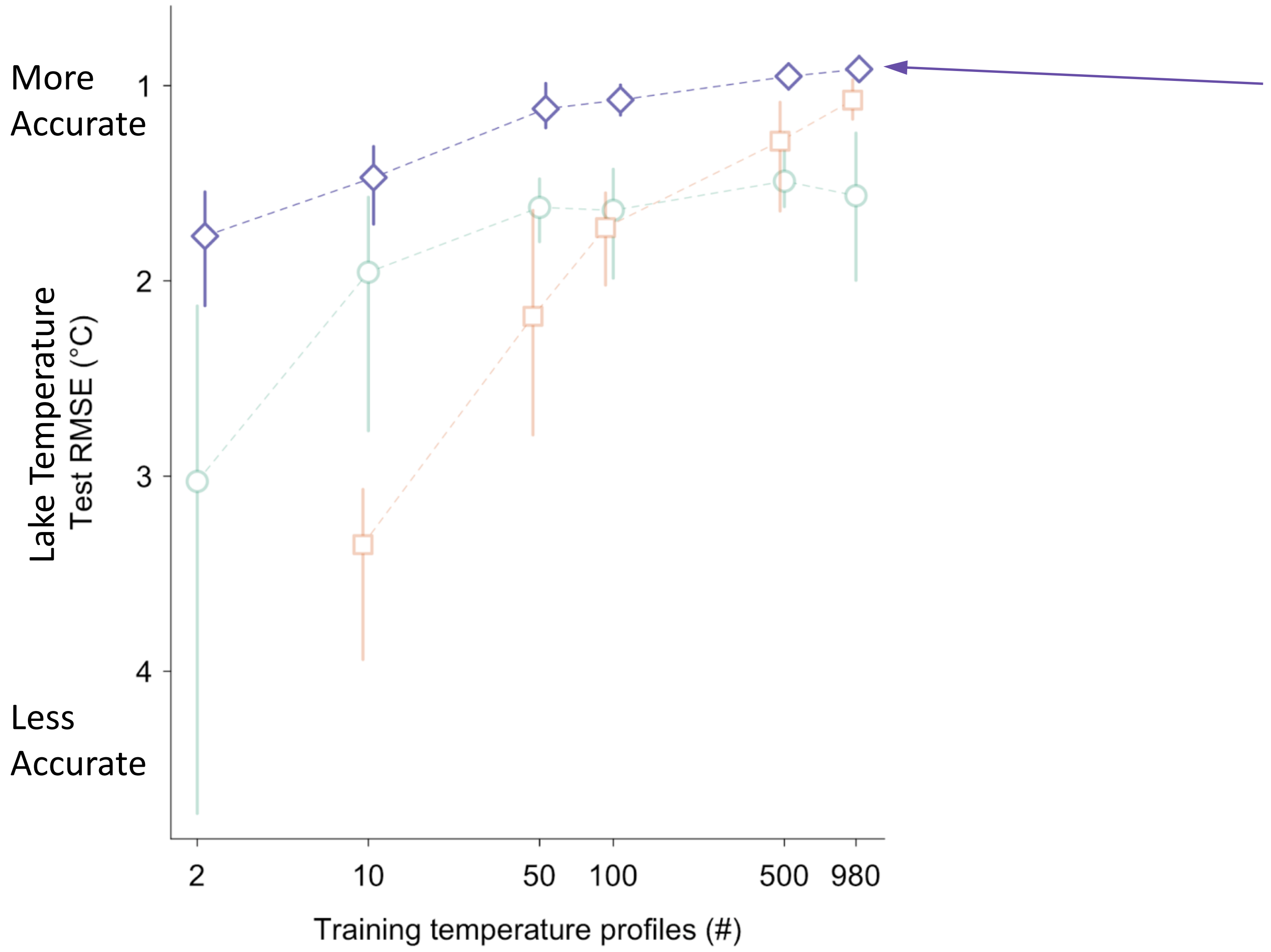
Theory vs. ML depends on data abundance



- When given enough data, **Deep Learning** methods can beat process-based models
- **Process-Based** models can be applied with more confidence to data-poor regions

Adapted from Read et al. 2019

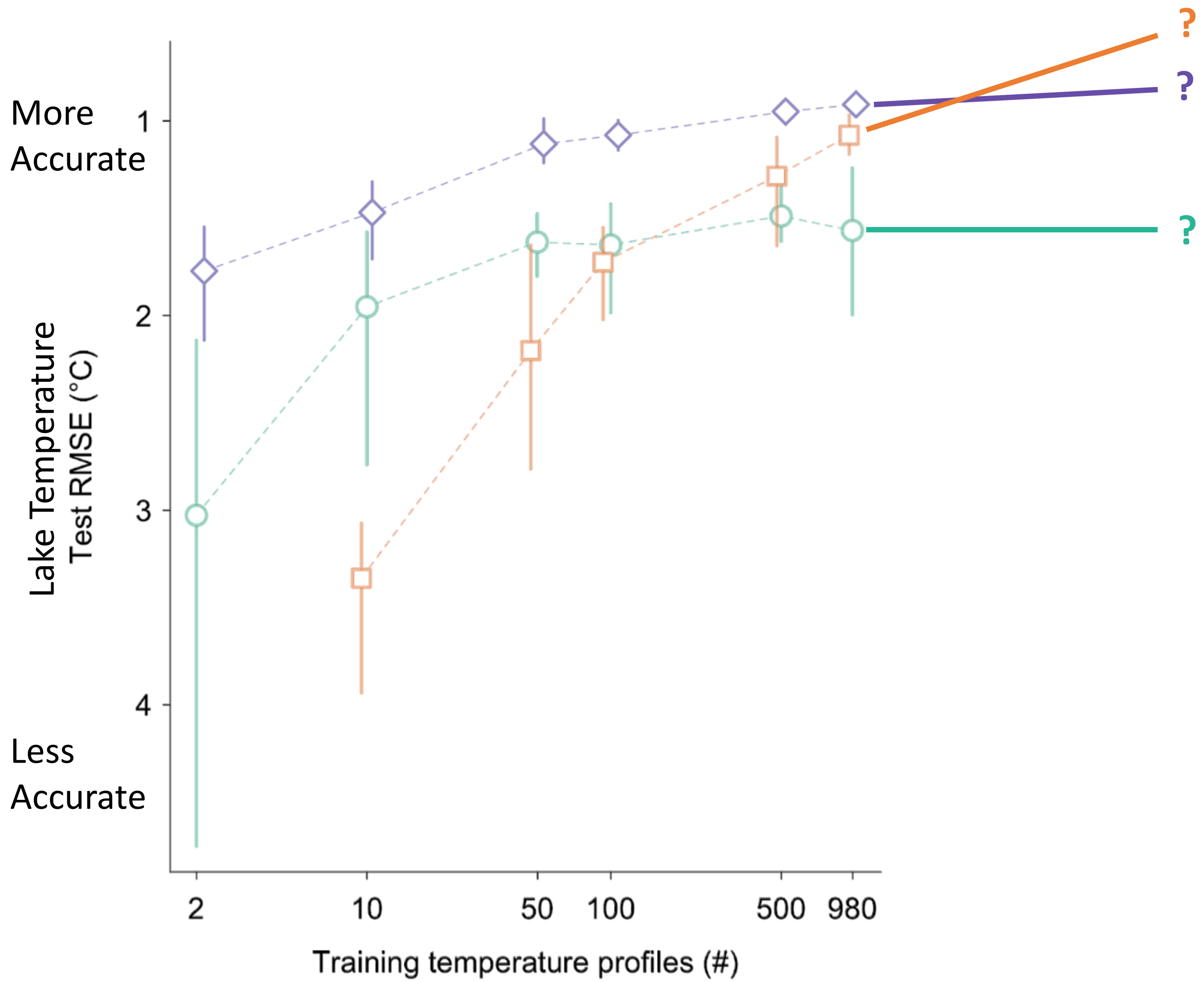
PGDL rises above theory and data



- **Process-Guided Deep Learning** performance was superior at all data densities
- When given enough data, **Deep Learning** methods can beat process-based models
- **Process-Based** models can be applied with more confidence to data-poor regions

Adapted from Read et al. 2019

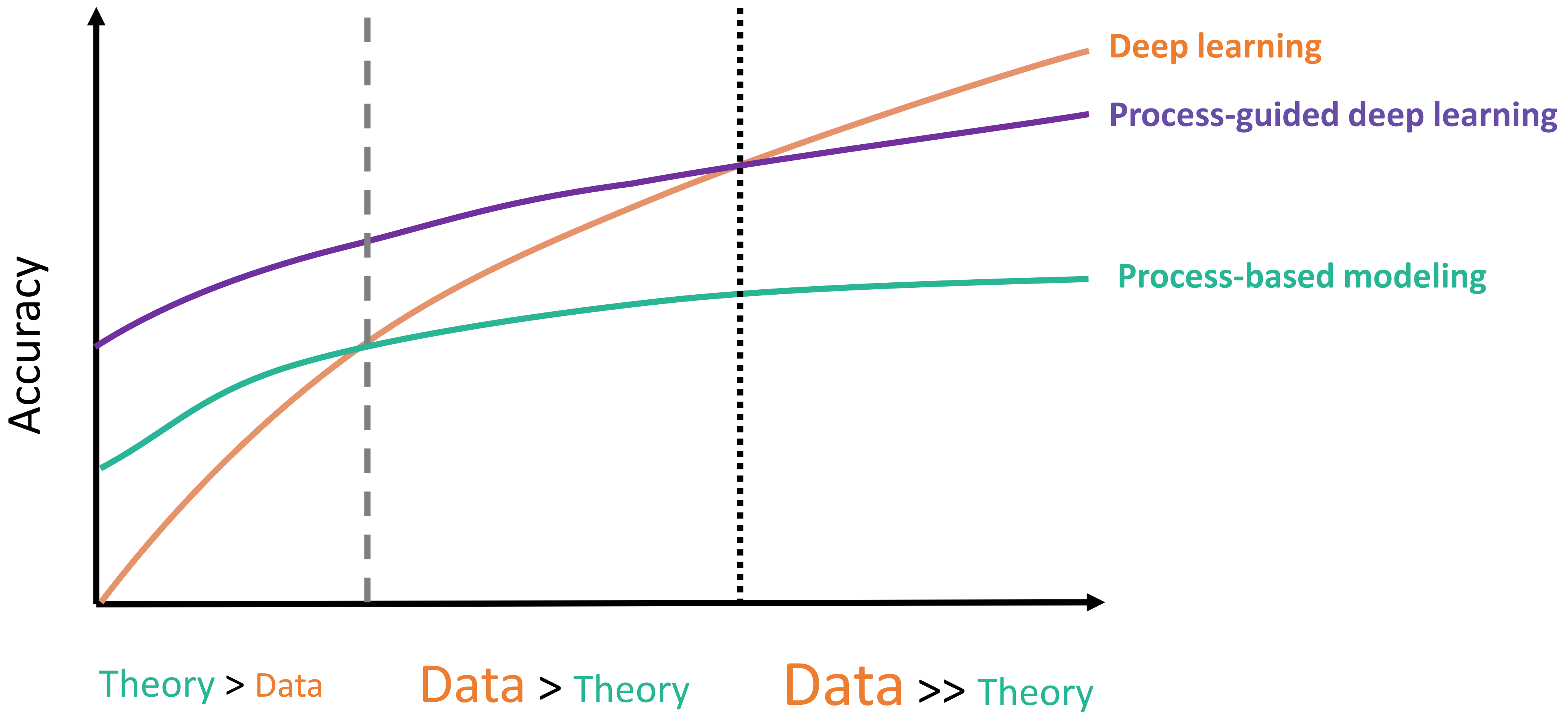
PGDL for abundant data: vanishing returns?



- Process-Guided DL
- Deep Learning
- Process-Based

Adapted from Read et al. 2019

Predictability regimes



Preliminary information – subject to revision. Not for citation or distribution.

Figure from Alison Appling

Predictability regimes

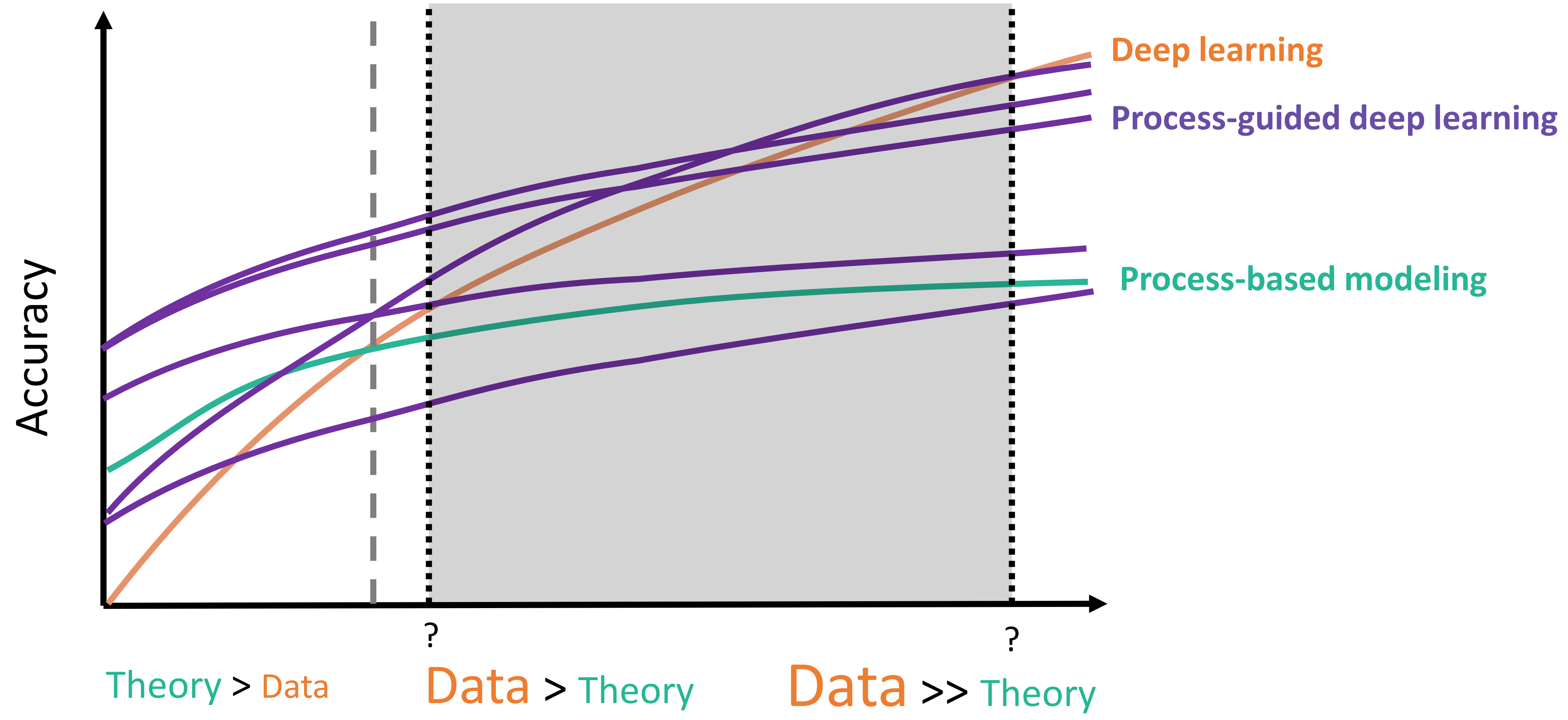


Figure from Alison Appling

What do we need for ecological forecasting?

1. Produce accurate predictions
2. Characterize prediction uncertainty
3. Make use of recent observations
4. Improve ecological understanding

Machine learning models are another tool in our arsenal

Uncertainty – Monte Carlo Dropout

- Randomly drops out portion of network during training phase
- Used as a regularization scheme to prevent overfitting
- When dropout used during prediction, gives an estimate of model uncertainty

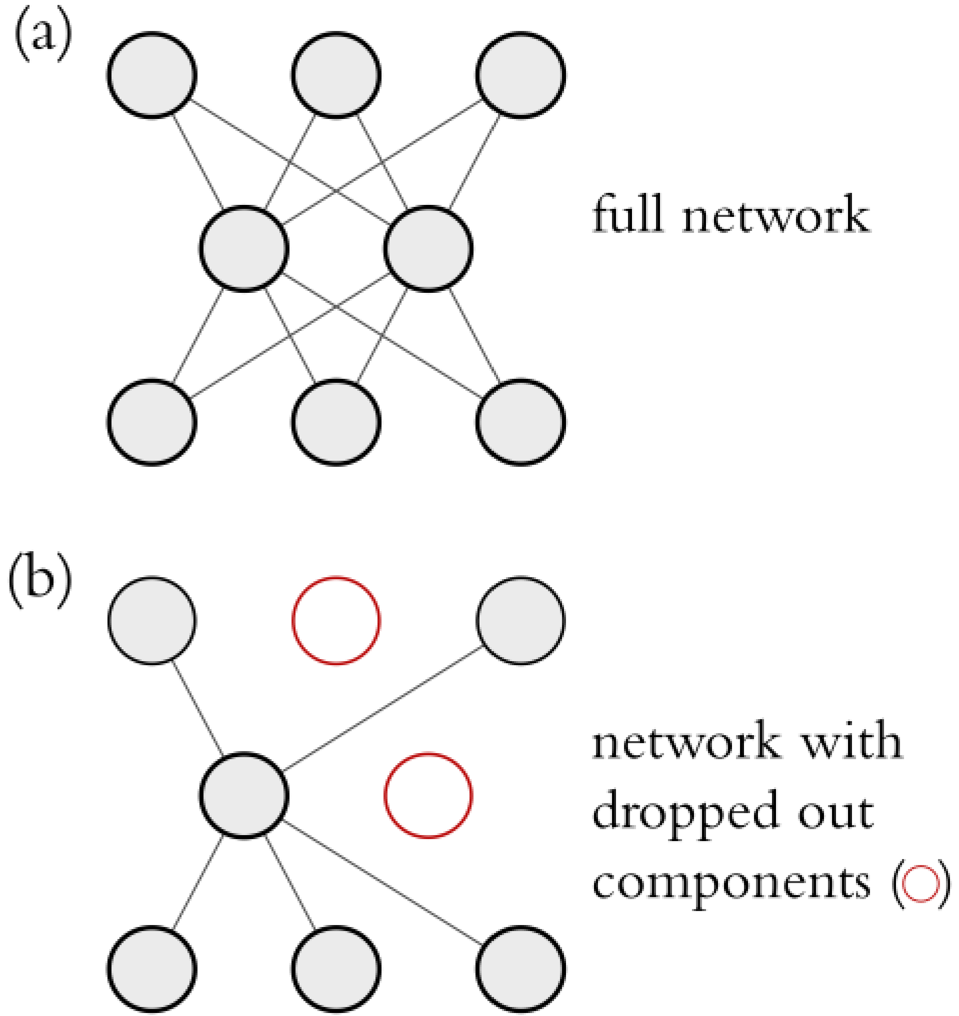
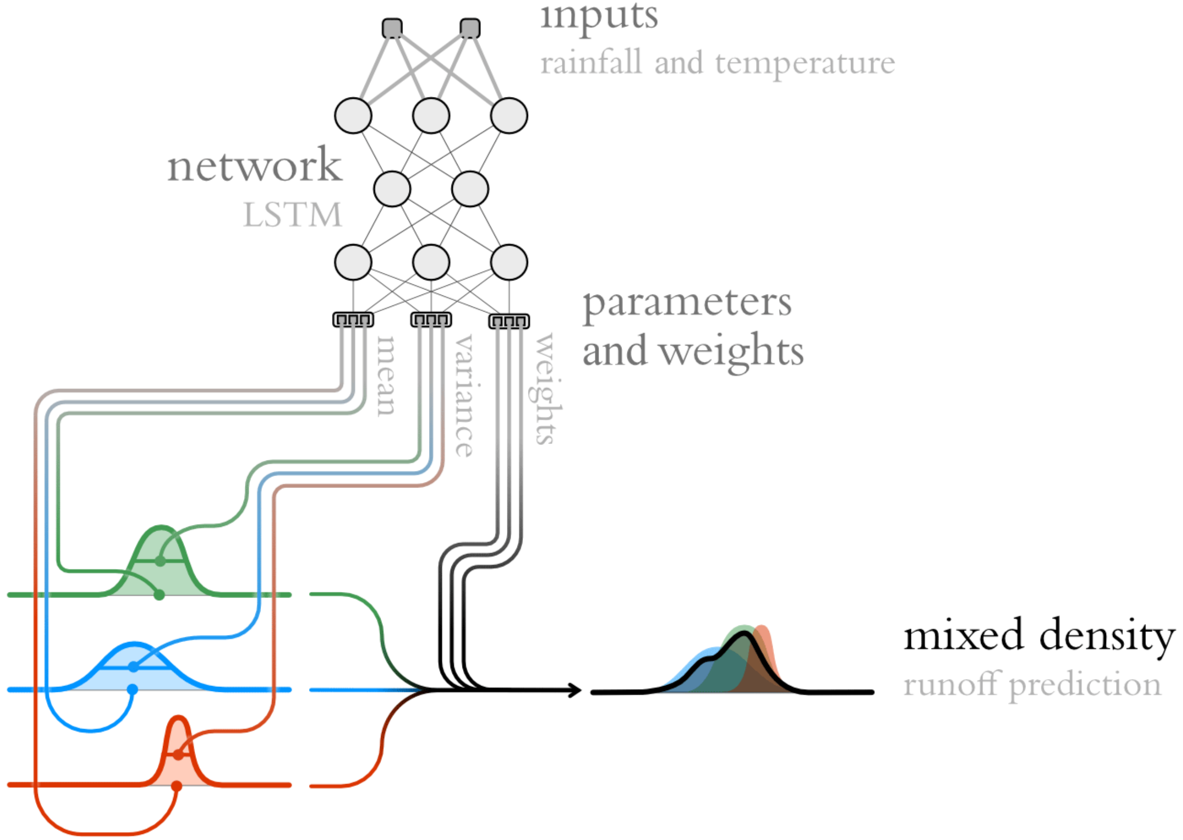
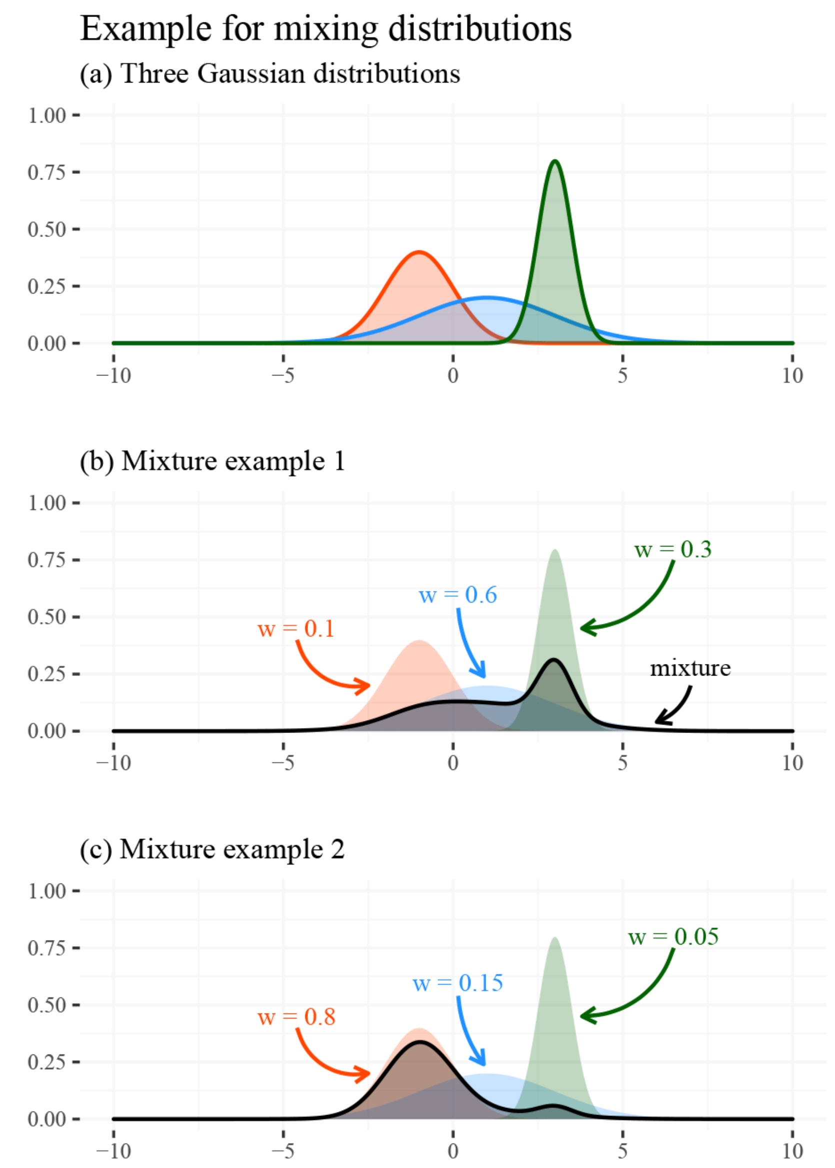


Figure from Klotz et al. 2022

Uncertainty – Mixture Density Networks

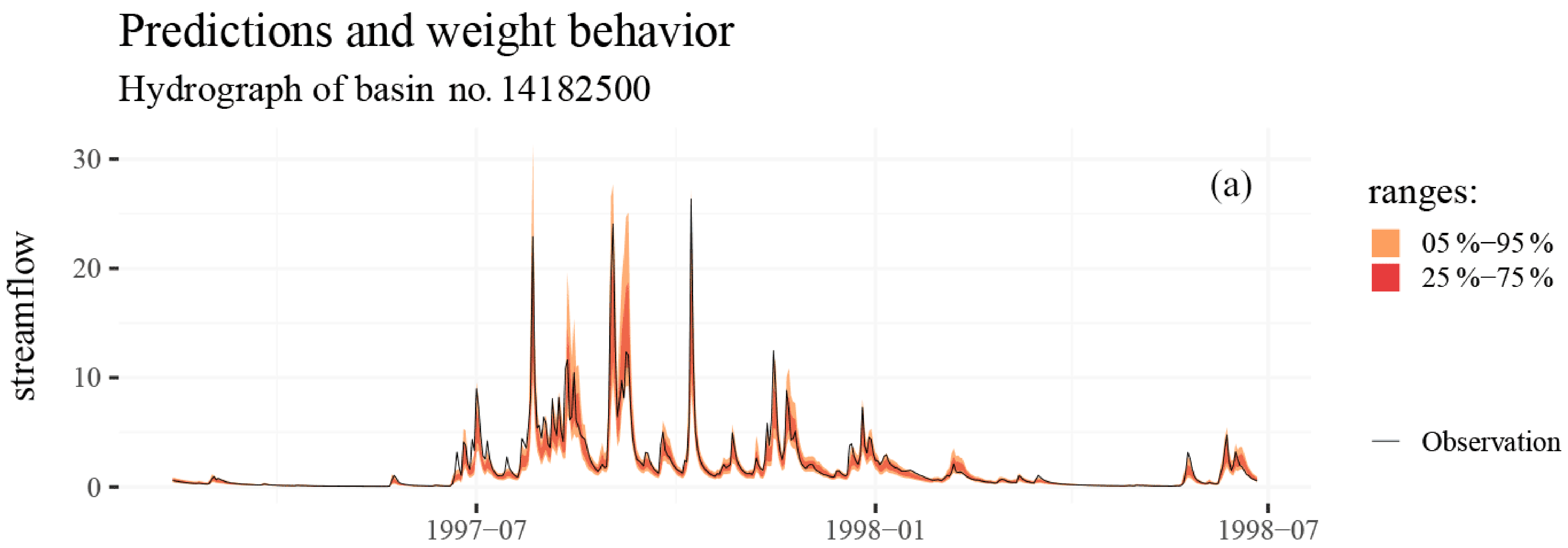




$$L(q \mid \mathbf{x}) = -\log \left[\sum_{k=1}^{K} \alpha_k(\mathbf{x}) \cdot \mathcal{N}\Big(q \mid \mu_k(\mathbf{x}), \sigma_k(\mathbf{x})\Big) \right]$$

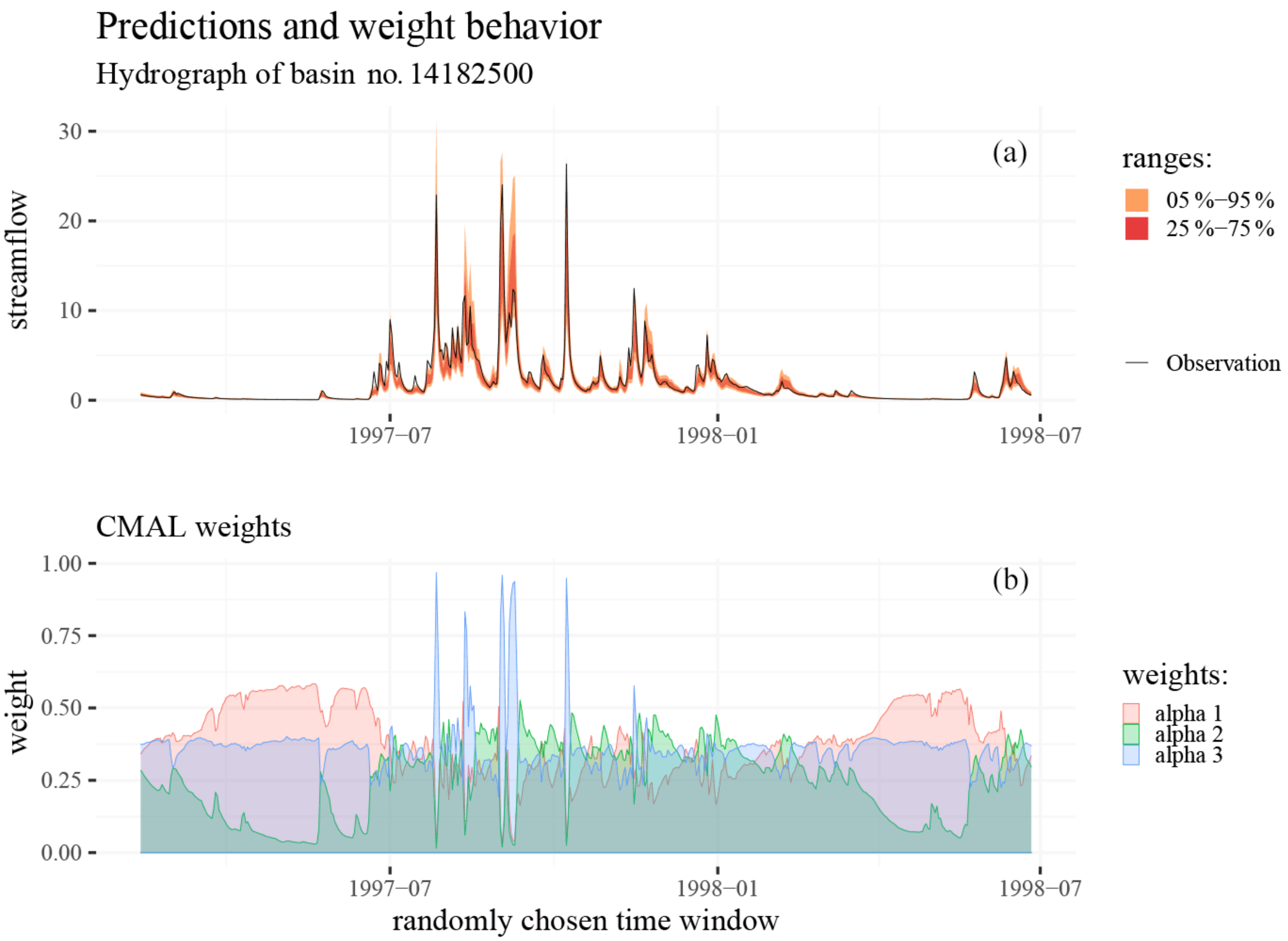
Klotz et al. 2022

Uncertainty – Mixture Density Networks



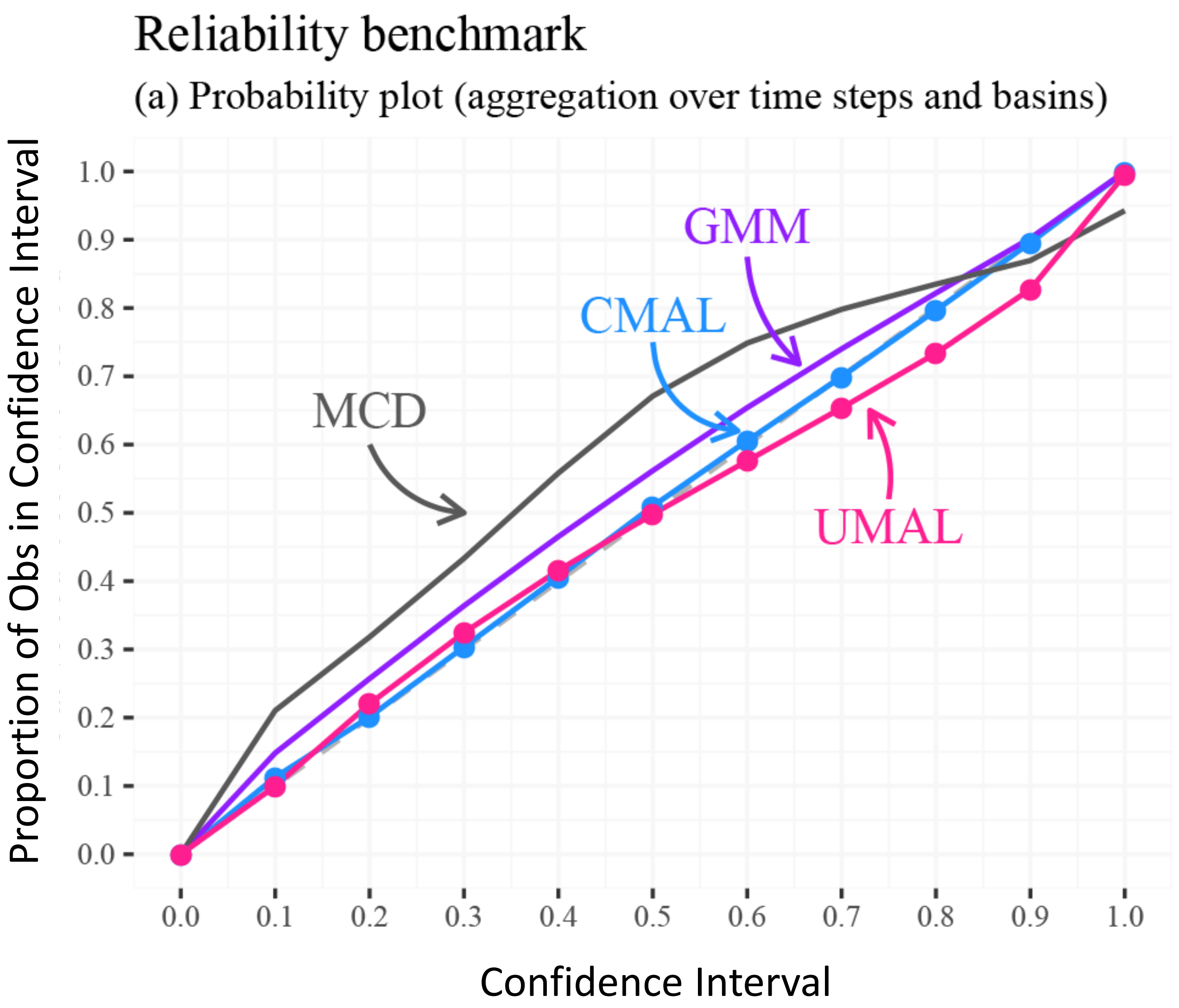
Klotz et al. 2022

Uncertainty – Mixture Density Networks



Klotz et al. 2022

Uncertainty – Mixture Density Networks



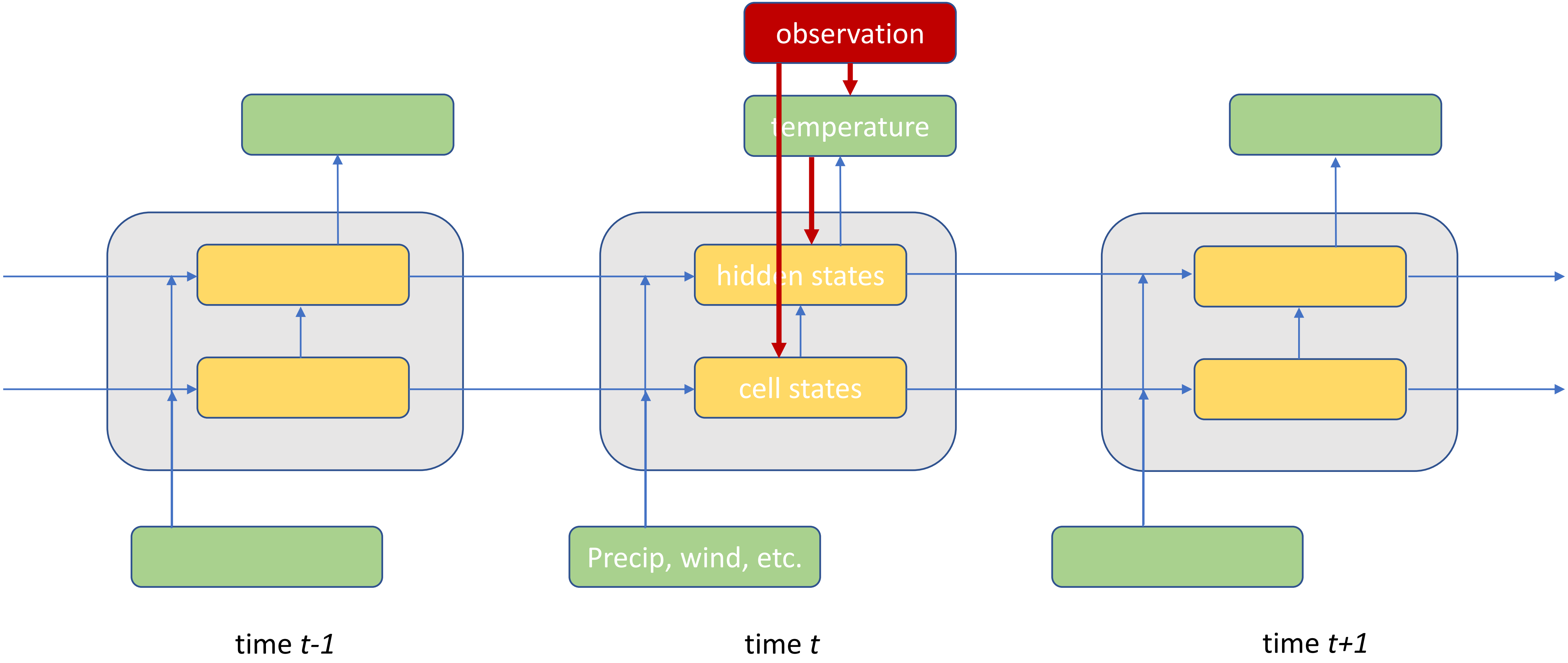
Klotz et al. 2022

What do we need for ecological forecasting?

1. Produce accurate predictions
2. Characterize prediction uncertainty
3. Make use of recent observations
4. Improve ecological understanding

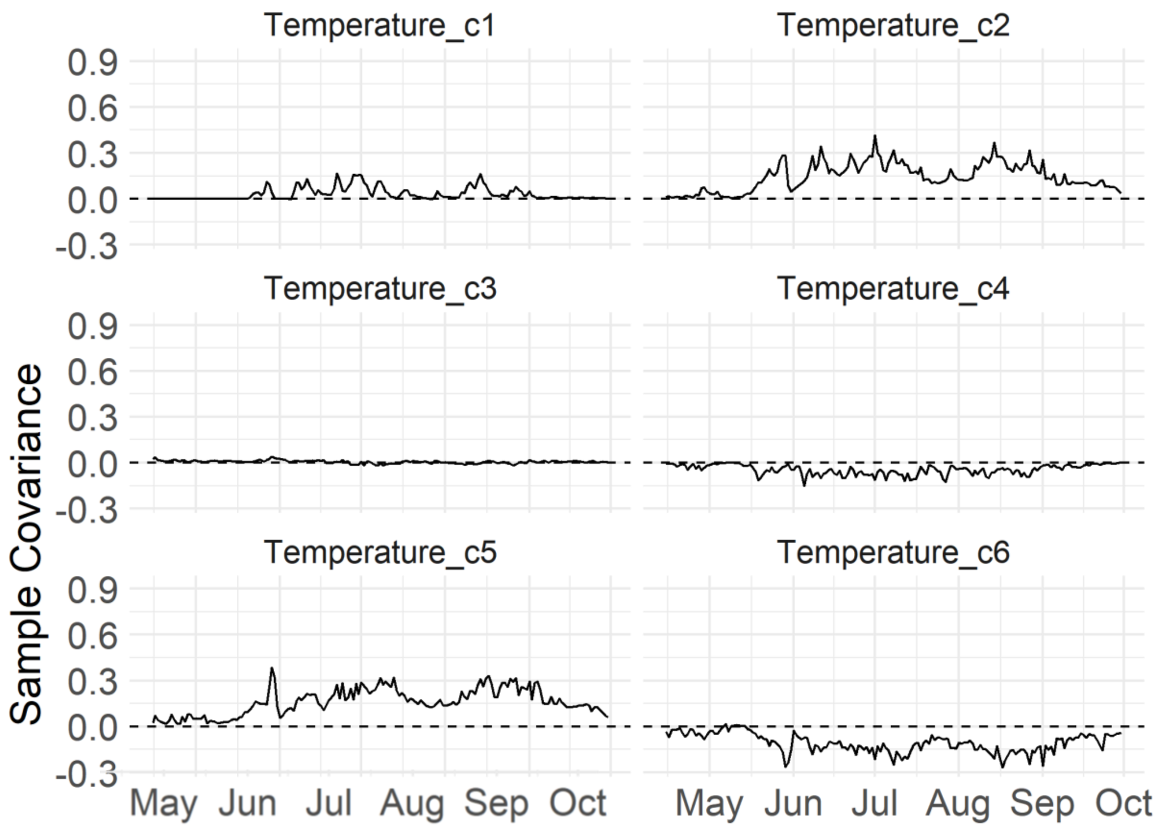
Machine learning models are another tool in our arsenal

LSTM with Data Assimilation (ensemble Kalman filter)



Zwart et al. 2021

LSTM with Data Assimilation (ensemble Kalman filter)

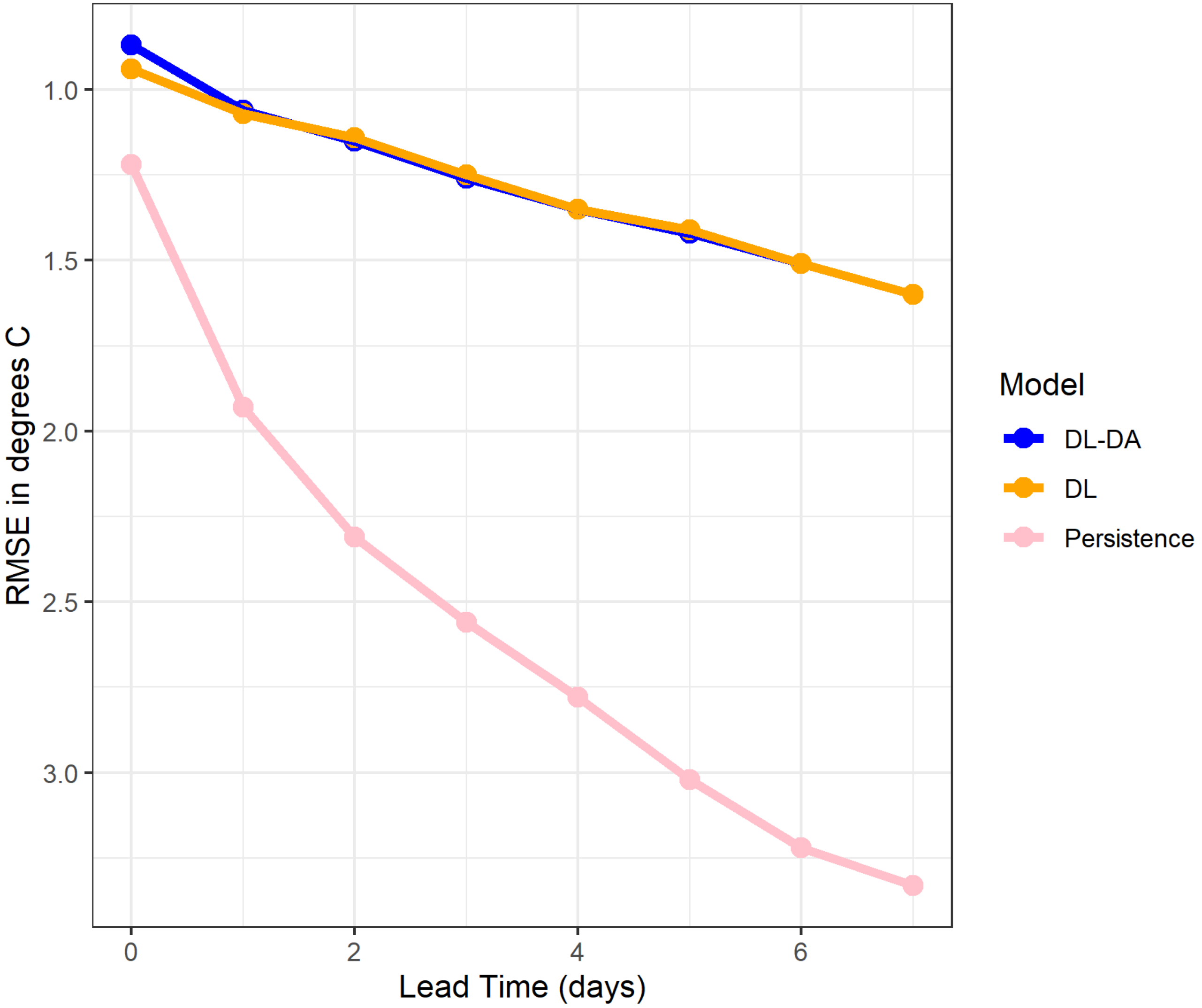


Use sample covariance to update LSTM states

Zwart et al. 2021

LSTM with Data Assimilation (ensemble Kalman filter)

More Accurate



Less Accurate

Zwart et al. 2021

Let the neural network figure it out

- Data integration kernels – use observations when available, predictions otherwise
- Invertible neural networks used to learn how to update cell states from observations
- Autoregressive techniques outperform more traditional data assimilation techniques

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

Heterogeneous Stream-reservoir Graph Networks with Data Assimilation

Shengyu Chen[1], Alison Appling[2], Samantha Oliver[2], Hayley Corson-Dosch[2], Jordan Read[2], Jeffrey Sadler[2], Jacob Zwart[2], Xiaowei Jia[1]

Technical Note: Data assimilation and autoregression for using near-real-time streamflow observations in long short-term memory networks

Grey S. Nearing[1,2], Daniel Klotz[3], Alden Keefe Sampson[4], Frederik Kratzert[5], Martin Gauch[3], Jonathan M. Frame[6,7], Guy Shalev[8], and Sella Nevo[8]

What do we need for ecological forecasting?

1. Produce accurate predictions
2. Characterize prediction uncertainty
3. Make use of recent observations
4. Improve ecological understanding

Machine learning models are another tool in our arsenal

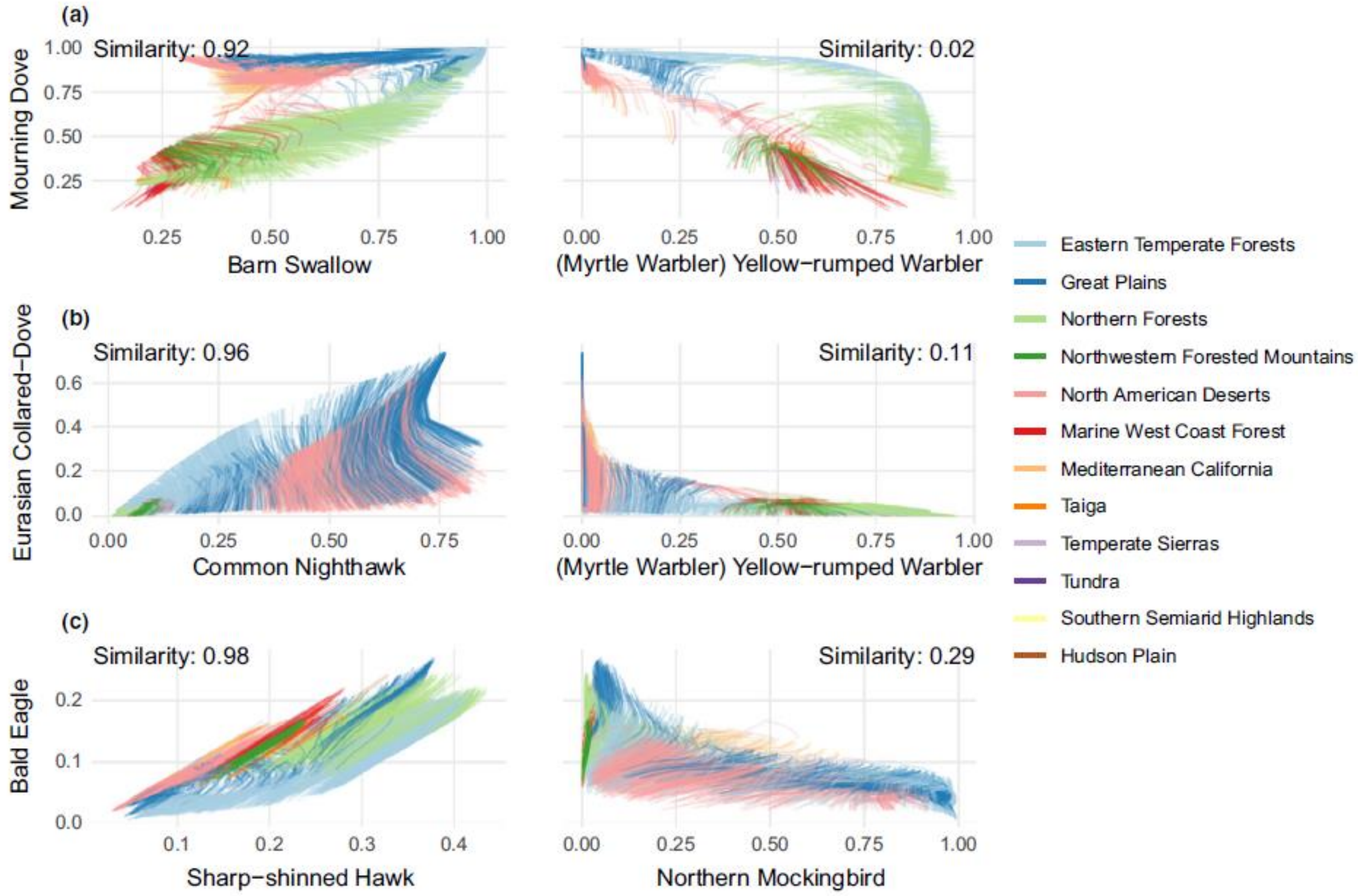
ECOLOGY LETTERS

Method | Full Access

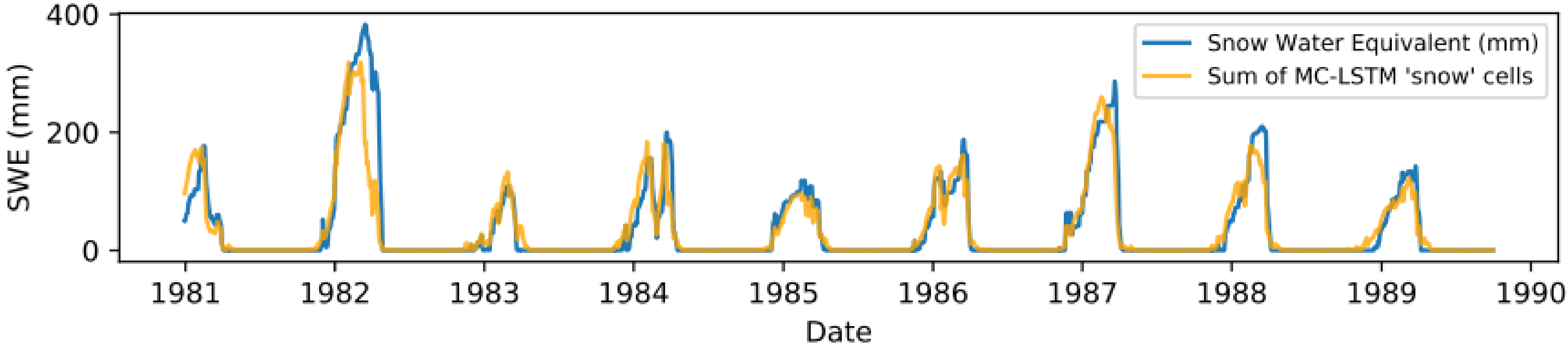
Neural hierarchical models of ecological populations

Maxwell B. Joseph

- Use neural networks to parameterize a species occupancy model
- Model structure is pre-defined



Deep Learning Interpretability



LSTM cells learn to track snow in memory cells without requiring snow data for training.

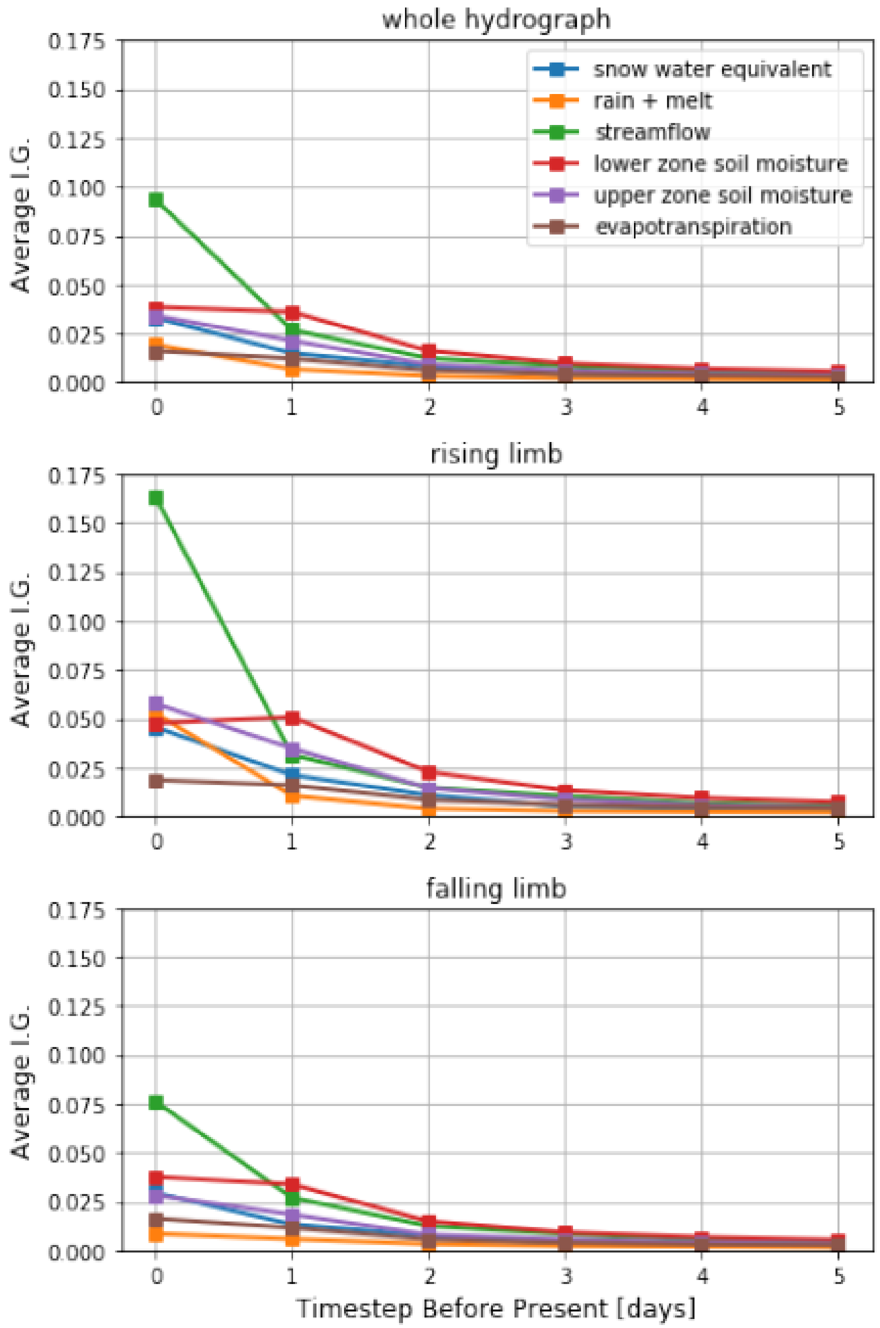
Can be applied for other difficult-to-observe states and fluxes (e.g. gas exchange, population abundance, biomass)

Kratzert et al. 2019; Hoedt et al. 2021

Deep Learning Interpretability

<u>Expected gradient:</u>

- Tells which drivers most influence cell states or model output



Nearing et al. 2020

Deep Learning Interpretability

How are our models capturing spatial relationships?

Are those relationships physically realistic?

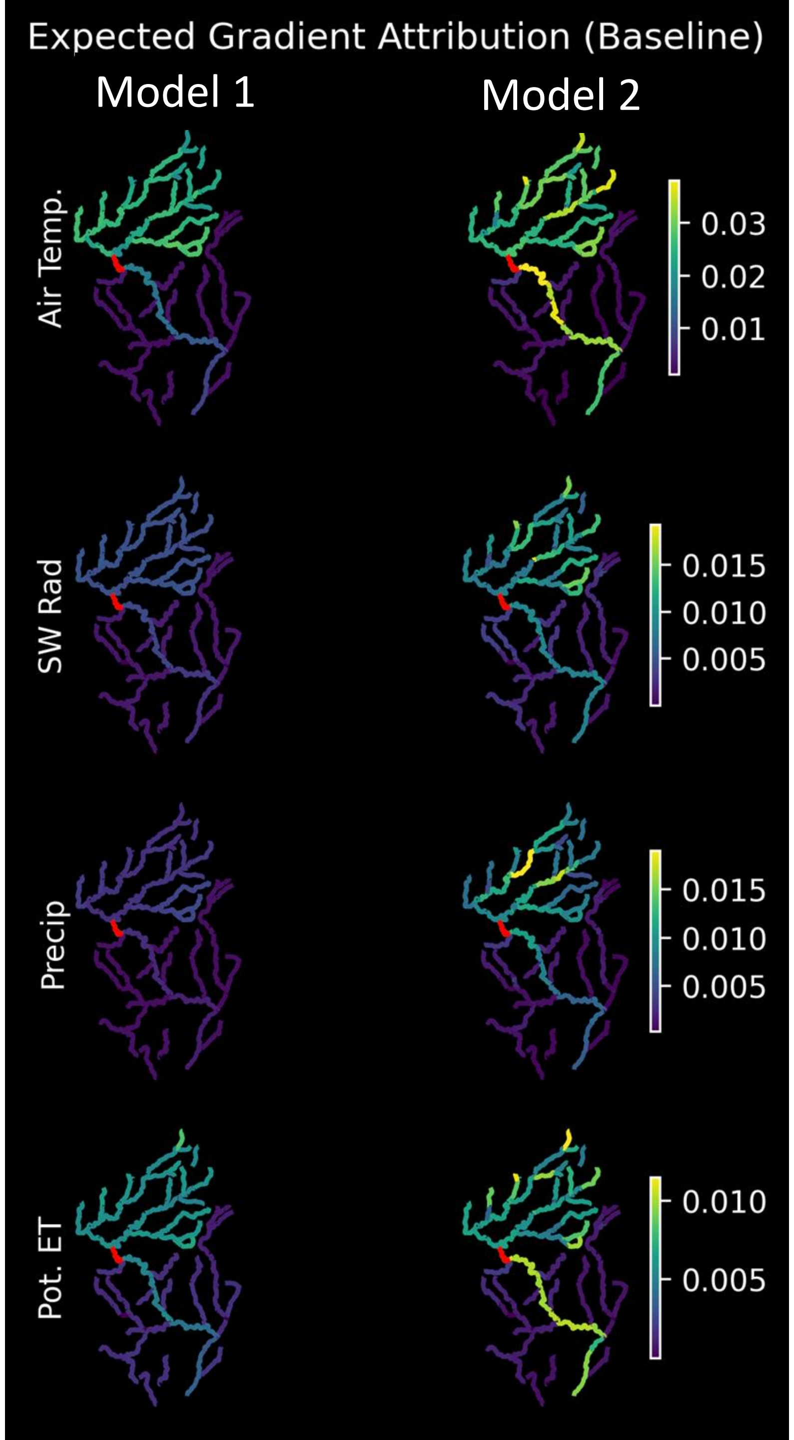


Figure from Simon Topp, Jeremy Diaz, and Lauren Koenig

Deep Learning
Forecast Interpretability

Mixture density networks + expected gradients = variance partitioning of forecast inputs

Summary: machine learning is highly suitable for ecological forecasting

1. Produce accurate predictions
2. Characterize prediction uncertainty
3. Make use of recent observations
4. Improve ecological understanding