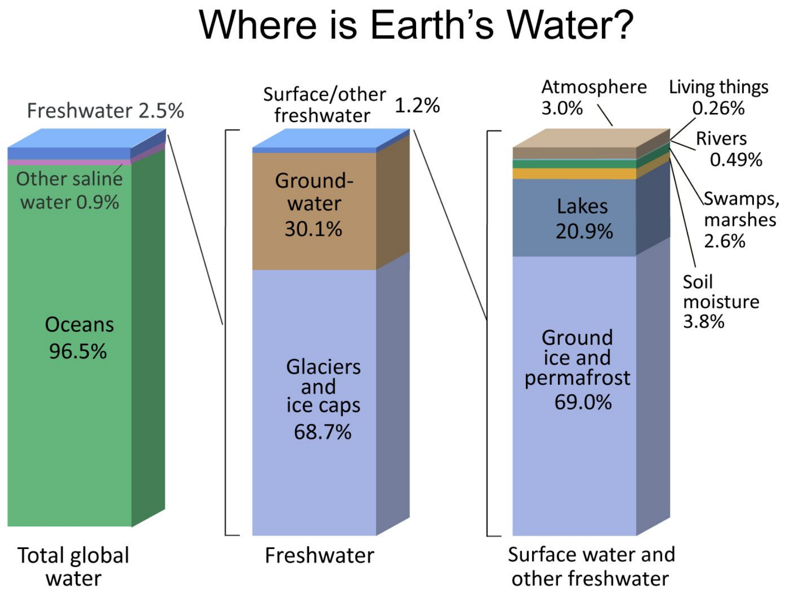
From Drop to Data

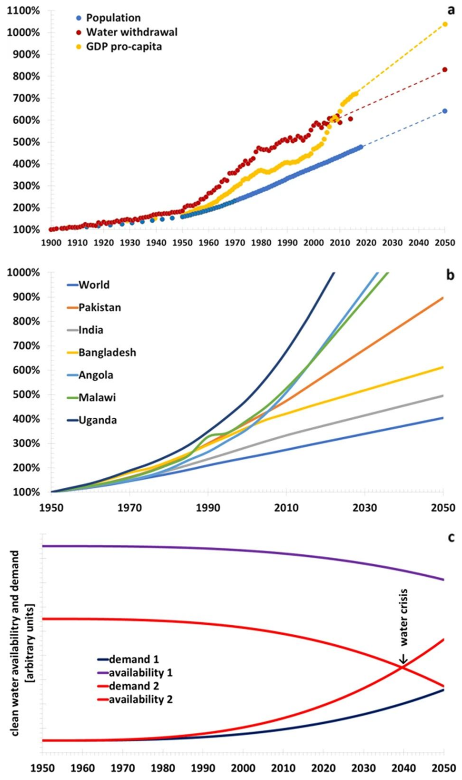
Synergizing AI/ML with Process-Based Understanding in Hydrological Sciences

Udit Bhatia, PhD
Indian Institute of Technology Gandhinagar

The Water Challenge: A problem well-known but not-so-well internalized



Credit: U.S. Geological Survey, Water Science School. https://www.usgs.gov/special-topic/water-science-school
Data source: Igor Shiklomanov's chapter "World fresh water resources" in Peter H. Gleick (editor), 1993, Water in Crisis: A Guide to the World's Fresh Water Resources. (Numbers are rounded).



a) Water withdrawal, GDP pro-capita, and world population. b) The population of the world and selected countries of Asia and Africa. c) Graphical concept of water scarcity, resulting from a more than linear growing demand and a similarly more than a linear reduction of clean water availability. (Boretti and Rosa, 2019)

Water Quantity & Quality: 2 Sides of Same Coin


BORGEN Magazine

THE GOOD NEWS | BUSINESS & NEW MARKETS | POLITICS | CELEB | TECH

YOU ARE AT: Home » Global Health » How Water Pollution in India Kills Millions

How Water Pollution in India Kills Millions

How Water Pollution in India Kills Millions


31°C

Hindustan Times

Education | India | World | Mumbai | Entertainment

HT Premium | Web Stories | Trending | Quiz | Videos | Photos | Tech

Home / India News / Floods, storms may have cost India $7.6bn last year alone: Re...

Floods, storms may have cost India $7.6bn last year alone: Report


INDIA TODAY

SIGN IN

Dark Mode

Premium

Home

News / India / 3,000 tourists stranded as flash floods, landslides hit north Sikkim

Flash floods hit north Sikkim's Pegong; NH10 blocked, around 3,000 tourists stranded

North Sikkim is experiencing flash floods amid heavy rains, with Lachung and Lachen being the most affected areas. With National Highway 10 closed due to landslides, approximately 3,000 tourists are believed to be trapped in these areas.


EDITION IN

THE TIMES OF INDIA

City | Guwahati | Mumbai | Delhi | Bengaluru | Hyderabad | Kolkata | Chennai | Agra | Agartala | Ahmedabad | Ajmer

CIVIC ISSUES | CRIME | POLITICS | SCHOOL AND COLLEGES | ASSAM ELECTIONS | WEATHER | EVENTS

NEWS / CITY NEWS / GUWAHATI NEWS / 34,000 Affected In Assam Floods, Lakhimpur Worst Hit Among 7 Di...

TRENDING | Cyclone in Gujarat | Gujarat Cyclone | West Bengal Panchayat Polls | Rajasthan Rains | Cyclone Biparjoy

34,000 affected in Assam floods, Lakhimpur worst hit among 7 dists

The 17 Sustainable Development Goals (SDGs): But where to start?

How dependent are the 17 SDGs on water sustainability?

THE DAVOS AGENDA 2021

If you want to make progress on all the major global challenges, start with water

Jan 28, 2021

Madeleine Bell

Strategy & Special Projects,

Desolenator

WORLD ECONOMIC FORUM

Clean water underpins the success or failure of every other challenge that we face.

25–29 January 2021

The Davos Agenda

The 17 SDGs: Clean water is the first step

The Sustainable Development Goals are all dependent, in one way or another, on clean water.

Madeleine Bell

Strategy & Special Projects,

Desolenator

25–29 January 2021

The Davos Agenda

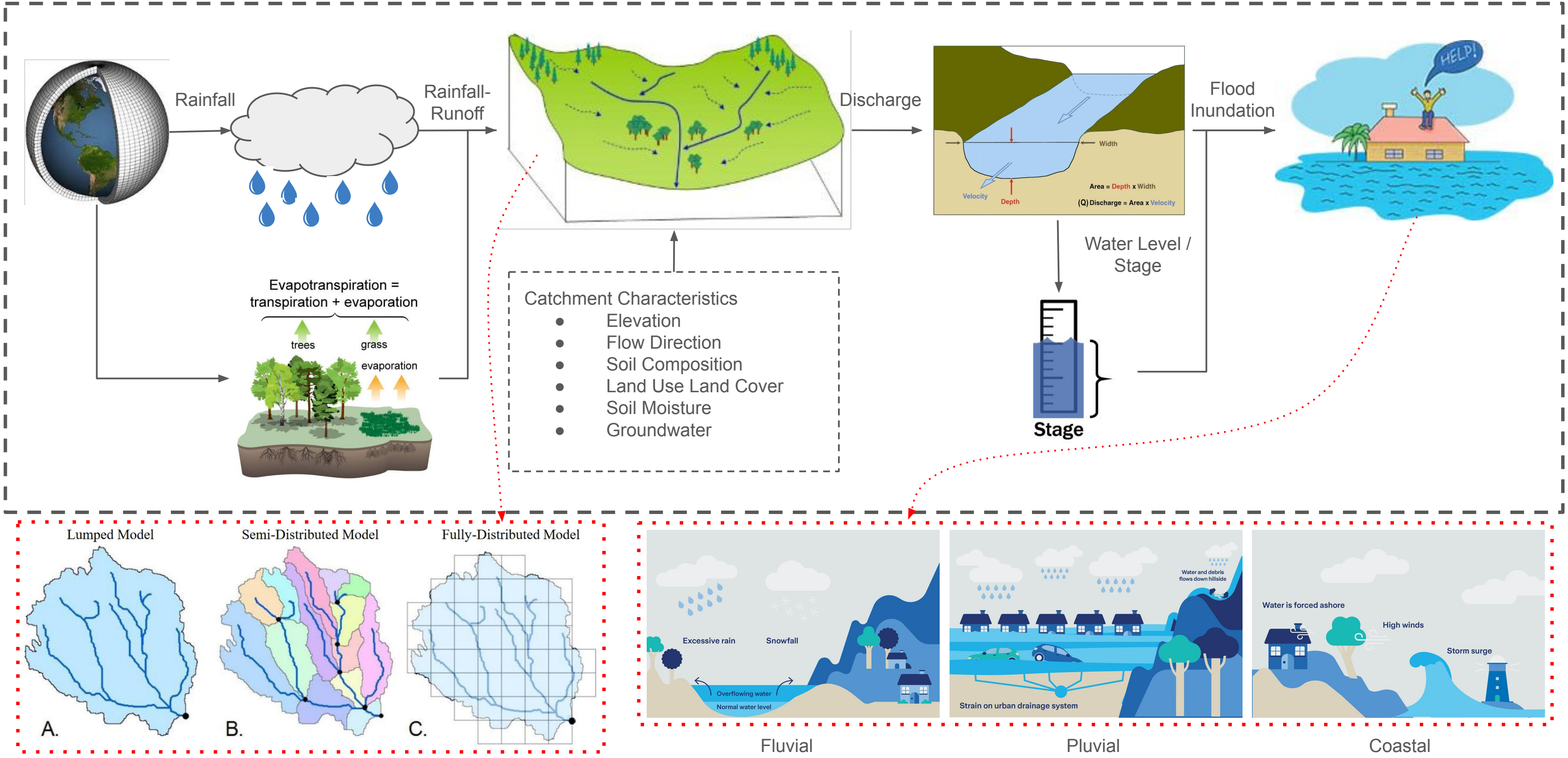
WORLD ECONOMIC FORUM

Environmental Matrix and Water

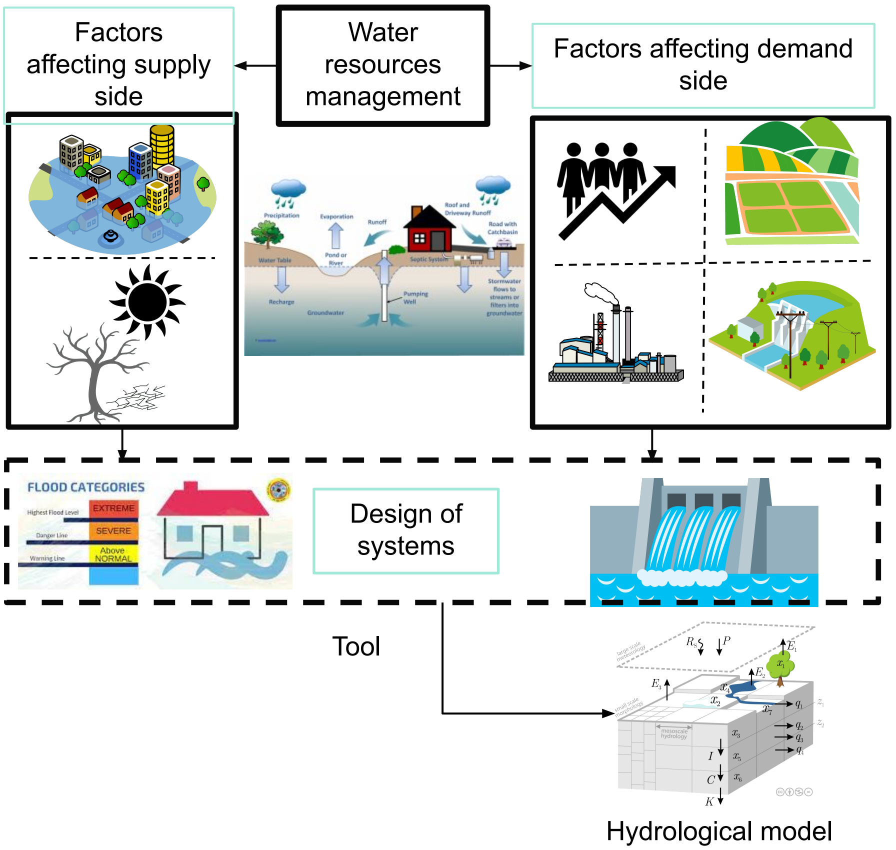
The term aqueous environmental matrix encompasses

1. **Precipitation**
2. **Surface water**
3. **Groundwater**
4. **Drinking water**
5. Wastewater, leachates, sediment pore water, and soil solutions.

Process Overview



Need for Hydrological Modeling



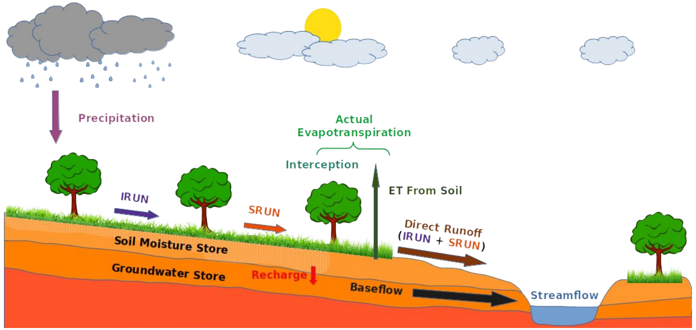


Figure : Rainfall-runoff process (unmanaged basins)

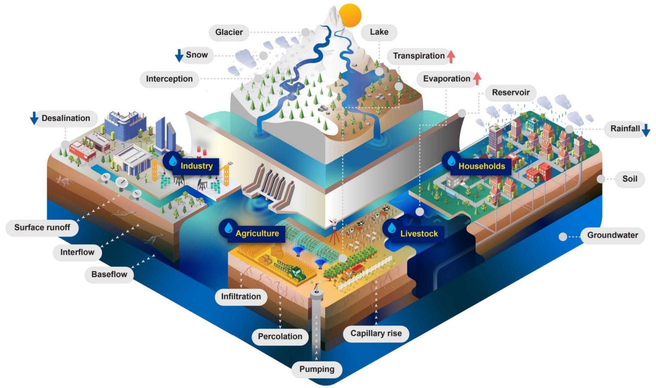


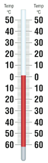
Figure : Hydrological modeling for managed basins

(Image source: https://github.com/iiasa/CWatM)

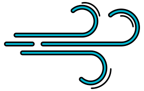
Data for Hydrological Modeling

Meteorological data

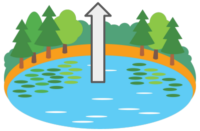
Precipitation



Temperature



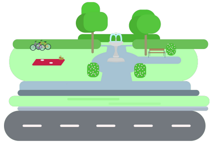
Wind speed



Evaporation

Topographical data

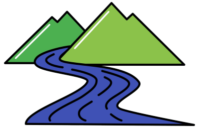
Digital Elevation Model



Land Use Land Cover

Soil types and properties

Hydrological data



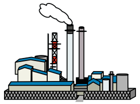
Streamflow and stage (depth)

Reservoir inflows, outflows, storage and water level

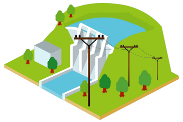
Other data

Drinking water

Irrigation



Industrial



Hydropower

Water demands

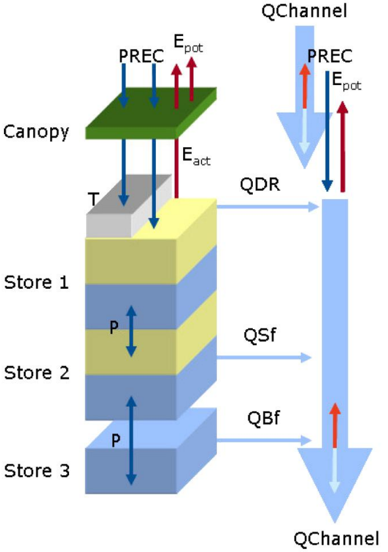
Water quality parameters

Historical flood extents

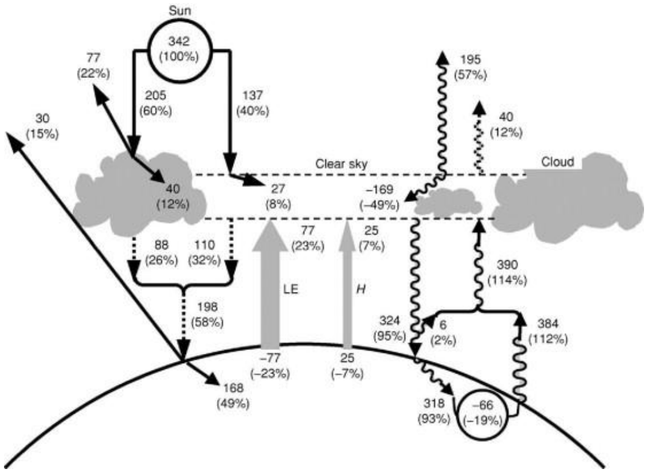
What we expect from models as Earth Scientists and Engineers?

- Interpretability
- Physical Consistency
- Preservation of complex relationship in space and time
- Reasonable predictability without compromising the interpretability

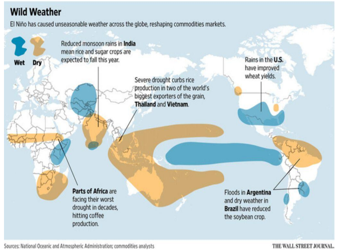
Interpretability

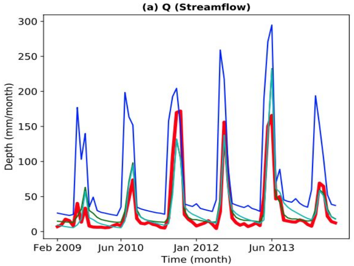


Physical Consistency



Complexity

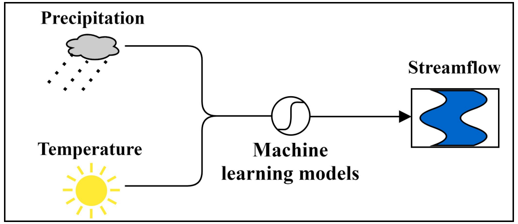




Why to Combine Physics and Data?

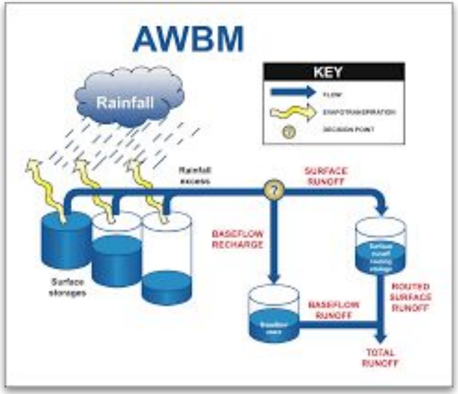
Hydrological models

Data driven models



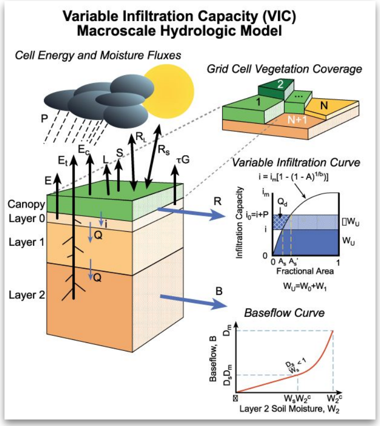
\+ Better predictive skills
\- No physical laws

Conceptual models



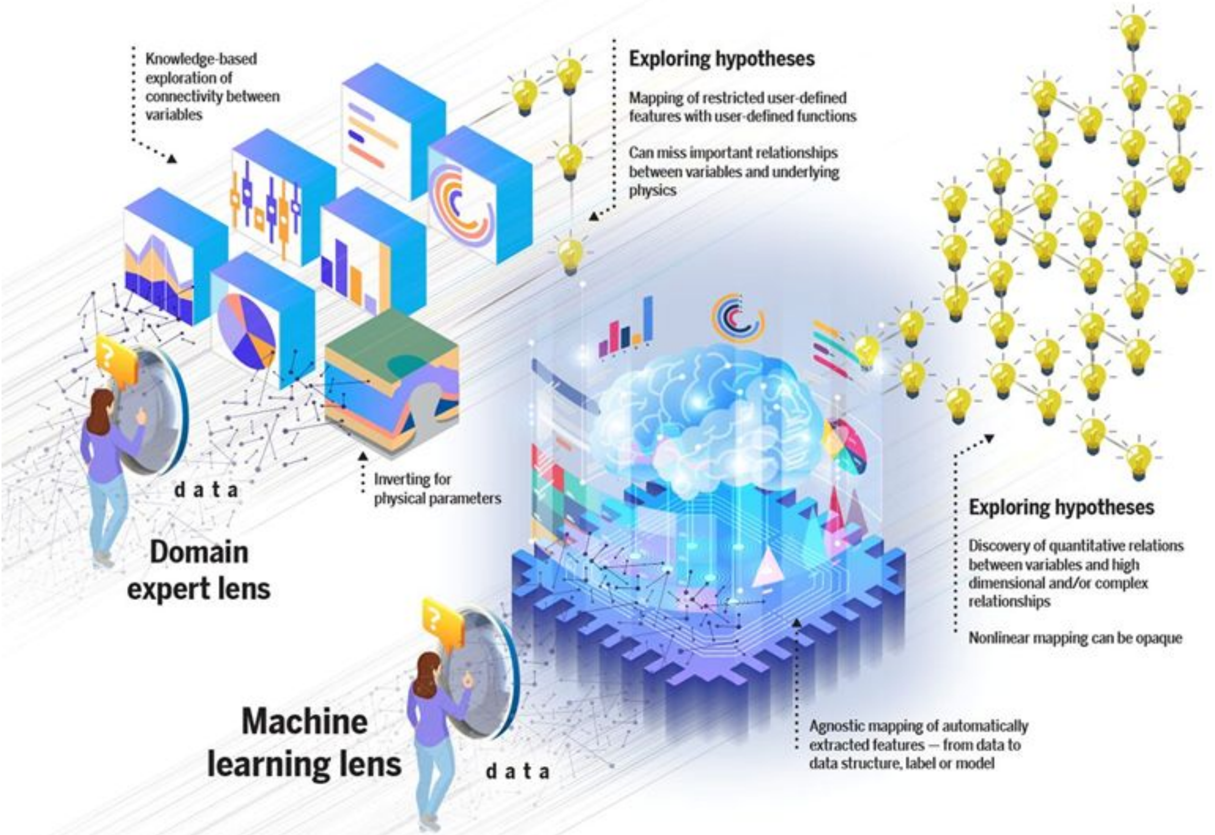
\+ Conceptual representation of various hydrological processes
\- Typically designed for specific region

Physically-based models

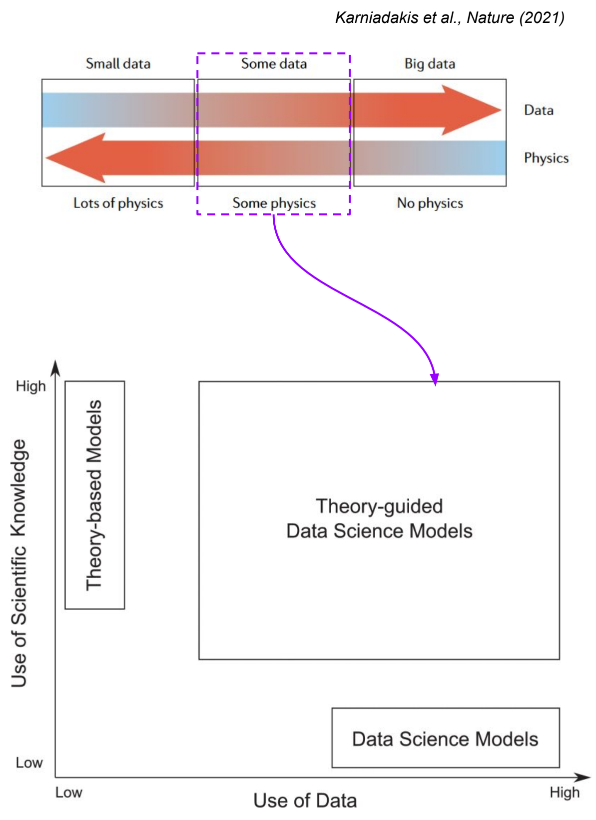


\+ Physical laws
\- Large data requirements, parameterizations

Process-based vs data-driven modeling lens

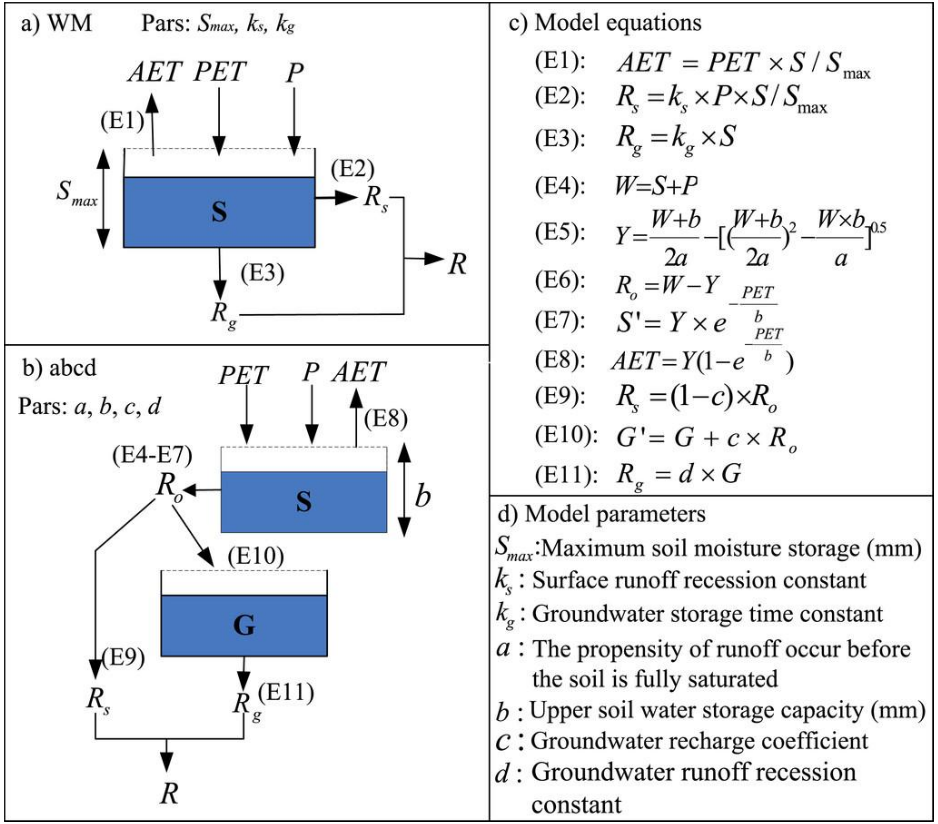


Bergen et al., Science (2019)

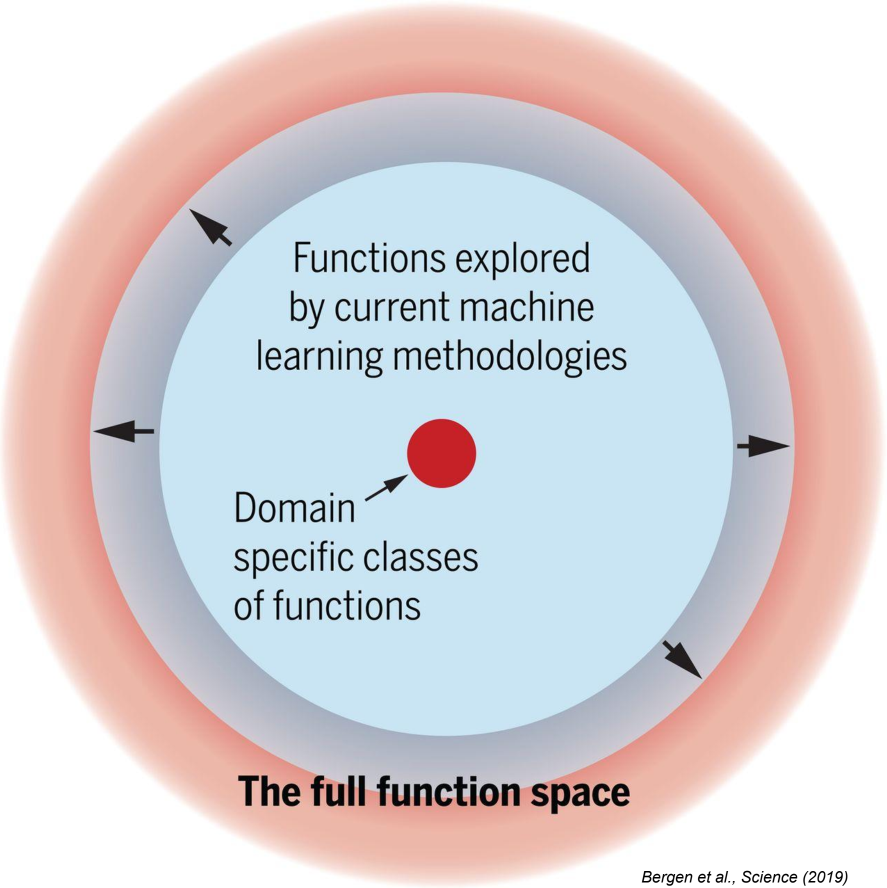


Karpatne et al., IEEE (2017)

Function space by process-based models

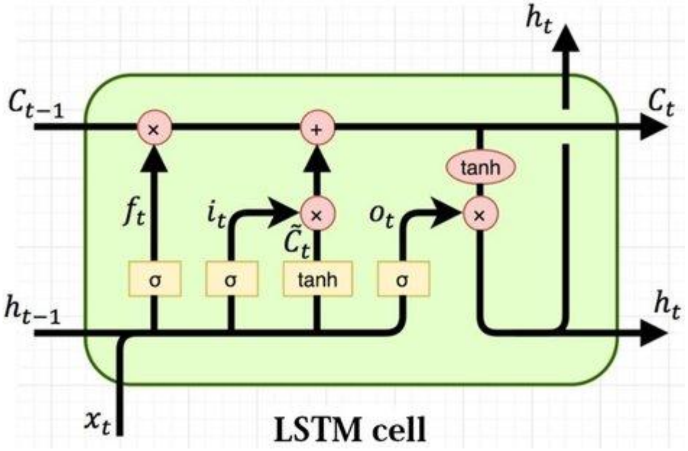


Peng et al., JHM (2016)

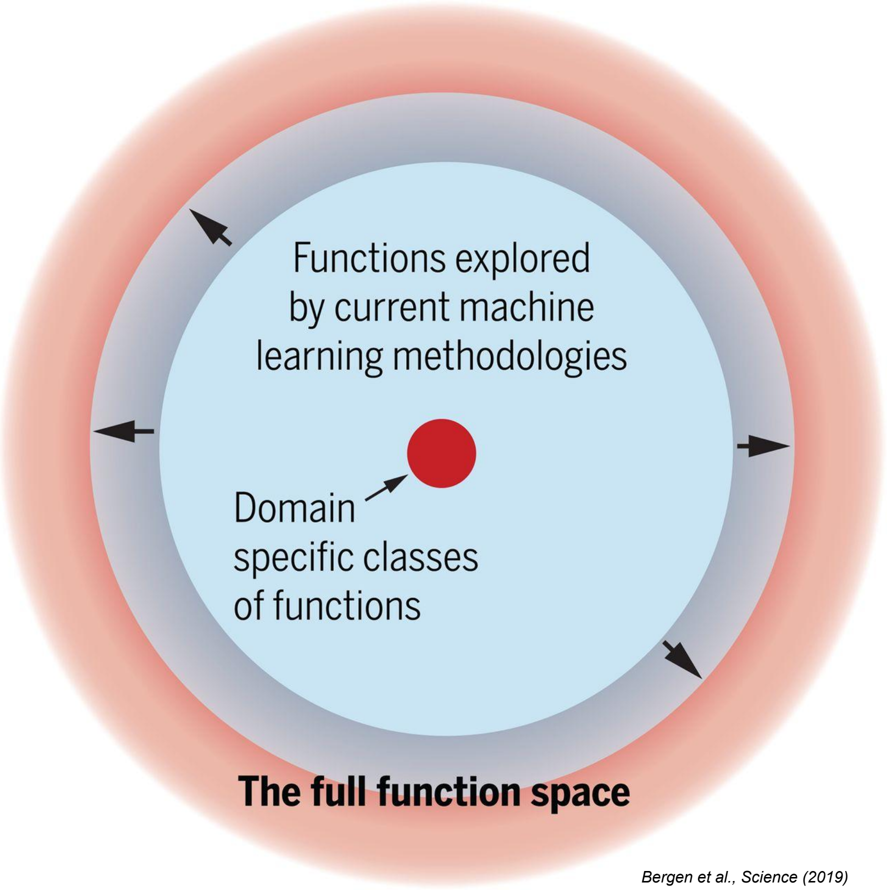


Bergen et al., Science (2019)

Functional space by data-driven models



$$i_t = \sigma\left(x_t U^i + h_{t-1} W^i\right)$$
$$f_t = \sigma\left(x_t U^f + h_{t-1} W^f\right)$$
$$o_t = \sigma\left(x_t U^o + h_{t-1} W^o\right)$$
$$\tilde{C}_t = \tanh\left(x_t U^g + h_{t-1} W^g\right)$$
$$C_t = \sigma\left(f_t * C_{t-1} + i_t * \tilde{C}_t\right)$$
$$h_t = \tanh(C_t) * o_t$$



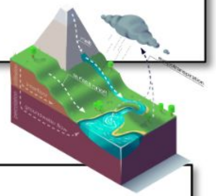
Bergen et al., Science (2019)

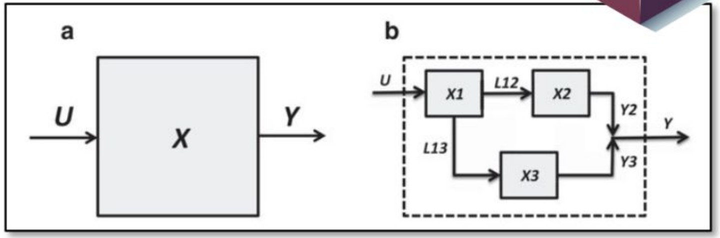
Drawing Parallels: *Parameterizations in Process-based vs data-driven modeling*

Process-based Modeling

State space model:

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

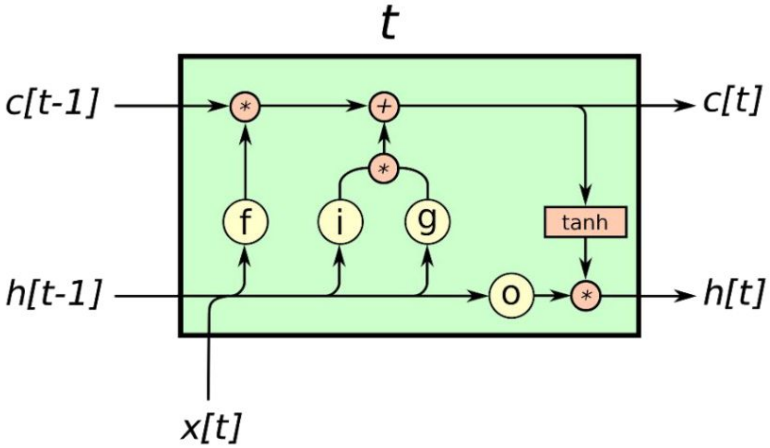




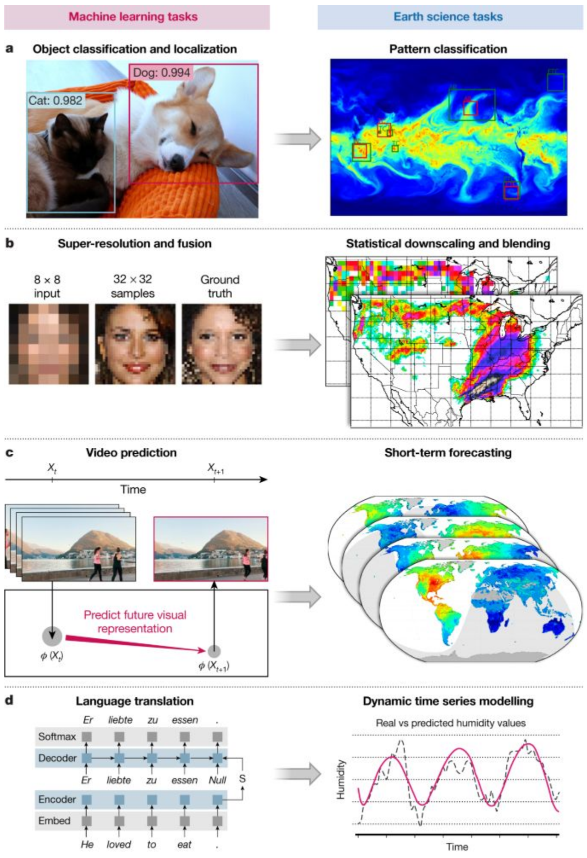
Data-driven Modeling

LSTM model:

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

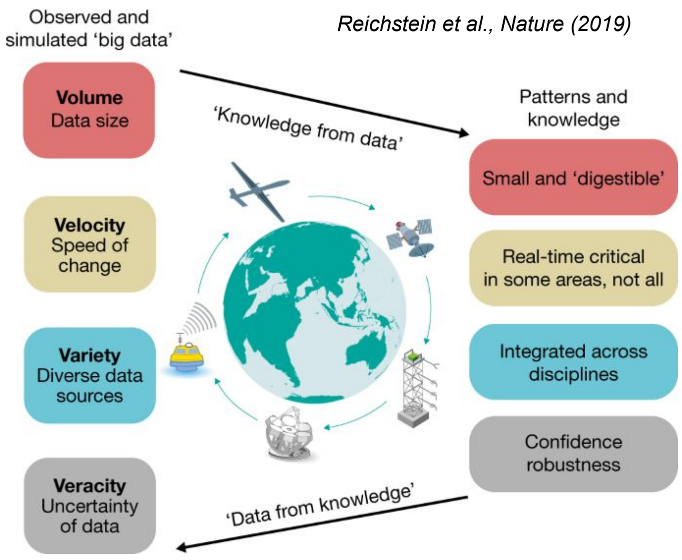


The Need and Challenges…

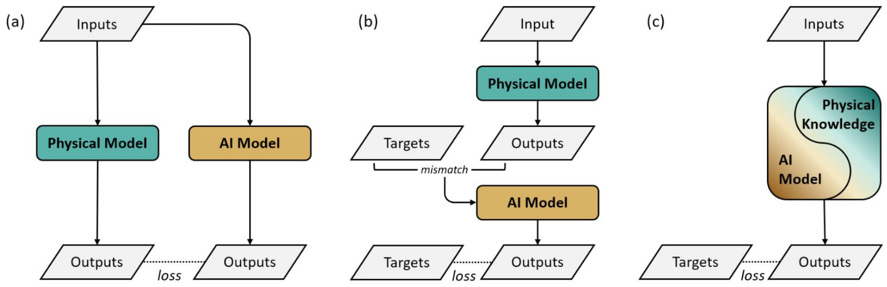


Reichstein et al., Nature (2019)

- Interpretability
- Physical Consistency
- Limited Labels
- Computationally Demanding
- Preservation of complex relationship in space and time

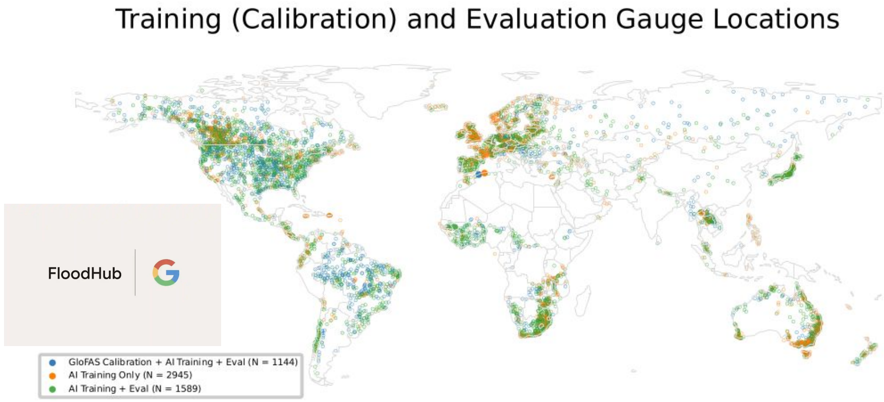


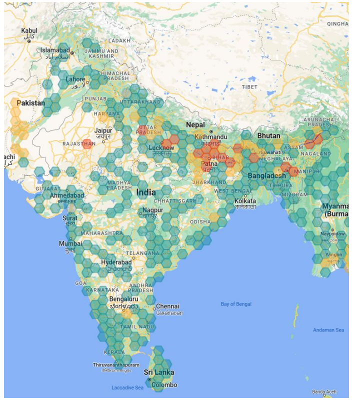
Reichstein et al., Nature (2019)



Jiang et al., GRL (2020)

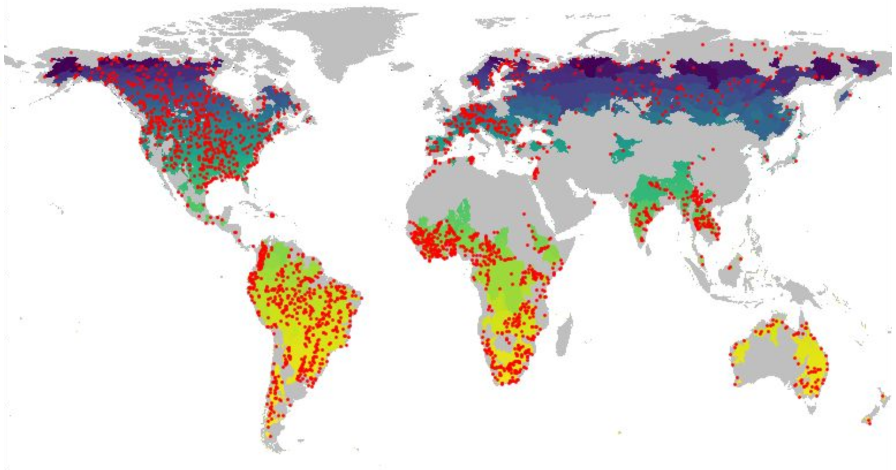
Challenges and limitations of existing models





- Data-driven model
- Coarser resolution of maps (which limits decision making at urban scale)
- Less representation and data availability of Indian cities
- Provides forecast wherever gauges are available (limited coverage)
- Doesn’t account pluvial floods
- Reservoir operations are ignored

GLOFAS
Global Flood Awareness System



- Hydrological Core: LISFLOOD at 3’ or 0.05º.
- Number of Calibration Sites: ~500 in CONUS and ~100 in India *(lower number of sites and shorter duration of hydrological observations for India)*
- Parameter Maps: ~100 (14 are calibrated parameters)
- Dynamic Input: ERA5 Surface Variables

Case Study-1: Lumped Physics Informed Machine Learning (PIML) model for monthly timestep

(Bhasme et al., 2022)

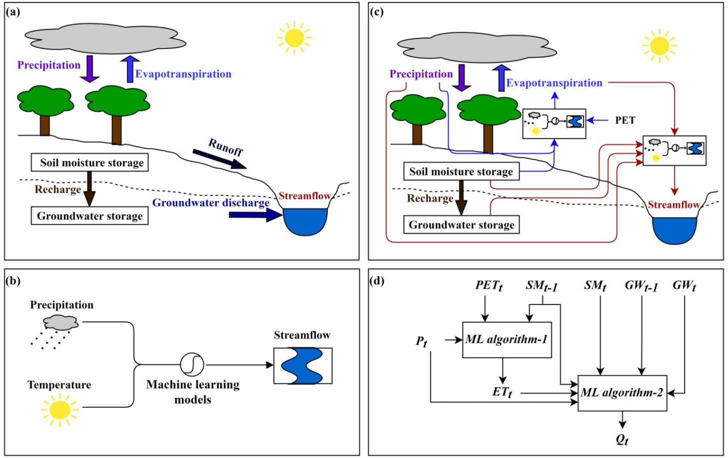


Figure: (a)The abcd model (Thomas et al., 1981)**; (b) Data-driven model; (c) PIML model; (d) PIML model framework**

$$ET_t = f(P_t, SM_{t-1}, PET_t)$$

$$Q_t = g(P_t, ET_t, SM_t, SM_{t-1}, GW_t, GW_{t-1})$$

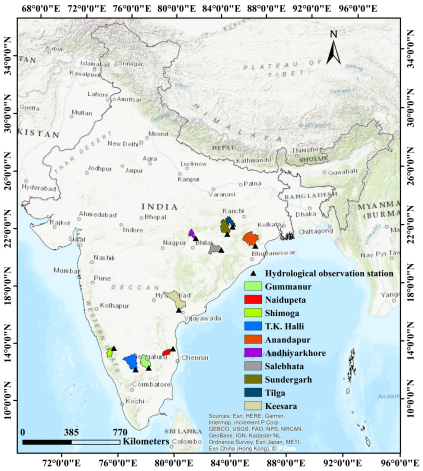


Figure: Study area (Ten subcatchments from peninsular India)

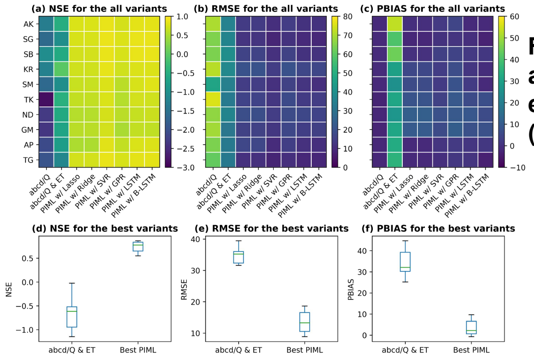


Figure: Results for actual evapotranspiration (ET) predictions

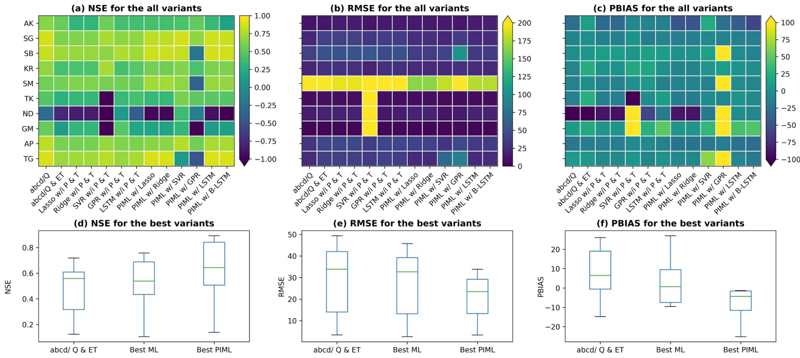


Figure: Results for streamflow (Q) predictions

Case Study-2: Improving the interpretability and predictive power of hydrological models: Applications for daily streamflow in managed and unmanaged catchments *(Bhasme and Bhatia, 2024)*

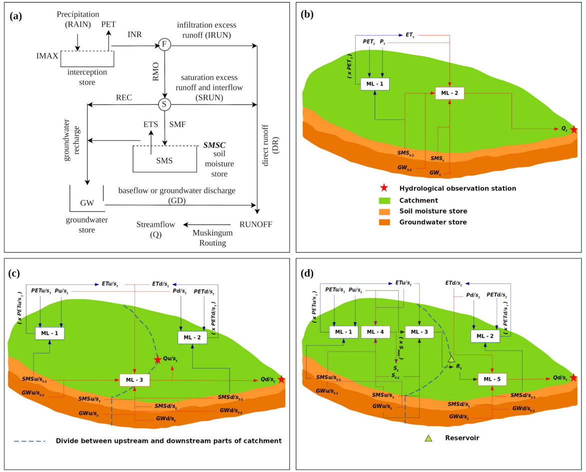


Figure: (a) The SIMHYD model (Chiew et al., 2002); **(b) Lumped PIML model; (c) Semi-distributed PIML model without reservoir; (d) Semi-distributed PIML model with reservoir**

$$\frac{ET_t}{PET_t} = a(P_t, PET_t, SMS_{t-1})$$

$$Q_t = b(P_t, ET_t, SMS_t, GW_t, ..., P_{t-j}, ET_{t-j}, SMS_{t-j-1}, GW_{t-j-1})$$

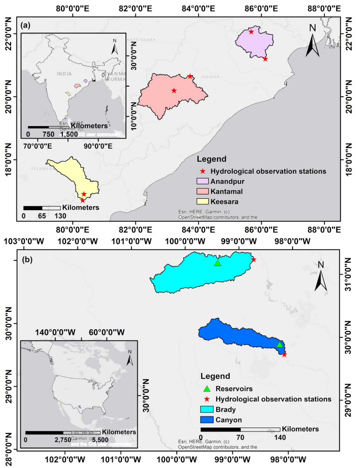


Figure: Study area (a) Semi-distributed PIML model without reservoir; (b) Semi-distributed PIML model with reservoir

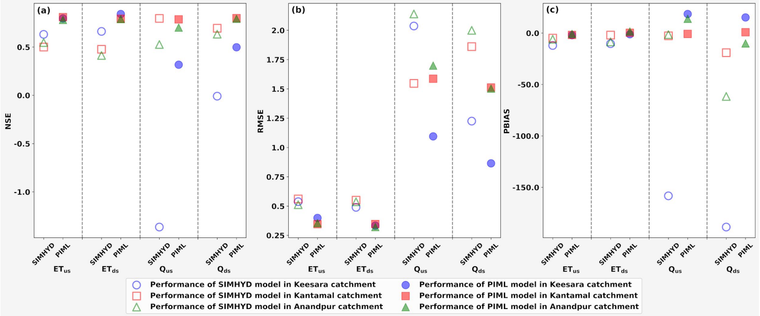


Figure: Results for semi-distributed PIML without reservoir

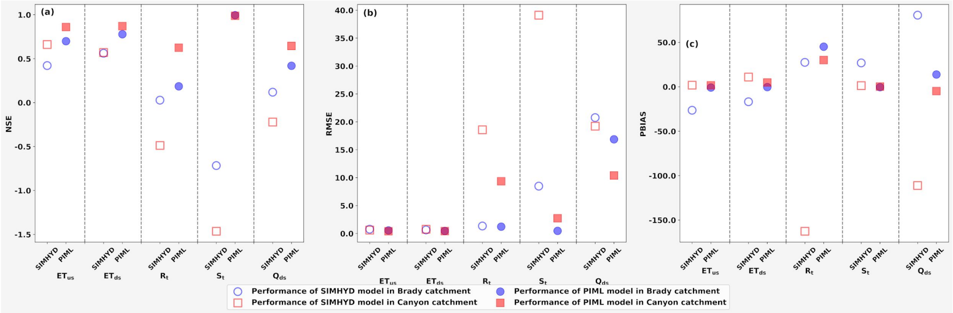


Figure: Results for semi-distributed PIML with reservoir

Case Study-3: Enhancing Fluvial Flood Predictions through Physics Informed Graph Neural Network

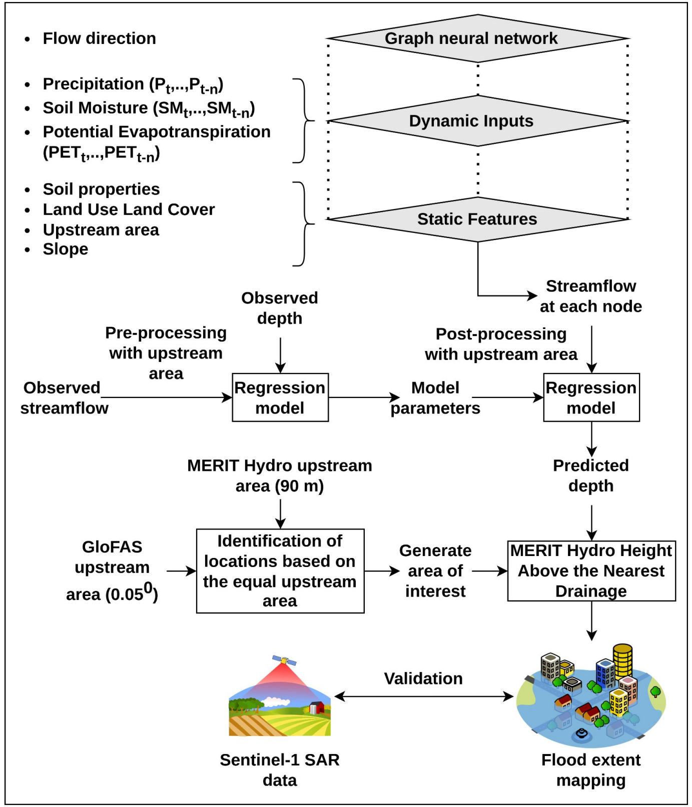


Figure: Distributed PIML for fluvial flood prediction

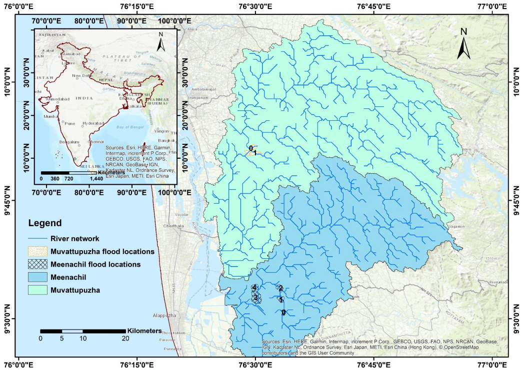


Figure: Study area for distributed PIML model for fluvial flood prediction

Table: Preliminary results (in %) for flood extent mapping for 9th and 21st August 2018 flood events

Catchment	Flood location Id	Flood extent mapping for 9th Aug 2018 (%)	Flood extent mapping for 21st Aug 2018 (%)
Muvattupuzha	0	60.65	56.09
	1	75.53	71.74
Meenachil	0	88.03	82.08
	1	82.2	71.02
	2	90.9	78.19
	3	97.7	94.72
	4	97.54	96.67

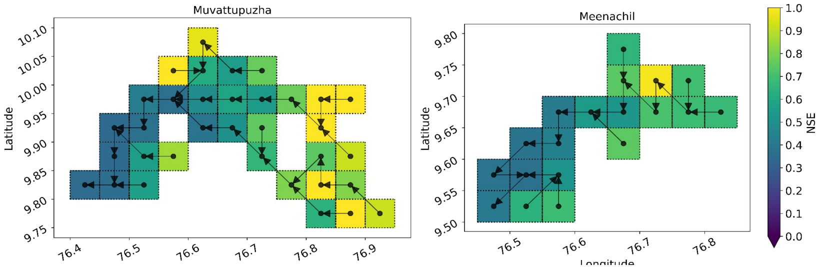
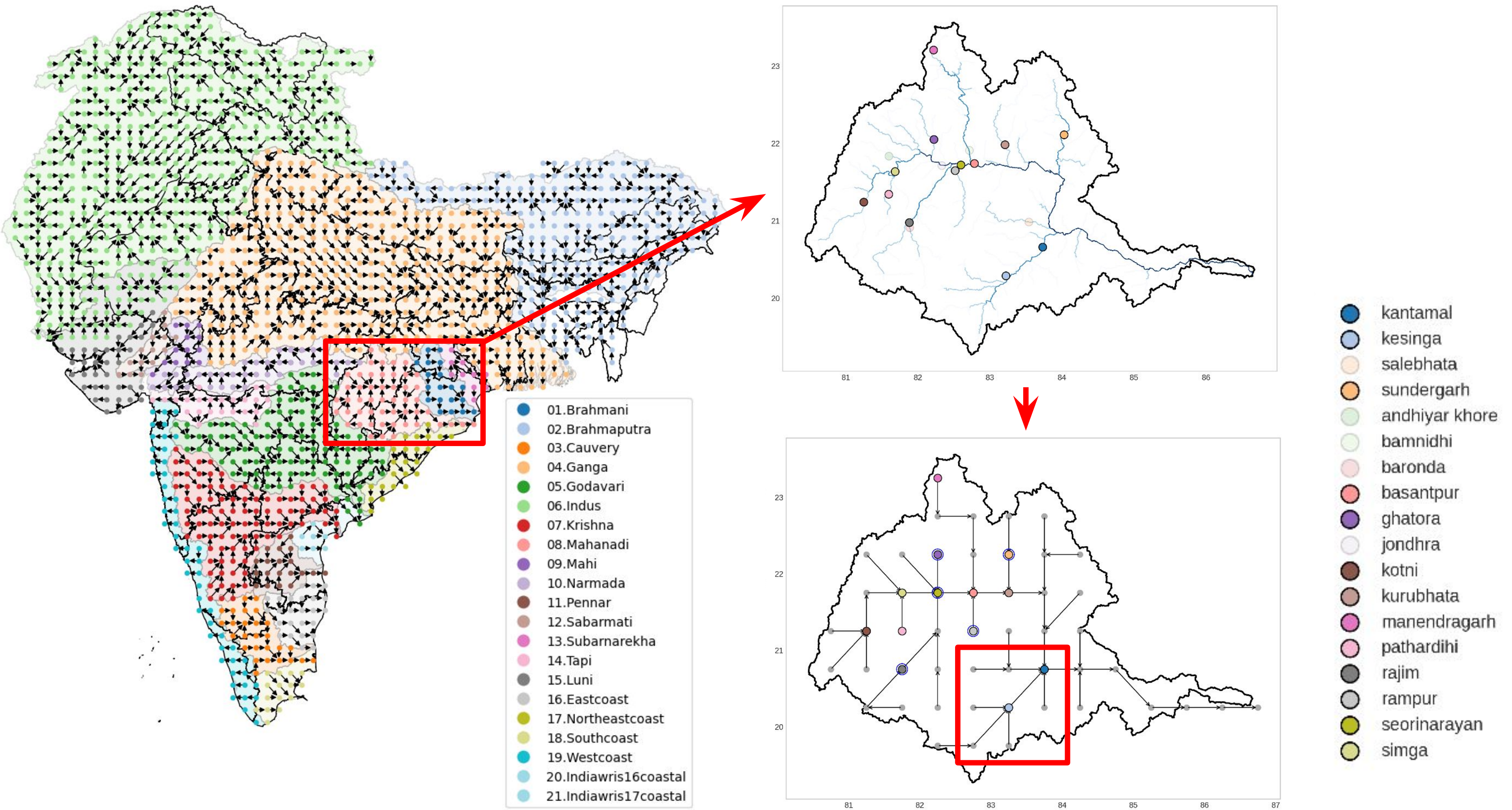
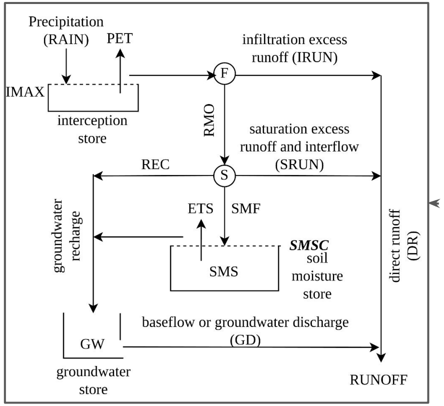


Figure: Performance of distributed PIML model (Network is based on the GloFAS Local Drain Direction datasets)

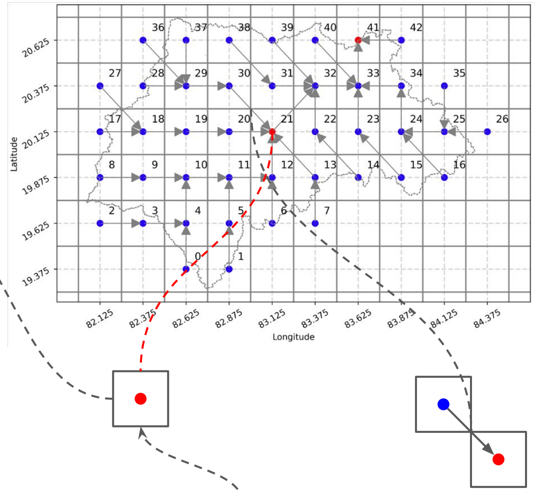
Case Study-4: Accounting Basin Heterogeneity - Towards Distributed Modeling

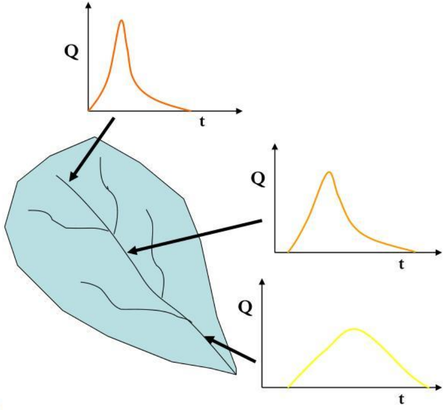


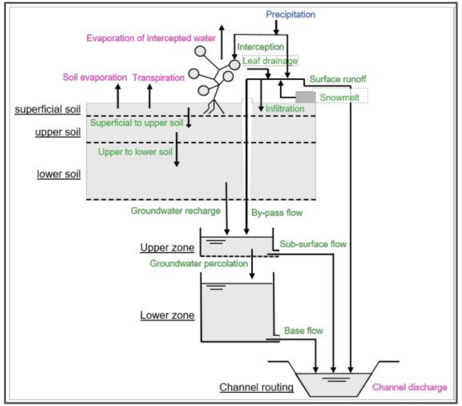
Understanding Nodes and Edges on a River Network



SIMHYD

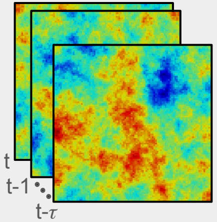






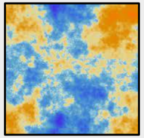
Lisflood

Dynamic Variables



1. ERA5 (Global)
2. GLEAM (Global)
3. IMD (India) / Daymet (US)

Static Variables



1. Elevation
2. Upstream Area
3. Soil properties
4. LULC

Muskingum Cunge

$$Q_{i+1}^{t+1} = C_1 Q_i^{t+1} + C_2 Q_i^t + C_3 Q_{i+1}^t$$

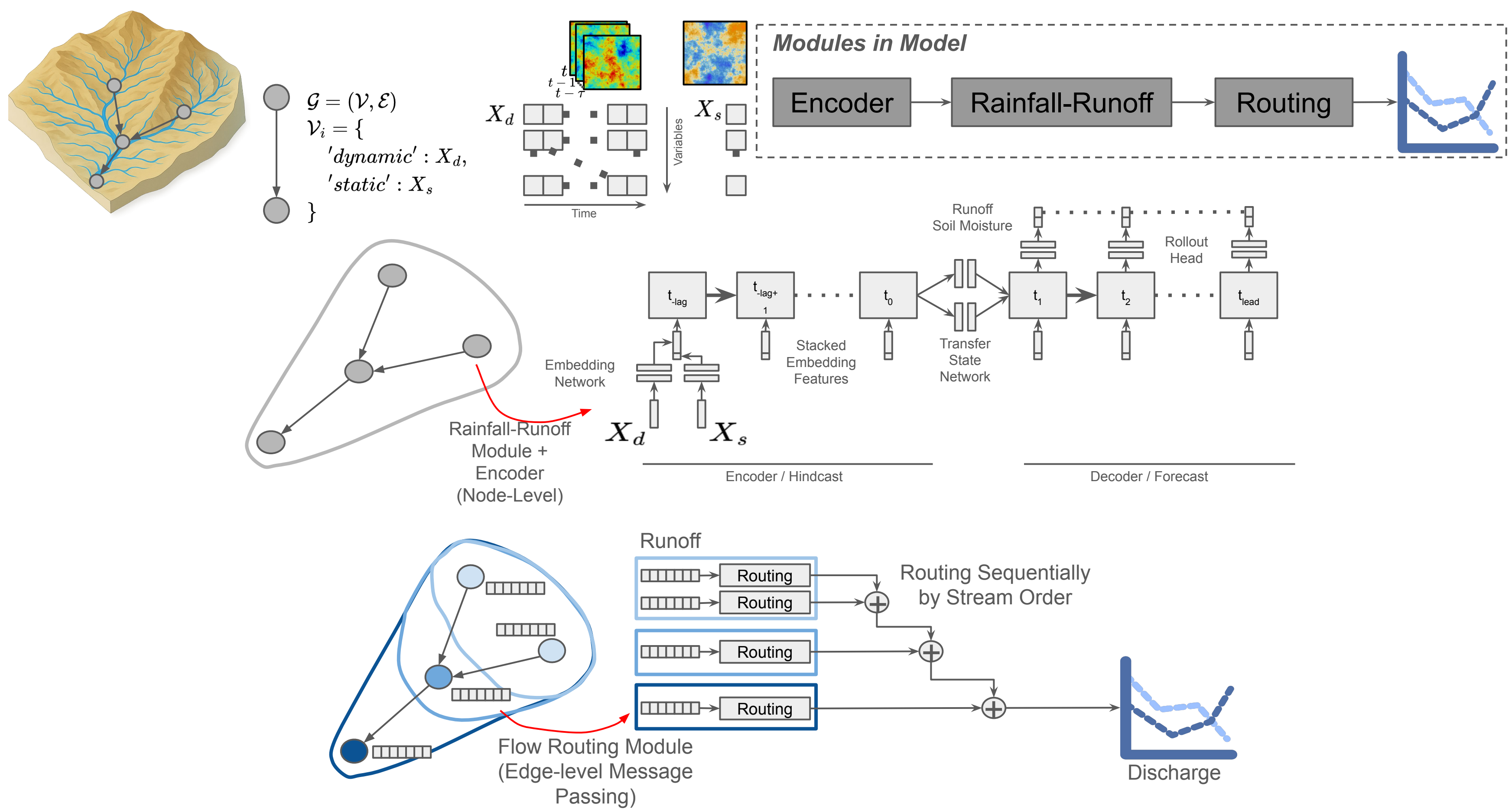
Kinematic Wave

$$\frac{\partial Q}{\partial x} + \alpha\beta Q^{\beta-1}\frac{\partial Q}{\partial t} = q$$

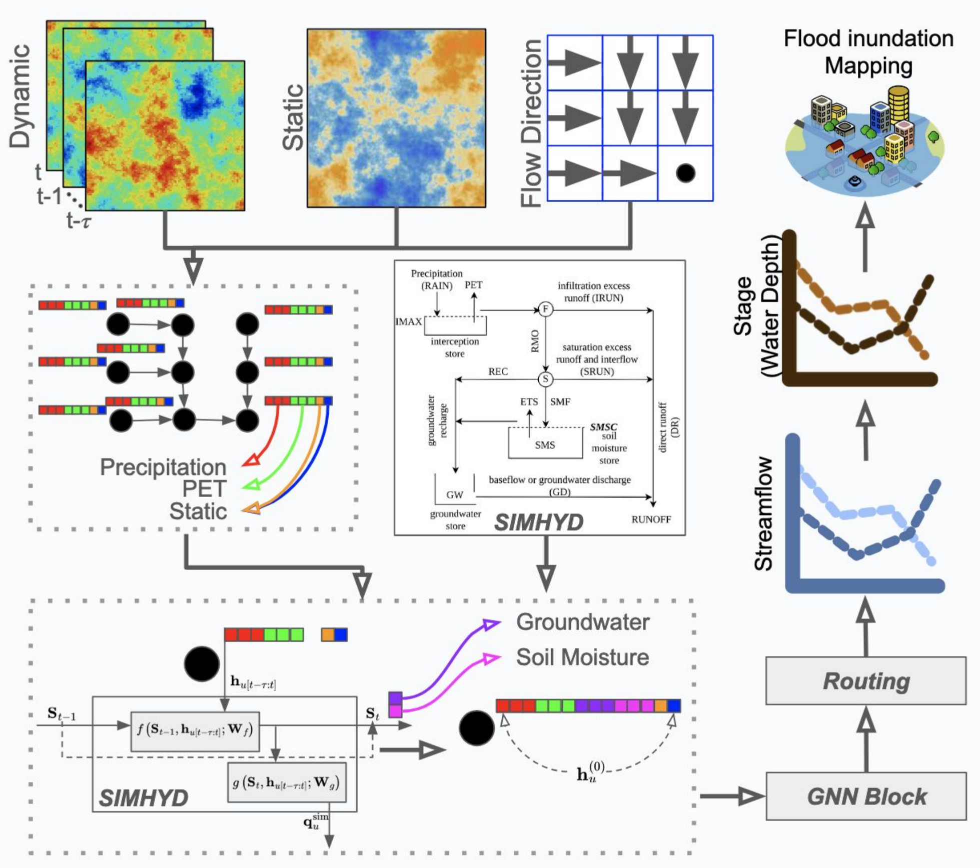
Data-driven

$$Q_{i+1}^{t+1} = f\left(Q_i^{t+1}, Q_i^t, Q_{i+1}^t\right)$$

Understanding Nodes and Edges on a River Network



Developing Hybrid Hydrological Models



Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

Node Features $\mathbf{h}_u \in \mathbb{R}^d, \forall u \in \mathcal{V}$

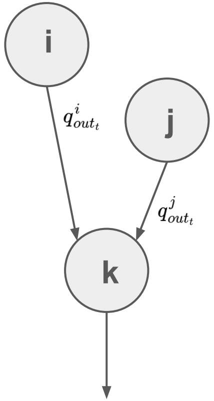
$$\mathbf{h}_{\mathcal{N}(u)}^{(k)} = AGGREGATE^{(k)}\left(\mathbf{h}_v^{(k)}, \forall v \in \mathcal{N}(u)\right)$$

$$\mathbf{h}_u^{(k+1)} = UPDATE^{(k)}\left(COMBINE\left(\mathbf{h}_u^{(k)}, \mathbf{h}_{\mathcal{N}(u)}^{(k)}\right)\right)$$

where, $k = 1, \dots, L$

Runoff $\mathbf{q}_u = ROLLOUT\left(\mathbf{h}_u^{(L)}\right)$

Streamflow $\mathbf{Q}_u = ROUTING\left(\mathbf{q}_u, \mathbf{q}_v \forall v \in \mathcal{N}(u)\right)$



$$q^k_{in_t} = q^i_{out_t} + q^j_{out_t}$$

$$q^k_{routed_t} = \frac{S^k_{river_t} + q^k_{in_t}}{(lag_{river} + 1)}$$

$$q^k_{out_t} = q^k_{routed_t} + runoff^k_t$$

$$\delta S^k_{river_t} = q^k_{in_t} - q^k_{routed_t}$$

$$S^k_{river_{t+1}} = S^k_{river_t} + \delta S^k_{river_t}$$

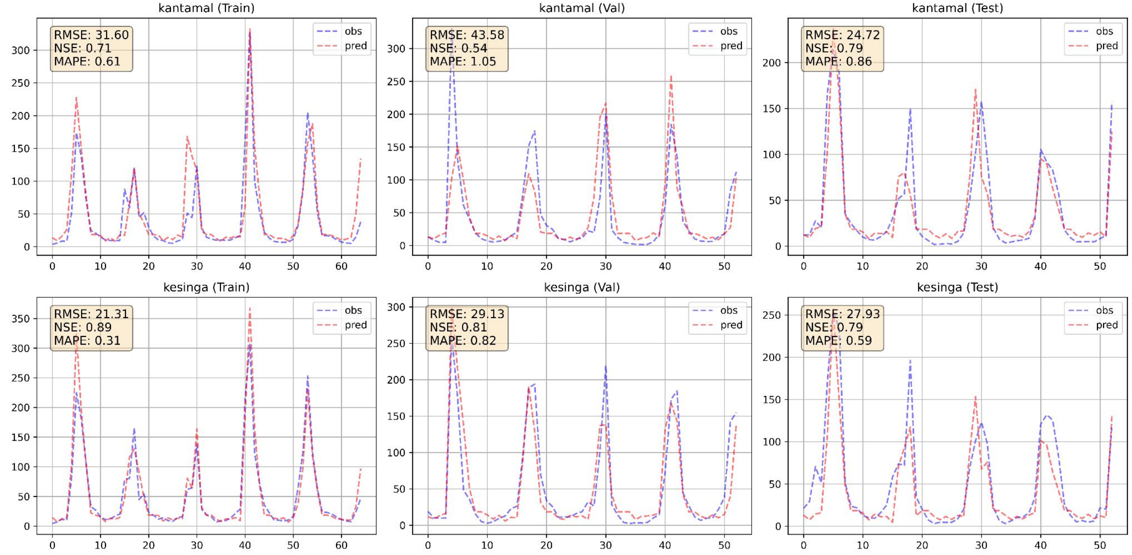
Assessing performance on an Indian Catchment

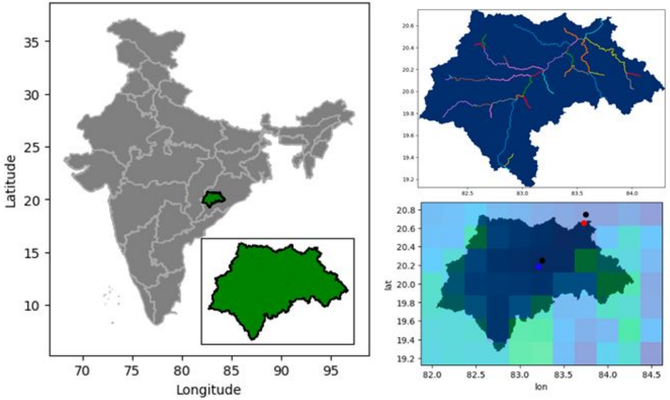
Study Area: Kantamal catchment (within Mahanadi basin)

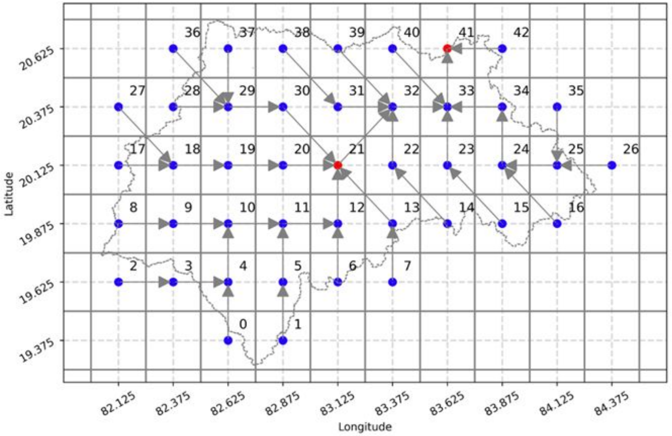
Frequency: Monthly
Spatial Resolution: 15 arcmins (0.25 degrees)

Inputs: Precipitation (IMD), PET (GLEAM), Groundwater and Soil Moisture (SIMHYD)
Outputs: Streamflow (IndiaWRIS)

Train | Val | Test: 2000-2007 | 2008-2012 | 2013-2018







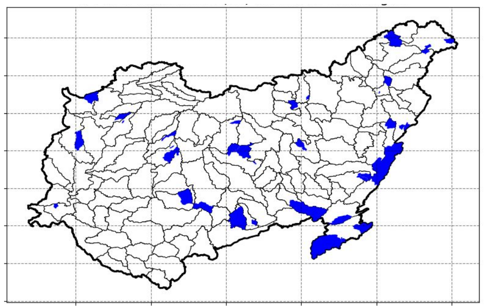
Assessing performance of catchments within one US eco-region

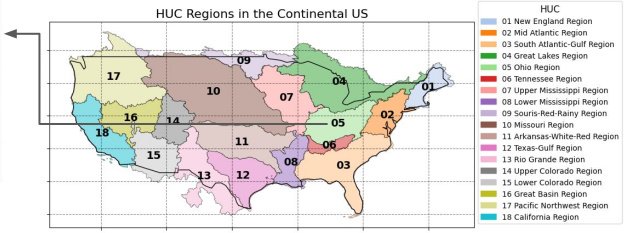
Study Area: 34 CAMELS (minimal human influence) Catchments in Ohio Region, US

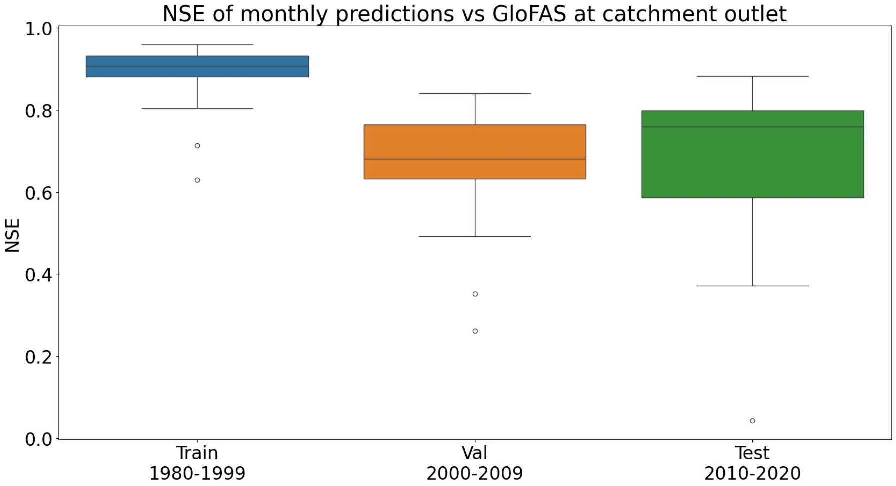
Frequency: Monthly and Daily
Spatial Resolution: 3 arcmins (0.05 degrees)

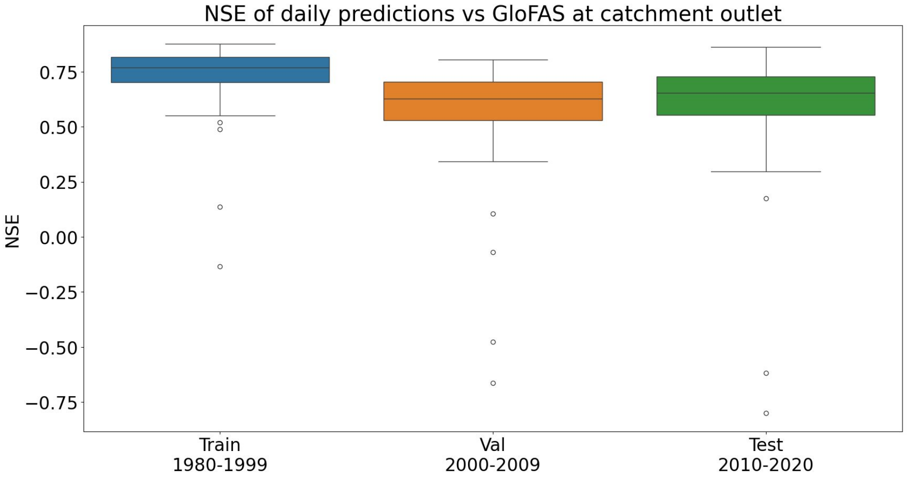
Inputs: Daymet, ERA5, Soil Composition, LULC
Outputs: Streamflow (GloFAS)

Train | Val | Test: 1980-1999 | 2000-2009 | 2009-2020









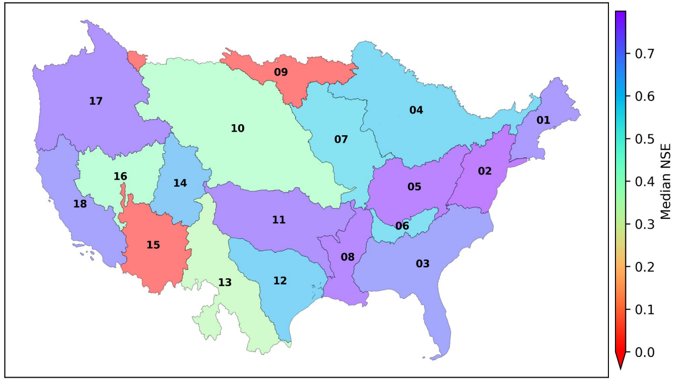
Training on data-rich basins…

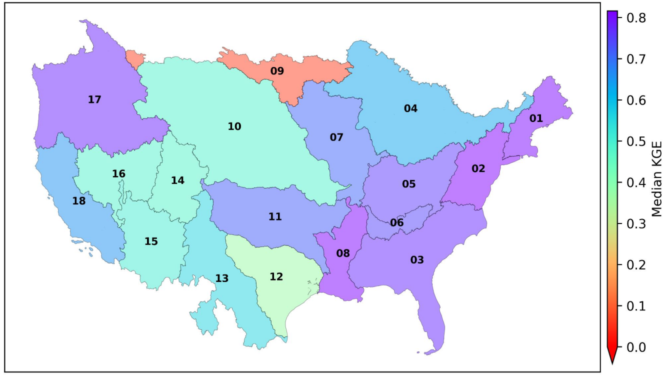
Study Area: 493 CAMELS (minimal human influence) Catchments in Continental US

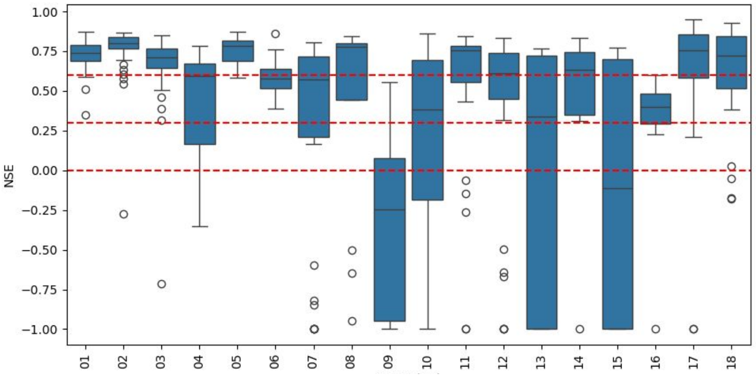
Frequency: **Monthly**
Spatial Resolution: 3 arcmins (0.05 degrees)

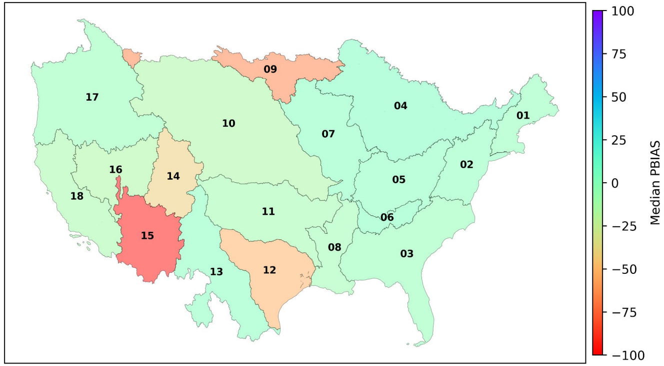
Inputs: ERA5, Soil Composition, LULC
Outputs: Streamflow (GloFAS)

Train | Test: 1999-2008 | 1989-1999









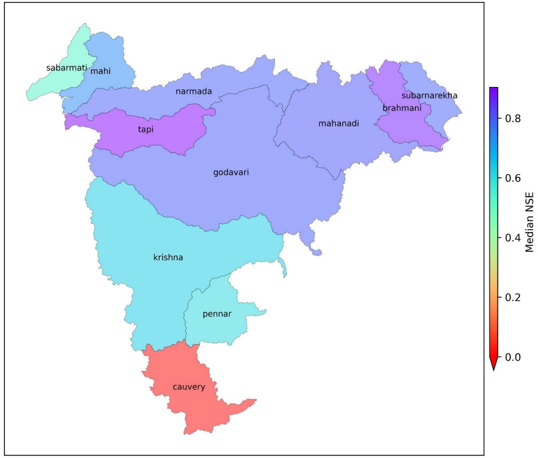
Testing on ungauged/basins

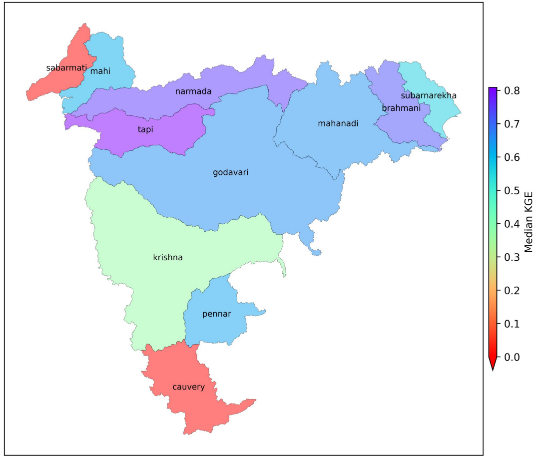
Study Area: 144 IndiaWRIS Gauges (>100 km2) Catchments in Indian mainland Basins

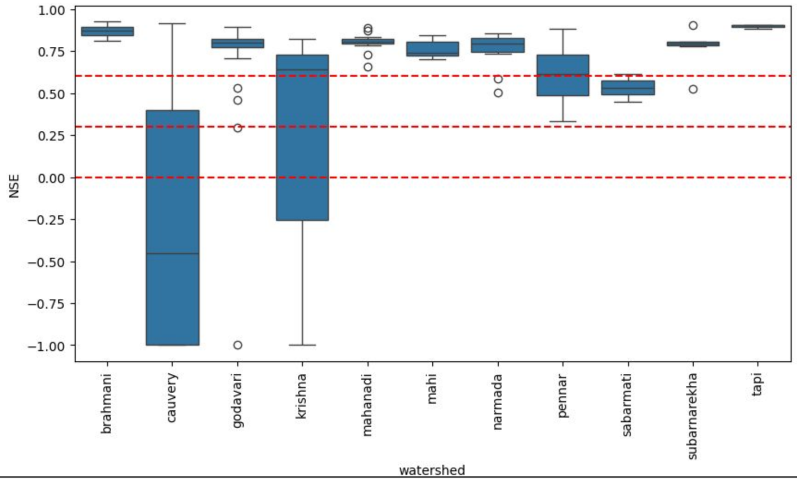
Frequency: Monthly
Spatial Resolution: 3 arcmins (0.05 degrees)

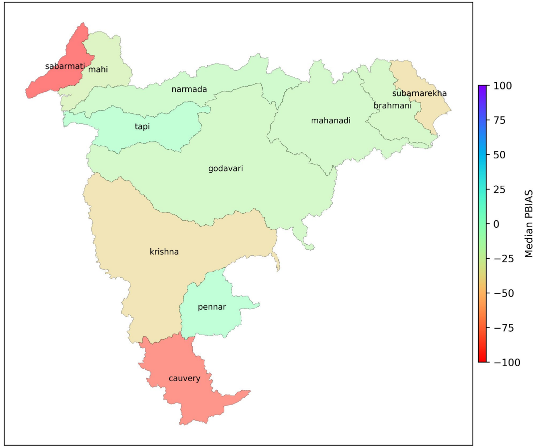
Inputs: ERA5, Soil Composition, LULC
Outputs: Streamflow (GloFAS)

Train | Test: 1999-2008 | 1989-1999









The Opportunities…

Data quality and availability

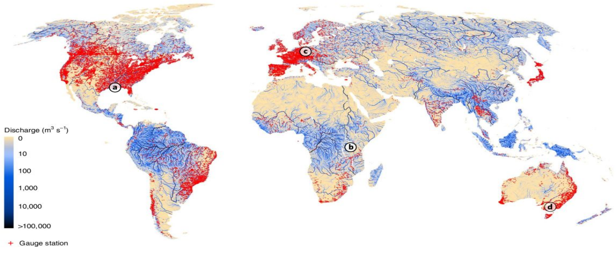


Figure: Global distribution of stream gauges (red crosses; N = 32,091) along the river network (blue) identified by GRADES. (Krabbenhoft et al., 2022)

Need of higher resolution datasets for better modeling

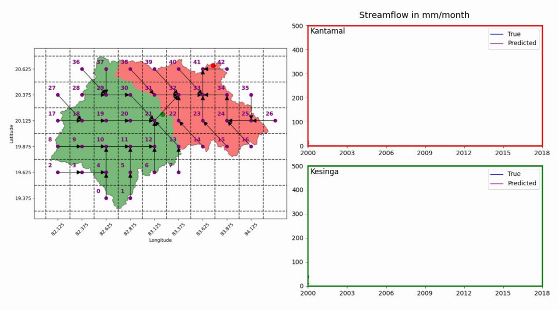


Figure: Streamflow output at outlet while inputs are at 0.25^0 degree resolution

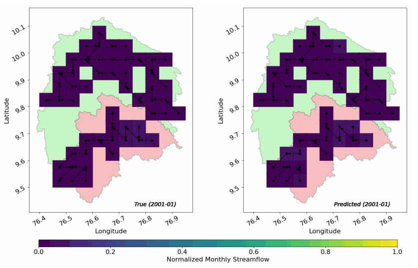
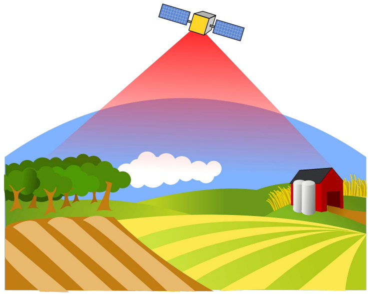


Figure: Streamflow output at all pixels when inputs are at 0.05^0 degree resolution

Integration of satellite products



Citizen Science

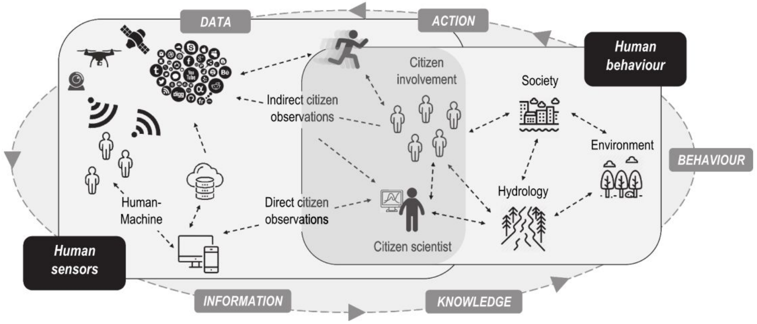
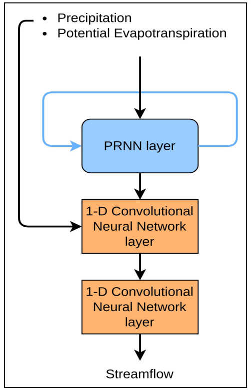


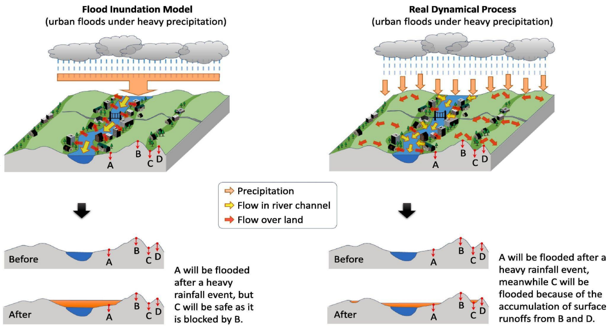
Figure: Data–information–knowledge–behaviour–action workflow characterizing citizen science projects for hydrological sciences. (Nardi et al., 2022)

floodResQ at Glance

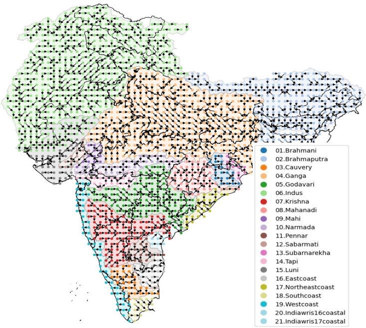
PIML



Boundary conditions

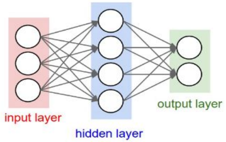


(Wang et al., 2019)



Process-based model for urban flooding simulation

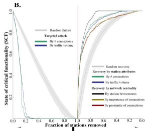
HydroGNN



(Figure Source: https://cs231n.github.io/neural-networks-1/)

Recovery strategies for resilient infrastructure system

(Bhatia et al., 2015)



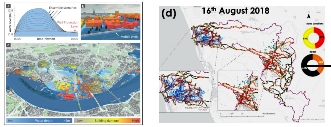
Deep Learning based emulator for urban flooding simulation

Flood risk maps

Impact

Critical Infrastructure network

Damage quantification

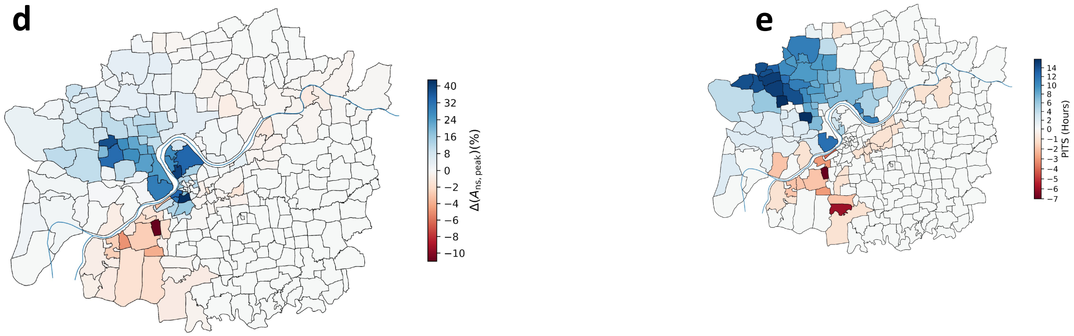


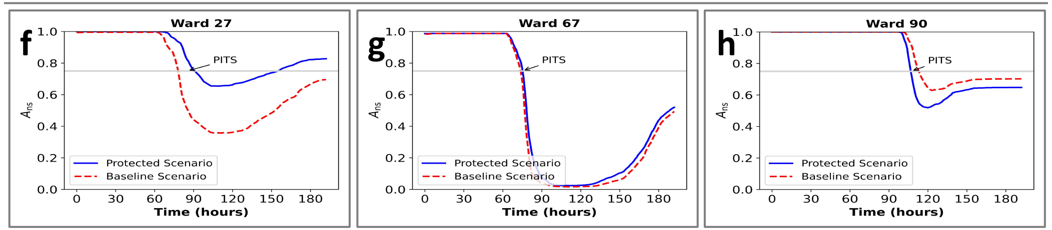
source:https://developer.nvidia.com/blog/simulating-real-world-floods-on-gpus/

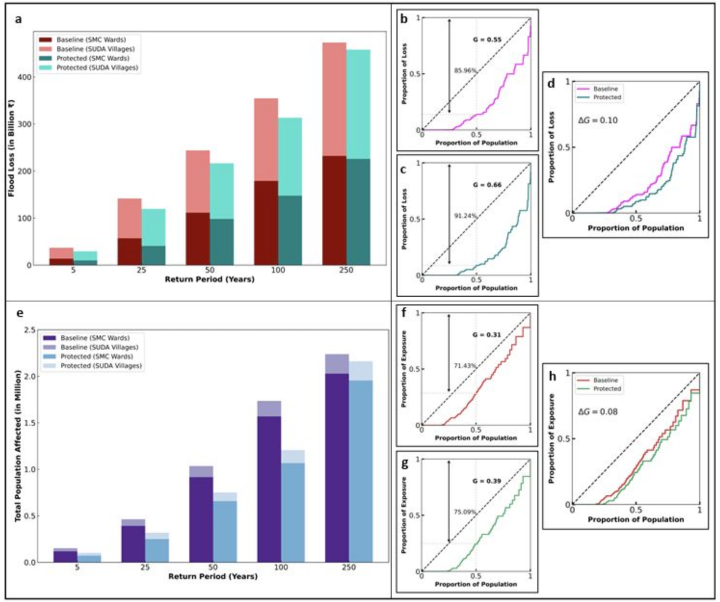
(Dave et al., 2021)

Decision support system

Side-effects of Adaptation

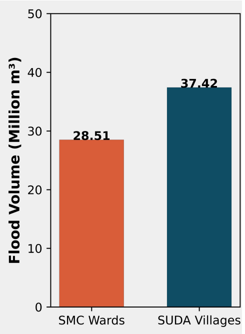


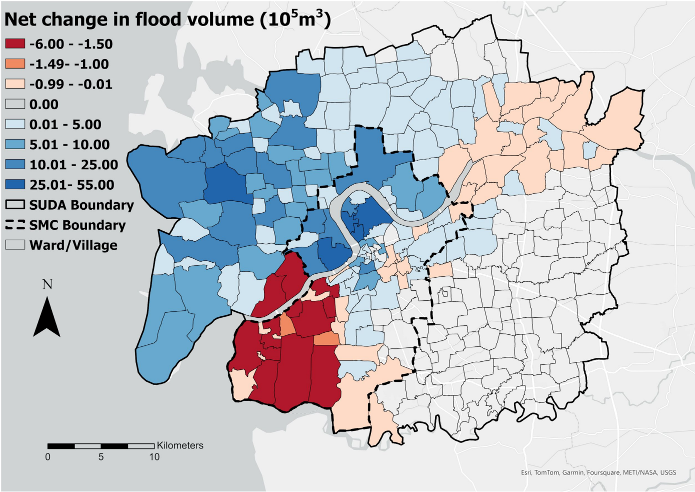


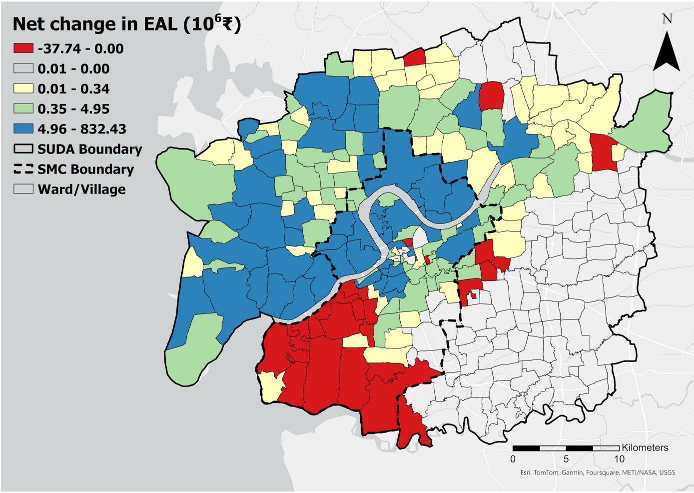


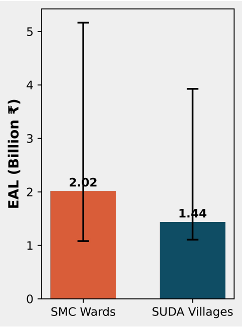
Kumar et al. (2025)

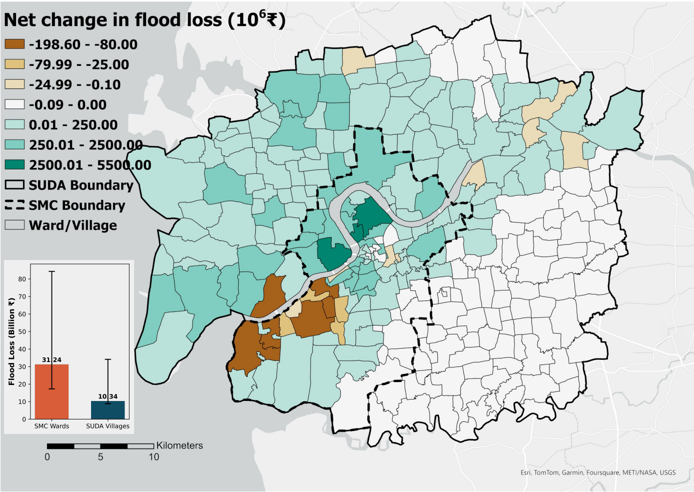
Flood beyond water

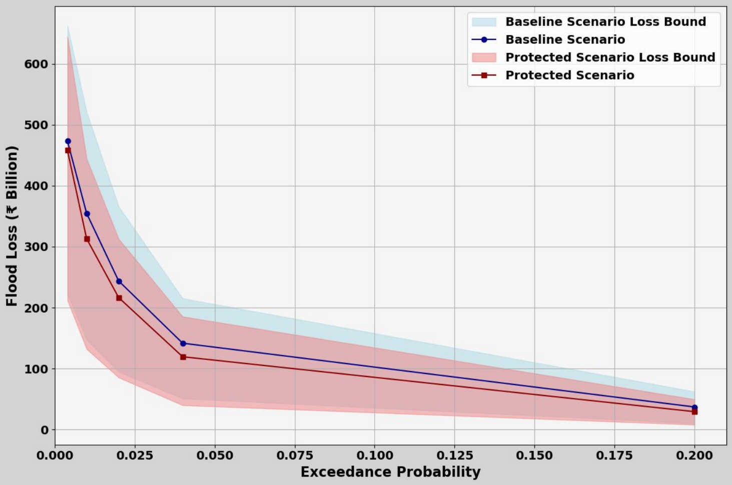


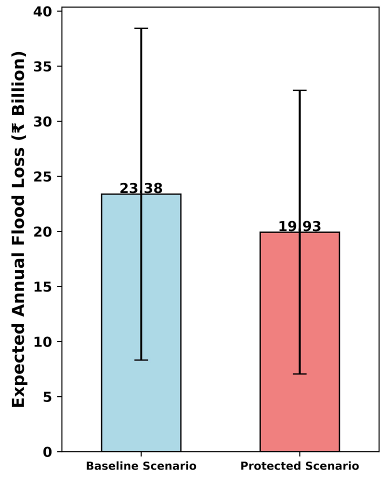












The Unmet Water Challenge: Dominance of Human Factors

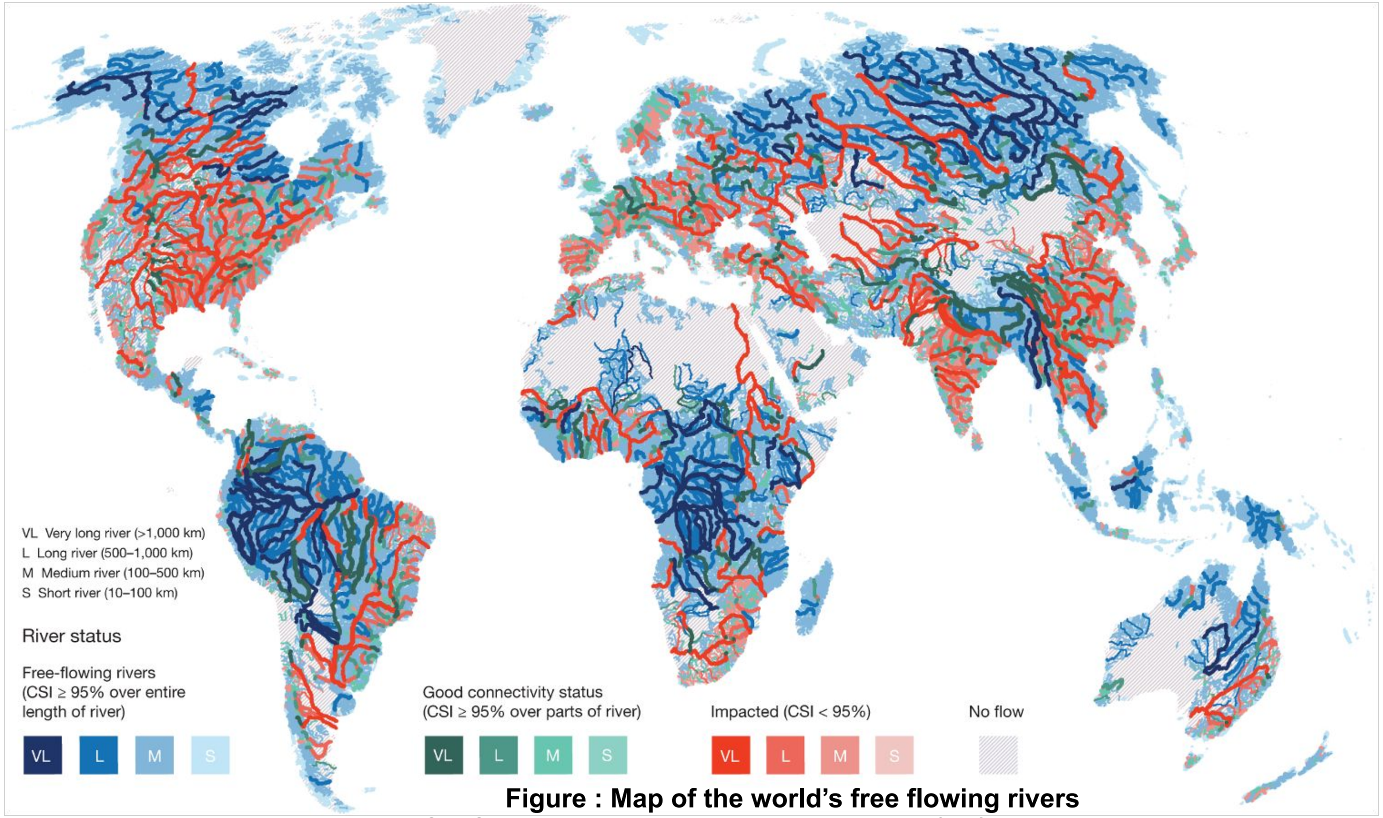


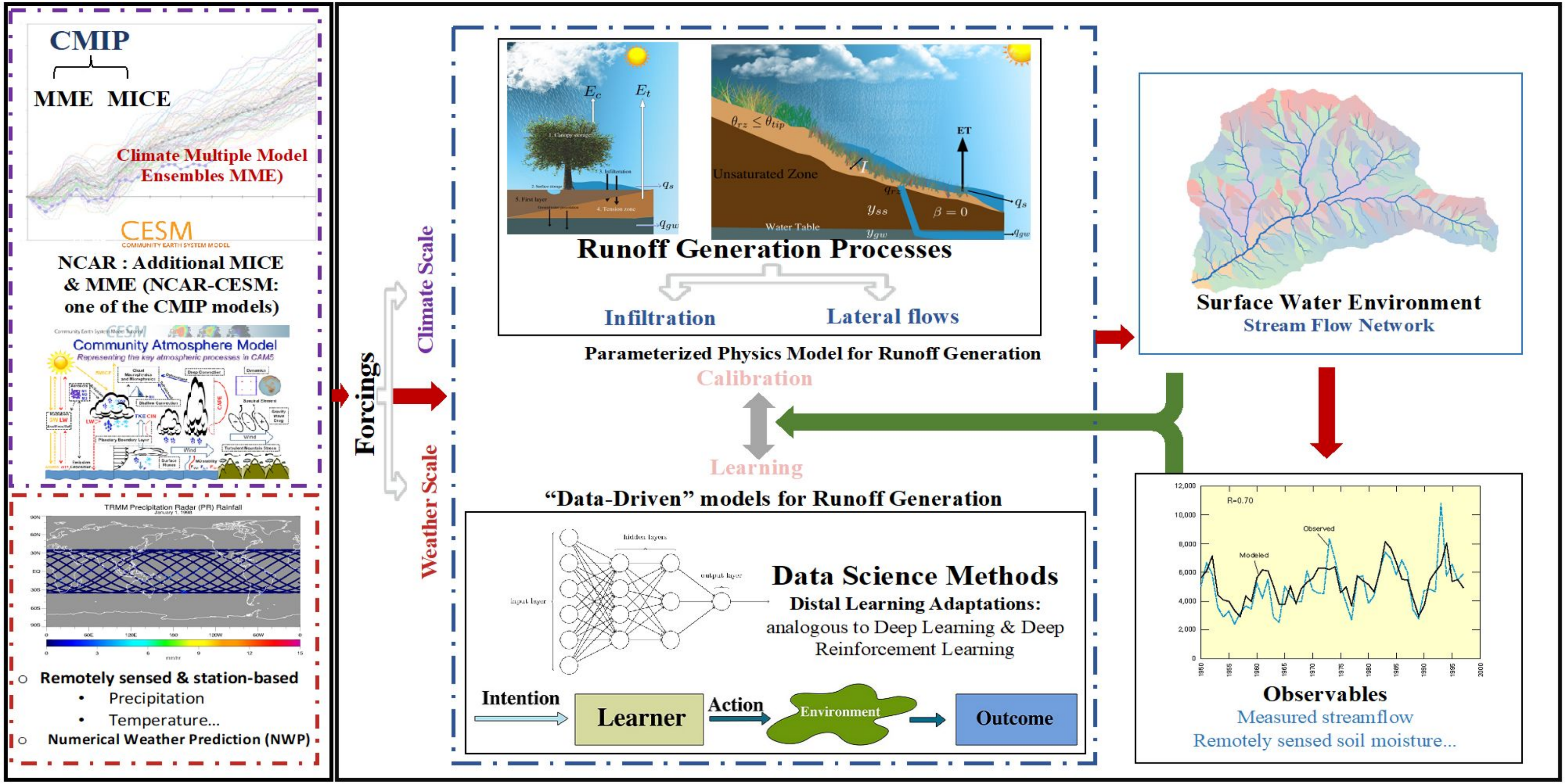
Figure : Map of the world's free flowing rivers

(Grill, G., Lehner, B., Thieme, M. et al. Mapping the world's free-flowing rivers. Nature 569, 215–221 (2019). https://doi.org/10.1038/s41586-019-1111-9)

Most of the largest rivers worldwide are managed but models do not handle human factors well

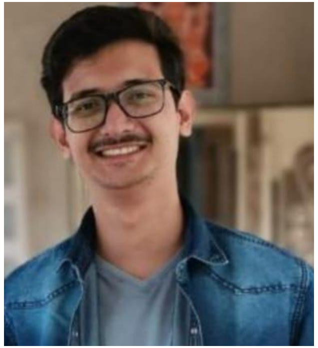
Only 37% of rivers longer than 1,000 kilometres remain free-flowing over their entire length and 23% flow uninterrupted to the ocean.

A Vision for Integrated Physics and Machine Learning in Hydrology Models



Key Contributors

Dr. Pravin Bhasme
Hydrological modeling, Physics informed machine learning



Sarth Dubey
Graph Neural Networks, Physics-Guided ML

References

- Boretti, A. and Rosa, L., 2019. Reassessing the projections of the world water development report. *NPJ Clean Water*, *2*(1), p.15.
- Krabbenhoft, C.A., Allen, G.H., Lin, P., Godsey, S.E., Allen, D.C., Burrows, R.M., DelVecchia, A.G., Fritz, K.M., Shanafield, M., Burgin, A.J. and Zimmer, M.A., 2022. Assessing placement bias of the global river gauge network. *Nature Sustainability*, *5*(7), pp.586-592.
- Chen, S., Zwart, J.A. and Jia, X., 2022, August. Physics-guided graph meta learning for predicting water temperature and streamflow in stream networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 2752-2761).
- Bhasme, P., Vagadiya, J. and **Bhatia, U.**, 2022. Enhancing predictive skills in physically-consistent way: Physics Informed Machine Learning for hydrological processes. *Journal of Hydrology*, *615*, p.128618.
- Bhasme, P. and **Bhatia, U.**, 2024. Improving the interpretability and predictive power of hydrological models: Applications for daily streamflow in managed and unmanaged catchments. *Journal of Hydrology*, *628*, p.130421.
- Thomas, H.A., 1981. Improved Methods for National Water Assessment. Report WR15249270, US Water Resource Council, Washington, DC
- Chiew, F., Peel, M., Western, A., et al., 2002. Application and testing of the simple rainfall-runoff model SIMHYD. In: Mathematical Models of Small Watershed Hydrology and Applications. Water Resources Publications, pp. 335–367.
- Nardi, F., Cudennec, C., Abrate, T., Allouch, C., Annis, A., Assumpcao, T., Aubert, A.H., Berod, D., Braccini, A.M., Buytaert, W. and Dasgupta, A., 2022. Citizens AND HYdrology (CANDHY): conceptualizing a transdisciplinary framework for citizen science addressing hydrological challenges. *Hydrological Sciences Journal*, *67*(16), pp.2534-2551.