UPSTREAM TECH

Improving Data-Driven Flow Forecasting in Large Basins using Machine Learning to Route Flows

David Lambl (david@upstream.tech)

EGU 2023
HS3.3
April 24, 2023

Our experienced and versatile team

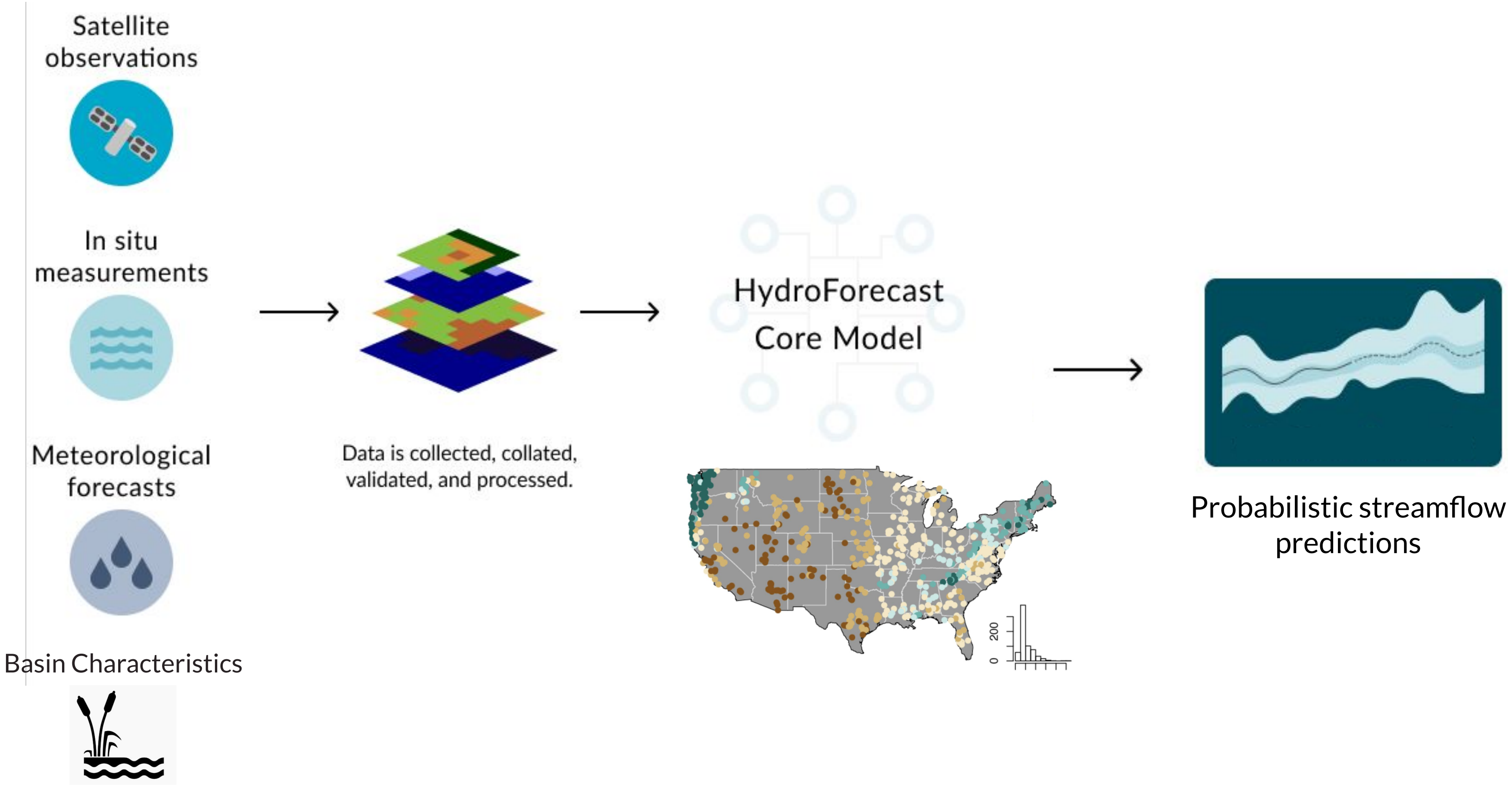
We build software to accelerate the pace, scale, and impact of environmental and climate work across the globe.

We are comprised of experts in software and machine learning, water resource management, and environmental science.

Outline

- HydroForecast: a theory-guided statistical prediction model
- Benefits of distributed models
- Modeling approach to runoff routing

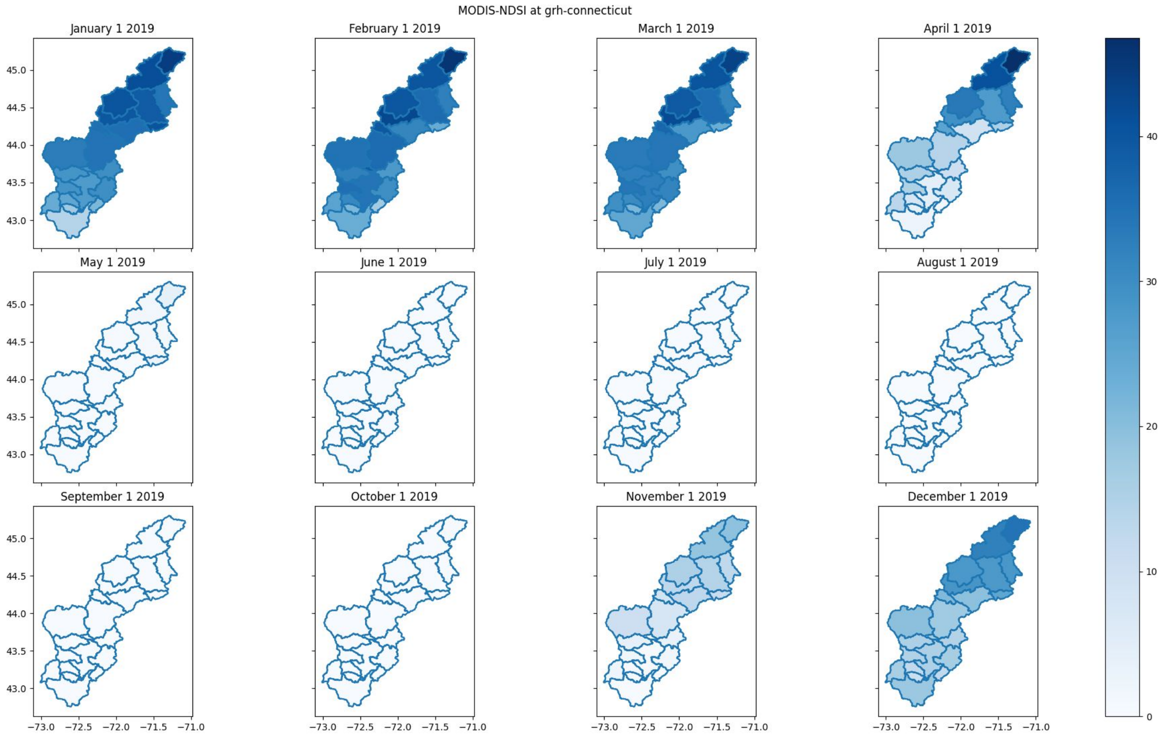
HydroForecast: a theory-guided statistical prediction model



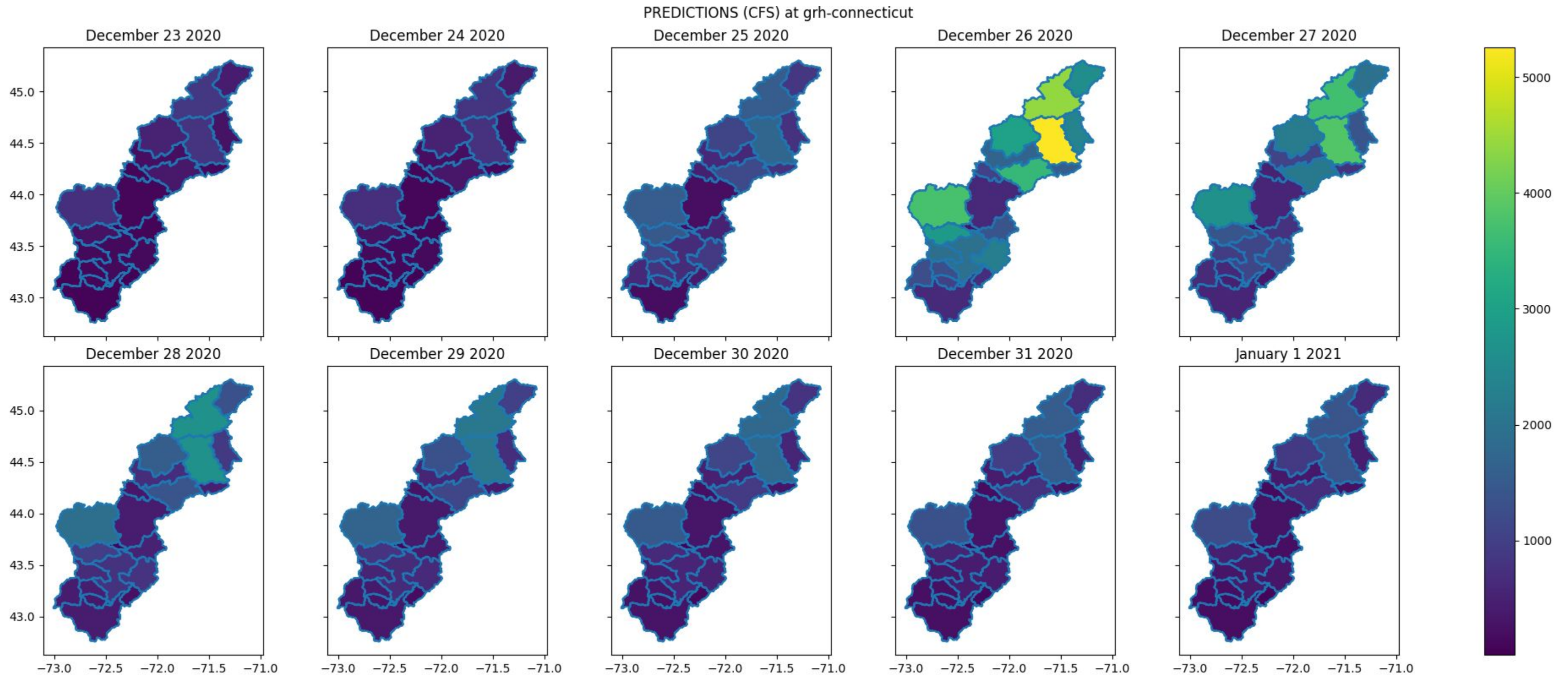
Addressing spatial differences with distributed model

The distributed structure of the model enables HydroForecast to see and react to spatial differences in weather and land surface conditions within a river basin.

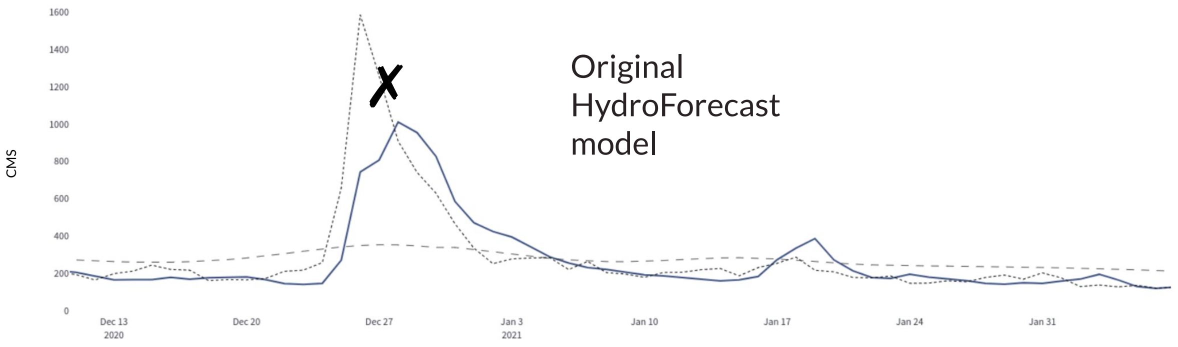
At right, snow cover differences between northern and southern parts of the basin are evident in monthly average Normalized Difference Snow Index (NDSI) data from NASA's MODIS.



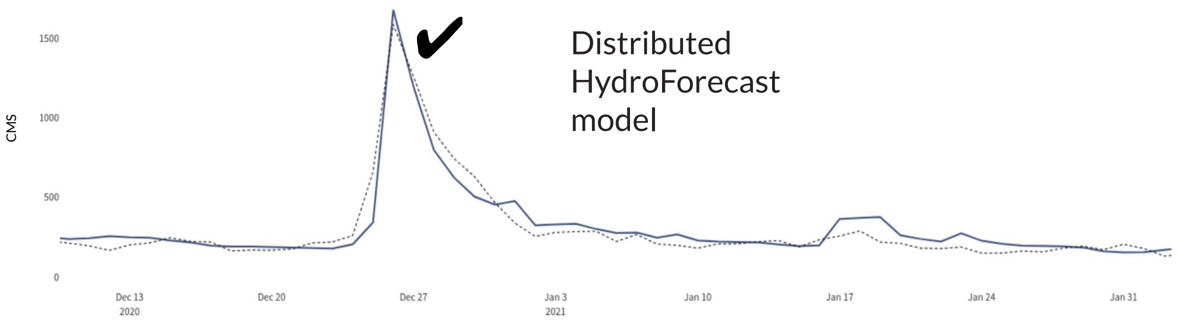
Improvement in hourly predictions in large basins



Improvement in hourly predictions in large basins

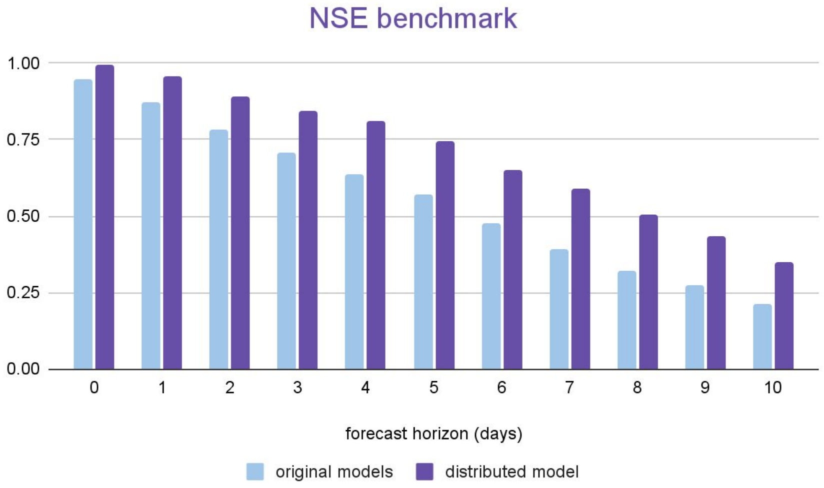


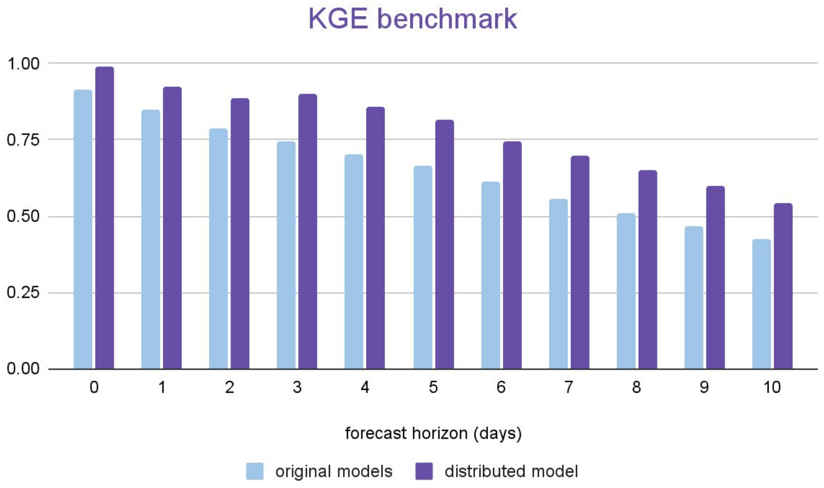
Original HydroForecast model



Distributed HydroForecast model

Improvement in metrics in large basins

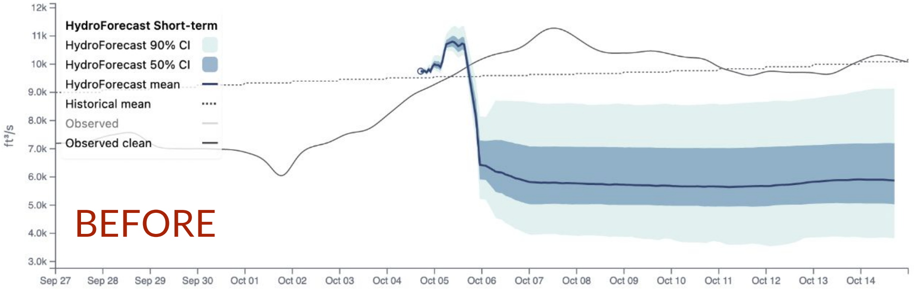


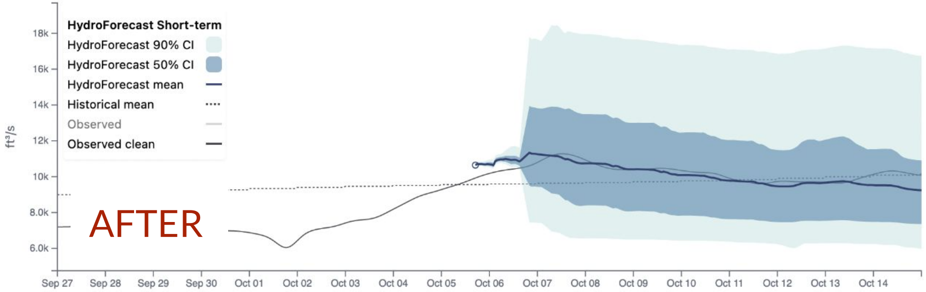


Handling alteration - **dynamically** in operations

1) Assimilating upstream gauges

2) Before (top) and after (bottom) imputing upstream dam release schedule into forecast.



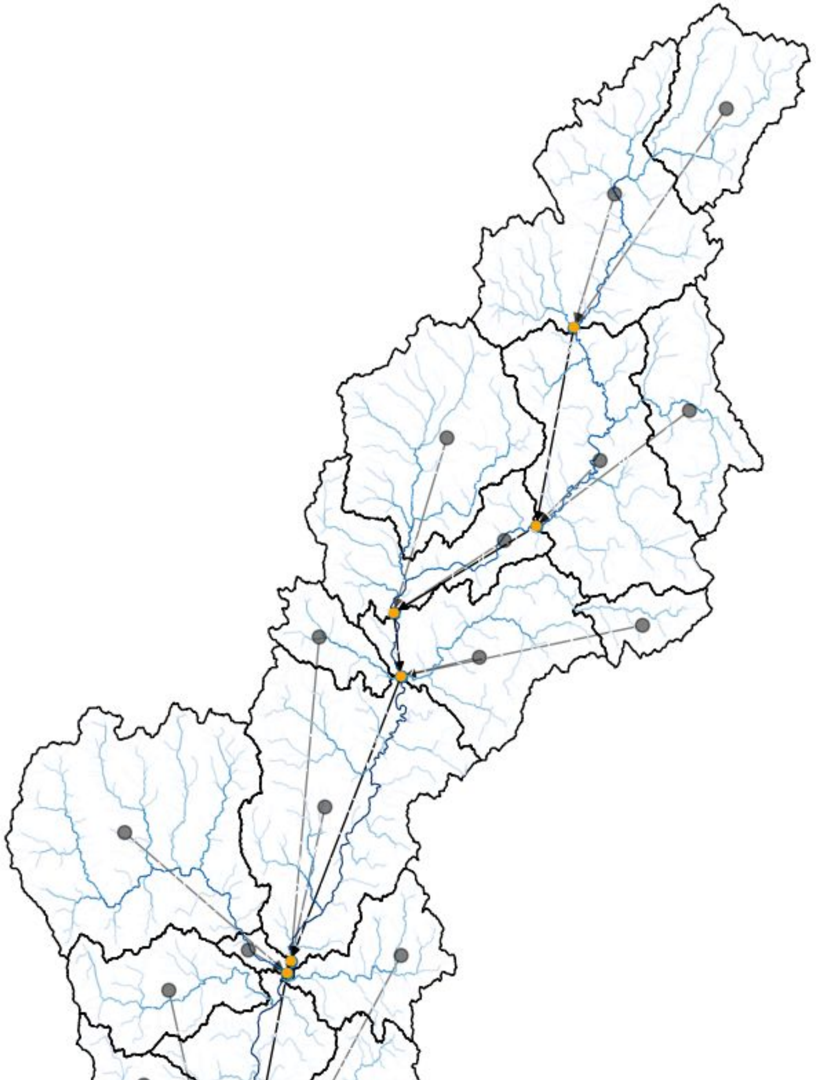


Modeling approach

1. **Delineate sub-basins**, split at gauges and major confluences

2. **Predict runoff** at each sub-basin with HydroForecast lumped model (LSTM)

3. **Route forecasts** through river network

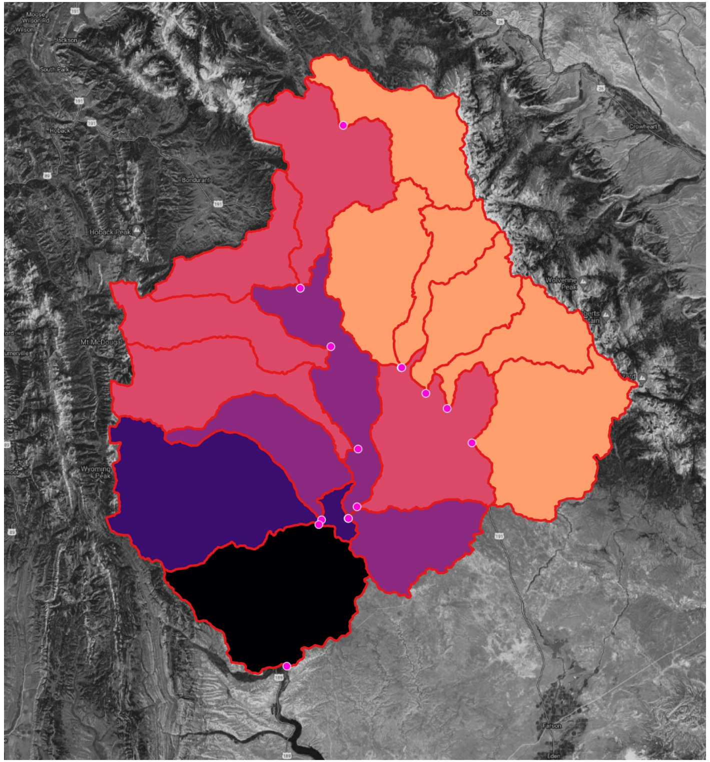


Runoff routing in layers

The model routes flows by “layer” in the river network. All reaches in a layer can be routed independently and in parallel, supporting rapid predictions in large basins. Each layer contains exactly one level of reaches.

Layers are created with a greedy tree algorithm with the trunk at the outlet.

At right, sub-basins are colored lightest to darkest by layer, showing how this basin can be efficiently processed in five layers.

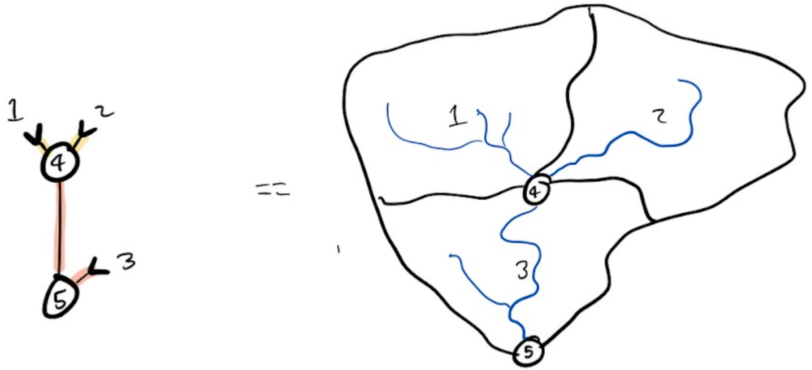


Step of runoff routing

1. **For each layer,** we route the flows down all reaches of that layer using unit hydrographs with mean and std obtained from a trained dense layer.

2. We **add routed flows** arriving at the bottom of each reach at a given time interval and add up reaches at confluences.

3. Those routed flows become the starting flows entering the next downstream layer.



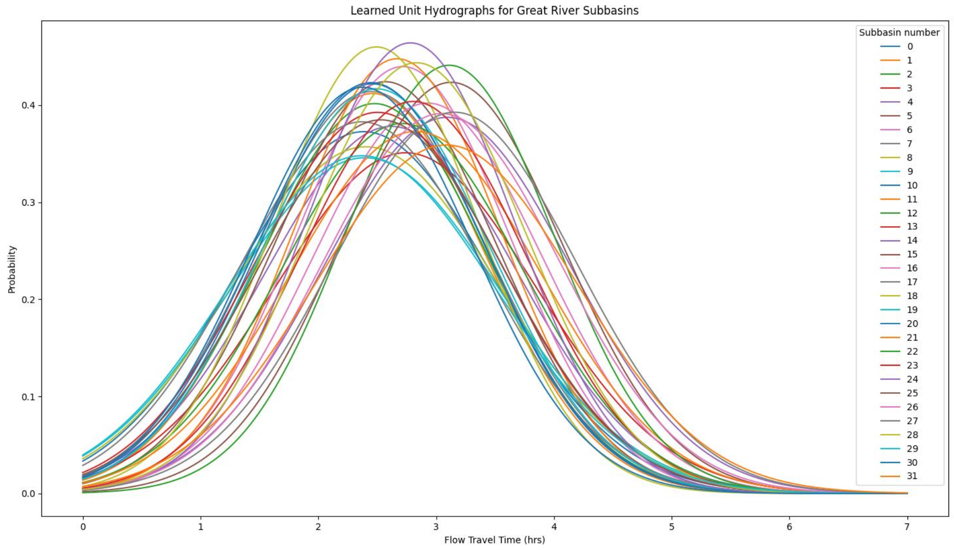
A graph (left) is used to represent basin and reach connectivity (right).

Estimating parameters of unit hydrograph

The model estimates **parameters of a unit hydrograph** at every time step for every reach:

- mean delay time
- delay spread
- volume gain/loss (optional).

The kernel of the dense layer are randomly initialized while biases are chosen to be physically realistic - the model shows significant accuracy improvements even before training of the dense layer.



Predicted unit hydrographs of flow delay at example reaches. Learned travel times can be easily inspected.

Summary of benefits of distributed model

Accuracy	*Model takes advantage of spatial heterogeneity within basin, which is particularly helpful in large basins.*
	Assimilation of upstream gauges and routing them downstream improves short-term accuracy.
Robustness	*Upstream flow modifications can be seamlessly assimilated.*
Scalability	*Model learns jointly runoff and routing times for large basins.*
	Forecasts can be created at internal nodes even without gauge data.
Speed	*Timely operational delivery of hundreds of forecast points thanks to fast matrix multiplications and parallelism.*

Thank you!

david@upstream.tech

UPSTREAM TECH