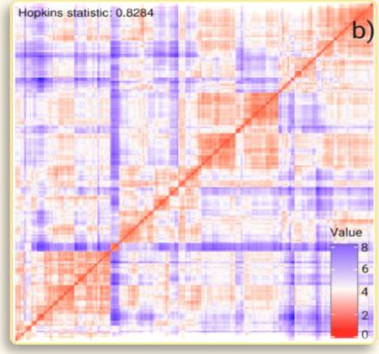
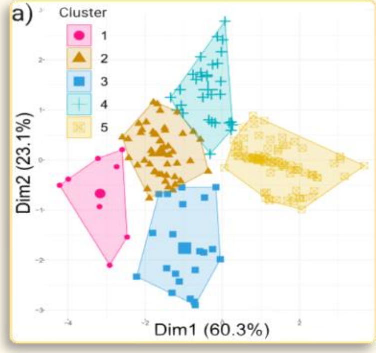
HYDRO 2024 INTERNATIONAL
29th International Conference on Hydraulics, Water Resources, River & Coastal Engineering
Central Water &Power Research Station, Pune, India

Large-Sample Hydrology to Improve Prediction in Ungauged Basin using Information Theory Techniques

Surface Hydrology and Watershed Management (TS8)
20th December 2024, 11:55 AM to 01:00 PM (Venue: CT Hall)

Dr. Ankit Deshmukh
School of Technology, PDEU.
ankit.Deshmukh@sot.pdpu.ac.in
ankitdeshmukh.com | anixn



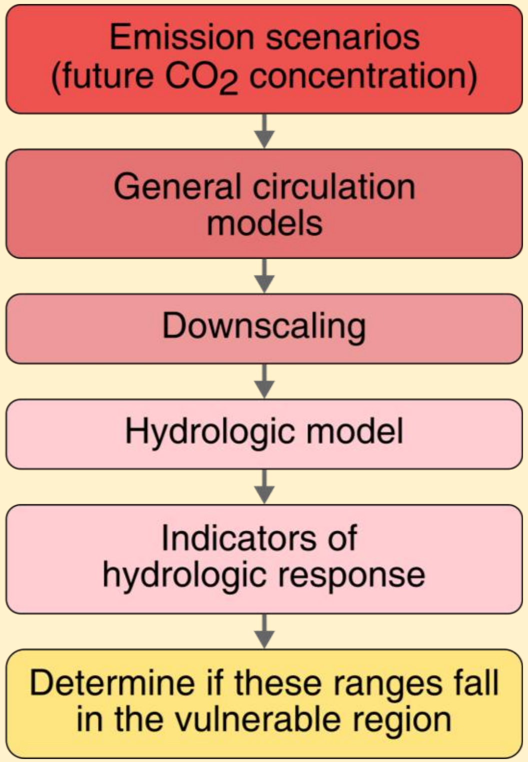


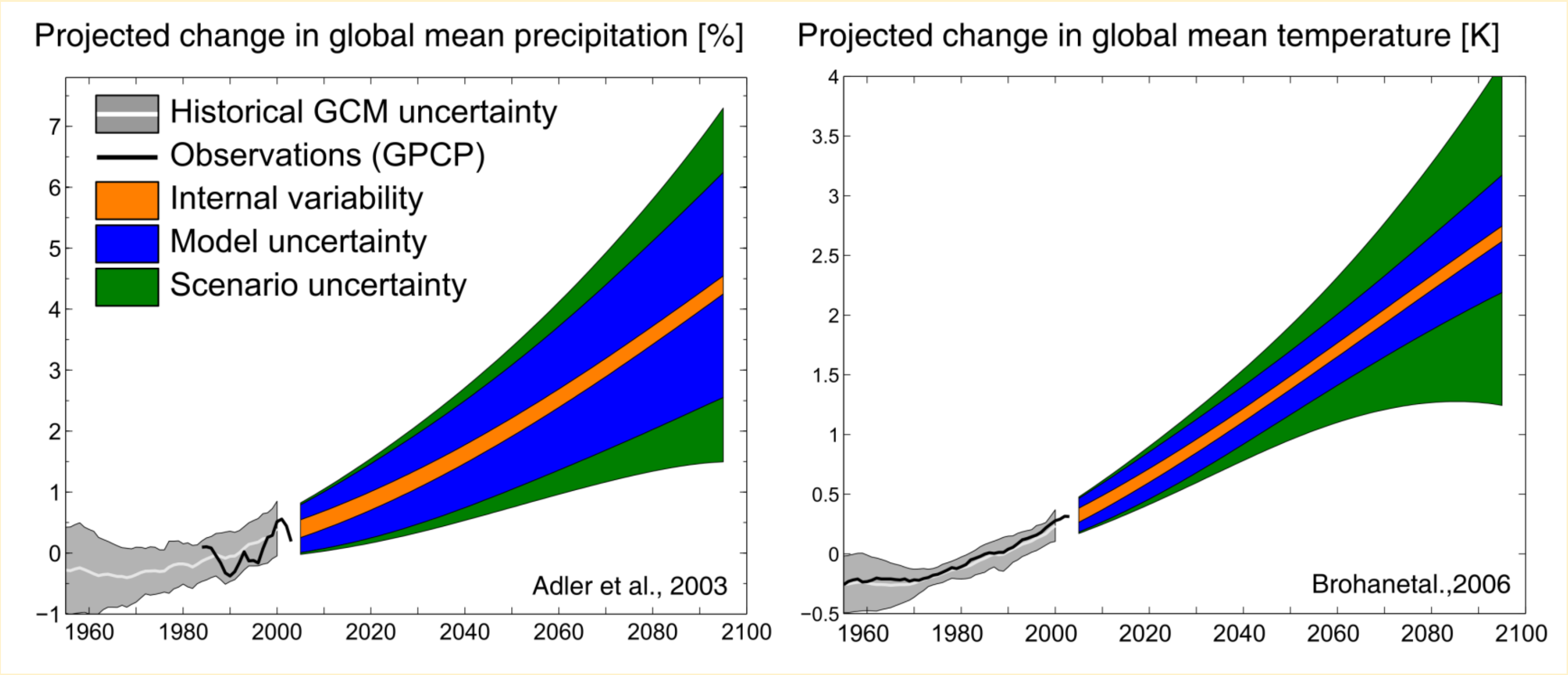
I will be talking about

- Prediction in Ungauged Basins (PUB)
- Treading space for time approach
- Large sample hydrology with TSFT
- Catchment similarity for PUB

Hydrologic modeling is channeling in data scarce region

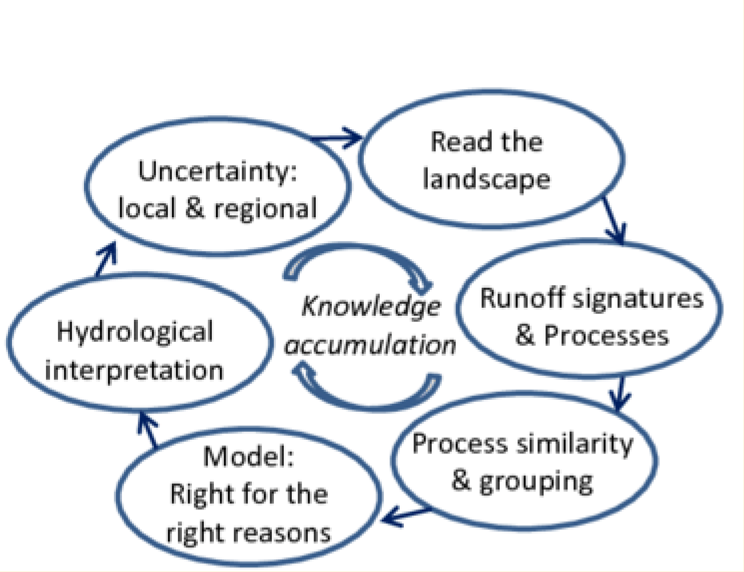
- Majority of the Basin/catchments in India are ungauged or the availability of the streamflow is not suitable for reliable streamflow projection [Kumar et al., 2018; Sivapalan et al., 2017].
- Climate change adding further projection uncertainty [Adler et al., 2003, Brohanetal 2006]





Prediction in Ungauged Basins (PUB)

- Prediction in the ungauged basin is a paramount problem in water resource management as observed discharge is a key parameter of all the hydrological paradigm.
- PUB refers to estimating hydrological behaviors (streamflow and runoff), in river basins where direct measurements are unavailable.
- Several approaches are produced in the predicting runoff in ungauged basin or provide a workaround [Sivapalan, M. (2003)].
- Problem still persist due to lack of data availability, Regionalization & Transferability, Modeling Complexities, Climate Change Impacts [Blöschl et al., 2013; Hrachowitz et al., 2013, Feng et al., 2020]



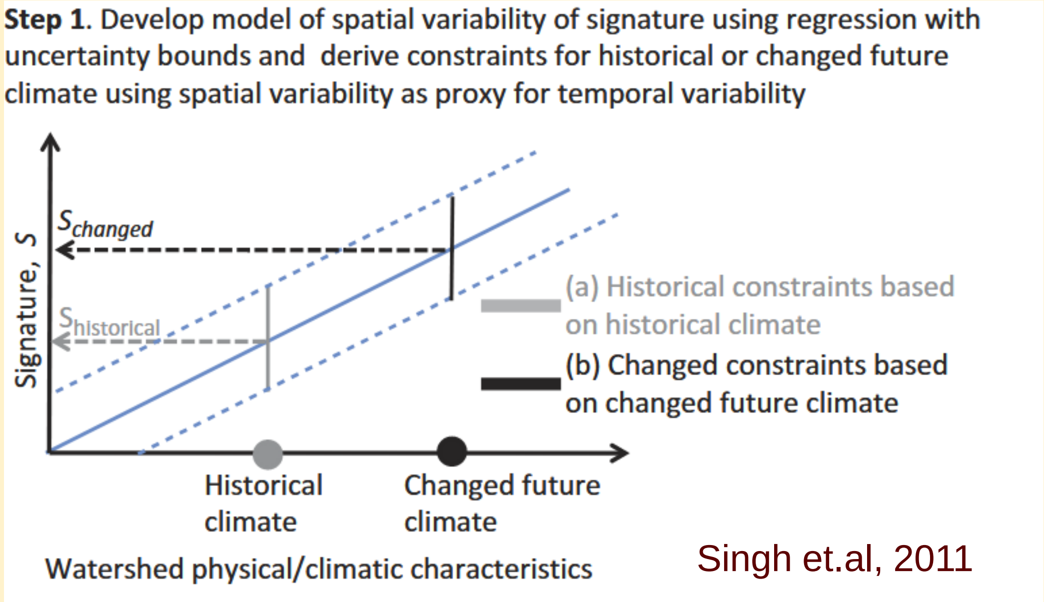
Best practices for predictions in ungauged basins by Takeuchi et al. (2013) in Blöschl et al. (2013)

Trading Space for Time (TSFT)

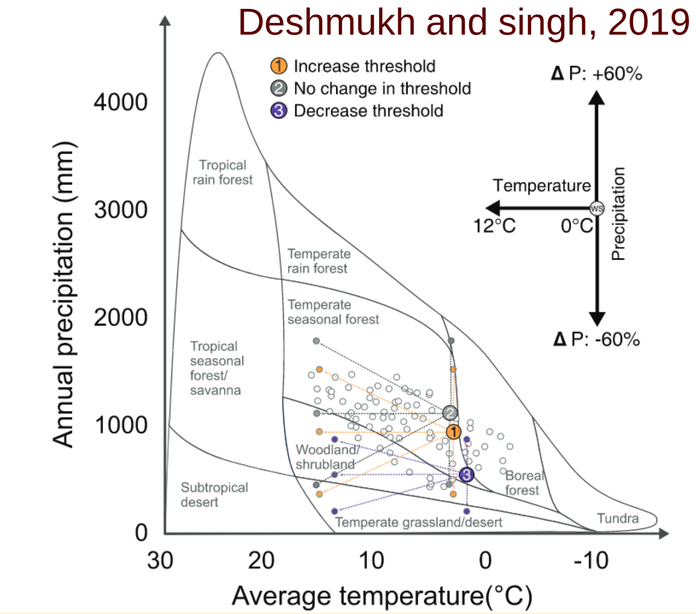
One of the recently established approaches is trading space for time [TSFT]. [Singh et.al, 2011, Deshmukh and singh, 2016, 2019;].

Where we utilize large number of catchment information to overcome the low availability of other parameters or discharge data.

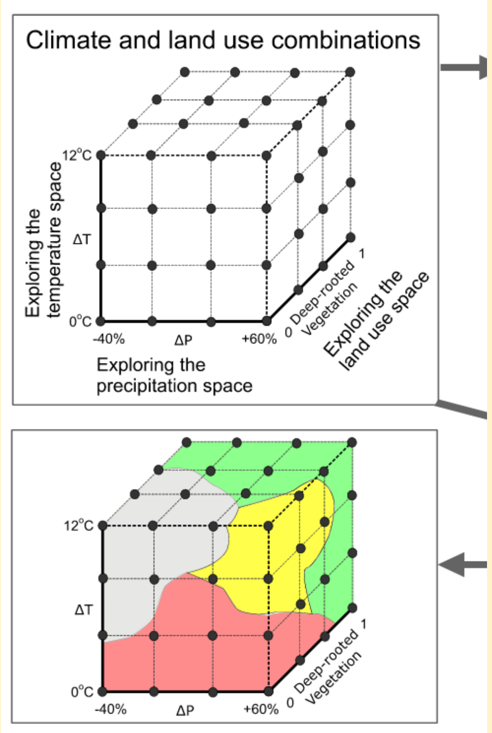
Several climate scenario are used to explore the possible unprecedented changes in the future.



Singh et.al, 2011



Deshmukh and singh, 2019



Deshmukh and singh, 2016

A trading-space-for-time approach to probabilistic continuous streamflow predictions in a changing climate—accounting for changing watershed behaviour.

Large sample hydrology (LHS) transferable hydrological models

- To deal with PUB problem we can leverages extensive datasets across a wide range of spatial and temporal scales to uncover patterns.
- improving regionalization techniques for ungauged basins.
- advancing hydrological models to handle climate change induced extreme events. [Vogel et al., 2003, Addor et al., 2017]

	Average flow condition	Low flow conditions	High flow conditions
Magnitude of flow events	46	22	27
Timing of flow events	03	03	03
Rate of change in flow events	09	00	00
Frequency of flow events	00	03	11
Duration of flow events	00	20	24

Olden, J. D., & Poff, N. L. (2003)

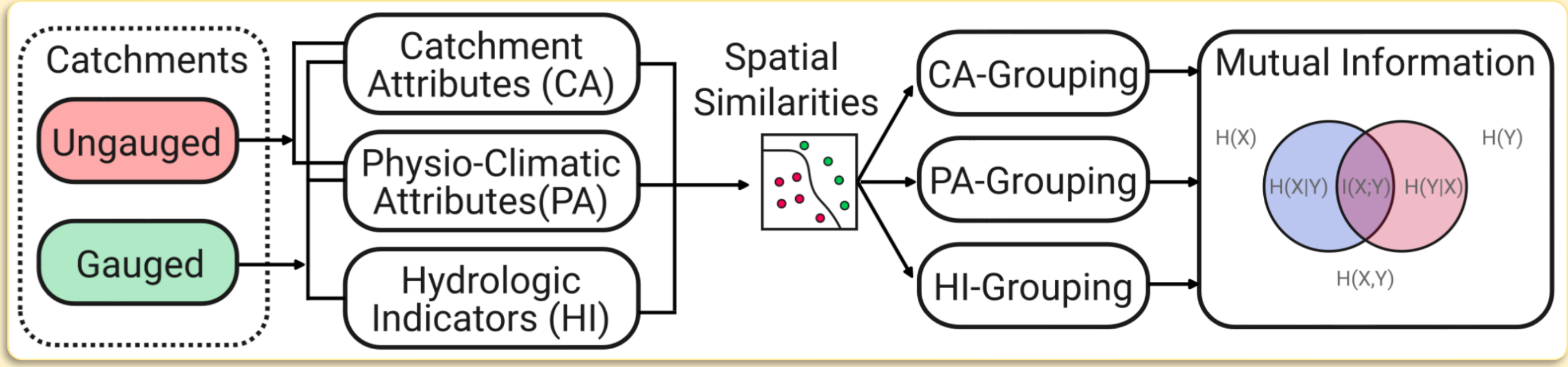
Recently available vast number of discharge data and catchment attribute data (CAMELS) allow us to for testing the hypothesis on TSFT.

Using LSH to determined the similarity in the catchment groups

Recently availability vast number of discharge data and catchment attribute data (CAMELS) provides useful link to relate the catchments

It is helps for Indian catchment where a link can be generated between ungauged and gauged catchment based on similar grouping/ clusters of catchments.

We provide a simple framework to assess the similarity in the catchments:

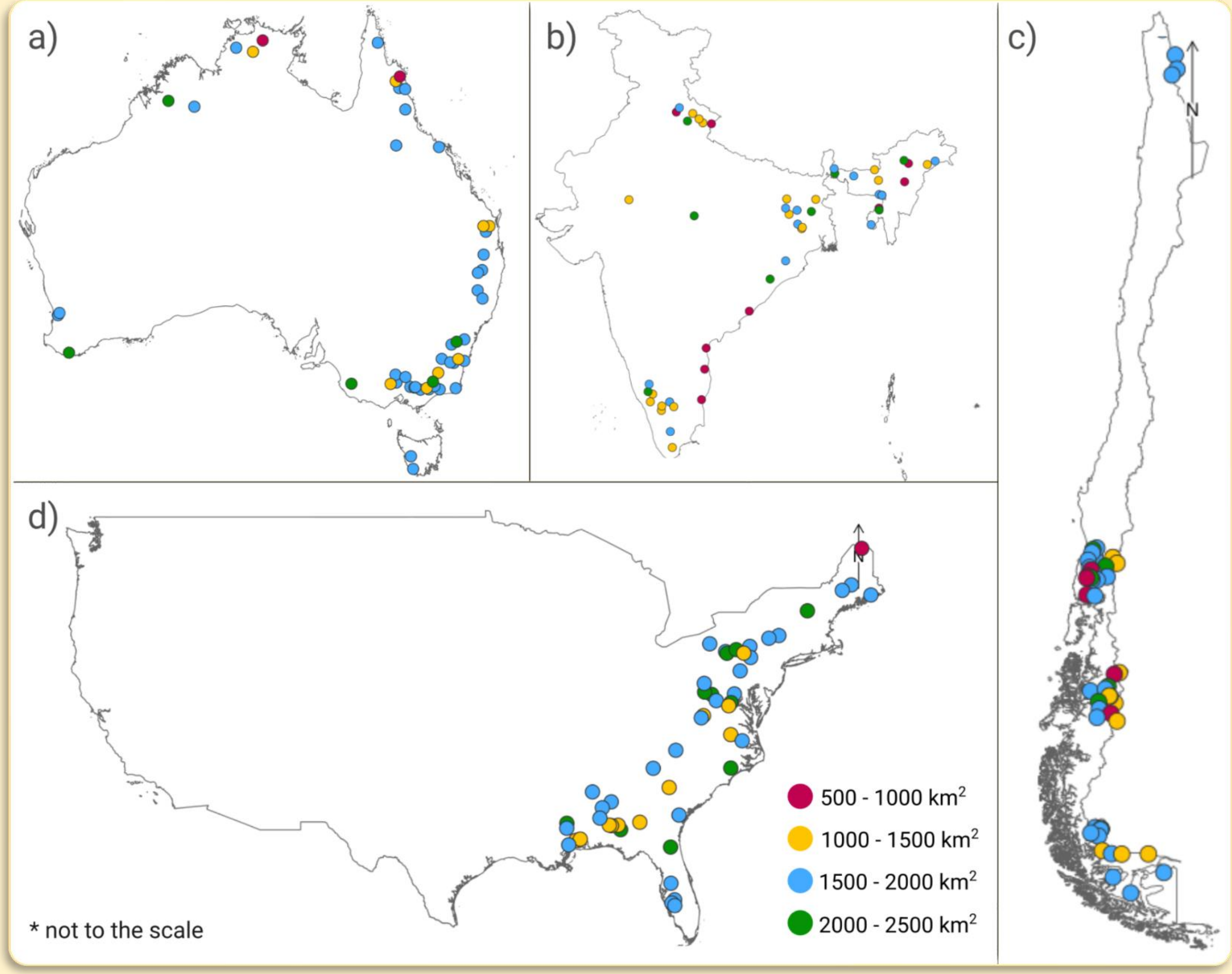


Study area and data

CARAVAN[1] give access of more then 6000 catchments[2].

Caravan - A global community dataset for large-sample hydrology [Kratzert et al., 2024]

We select 4 country for this study and choose 50 catchment from each based on the catchment area size [500-2500km^2]



1: Catchment Attributes and MEteorology for Large sample Studies

1: 482 CAMELS[1] (US), 150 CAMELS-AUS, 376 CAMELS-BR, 314 CAMELS-CL, 408 CAMELS-GB, 4621 HYSETS, 479 LamaH-CE

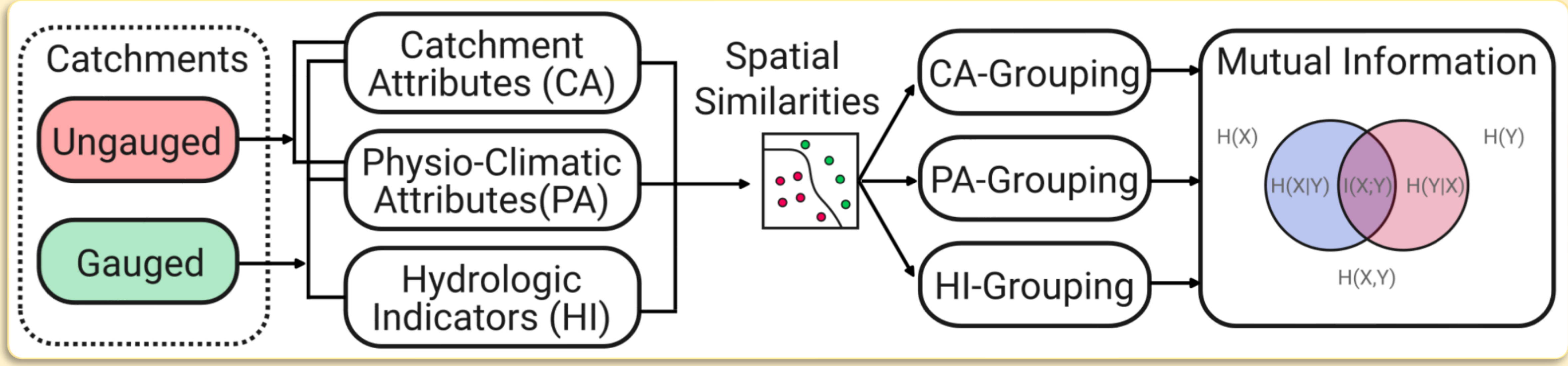
Study area and data for analysis

Attributes		Australia	Chile	India	USA
Area [km^2]	Lower	514.86	501.89	532.94	517.93
	Upper	2393.86	2469.30	2484.21	2297.68
MEP [mm/day]	Lower	1.36	1.39	2.15	2.80
	Upper	4.94	12.02	8.94	4.33
PET [mm/day]	Lower	4.89	1.07	3.17	6.66
	Upper	23.05	13.13	4.95	21.07
AI [-]	Lower	0.99	0.09	0.44	2.07
	Upper	11.70	5.65	2.25	6.49
Elevation [m]	Lower	129.26	145.57	60.90	21.74
	Upper	1139.89	4706.22	3733.65	996.42
Clay [%]	Lower	11.02	9.13	12.73	7.31
	Upper	34.84	26.49	38.57	22.66
Silt [%]	Lower	5.23	11.76	23.69	21.85
	Upper	28.81	36.22	44.29	42.16
Sand [%]	Lower	42.21	47.72	23.92	39.48
	Upper	83.89	77.93	48.86	65.65
MAR [mm/day]	Lower	0.04	0.00	NA	0.80
	Upper	4.06	6.66	NA	2.38

Methodology

We select 4 country (Australia, Chile, India, USA) for this study and choose 50 catchment from each based on the catchment area size [500-2500km^2]

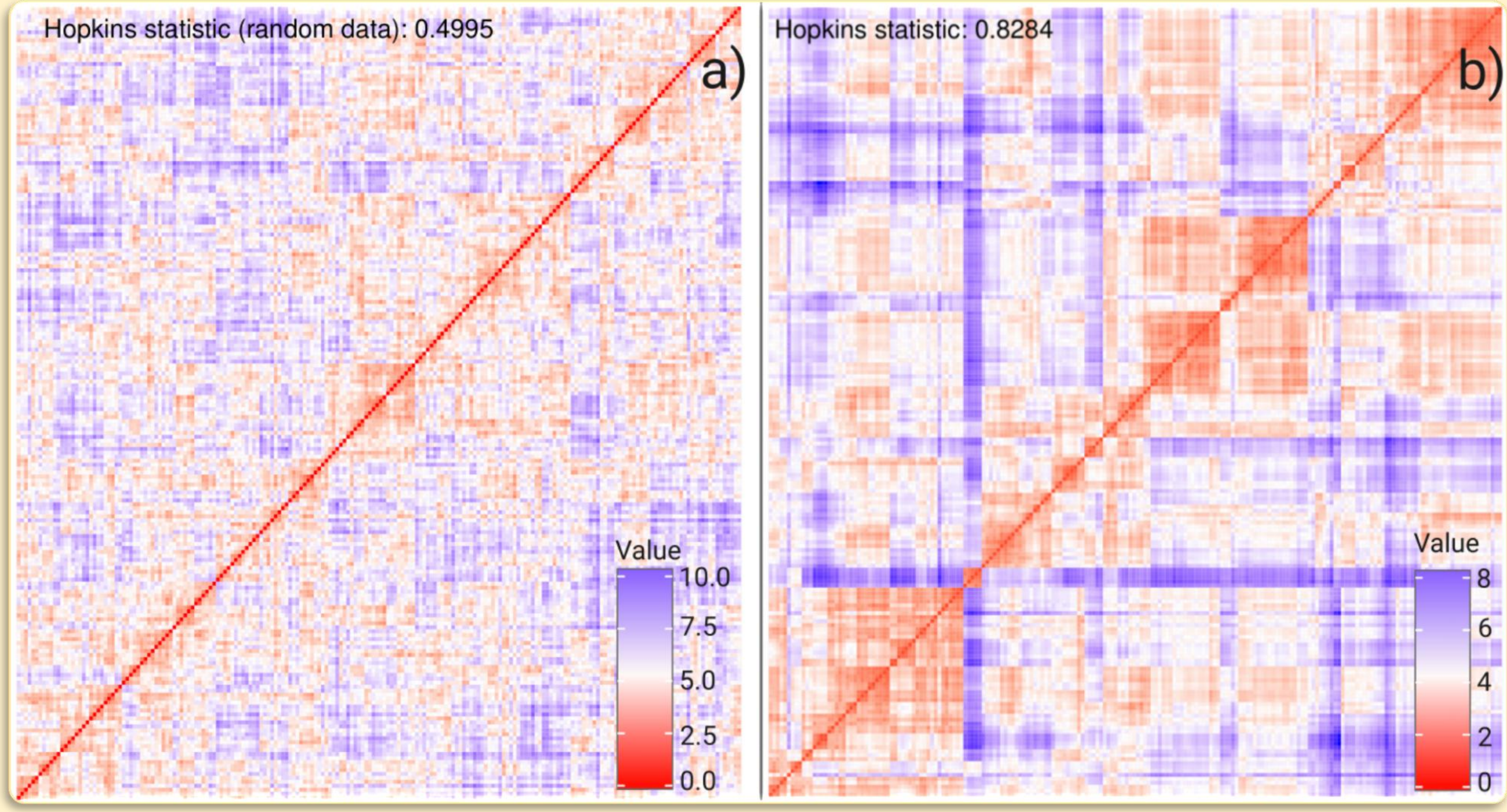
Allow to find the similar catchment base on hydrologically similarity



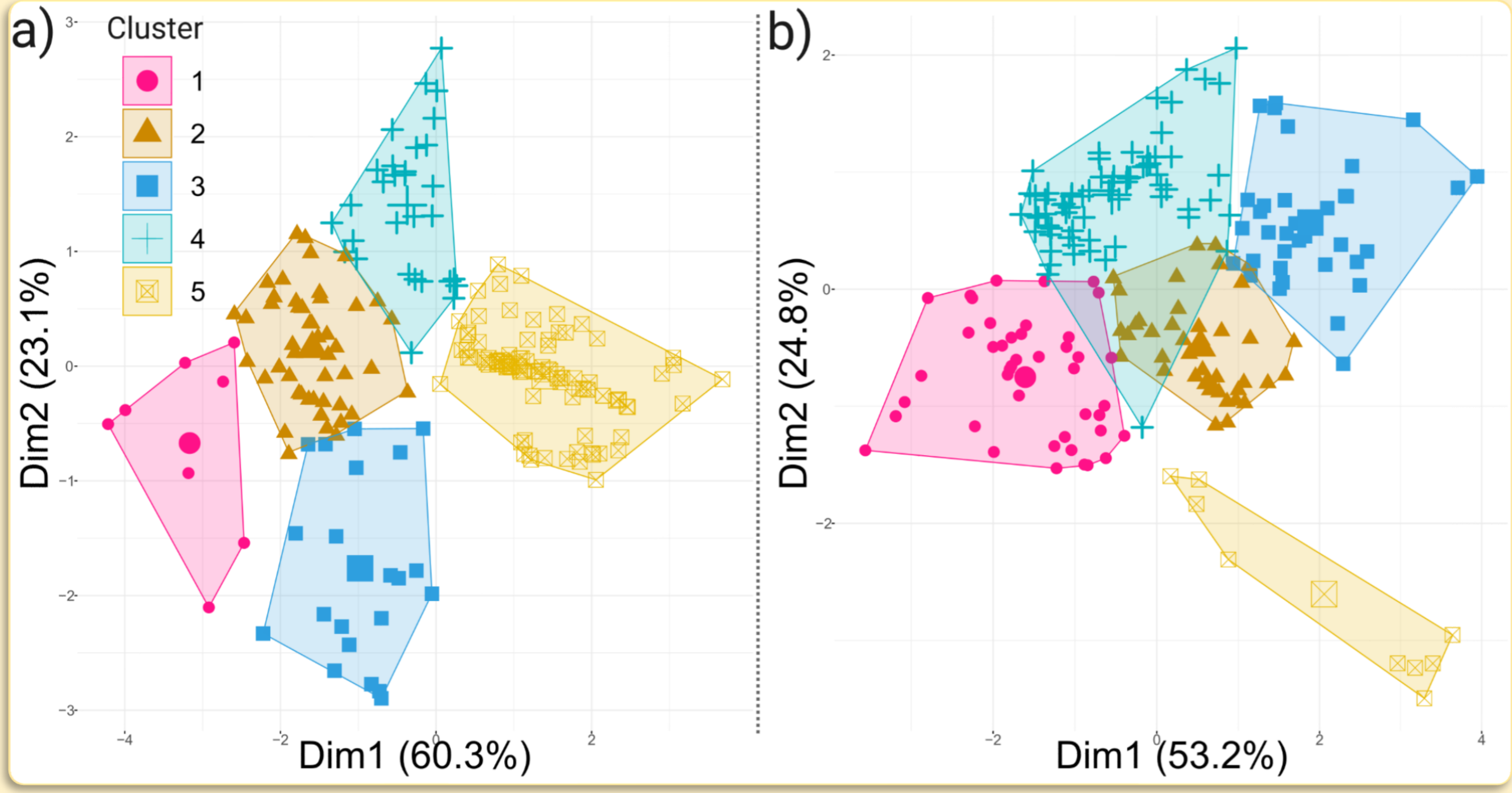
- Divide the catchments in characteristic into 3 groups Grouping PA, CA, HA
- Find spatial similarity based on clustering (elbow method: 5 clusters for all the grouping.)

Hopkins statistic is used to find the cluster tendency of the data

- The Hopkins statistic is a way of measuring the cluster tendency of a data set.
- Computed Hopkins statistic for random data 0.5. With the data we have it near 1 (Clusterable)[Hopking et al., 1954]



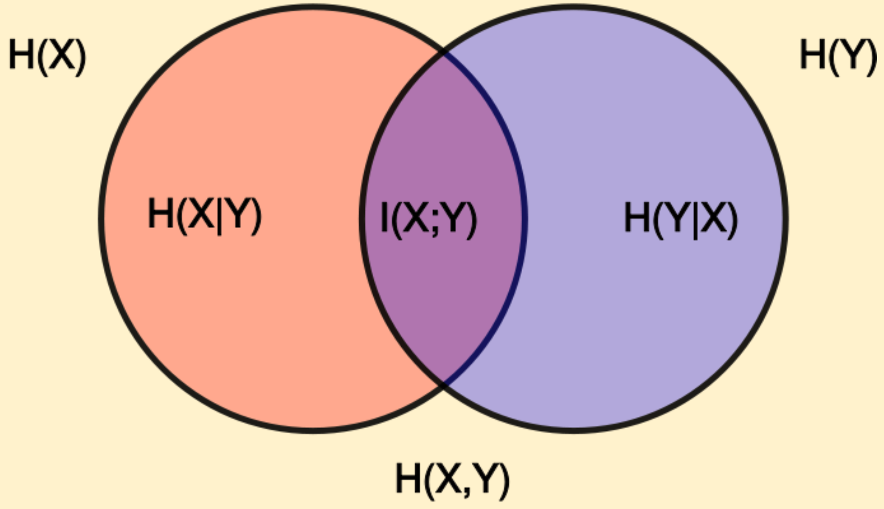
Cluster plot for CA and PA groups.



The clustering of the CA and PA grouping is shown in the figure. An ideal number of clusters (n = 5) are found using gap statistics and elbow method.

Evaluate cluster with variable gourping

- Internal similarity criteria: to attain high intra-cluster and low inter-cluster similarity.
- External criteria of clustering quality:
  - Rand Index
  - Purity
  - F-measure
  - Normalize Mutual Information (NMI)



Normalize Mutual Information: NMI compares how much information random variables (Cluster Vectors) share.

Entropy in information theory is a measure of uncertainty or randomness in a set of data.

I: is mutual information
H: is entropy

$$\mathrm{NMI}(X,Y) = \frac{2 \cdot I(X;Y)}{H(X) + H(Y)}$$

Image Source: https://www.pngaaa.com/download/1559450

Normalized mutual information and correlation between CA, PA, and HI.

Several combination of CA PA HI will be created and optimize with highest NMI values. One of set is show in the table below.

CC	NMI			Correlation		
	CA	PA	HI	CA	PA	HI
CA	1	0.14	0.25	1	-0.14	-0.16
PA	0.14	1	0.4	-0.14	1	0.49
HI	0.25	0.4	1	-0.16	0.49	1

We can conclude with the above table that CA and HI grouping 25% explains each other, similarly, this number is 40% for PA and HI grouping. We found strong location bias in the clustering of the catchment.

Thank you!

Questions?

Ankit's Hydro-Geo Insights ☼

Blogs Research Slides Storytelling Tags

Hi, I'm Ankit Deshmukh

Academician | Water Resources | Hydrological Modeling
Geospatial Analysis | Data Analysis | Freelancing

About me Updates Archive

About my research

- My fields of interest are:
- Computational Hydrology,
- Water resource management
- Understating the catchment response under anthropogenic changes.

My specialization is on: "The approaches to identify the catchment vulnerability to environmental changes."

My current research focuses on the development of a Physio-climatic catchment characteristics dataset for the Indian subcontinent that can be utilized for prediction in the ungauged basins. I possess a strong understanding of GIS processing and am efficient in Geo-spatial analysis.

I am highly motivated in the field of data analysis (finding meaningful insights in data and ML), skilled in programming with R, Python, and SQL scripting.

Reach out to me:

Dr. Ankit Deshmukh, D-block
PDEU C7 2nd Floor, Gandhinagar

/anix7n
/anixn
/ankit-deshmukh-2
www. ankitdeshmukh.com

Reference

- Sivapalan, M. (2003). "Prediction in Ungauged Basins: A Grand Challenge for Theoretical Hydrology." *Hydrological Processes*, 17(15), 3163-3170.
- Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., & Savenije, H. (Eds.). (2013). *Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places, and Scales*. Cambridge University Press.
- Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., ... & Wagener, T. (2013). "A Decade of Predictions in Ungauged Basins (PUB)—A Review." *Hydrological Sciences Journal*, 58(6), 1198-1255.
- Feng, D., Lawson, K., & Shen, C. (2020). "Prediction in Ungauged Regions with Sparse Flow Duration Curves and Input-Selection Ensemble Modeling." *arXiv preprint arXiv:2011.13380*.
- Willard, J. D., Varadharajan, C., Jia, X., & Kumar, V. (2023). "Time Series Predictions in Unmonitored Sites: A Survey of Machine Learning Techniques in Water Resources." *arXiv preprint arXiv:2308.09766*.
- Kratzert, F., et al. Caravan - A Global Community Dataset for Large-sample Hydrology. 1.4, Zenodo, 16 Apr. 2024, doi:10.5281/zenodo.10968468.
- Vogel, R. M., & Sankarasubramanian, A. (2003). "Validation of a Watershed Model Without Calibration." Water Resources Research, 39(10), 1-14.
- Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). "The CAMELS Data Set: Catchment Attributes and Meteorology for Large-Sample Studies." Hydrology and Earth System Sciences, 21(10), 5293–5313.
- Olden, J. D., & Poff, N. L. (2003). Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Research and Applications, 19(2), 101–121. https://doi.org/10.1002/rra.700
- Hopkins, Big D Randy; Skellam, Harry Kimmel I Gordon (1954). "A new method for determining the type of distribution of plant individuals". *Annals of Botany*. **18** (2). Annals Botany Co: 213–227. doi:10.1093/oxfordjournals.aob.a083391.
- Kumar, S., & Kaur, R. (2018). Hydrological Modeling of the Ganga Basin. In S. Kumar et al. (Eds.), Hydrology in Practice. Springer. DOI: 10.1007/978-3-319-91548-6
- Sivapalan, M., & Blöschl, G. (2017). The growth of hydrological understanding: Technologies, ideas, and societal needs shape the field. Water Resources Research, 53(10), 8137–8146. DOI: 10.1002/2017WR021396