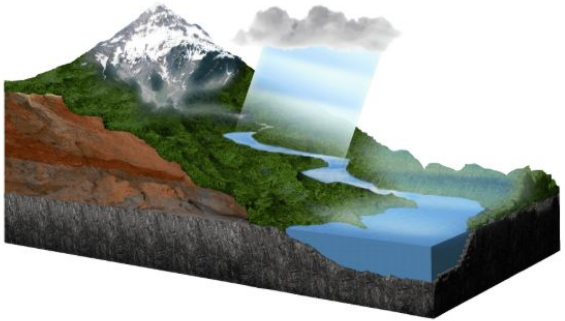
Deep Learning for Streamflow Forecasting: Tricks of the Trade

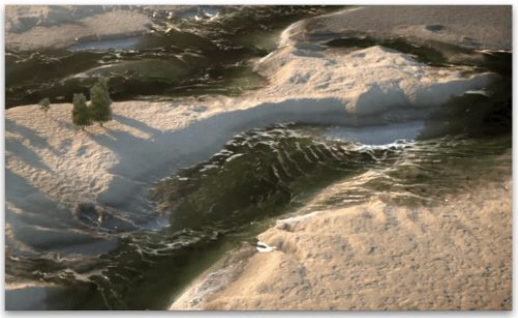
Martin Gauch
gauch@google.com

Google Research

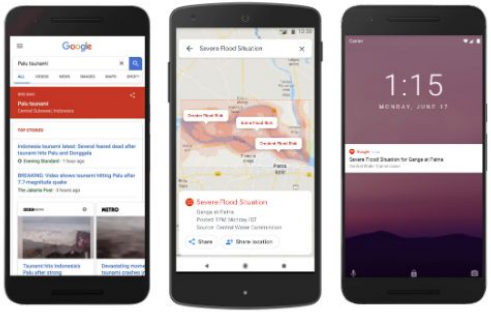
Flood forecasting at Google



Hydrologic model

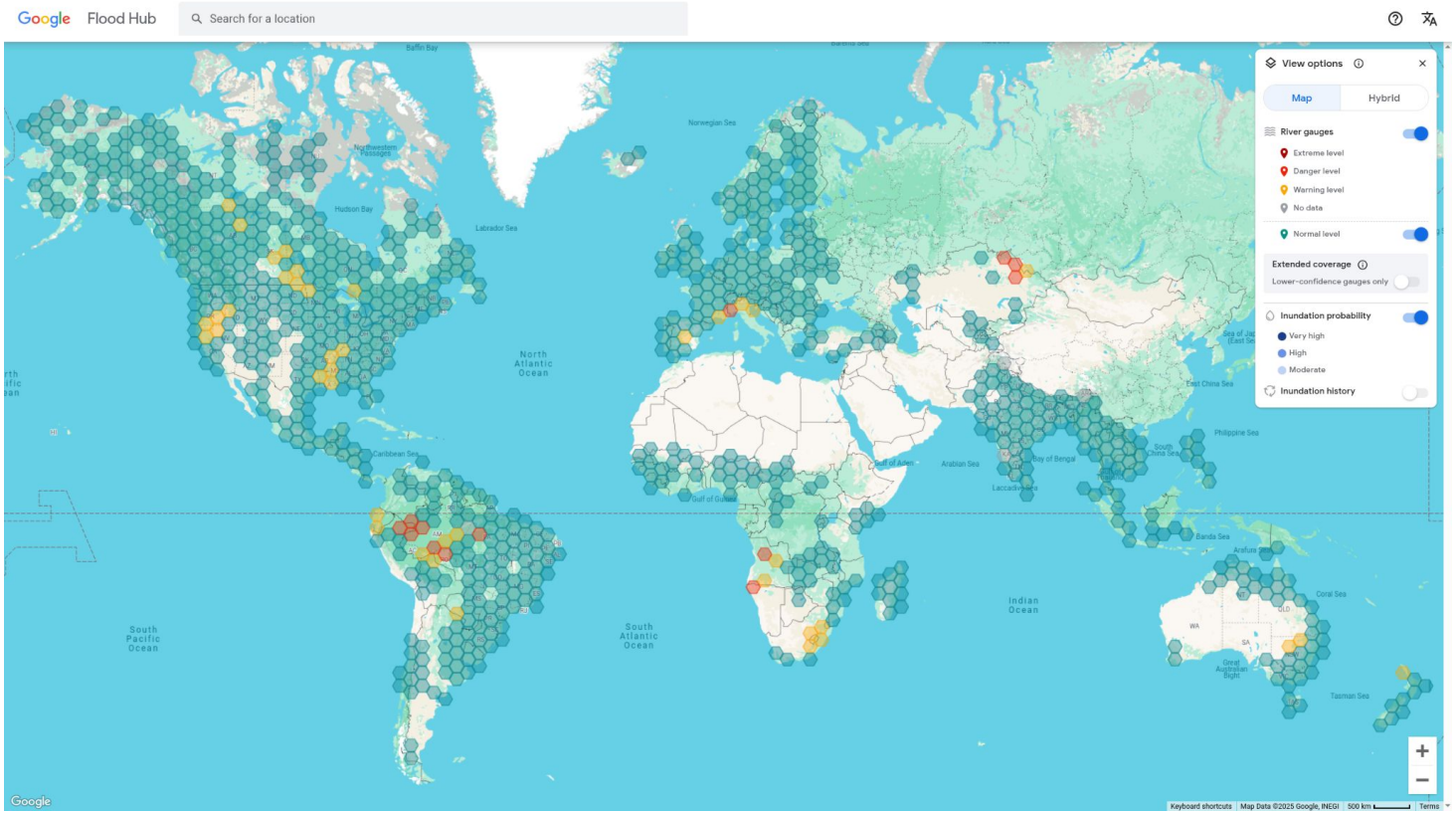


Inundation model



Warning distribution

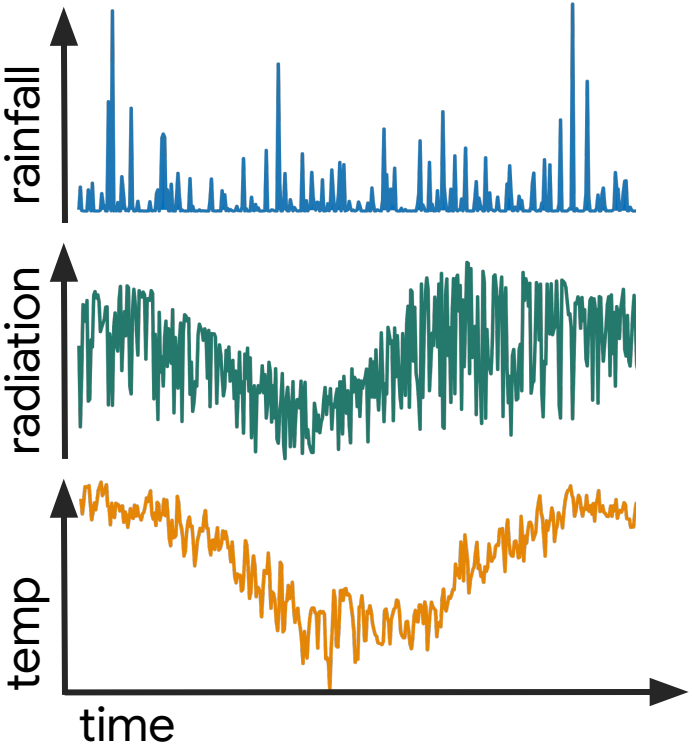
Flood forecasting at Google



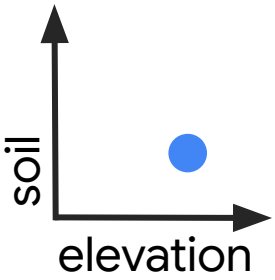
https://g.co/floodhub

Google

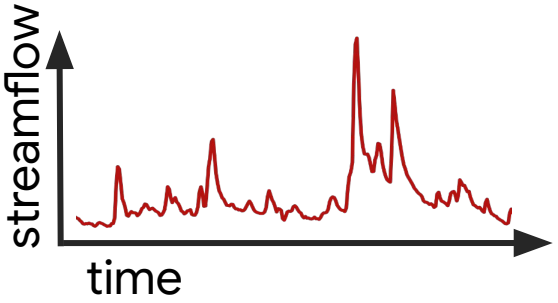
Streamflow prediction



+

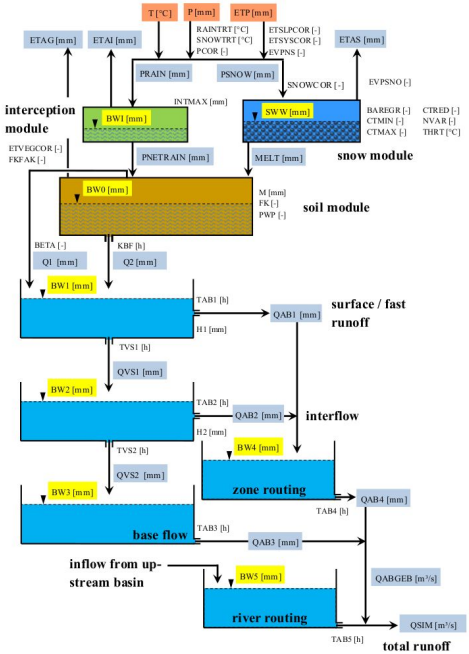


→

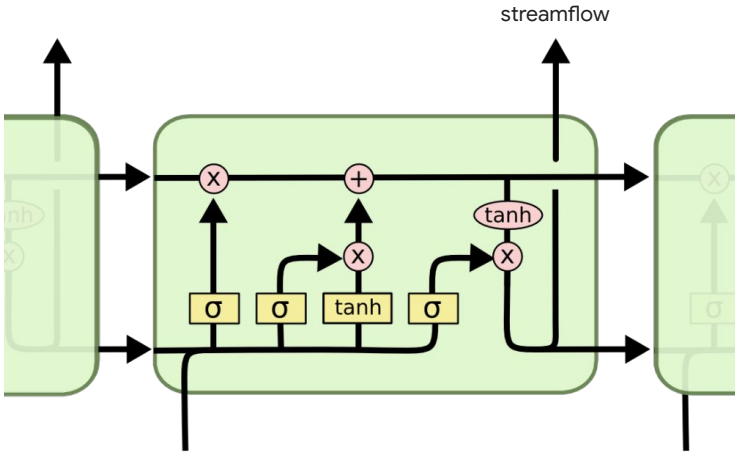


It's not that different!

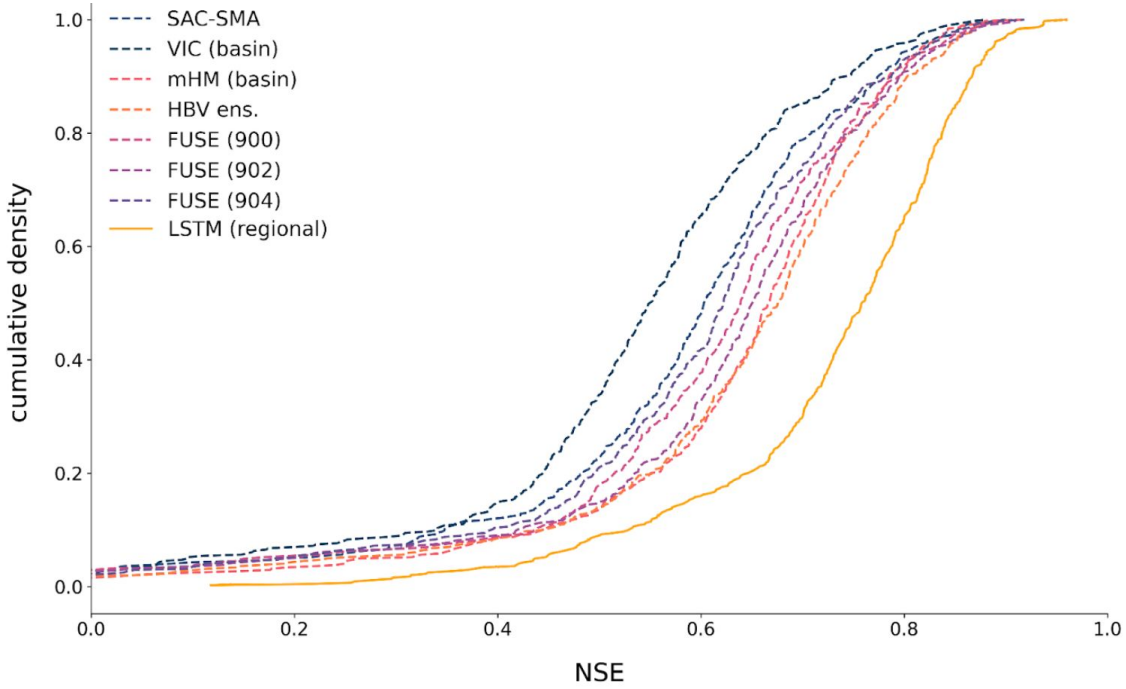
$$\mathbf{S}[t] = f(\mathbf{I}[t], \mathbf{S}[t-1]; \Theta_i)$$
$$\mathbf{O}[t] = g(\mathbf{S}[t]; \Theta_j)$$



$$\{\mathbf{c}[t], \mathbf{h}[t]\} = f(\mathbf{x}[t], \mathbf{c}[t-1], \mathbf{h}[t-1]; \theta_i)$$
$$\widehat{y}[t] = g(\mathbf{h}[t]; \theta_j)$$



It works!



Kratzert et al.: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, Hydrol. Earth Syst. Sci., 2019.

Tricks of the trade

Tricks of the trade

Basics

Missing data

Forecasting

Real-time streamflow

Temporal resolution

Tricks of the trade

Basics

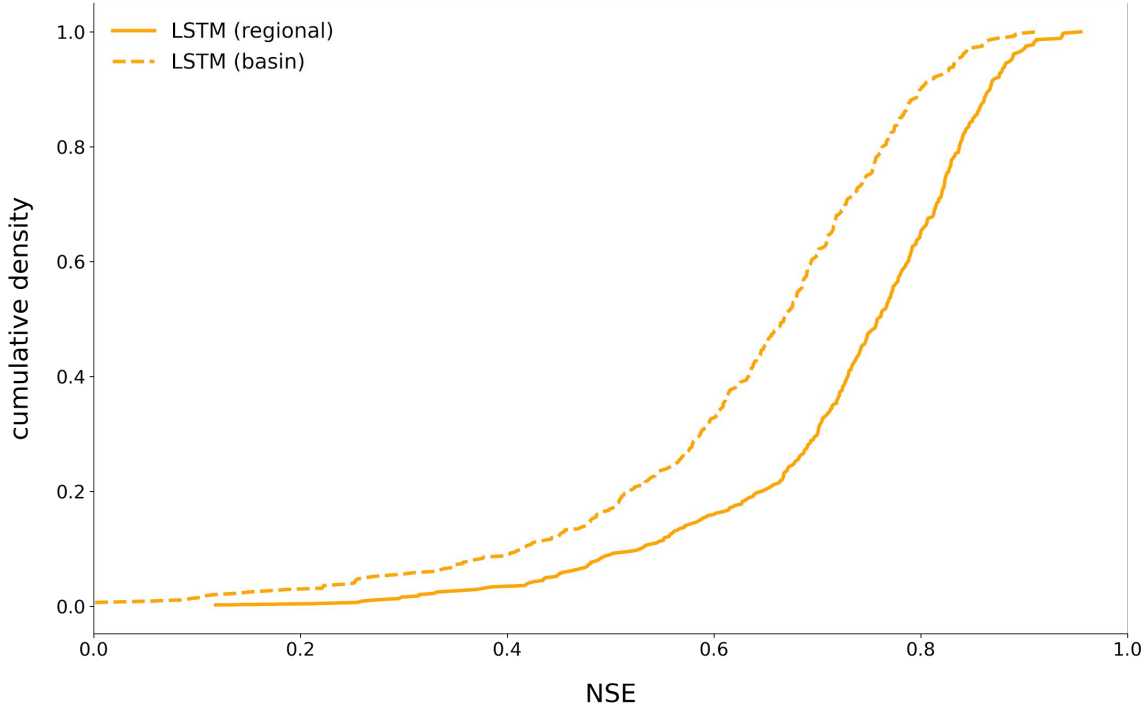
Missing data

Forecasting

Real-time streamflow

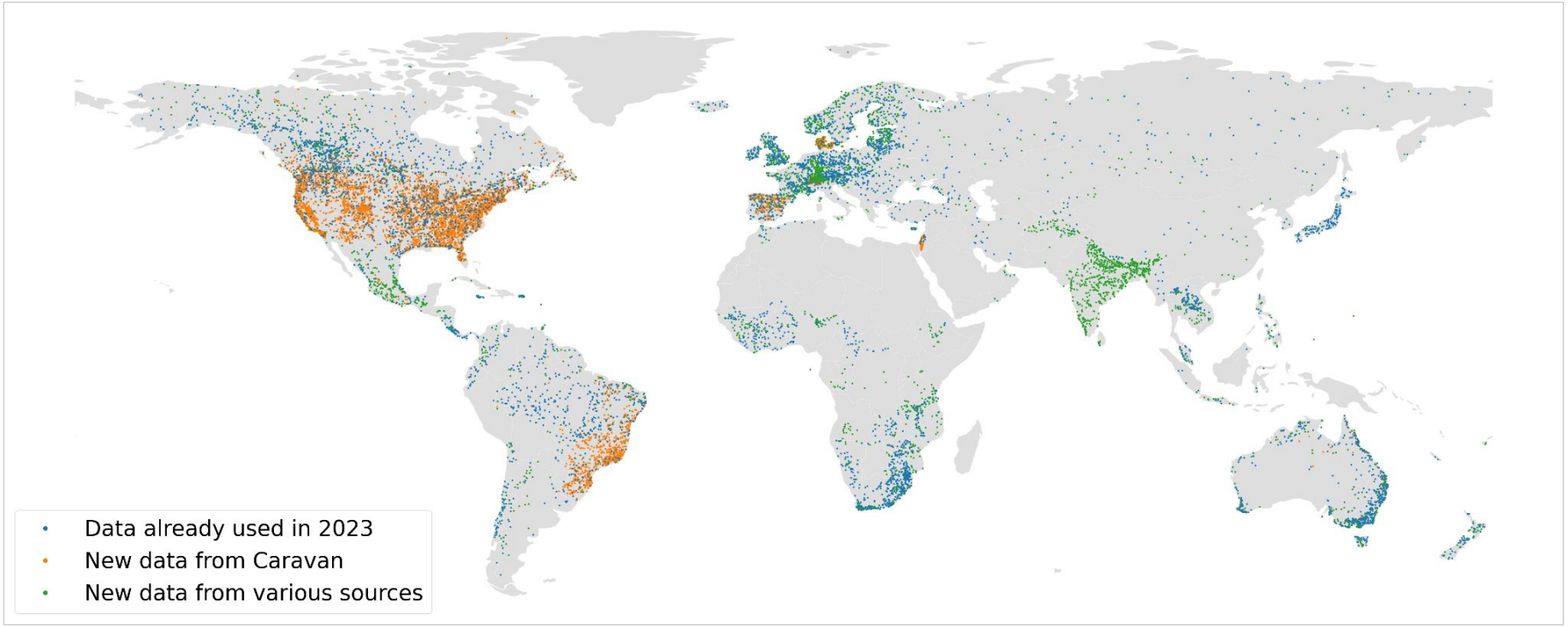
Temporal resolution

More data: more gauges

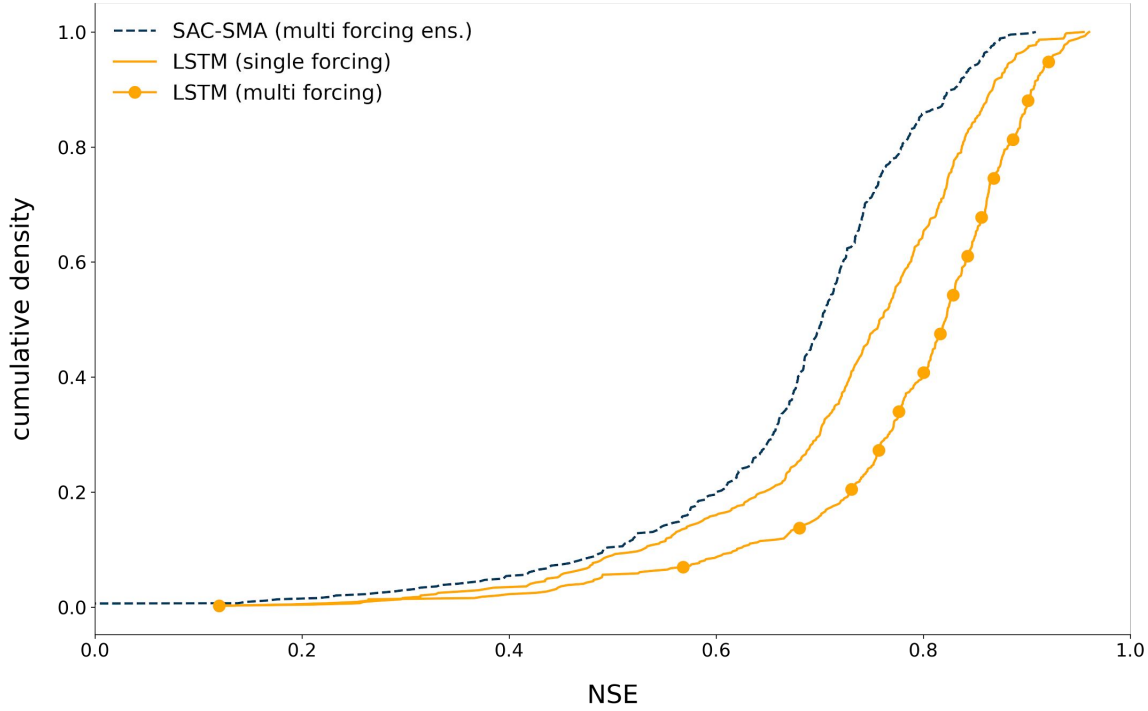


Kratzert et al.: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, Hydrol. Earth Syst. Sci., 2019.

More data: more gauges

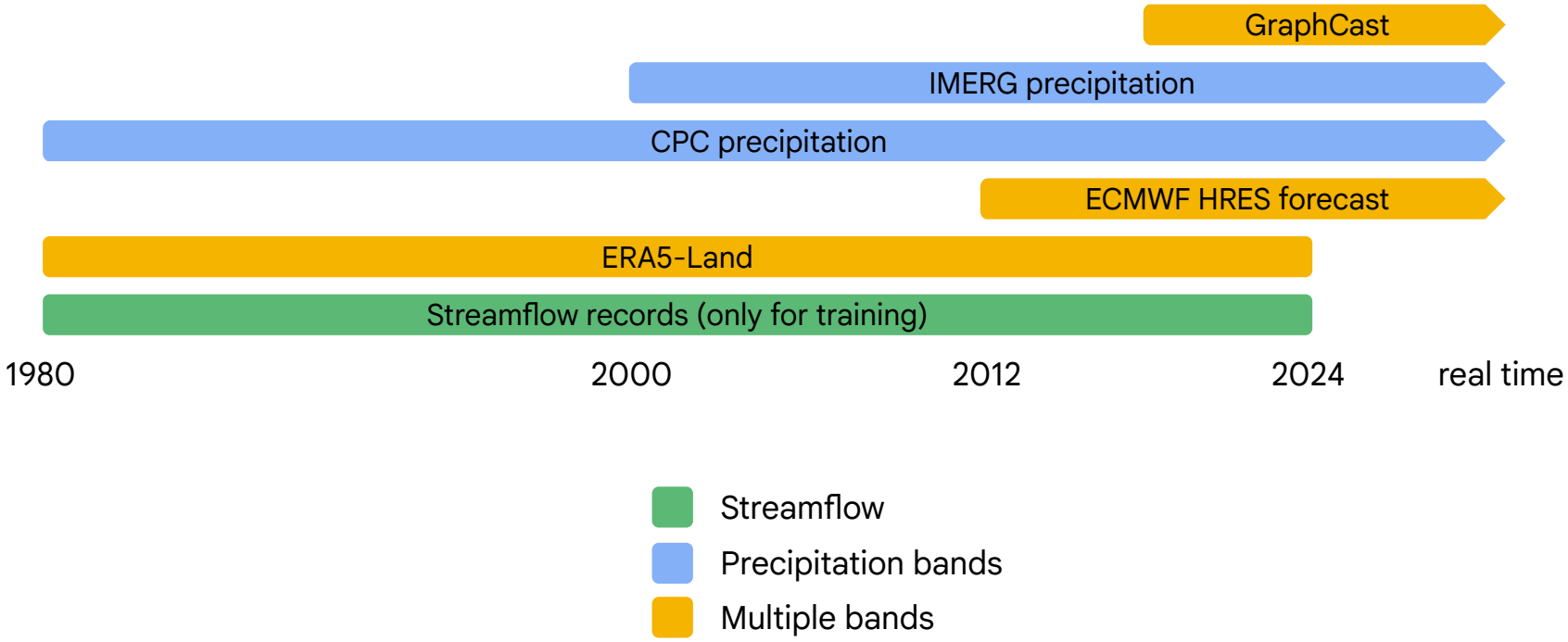


More data: more forcings

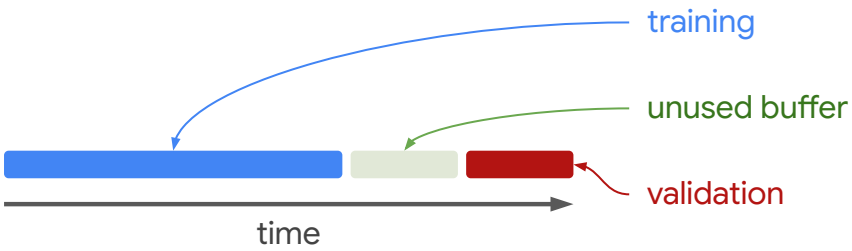


Kratzert et al.: A note on leveraging synergy in multiple meteorological data sets with deep learning for rainfall–runoff modeling, Hydrol. Earth Syst. Sci., 2021.

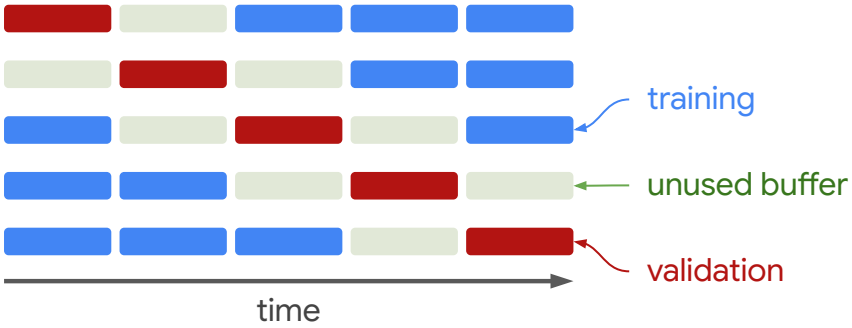
More data: more forcings



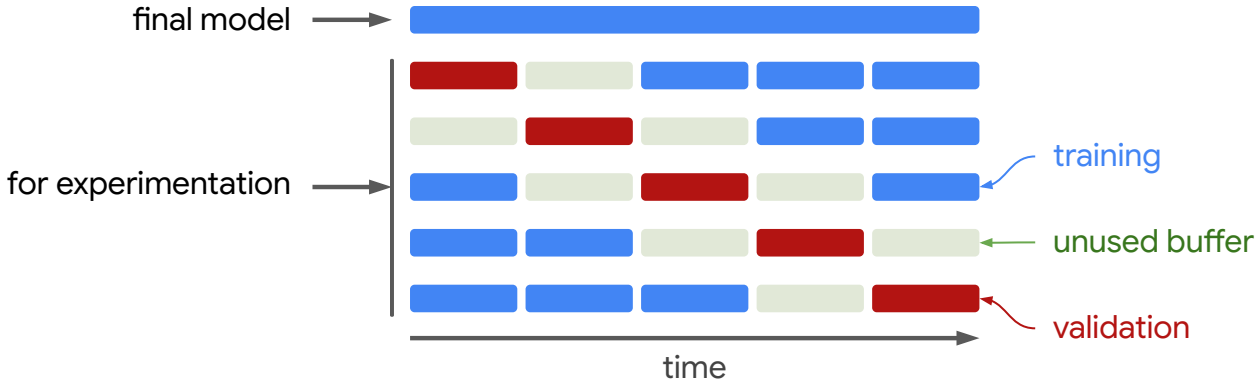
More data: more time



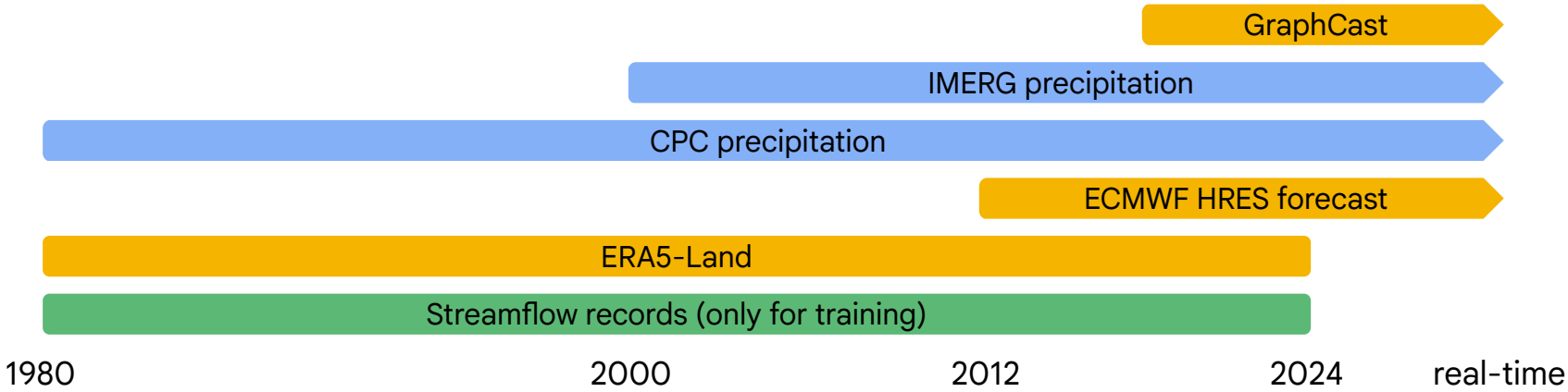
More data: more time



More data: more time



More data: at Google



Streamflow
Precipitation bands
Multiple bands

Tricks of the trade

Basics

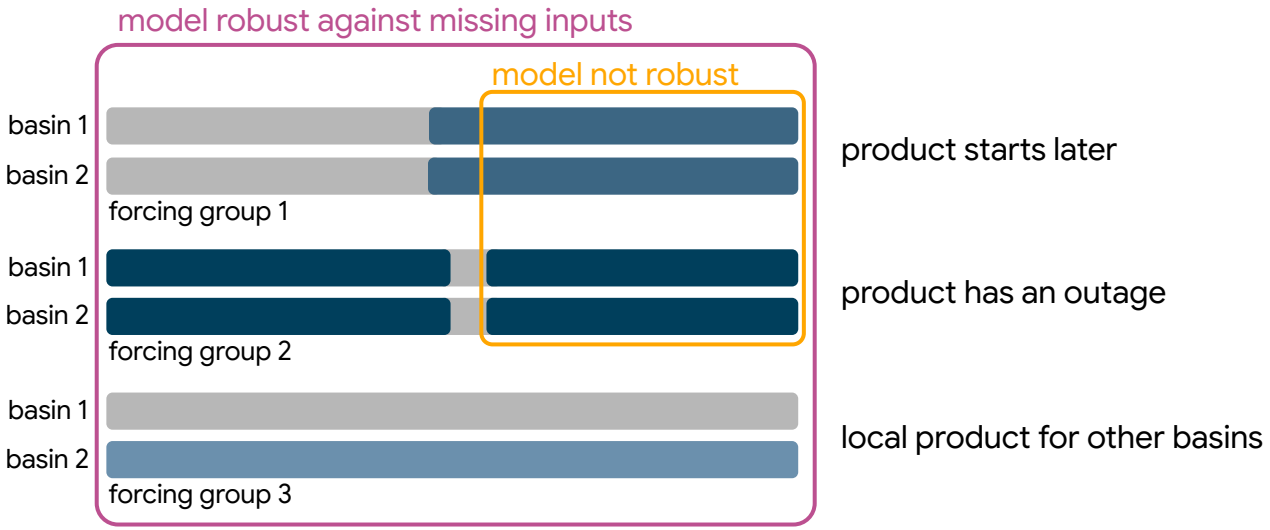
Missing data

Forecasting

Real-time streamflow

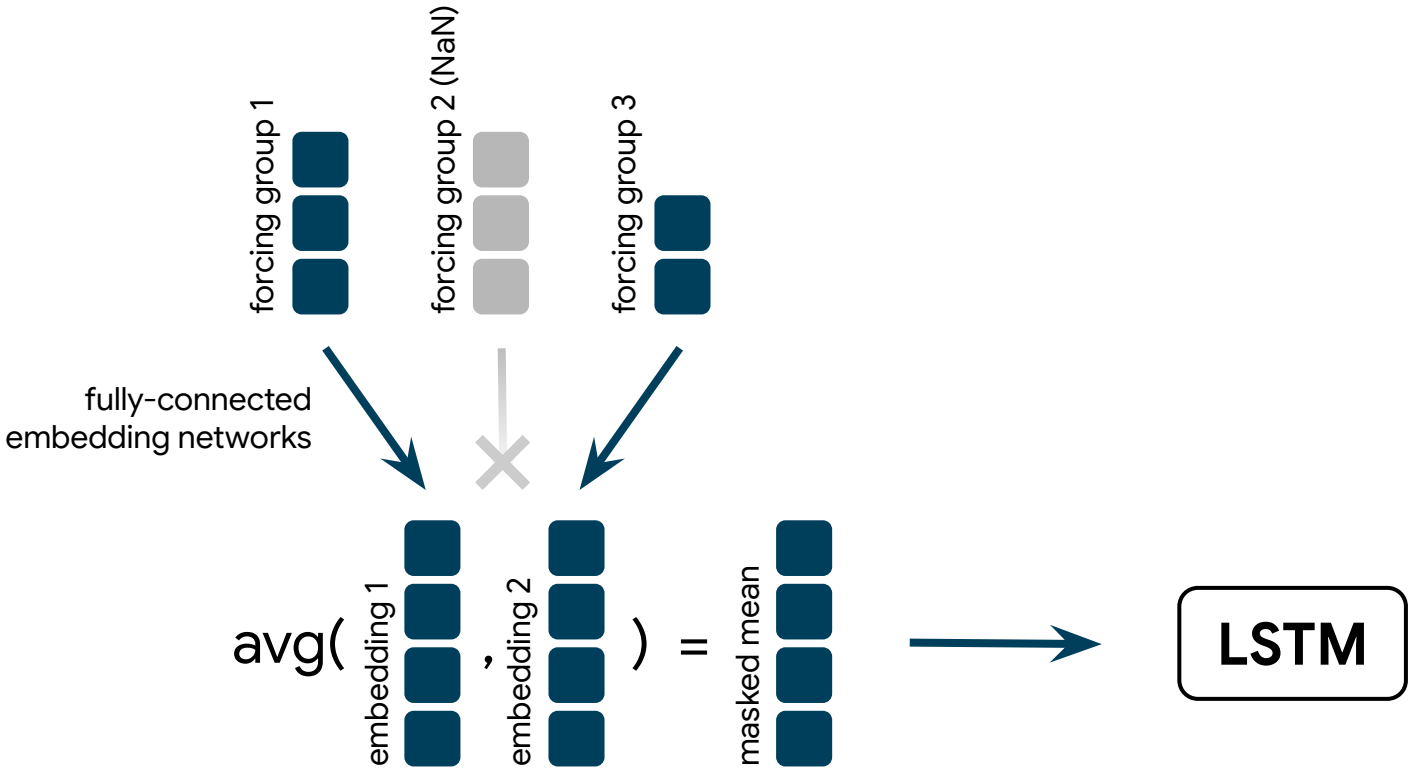
Temporal resolution

Dealing with missing data



Gauch et al.: How to deal w___ missing input data, EGUsphere [preprint], 2025.

Masked mean embedding



Tricks of the trade

Basics

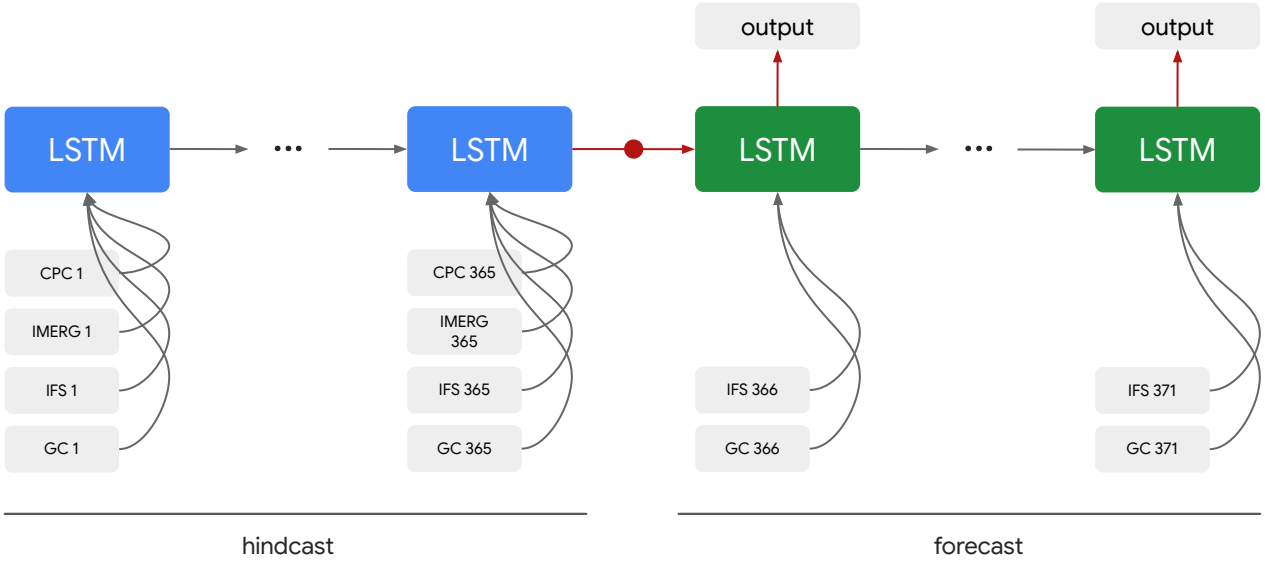
Missing data

Forecasting

Real-time streamflow

Temporal resolution

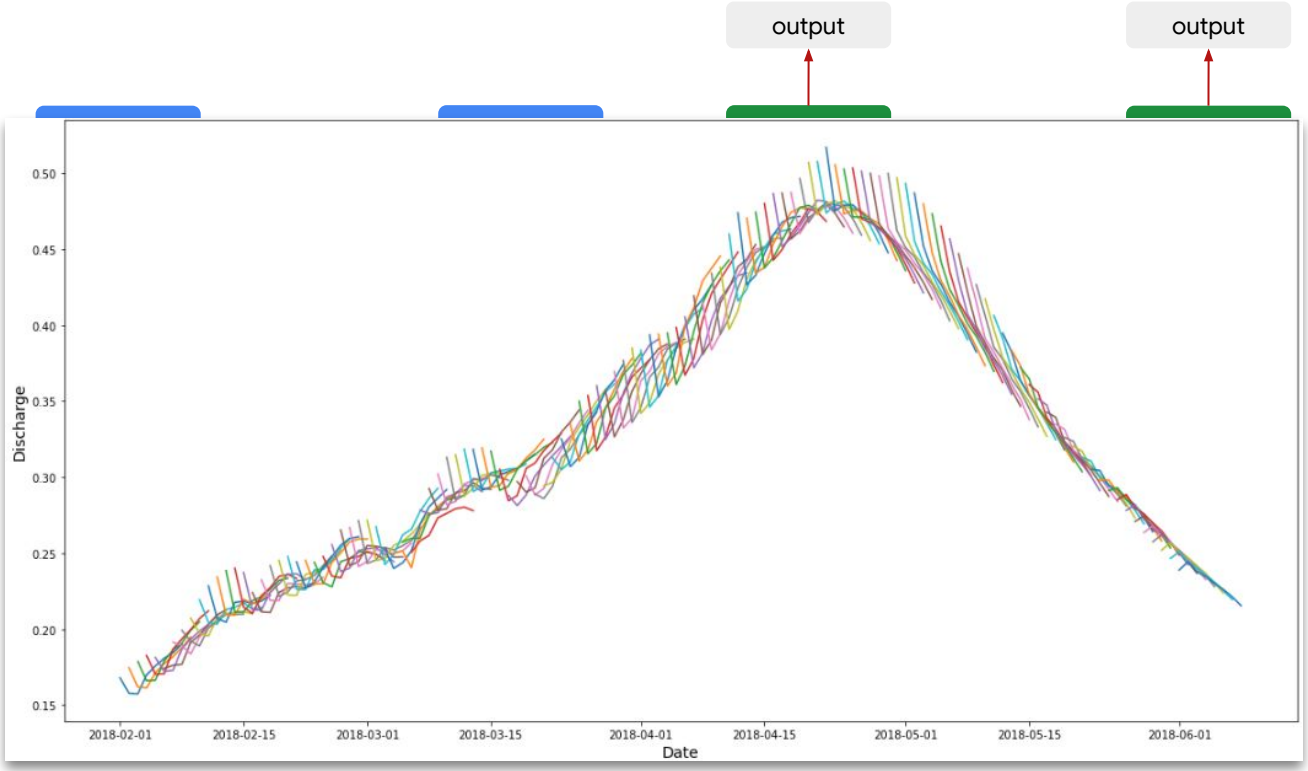
An early deep learning hydro model (~2021)



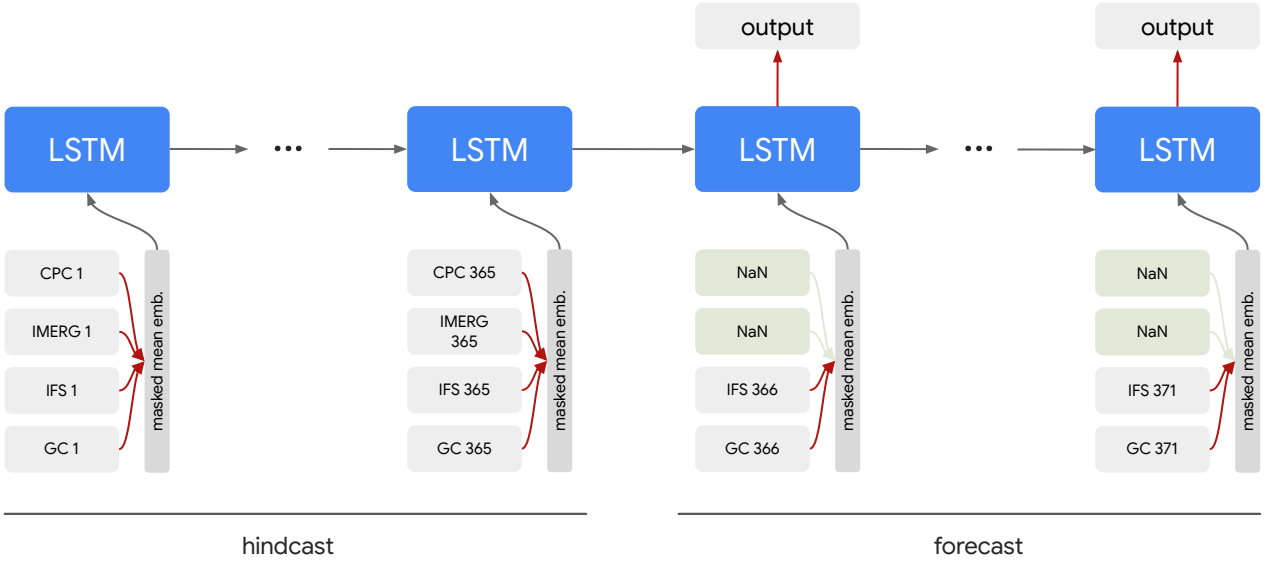
= fully-connected network or linear layer

* visualization is slightly simplified and not showing static inputs

An early deep learning hydro model (~2021)



Google's hydro model today



= fully-connected network or linear layer

* visualization is slightly simplified and not showing static inputs

Tricks of the trade

Basics

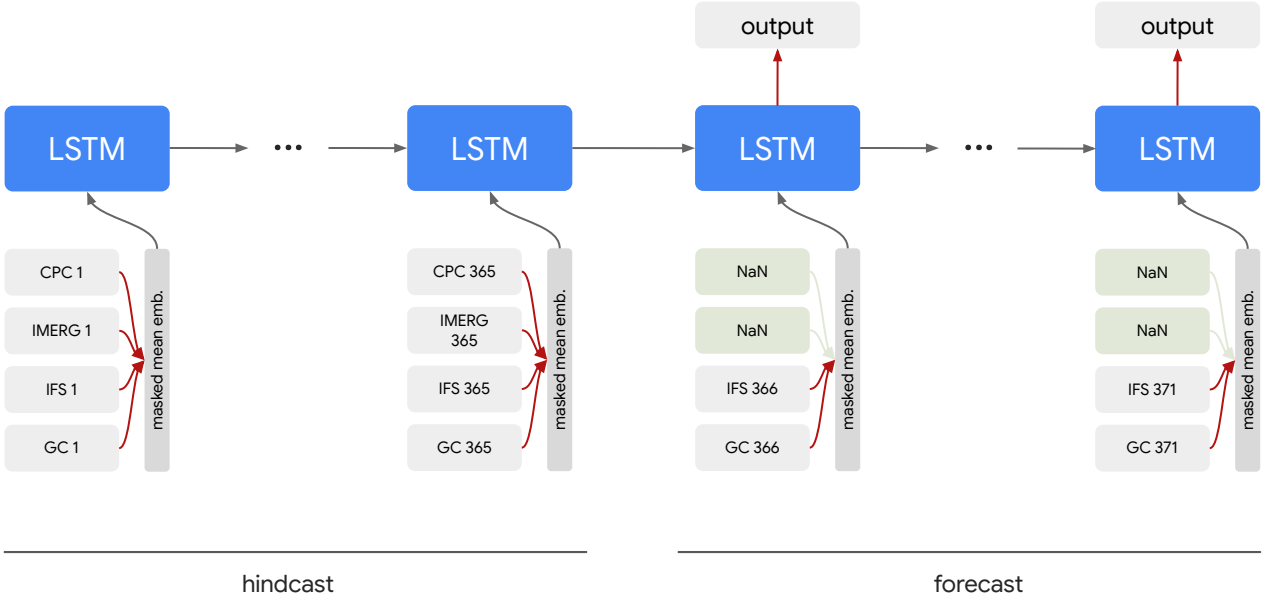
Missing data

Forecasting

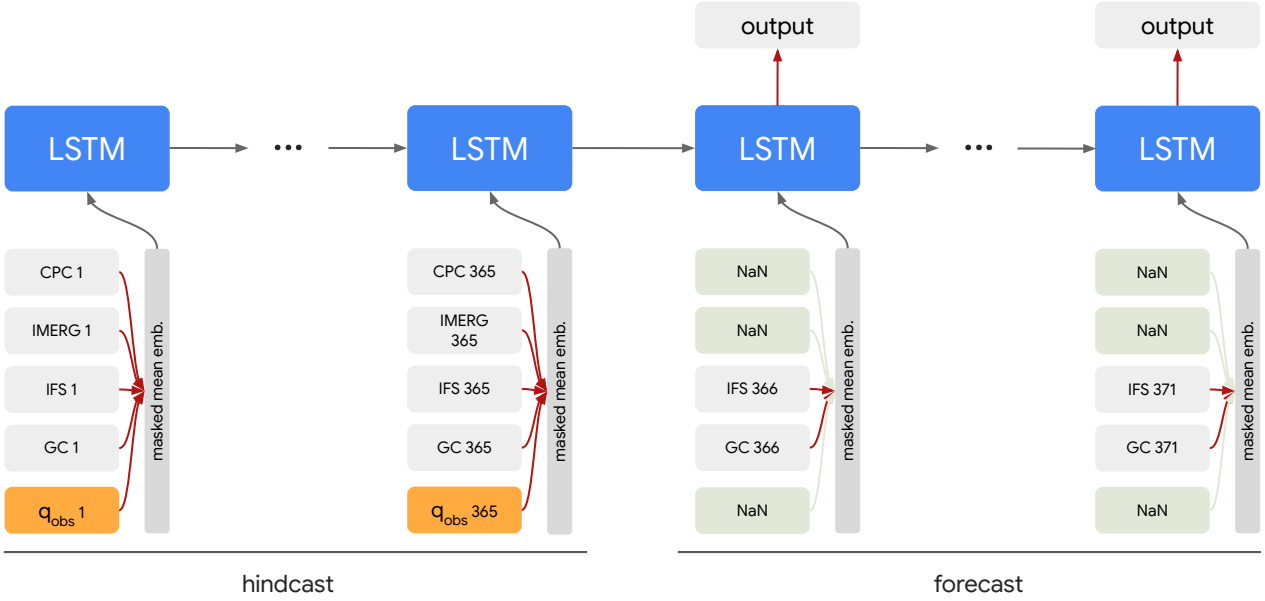
Real-time streamflow

Temporal resolution

Using real-time streamflow



Using real-time streamflow



Tricks of the trade

Basics

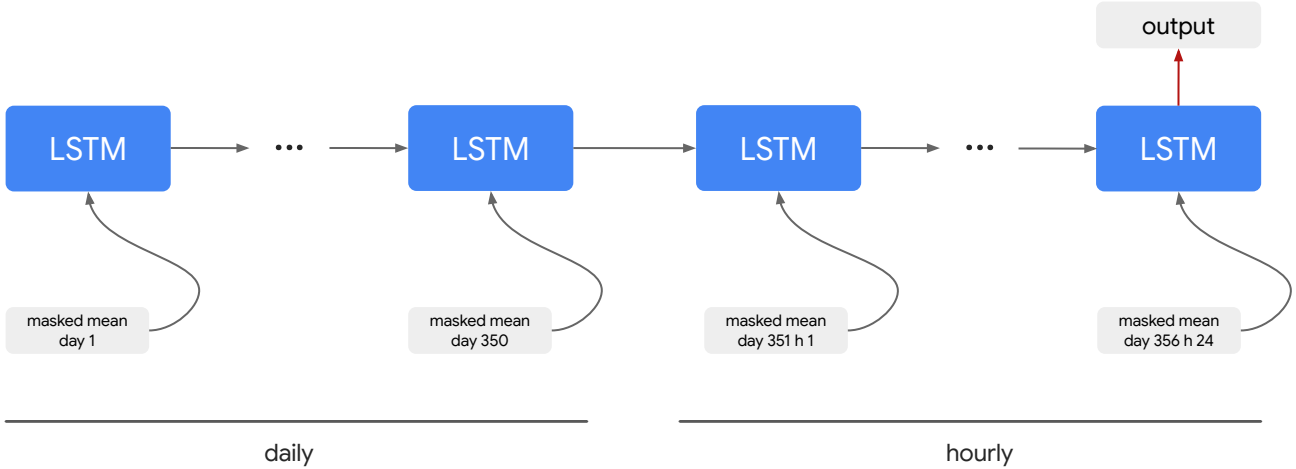
Missing data

Forecasting

Real-time streamflow

Temporal resolution

Efficient hourly predictions: multi-timescale LSTM



Gauch et al.: Rainfall–runoff prediction at multiple timescales with a single Long Short-Term Memory network, Hydrol. Earth Syst. Sci., 2021.
Acuña Espinoza et al.: An approach for handling multiple temporal frequencies with different input dimensions using a single LSTM cell, EGUsphere [preprint], 2024.

Tricks of the trade

Basics:
More data

Missing data:
Masked mean embedding

Forecasting:
Avoiding “hairs”

Real-time streamflow:
Using observations as inputs

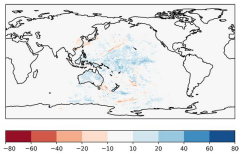
Temporal resolution:
Multi-timescale LSTM

Streamflow prediction

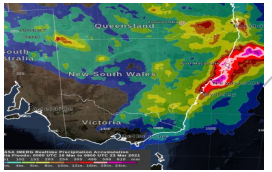
Hydrologic model: Long-Short Term Memory (LSTM)

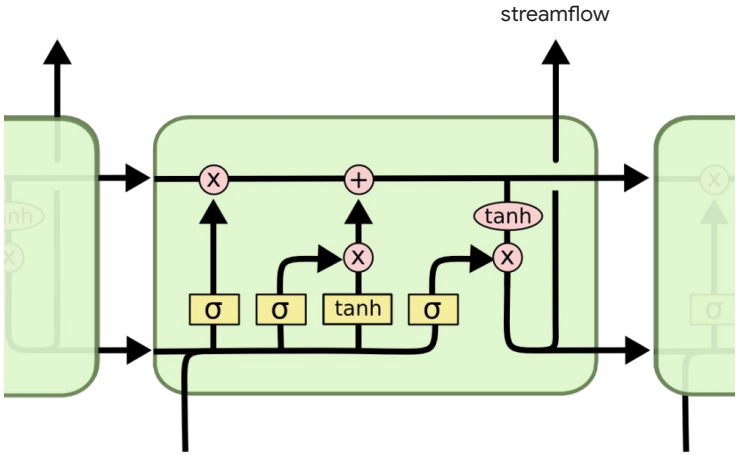
Geophysical catchment characteristics (HydroSheds)

Weather forecasts (ECMWF, GraphCast)

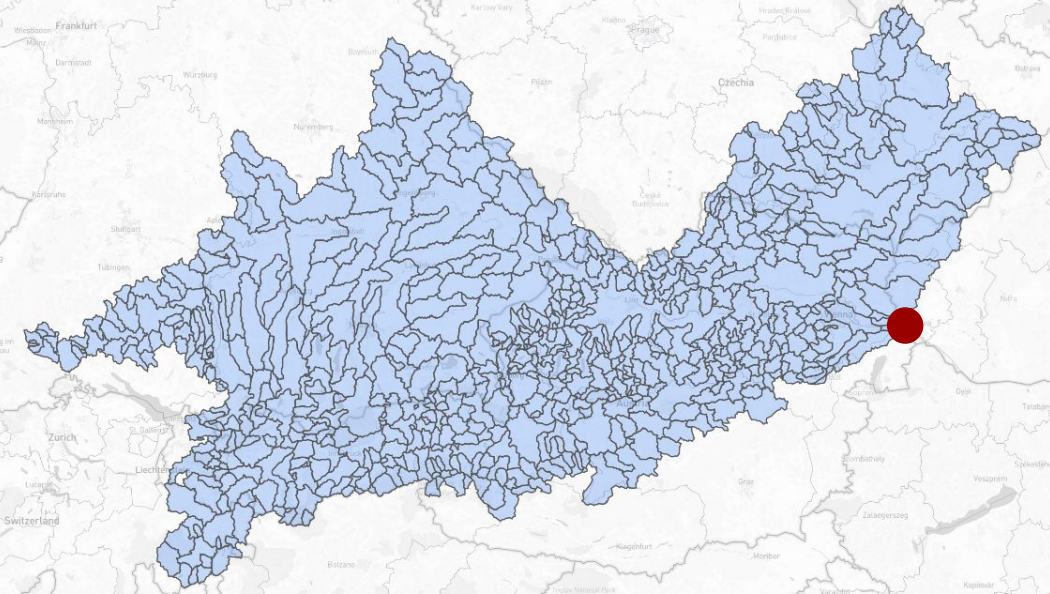


Real-time precipitation estimates (NASA, NOAA)

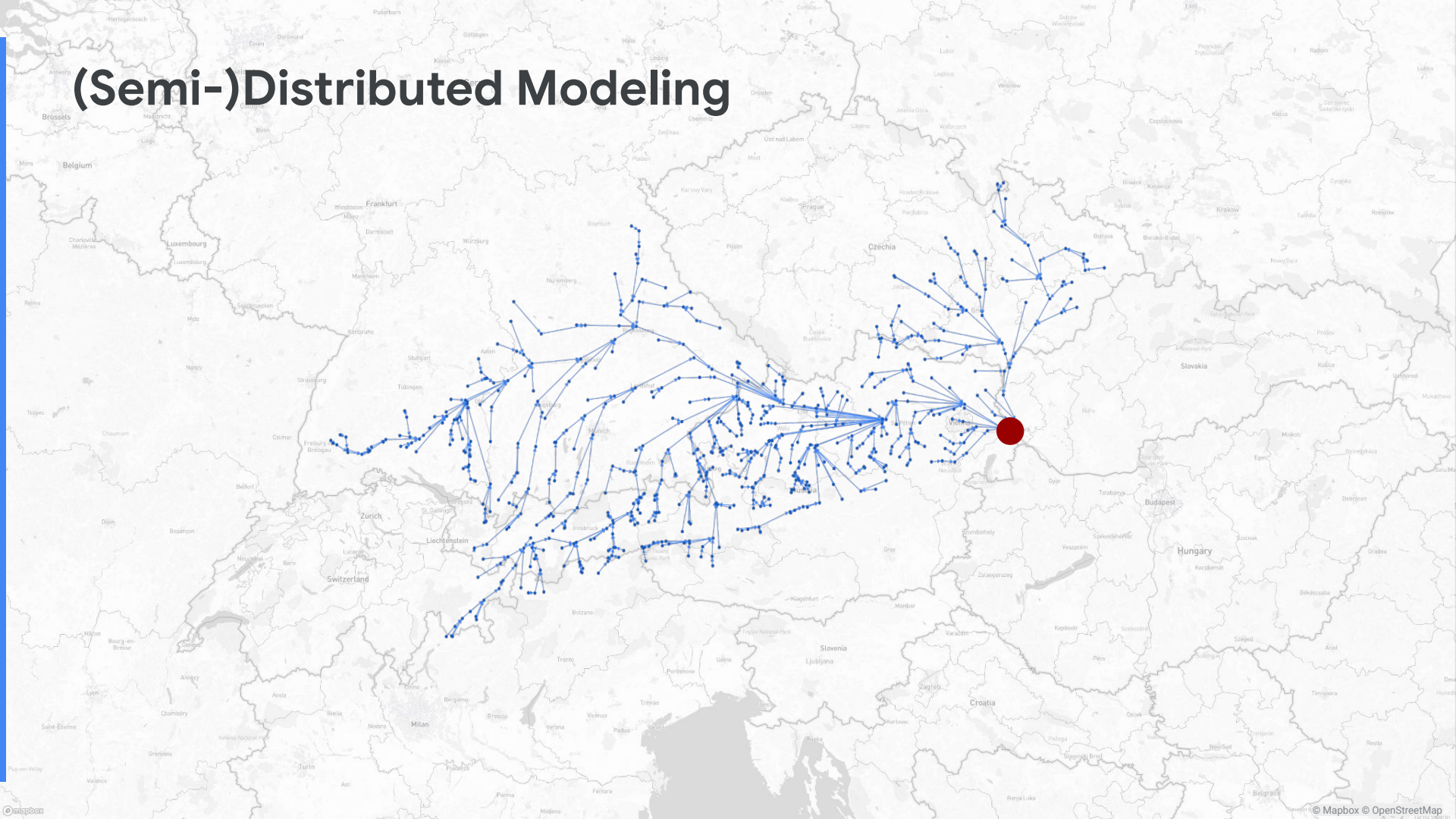




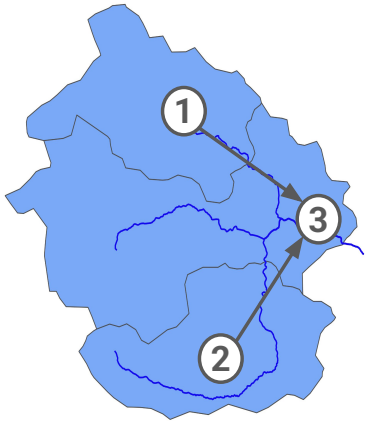
(Semi-)Distributed Modeling

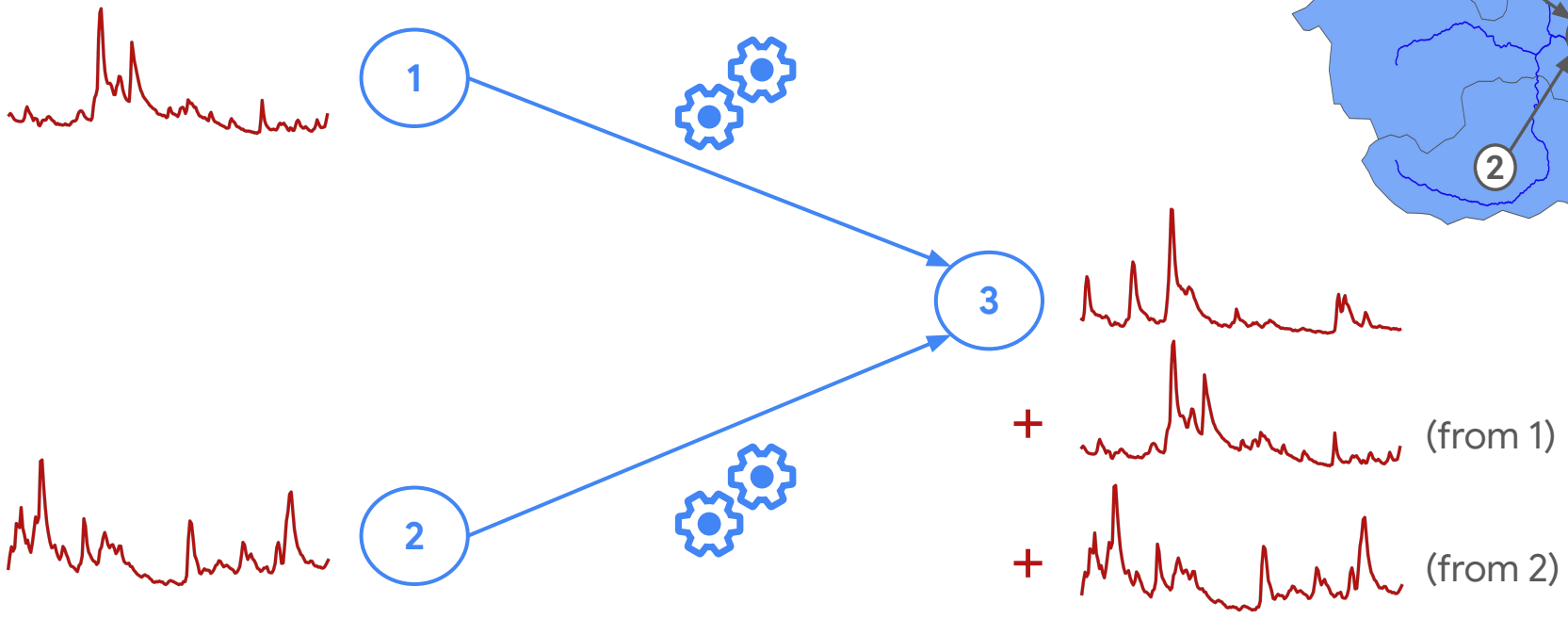


(Semi-)Distributed Modeling



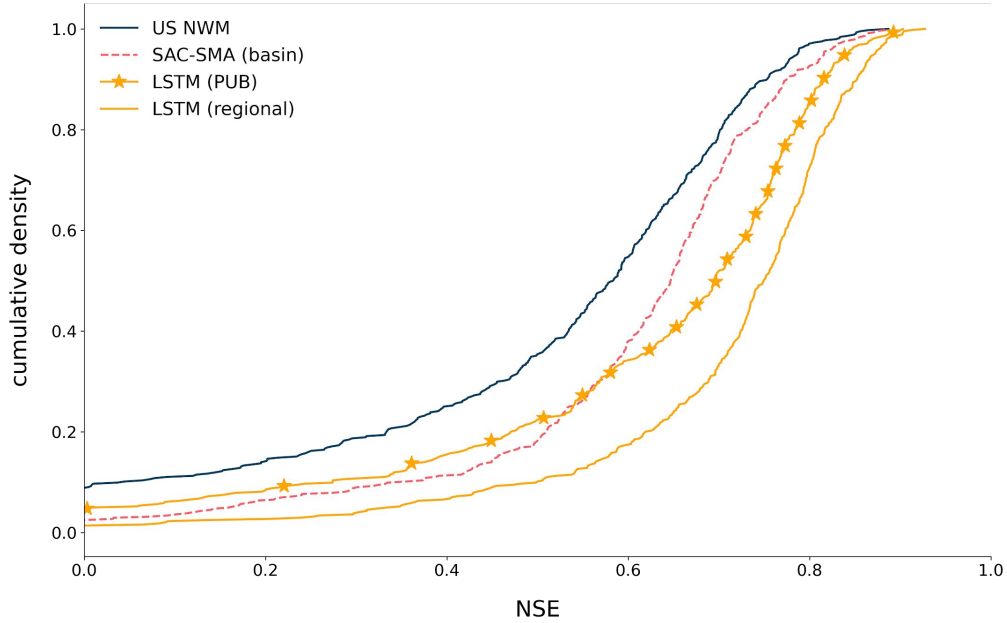
Routing





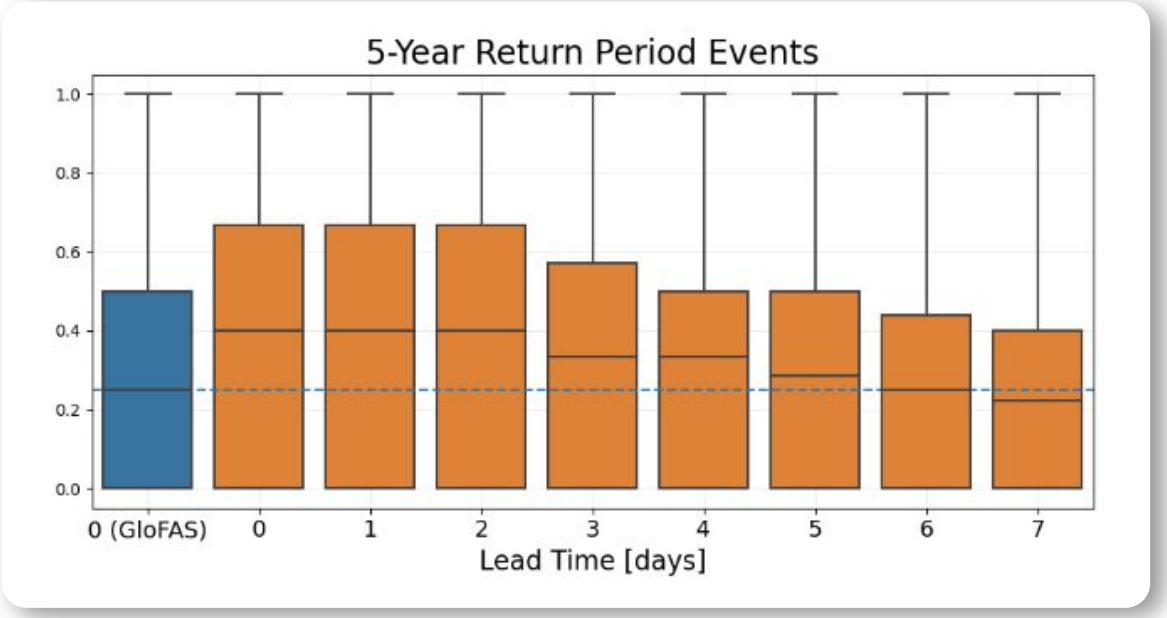
Local model → Edge model → Aggregator model

Prediction in ungauged basins



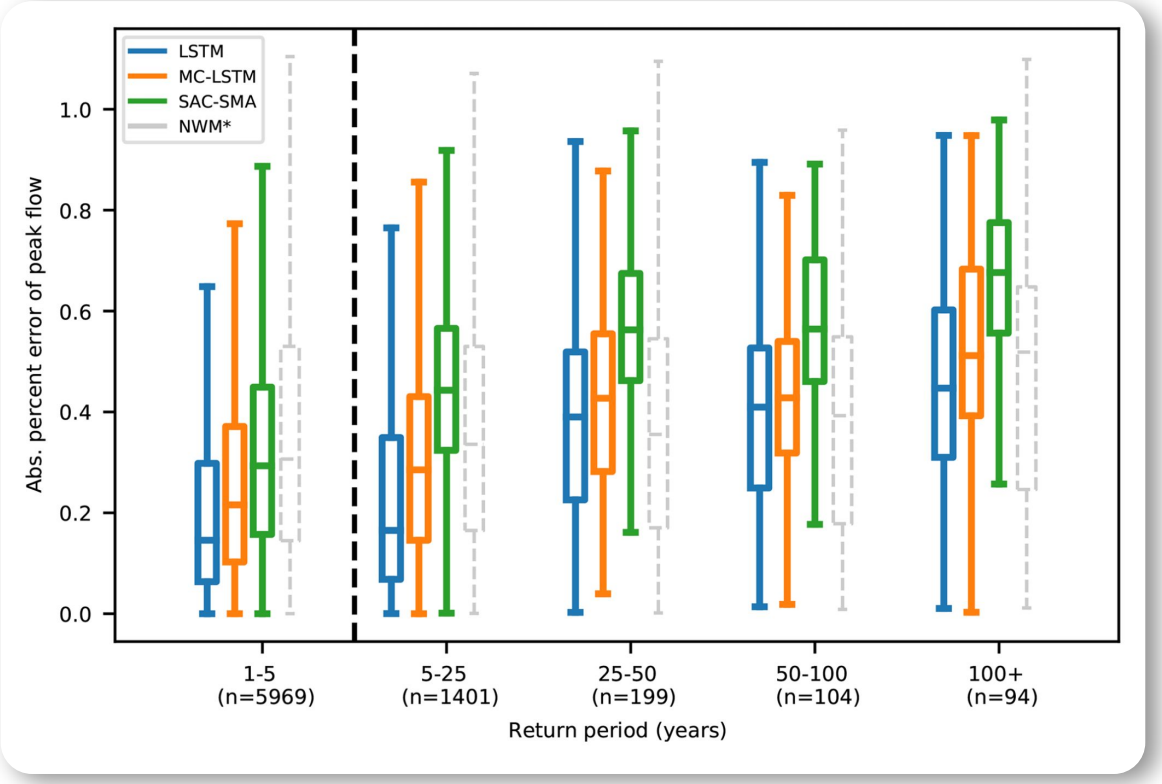
Kratzert et al.: Toward improved predictions in ungauged basins: Exploiting the power of machine learning, Water Resources Research, 2019.

Comparison to the EU's GloFAS



Nearing et al.: Global prediction of extreme floods in ungauged watersheds, Nature, 2024.

Predicting extreme events



Frame et al.: Deep learning rainfall–runoff predictions of extreme events, Hydrol. Syst. Sci., 2022.