EGU General Assembly 2019
Vienna | Austria | 7–12 April

Exploratory insights into hydrological applications of data clustering

Nilay Dogulu

nilay.dogulu@gmail.com

DoguluNilay

CITATION:

Dogulu, N. (2019) *Exploratory insights into hydrological applications of data clustering.*
Geophysical Research Abstracts, Vol. 21, **EGU2019-12099**,
European Geosciences Union General Assembly,
Vienna, Austria, 7 – 12 *April* 2019.

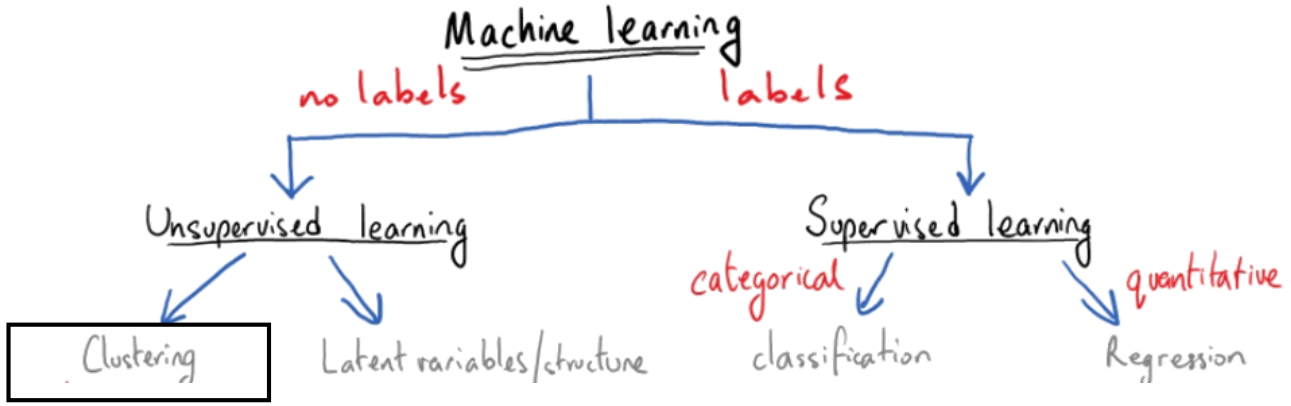
Clustering?
Algorithm?

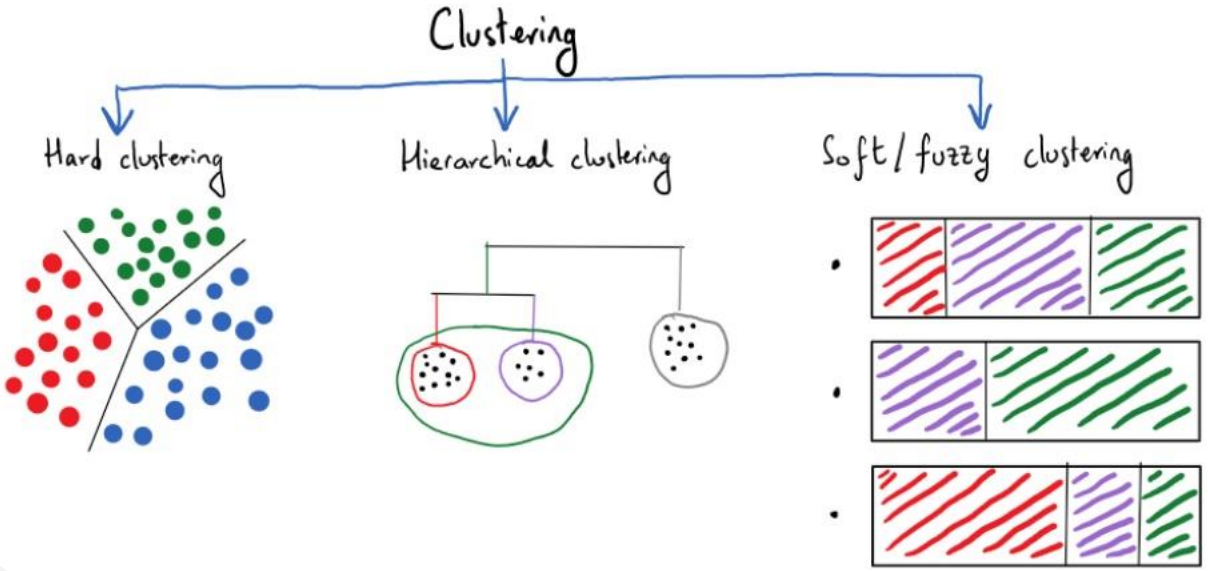
Art by Oliver Flores

Yes, I used !

Yes, I will use !

Clustering?
Algorithm?





http://2016.igem.org/Team:Kent/Model

Clustering algorithms

Wide range of applications in diverse sciences

ASTRONOMY

Photo by Miguel Claro

> “The cosmos is within us. We are made of star-stuff. We are a way for the universe to know itself.”
>
> — Carl Sagan

BIOLOGY

Art by Natalie Spencer

> “Biology is the study of the complex things in the Universe. Physics is the study of the simple ones.”
>
> — Richard Dawkings

GEOSCIENCES

Image by NASA Earth Observatory

> “Nature brings us back to absolute truth whenever we wander.”
>
> — Louis Agassiz

Clustering algorithms

Wide range of applications in diverse sciences

HYDROLOGY

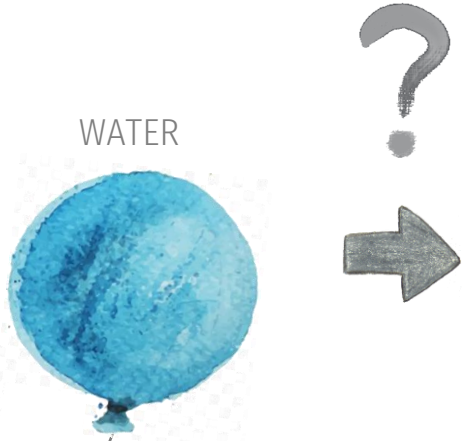
the function and behaviour
of WATER on EARTH

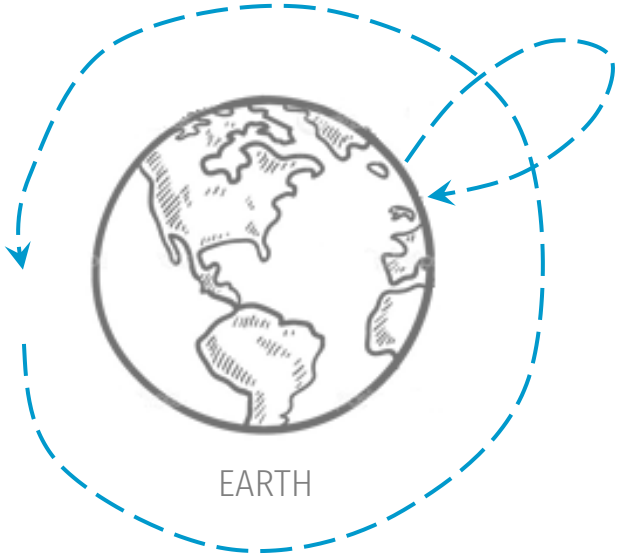
YOU

HOW ?

WHEN ?

WHERE ?





Clustering of hydrological data offers rich insights into diverse concepts and relations that underline heterogeneity and complexity inherent in catchment systems.

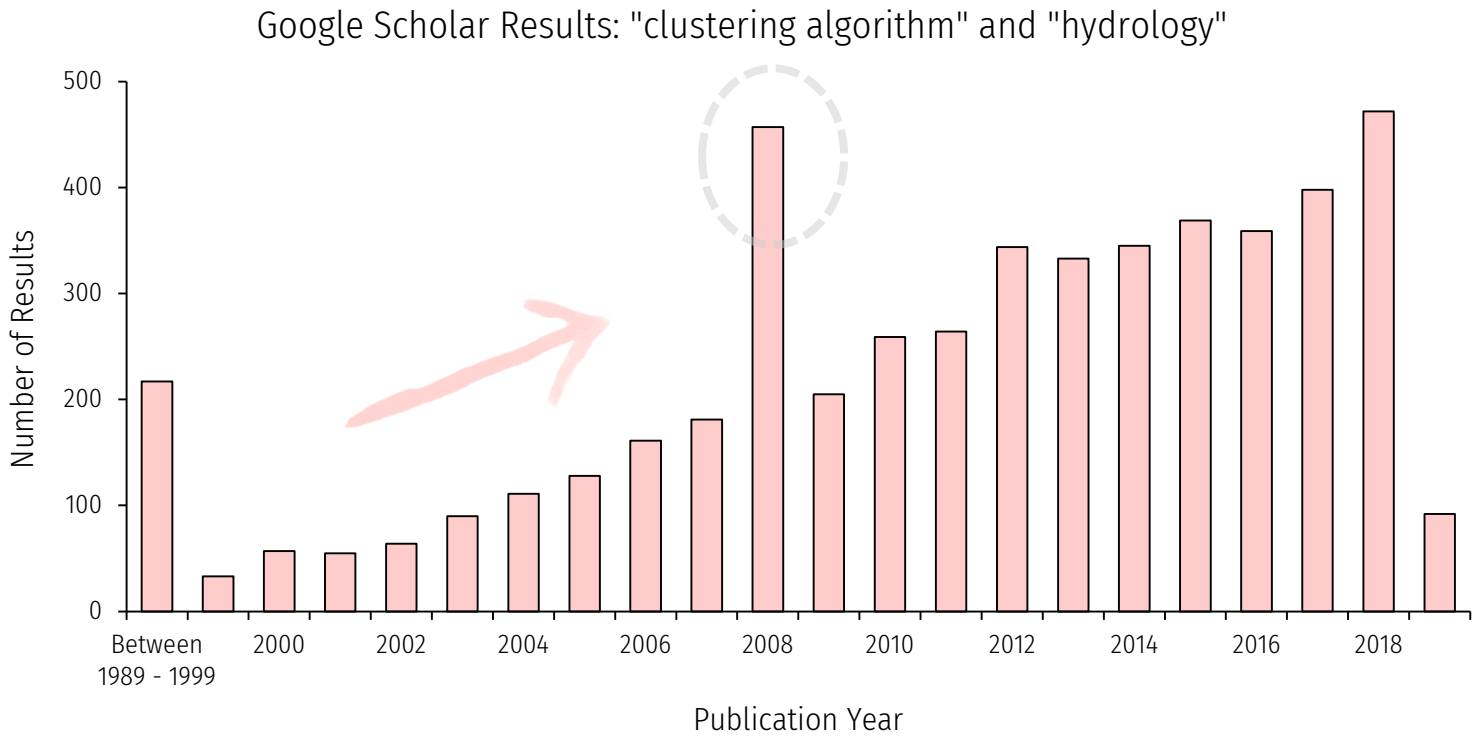
CLUSTERING ALGORITHMS IN HYDROLOGY

CITATION:

Dogulu, N. (2019) *Exploratory insights into hydrological applications of data clustering.*
Geophysical Research Abstracts, Vol. 21, **EGU2019-12099**,
European Geosciences Union General Assembly,
Vienna, Austria, 7 – 12 *April* 2019.

EGU General Assembly 2019
Vienna | Austria | 7–12 April

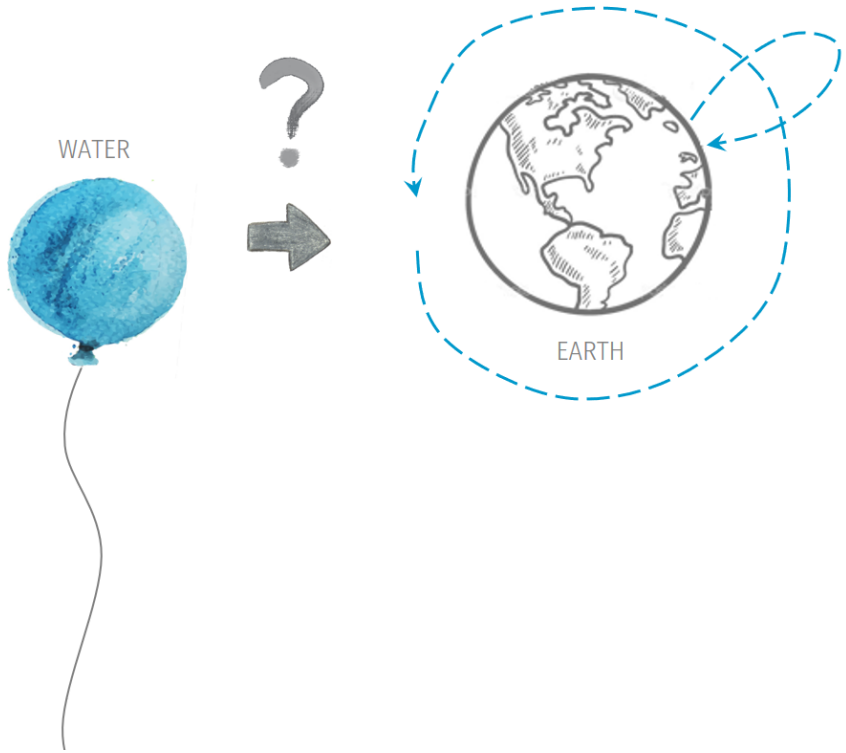
Clustering algorithms in Hydrology



High increase,

possibly due to higher availability of computational power as well as (large) open datasets.

Clustering algorithms in Hydrology



DATA	Clustering of catchments Clustering of Q time series Clustering of precipitation …
CONTEXT	Catchment similarity Regionalization Data-driven hydrologic modelling …
METHOD	K-means clustering Hierarchical clustering …

Clustering algorithms in Hydrology

LITTLE RESEARCH ON USE OF CLUSTERING ALGORITHMS FOR HYDROLOGICAL APPLICATIONS

Appendix 2

Protocols for hydrologic classification and a review of Australian applications

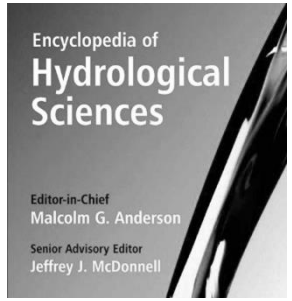
J.D. Olden[1], C.A. Reidy Liermann[1], B.J. Pusey[2] and M.J. Kennard[2]

2009

19: Data-driven Modeling and Computational Intelligence Methods in Hydrology

DIMITRI P SOLOMATINE

"In hydrology, clustering and classification are used much less frequently than prediction algorithms."



2006

International Journal of Scientific Engineering and Technology
Volume No.3 Issue No.4, pp : 359-363
(ISSN : 2277-1581)
1April 2014

Application Of Clustering Data Mining Techniques In Temporal Data Sets Of Hydrology: A Review

Rakesh Purviya[1], H.L. Tiwari[2] and Satanand Mishra[2]

2014

International Journal of Computer Applications (0975 – 8887)
Volume 117 – No. 23, May 2015

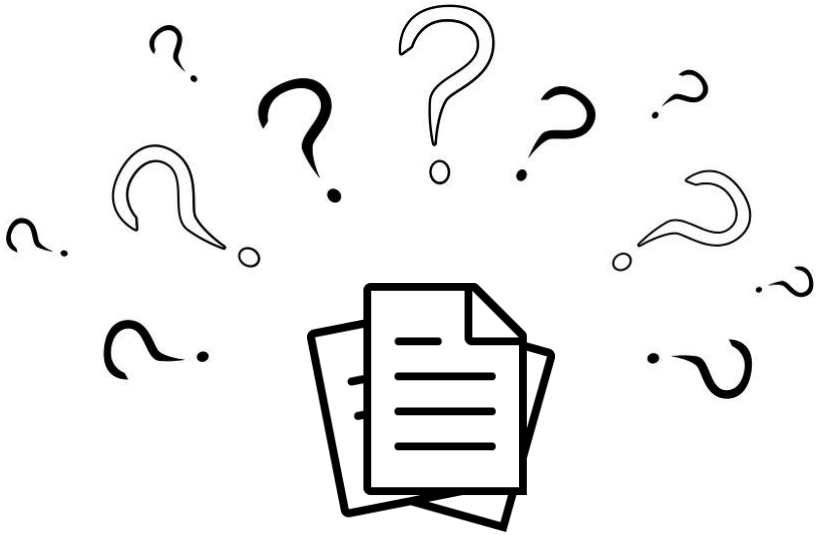
Study of Time Series Data Mining for the Real Time Hydrological Forecasting: A Review

Satanand Mishra, C. Saravanan, V. K. Dwivedi,

2015

Clustering algorithms in Hydrology

RAPID GROWTH OF THE LITERATURE, YET THE EXTENT AND IMPLICATIONS OF SUCH GROWTH IS UNKNOWN.



NEED FOR A COMPREHENSIVE REVIEW

METHODOLOGY

CITATION:

Dogulu, N. (2019) *Exploratory insights into hydrological applications of data clustering.*
Geophysical Research Abstracts, Vol. 21, **EGU2019-12099**,
European Geosciences Union General Assembly,
Vienna, Austria, 7 – 12 *April* 2019.

EGU General Assembly 2019
Vienna | Austria | 7–12 April

Methodology

A comprehensive analysis and synthesis of selected journal articles from the scientific literature (1999 – 2019)

Journal Articles

Database

Analysis

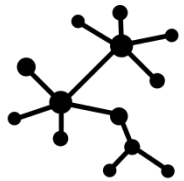
206 — Number of articles selected for review (as of April 1, 2019)

Article — Year, Journal, Research theme, Keywords, Research objective

Clustering — Purpose, Algorithm(s) used, Justification for method selection, # of clusters and how it is determined, Variables, Cluster validity

Case study — Country, Number of catchments, Data period

Quantitative
Qualitative


Database

Journals

57 Journals
with a variety of
focus areas

HYDROLOGY (20)

- Journal of Hydrology
- Hydrological Processes
- Water Resources Research
- Hydrology and Earth System Sciences
- Hydrological Sciences Journal
- Journal of Hydroinformatics
- Journal of Hydrologic Engineering
- Water
- Hydrology Research
- Water Resources Management
- Journal of River Basin Management
- River Research and Applications
- Ecohydrology
- Advances in Water Resources
- Journal of the American Water Resources Association
- Journal of Water Resources Planning and Management
- Journal of Applied Water Engineering and Research
- Canadian Water Resources Journal
- Journal of Hydrology and Hydromechanics
- Water Practice & Technology

CLIMATE (9)

- International Journal of Climatology
- Theoretical and Applied Climatology
- Journal of Climate
- Climate Research
- Climate
- Weather and Climate Extremes
- Atmosphere
- Atmospheric Research
- Monthly Weather Review

OTHERS (9)

- Journal of Arid Environments
- Journal of Mountain Science
- Journal of Glaciology
- Limnology and Oceanography
- Natural Hazards

- Remote Sensing of Environment
- Remote Sensing Applications

- PLOS ONE
- Arabian Journal for Science and Engineering

INFORMATICS (7)

- Environmental Modelling & Software
- Computer & Geosciences
- Journal of Environmental Informatics
- Neural Networks
- Stochastic Environmental Research and Risk Assessment
- Applied Soft Computing
- Entropy

ENVIRONMENT (12)

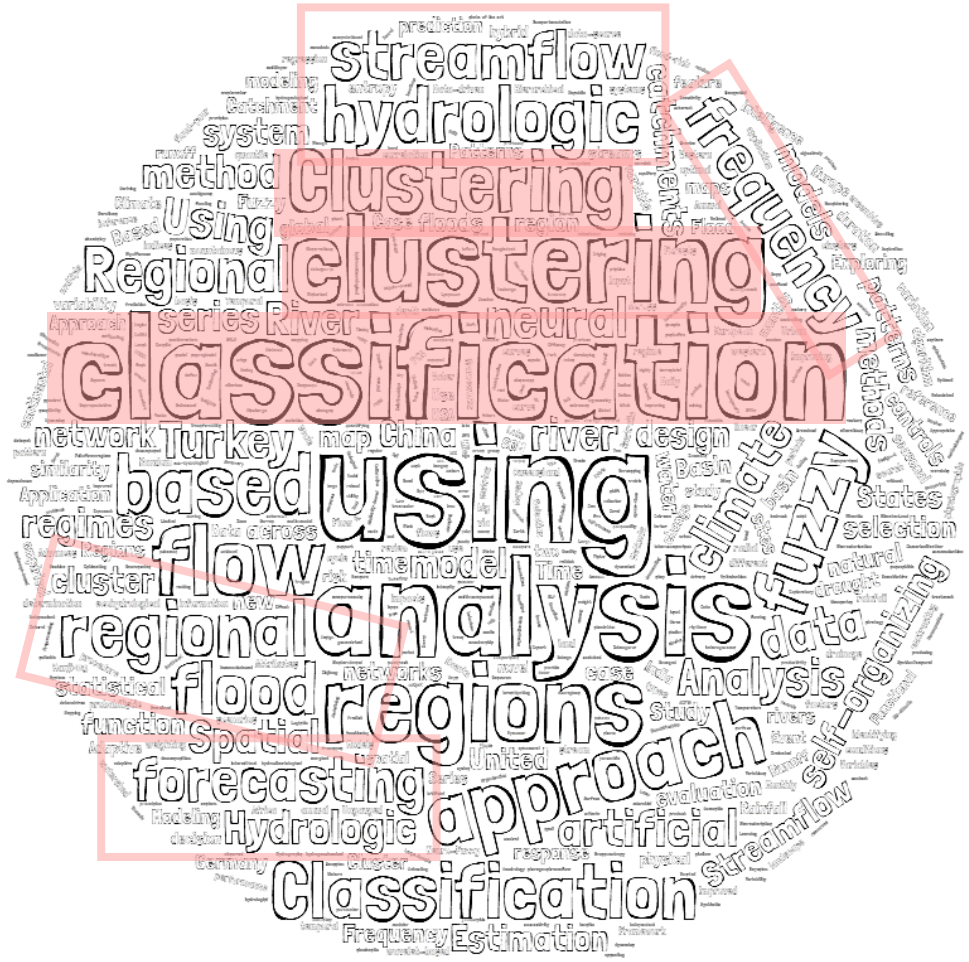
- Environmental Management
- Environmental and Ecological Statistics
- Earth Sciences Research Journal
- Freshwater Science
- Freshwater Biology
- Physics and Chemistry of the Earth
- Physical Geography
- Geographical Review
- Geographical Research Letters
- Ecological Indicators
- Ecography
- Conservation Biology

Database

Journal articles

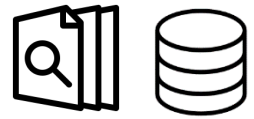
KEYWORDS

TITLES



Database

RESEARCH THEMES



uncertainty quantification
input variable selection

RESULTS

CITATION:

Dogulu, N. (2019) *Exploratory insights into hydrological applications of data clustering.*
Geophysical Research Abstracts, Vol. 21, **EGU2019-12099**,
European Geosciences Union General Assembly,
Vienna, Austria, 7 – 12 *April* 2019.

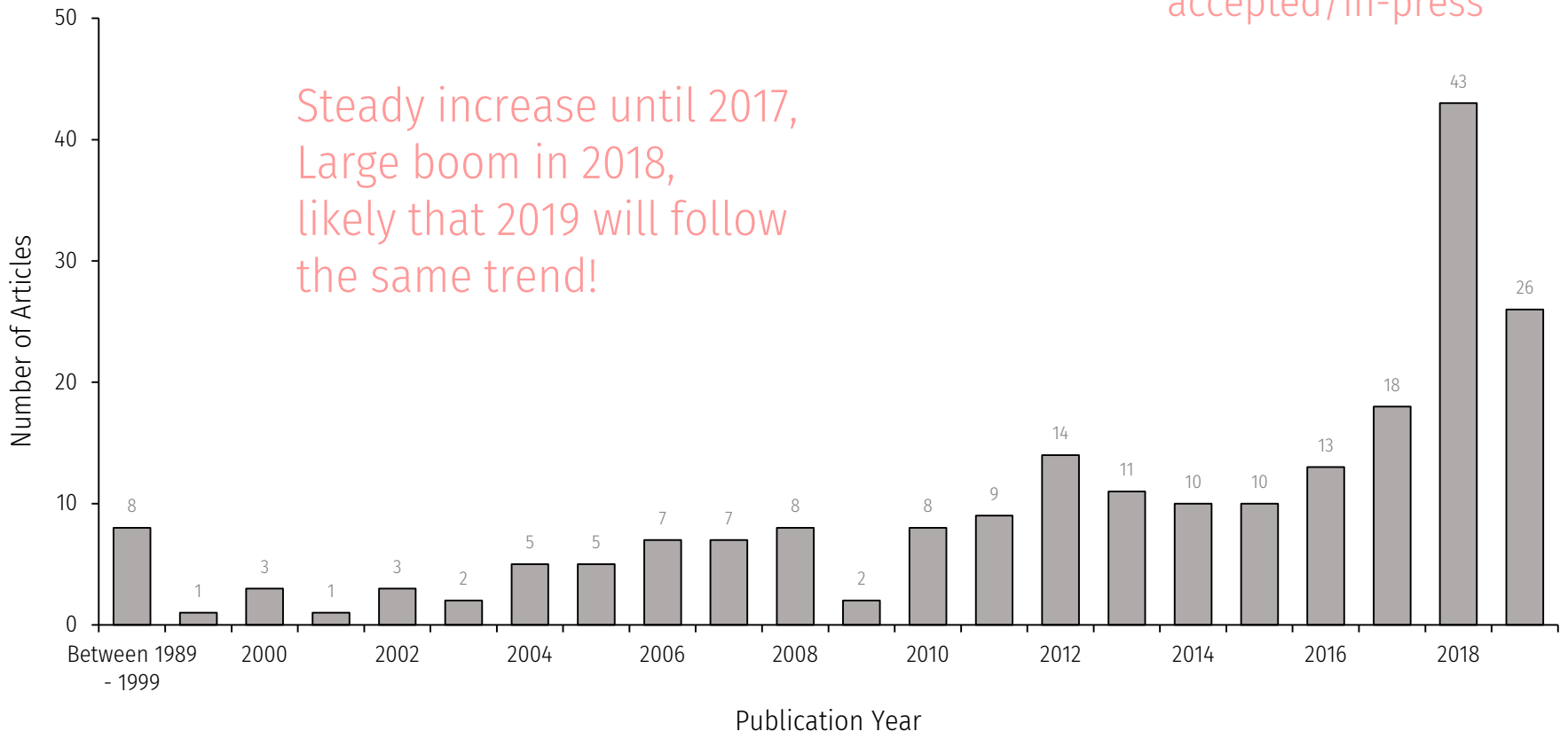
EGU General Assembly 2019
Vienna | Austria | 7–12 April

Results

Journal articles – BY YEAR

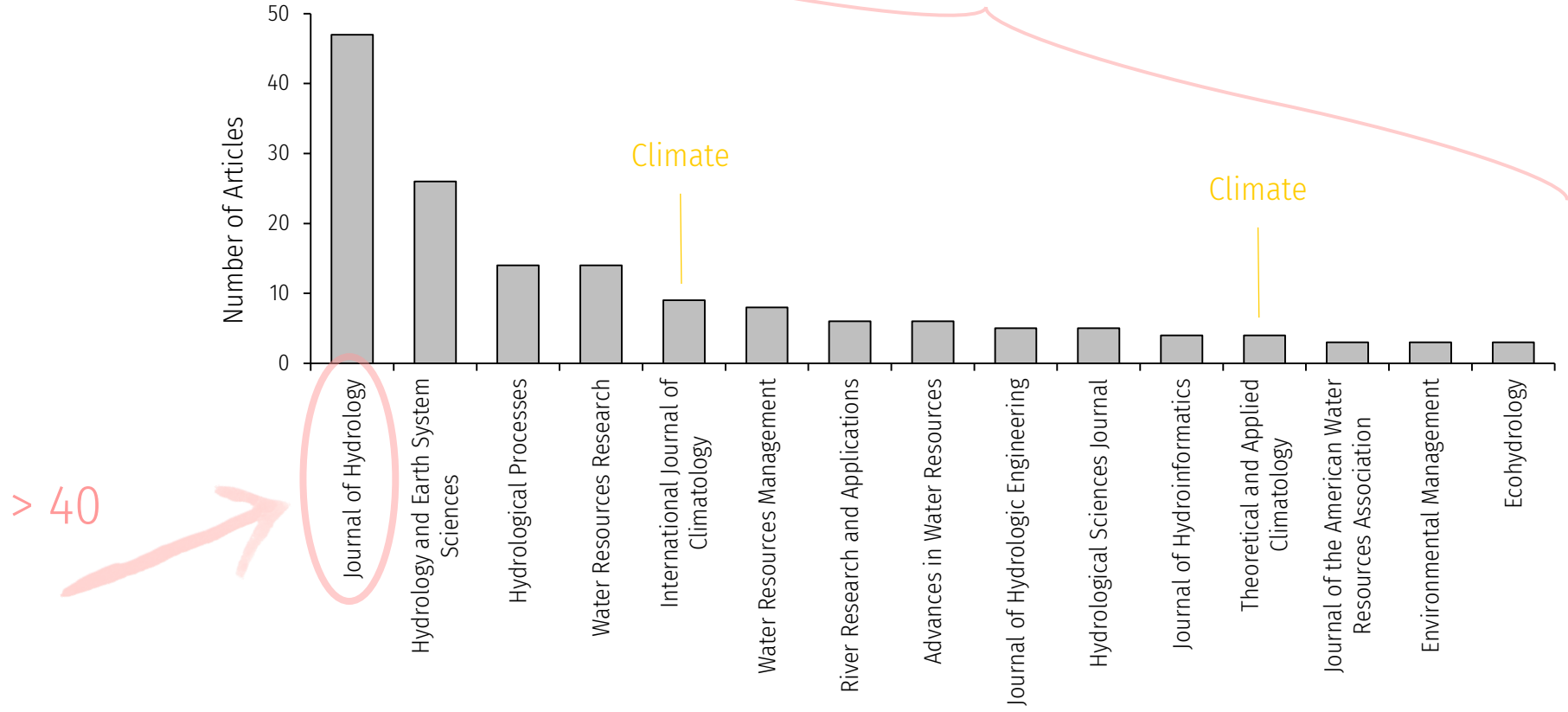
2019:
Only first three months, including accepted/in-press

Steady increase until 2017,
Large boom in 2018,
likely that 2019 will follow the same trend!



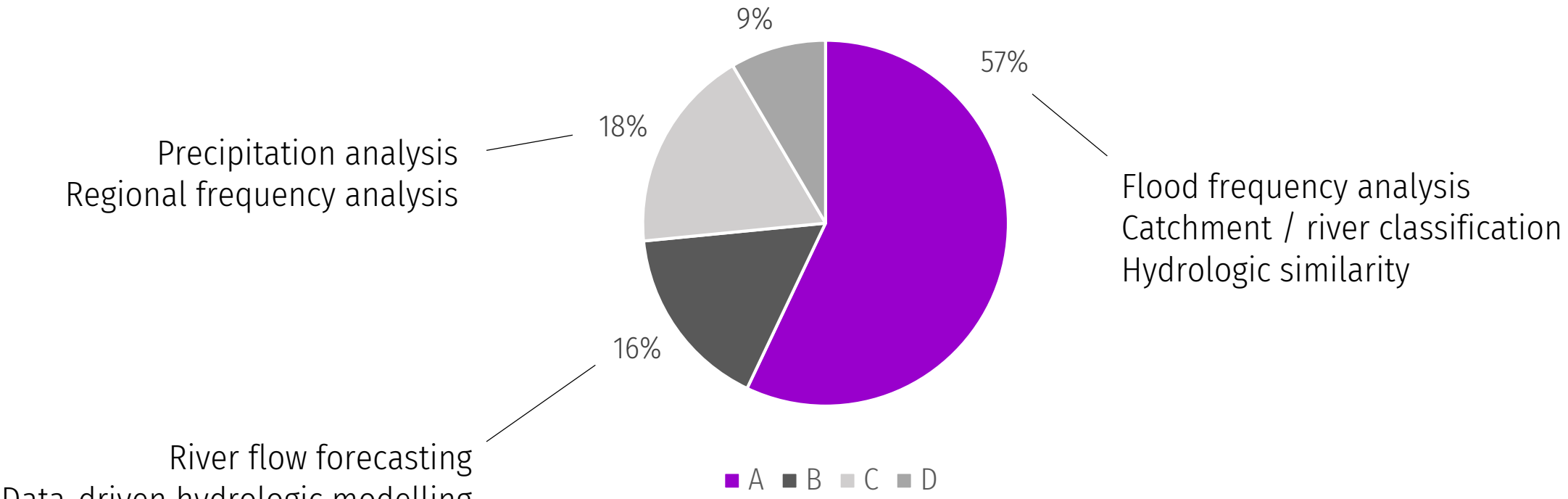
Results

Journal articles – BY JOURNAL (TOP 15)



Results

Journal articles – BY RESEARCH THEME



Results

Clustering methods

Which methods?

Hierarchical clustering
Agglomerative, Ward's minimum variance

K-means

Fuzzy c-means

SELF-ORGANIZING MAP (SOM)

Bayesian mixture clustering

Consensus clustering

Subtractive clustering

Early 2000

2018

Results

Clustering methods

Why method “X”?

No direct statement in most articles

Reference to previous literature

Comparative studies

“X” method has not been explored
“X” method is better than “Y” method

Results

Cluster validity & (optimal) number of clusters

Number of clusters?

Case specific:

Depends on purpose, context, level of detail required

How is optimal number of clusters determined?

Trial-and-error
User-defined
Data replication
Heterogeneity test
Dendogram
Using SOM
Hierarchical clustering
Silhouette coefficients
Cluster validation measures (partition index, separation index, Xie-Beni index)

Results

Clustering methods: comparative studies

Reported performance of the methods used

N/A
No definitive conclusion is drawn

SOM

Results

Clustering methods: case studies

Which countries?

USA
UK
Taiwan
China
Turkey
New Zeeland
Australia
Mexico
Canada
Italy

How many catchments?

Around 85-90 on average

How many raingauges?

Changes between 100 - 10000

How long is the data?

Depends on the purpose and context

CONCLUSIONS

CITATION:

Dogulu, N. (2019) *Exploratory insights into hydrological applications of data clustering.*
Geophysical Research Abstracts, Vol. 21, **EGU2019-12099**,
European Geosciences Union General Assembly,
Vienna, Austria, 7 – 12 *April* 2019.

TAKE HOME LESSONS

Challenges and opportunities

Selecting the variables/features to cluster upon (Guyon & Elisseeff, 2003; Andrews & McNicholas, 2014; Chandrashekar & Sahin, 2014)

Scaling of data (Doherty et al., 2007; Tanioka & Yadohisa, 2012)

Determining the optimal number of clusters, i.e. evaluation of clustering algorithms (Rand, 1971)
— validity indices for cluster analysis (Arbelaitz et al., 2013)

Handling different types of data (Gan et al., 2007a)
Fuzzy clustering (Gan et al., 2007b; Ahani & Nadoushani, 2016)
Multi-objective clustering (Handl & Knowles, 2007)
Clustering for not purely random processes (Angelov, 2014)
— distribution of hydrological data are often strongly non-normal (Kundzewicz & Robson, 2004)

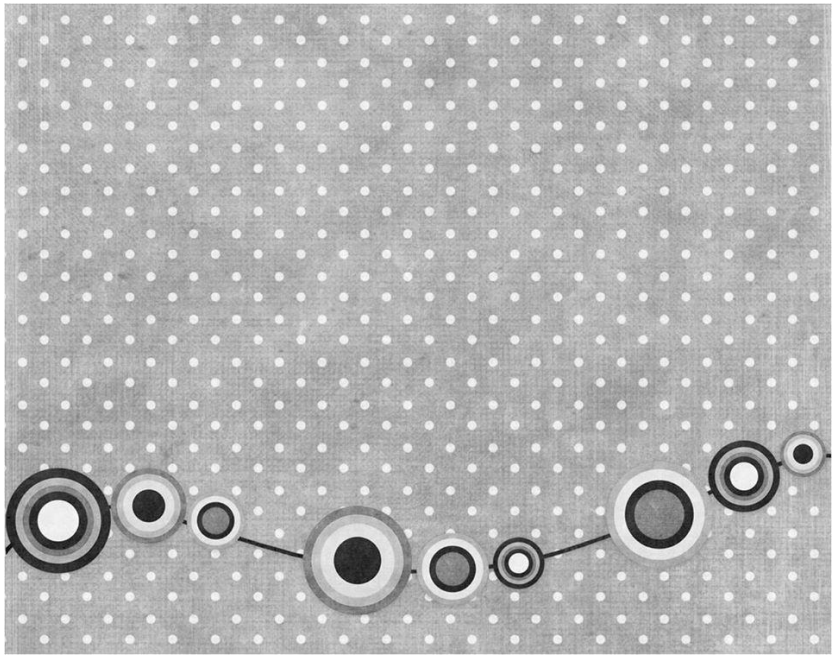
Exclusive vs. overlapping clustering (De Mulder, 2013; N'Cir, 2015)

Clusters with arbitrary shapes (Ertoz et al., 2013; Huang et al., 2014)

Time series (seasonality, noise) (Lin & Chen, 2002)
Sequential clustering (Inniss, 2006)

TAKE HOME LESSONS

Future research directions



Source: Unknown

Hydrological research studies so far are **disconnected**, and their scope is **limited**.

A **clear and integrative vision** on clustering algorithms is lacking in the **hydrological sciences**.

Hydrology can contribute to **benchmarking** in the field of **cluster analysis.**

We need **a systematic initiative** from the **hydrological domain** to enable expansion of knowledge in **use of clustering algorithms.**

Let's **build a community effort** benefiting from availability of **large hydrological datasets** across continents.

Contact me via  nilay.dogulu@gmail.com

if you are interested!

Thanks for your attention !

CITATION:

Dogulu, N. (2019) *Exploratory insights into hydrological applications of data clustering.*
Geophysical Research Abstracts, Vol. 21, **EGU2019-12099**,
European Geosciences Union General Assembly,
Vienna, Austria, 7 – 12 *April* 2019.

EGU General Assembly 2019
Vienna | Austria | 7–12 April

"Absence of evidence is **not** evidence of absence." — Carl Sagan

Questions ?

CITATION:

Dogulu, N. (2019) *Exploratory insights into hydrological applications of data clustering.*
Geophysical Research Abstracts, Vol. 21, **EGU2019-12099**,
European Geosciences Union General Assembly,
Vienna, Austria, 7 – 12 *April* 2019.

EGU General Assembly 2019
Vienna | Austria | 7–12 April