

Mattia Neri<sup>1</sup>, Paulin Coulibaly<sup>2</sup>, and Elena Toth<sup>1</sup>

<sup>1</sup>University of Bologna, Bologna, Italy (mattia.neri5@unibo.it)

<sup>2</sup>McMaster University, Hamilton, Canada

## Study aim

This work aims to provide an **additional signature** for the identification of similarity in hydrological dynamics, **based on the interaction (i.e. information flow)** between:

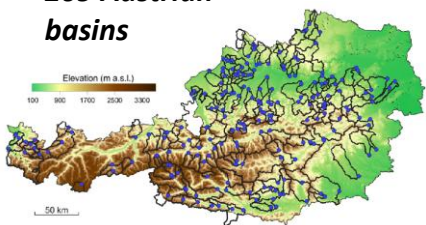
- meteorological forcing data (e.g. precip.  $P$ )
- streamflow ( $Q$ )

## The idea

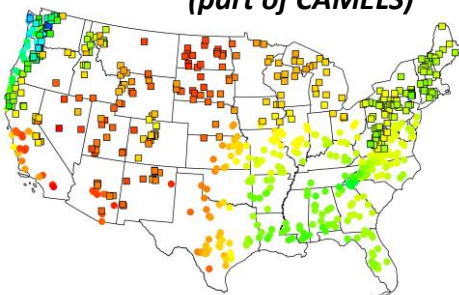
Relate **catchment similarity** to the **amount of information transferred from forcings to runoff**, using of the concepts of Information Theory, in particular of **transfer entropy** (Schreiber, 2000).

## Two large study regions

209 Austrian  
basins



504 US basins  
(part of CAMELS)



## The approach

### Step 1)

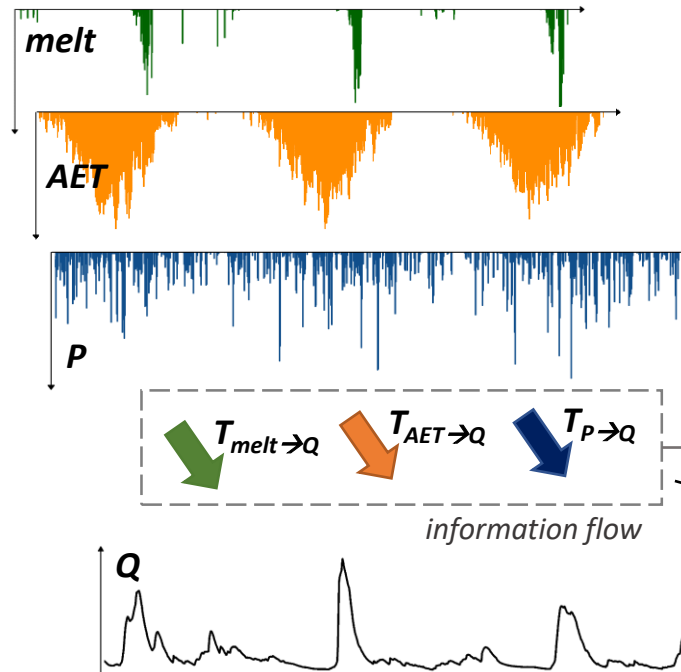
Identification (and/or estimation) of main meteorological forcings

- Precipitation ( $P$ )
- Actual evapotranspiration ( $AET$ )
- Snow melt ( $melt$ )

Estimated with a rainfall-runoff model

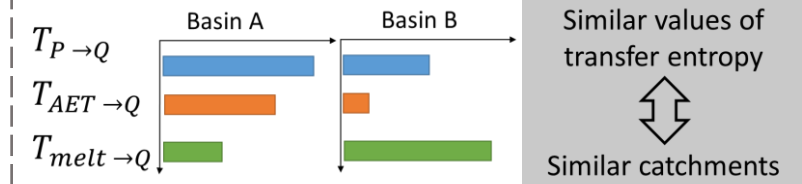
### Step 2)

Computation of the information flow from each forcing variable to runoff using **1-day lag transfer entropy** ( $T_{x \rightarrow Q}$ )

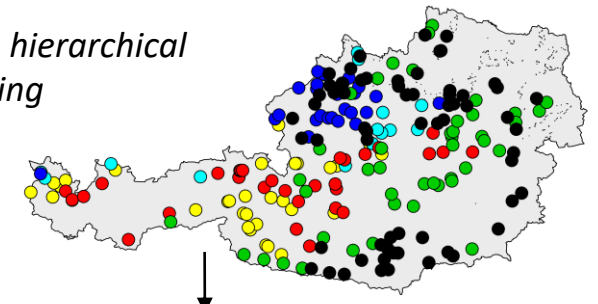


### Step 3)

$T_{x \rightarrow Q}$  are used as **similarity metrics** to identifying dominant hydrological processes and classify catchments.



Ward's hierarchical  
clustering



**Analysed against catchment attributes and typical runoff signatures**

## Purpose and reasons of the study

The delineation of similar groups of basins and the identification of dominant hydrological processes is required for several regionalisation applications.



The assessment of **hydrological similarity** strongly **depends on the specific purpose and features** to be regionalised.



For the **transfer of information** regarding rainfall-runoff processes **at fine temporal scale** (i.e. for the regionalisation of R-R model parameters), we must consider the **features of the entire runoff hydrograph and its forcings**, accounting for dominant streamflow generation and propagation dynamics

Previous literature on this topic grouped similar basins based on:

- **Similarity of runoff response** (e.g. Singh et al., 2016; Pérez Ciria and Chiogna 2020; Brunner et al., 2020)
- **Similarity** of both i) **runoff response** and ii) **climate** (Toth, 2013)

**BUT** no studies have so far focused on the **interaction** between climatic forcings and runoff time-series



The **aim** is to provide an **additional signature** based on the interaction (**i.e. information flow**) between:

- i) meteorological forcing data (e.g. precip.  $P$ )
- ii) streamflow ( $Q$ )

# How do we measure the interaction between forcings and runoff?

**High nonlinearity** of rainfall-runoff relationships



**Need for a metric** able to look **behind linear correlation analysis** between meteorological inputs and streamflow

**Shannon's information theory** (Shannon, 1948) may be a solution:

based on notion of *entropy*, i.e. the content of information of a signal, or the content of information shared between more variables (i.e. *mutual information*)



Already widely used in hydrology (Singh, 1997) and more in specific to define similarity (see e.g. Rajsekhar et al., 2013; Loritz et al., 2018)



## **Transfer entropy** (Schreiber, 2000)

- Entropy-based quantity
- time-asymmetric
- analyses the interaction between signals

Already demonstrated to be able to quantify the active transfer of information between hydrologic processes at various timescales (Bennett et al., 2019)

## **Research question**

Are we able to quantify the interaction between meteorological forcing and river discharge with the concepts of Information Theory?  
Is such measure representative of the hydrological behaviour of a watershed?

# What is Transfer Entropy?

A particular case of *conditional mutual information*.

Given to signals X and Y:

In general, Transfer Entropy  $T_{X \rightarrow Y}$  can be thought as the additional knowledge we gain about the variable Y at time  $t$  by measuring X at the previous  $l_x$  time steps, with respect to the information already given by knowing the previous  $l_y$  states of Y.

It **quantifies** the **effective information flow** from X to Y, given fixed time lags.

## Pros:

- able to **look beyond linear relationship** between variables

## Cons:

- probability distribution functions are generally unknown
- variables need to be **discretised** (quantised)

$$T_{X \rightarrow Y}(l_x, l_y) = I(Y_t, X_{t-1}, \dots, X_{t-l_x} | Y_{t-1}, \dots, Y_{t-l_y})$$

**Click here for  
more details about  
information theory and  
entropy terms**

# The approach

1) Identification of the **main hydrological forcings**, in this case:

- **Precipitation (P)**
- **Actual evapotranspiration (AET)**
- **Snow melt (*melt*)**

simulated with R-R model  
(TUWmodel, Viglione and Parajka 2018)

Exploring the potential of transfer entropy for  
identifying similarity of catchment dynamics

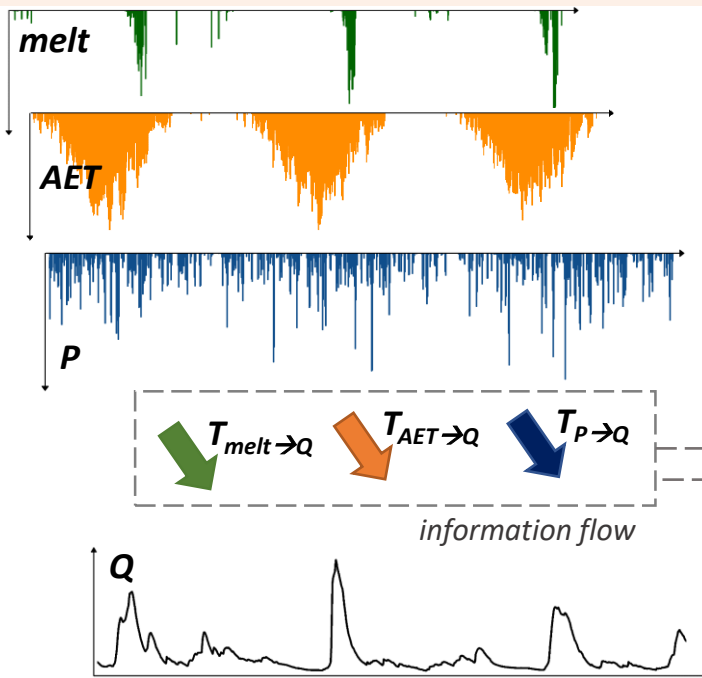
Mattia Neri, Paulin Coulibaly, and Elena Toth

2) Computation of the **information flow**  
from each forcing variable to runoff  
using **1-day lag transfer entropies**

- $T_{P \rightarrow Q}(1,1) = I(Q_t, P_{t-1} \mid Q_{t-1})$
- $T_{AET \rightarrow Q}(1,1) = I(Q_t, AET_{t-1} \mid Q_{t-1})$
- $T_{melt \rightarrow Q}(1,1) = I(Q_t, melt_{t-1} \mid Q_{t-1})$

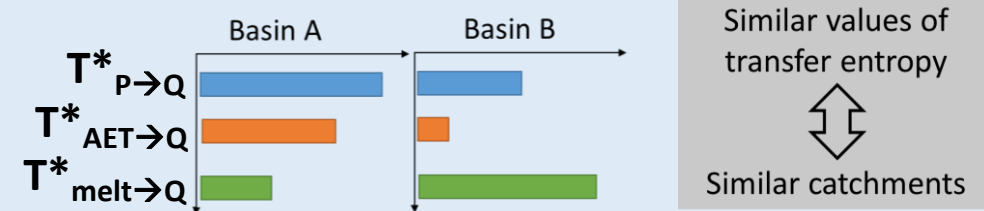
We consider only the previous time  
steps (time lags  $l_x = l_y = 1$ ):

- Avoid unreliability of the estimates
- Use simplest methodology to evaluate potential of transfer entropy

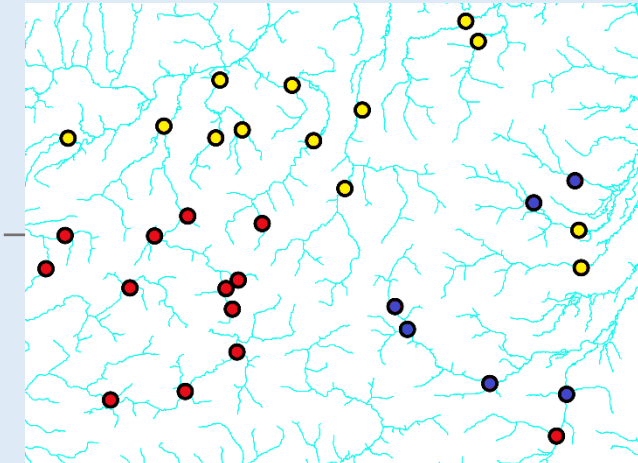


Each basin is **characterised by 3 values of Transfer Entropy** estimating information flow from *P*, *AET* and *melt* to *Q*

3) Standardised  $T^*_{x \rightarrow Q}$  are used as **similarity metrics** to identifying dominant hydrological processes and classify catchments



Application of a  
**clustering algorithm:**  
*Ward's hierarchical clustering*

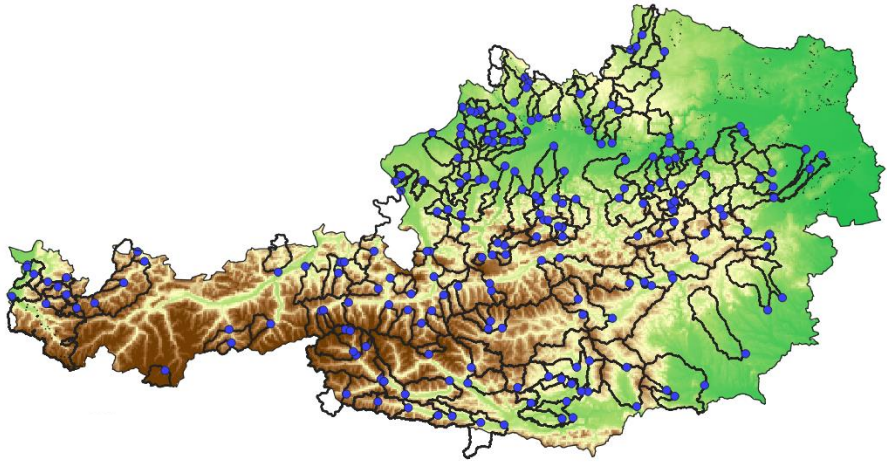


Analysed against:

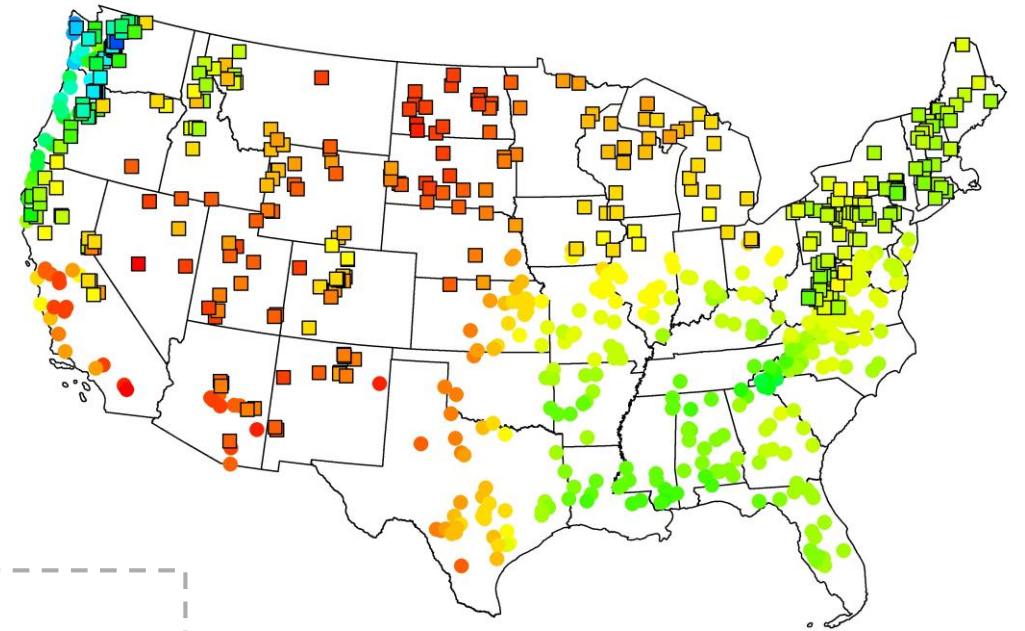
- catchment **attributes**
- **typical runoff signatures**

## Two study regions

- 1) Very densely gauged set of 209 catchments across Austria



- 2) 504 US watersheds (part of the CAMELS dataset) including wider variety of hydrological conditions and catchment characteristics



### Main attributes available for comparison in both regions

Elevation (m a.s.l.)  
Slope (m/m)  
Area (km<sup>2</sup>)  
Precipitation (mm/yr)  
Aridity (-)  
Snow fraction (*SnowF*)  
Snow depth / SWE (cm)  
Land cover types  
Irradiation (KWh/m<sup>2</sup>/day)  
...

### Main runoff signatures used to evaluate results

1) Mean annual runoff (mm/yr)  $\overline{Q}_y = 365 \overline{Q}_d = \frac{365}{T} \sum_{t=1}^T Q_d(t)$

- 2) Seasonality: Parde's range (-)

$$Par_i = \frac{\overline{Q}_i}{\overline{Q}_y} \longrightarrow \text{mean monthly runoff}$$

$$\Delta Par = \max(Par_i) - \min(Par_i)$$

- 3) Low flow statistics

$$q_{95} = \frac{Q_{95\%}}{\overline{Q}_d}$$

- 4) High flow statistics

$$q_{05} = \frac{Q_{05\%}}{\overline{Q}_d}$$

- 5) Hydrograph memory: integral scale

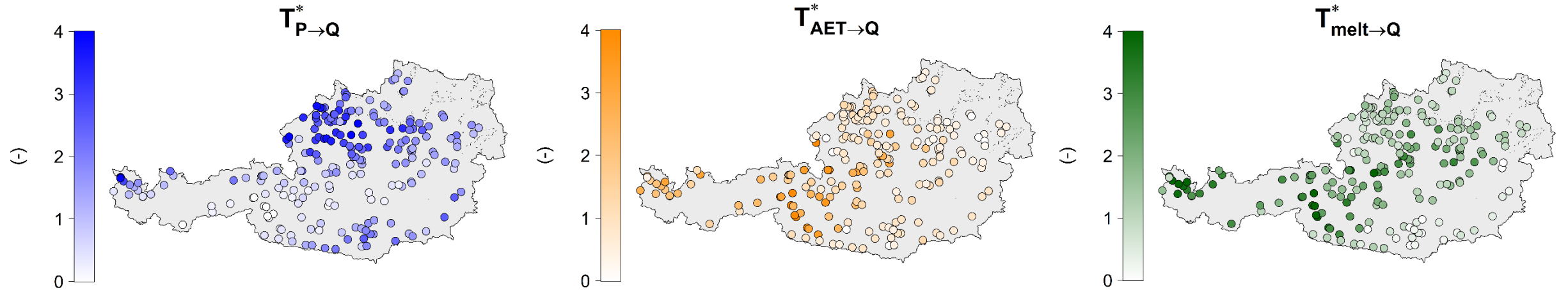
$$\tau_{1/e} \text{ (days)}$$

i.e. the time lag at which the autocorrelation function drops below  $1/e \sim 0.368$

# Example of the results (1): Standardised Transfer Entropy values (Austria)

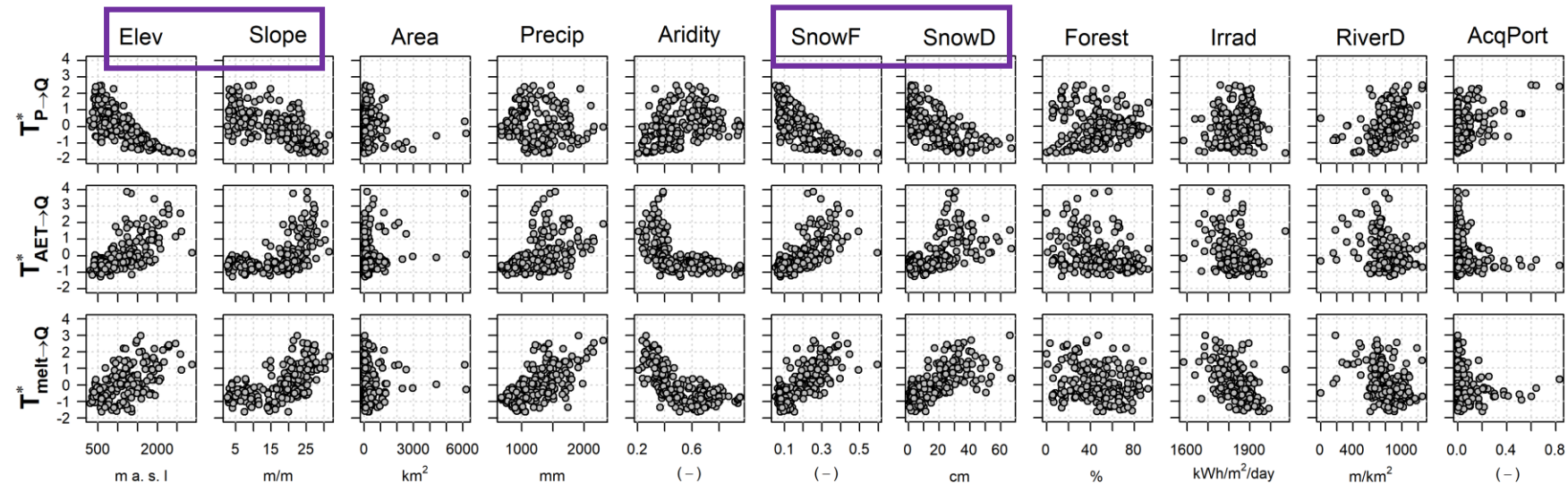
Exploring the potential of transfer entropy for  
identifying similarity of catchment dynamics

Mattia Neri, Paulin Coulibaly, and Elena Toth



## Analysed against catchment attributes:

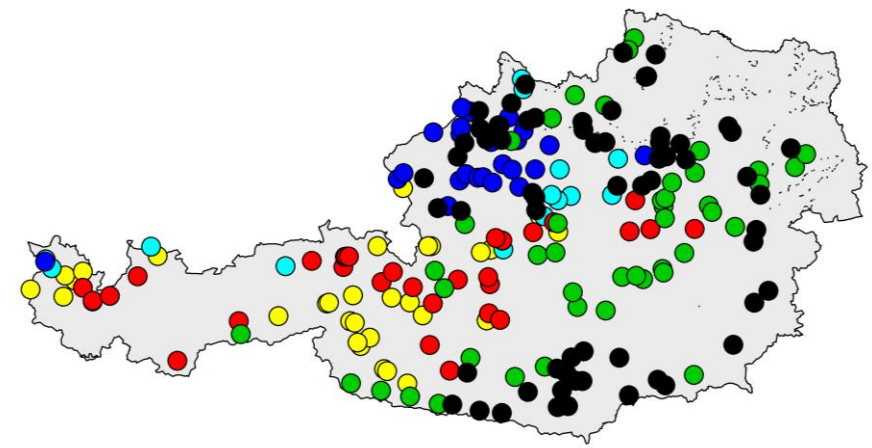
- In Austria, differences in  $T^*$  values tend to be explained mainly by attributes linked to elevation (slope, snow, ...)
- Impact of AET and melt are linked also to aridity



# Example of the results (2): Clusterisation based on Transfer Entropy values (Austria)

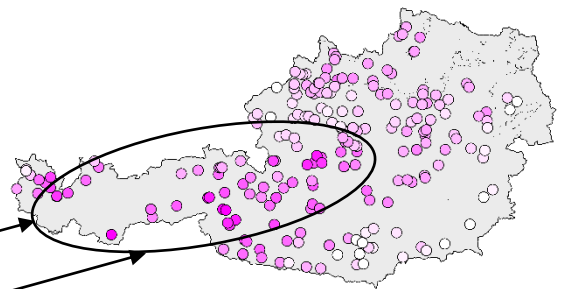
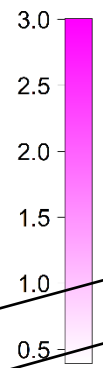
Analysed against runoff signatures

6 clusters

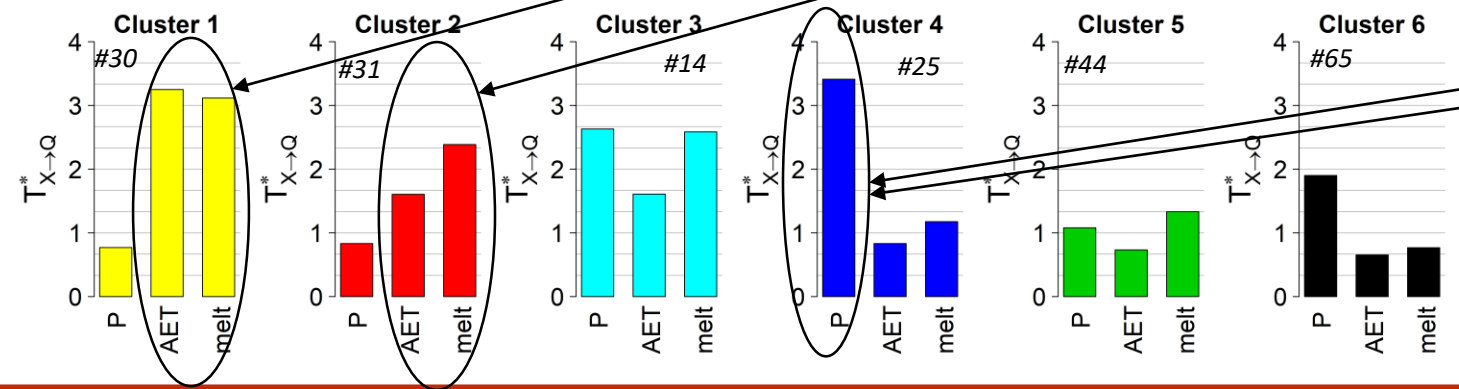
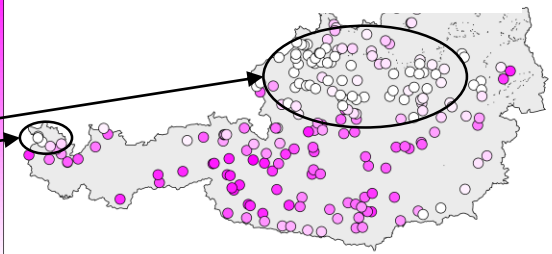
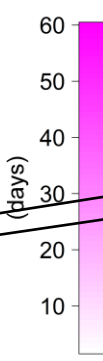


Consistently, mountainous catchments with higher seasonality show high information flow (i.e. impact on streamflow) from snow melt and AET ☺

Seasonality (Parde's range)



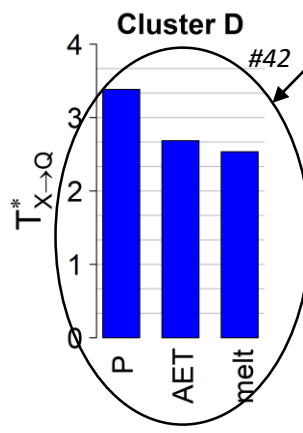
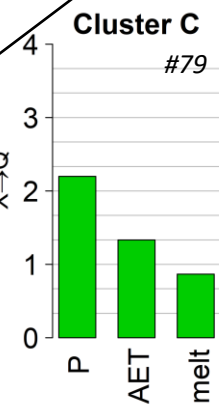
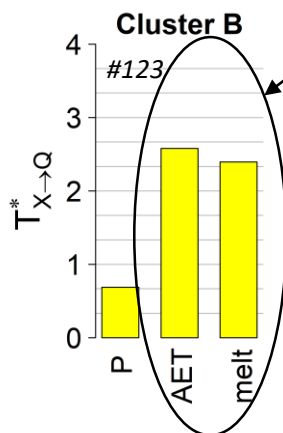
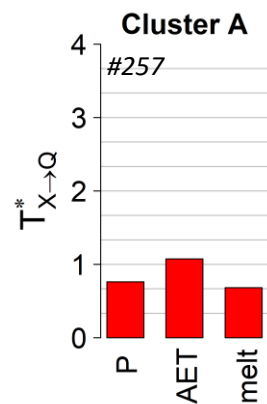
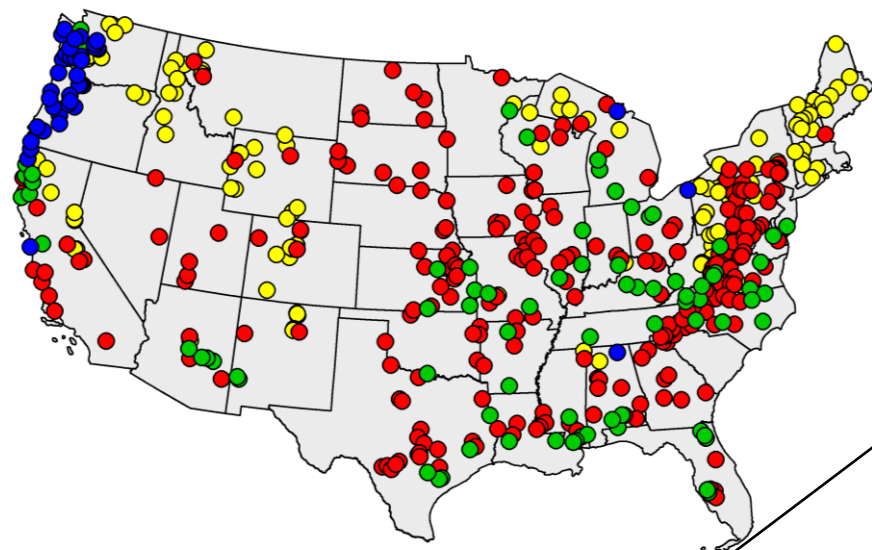
Hydrograph memory (integral scale)



Impact of 1day-lag rainfall is maximum for basins with lowest hydrograph memory  
(Consistent with what observed by, e.g., Viglione et al., 2013)

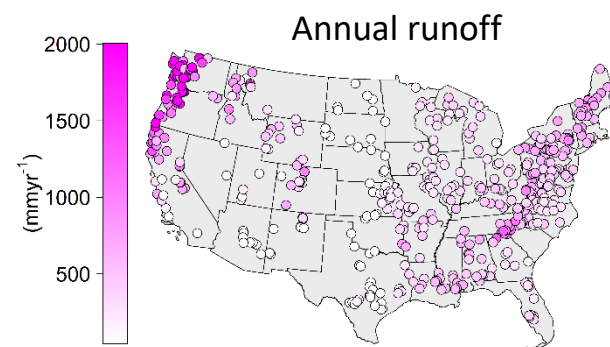
## Example of the results (3): Clusterisation based on Transfer Entropy values (USA)

4 clusters



**Snow dominated catchments** with high impact of *AET* and *melt* and low impact of 1-day-lag *P*

**Well delimited region** (also identified by previous classifications, e.g. Jehn et al. 2020, Brunner et al. 2020), characterised by high annual precipitation and runoff: we observe **high information flow from all the components** and highest flow from *P*.



Clusters **A** and **C** are more confused:  
we need to investigate deeper

→ **Ongoing analyses....**

## First conclusions

- Despite the limitations, the present analysis **demonstrates the potential of transfer entropy**, as further instrument for assessing hydrological similarity and for quantifying the connection between different processes.
- It should be seen as a **complementary approach** to classification techniques based on consolidated streamflow signatures.
- More in detail: **in Austria**, i) the method is able to **distinguish the predominant or partial role of snow melt and evapotranspiration**, and ii) the amount of **information flow transferred from the precipitation** to the runoff hydrograph can help to assess **differences in catchment response time**.

## Next steps

- **Go deeper in the analysis of CAMELS dataset**: such wide dataset generally needs to be divided in homogeneous groups of basins for most of regionalisation approaches → we want to **verify if transfer entropy can help to classify catchment**, supporting more typical approaches based on classical signatures
- ↓
- **Couple** the use of **transfer entropy** in catchment **classification** to the application of **rainfall-runoff model regionalisation techniques**
  - **Test** the impact of **different time lags** for transfer entropy computations

## Main references

- Addor, N., Newman, A. J., Mizukami, N., and 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.
- Bennett, A., Nijssen, B., Ou, G., Clark, M., and Nearing, G. (2019). Quantifying Process Connectivity With Transfer Entropy in Hydrologic Models. *Water Resources Research*, 55(6):4613–4629.
- Brunner, M. I., Melsen, L. A., Newman, A. J., Wood, A. W., and Clark, M. P. (2020): Future streamflow regime changes in the United States: assessment using functional classification, *Hydrology and Earth System Sciences*, 24, 3951–3966.
- Cover, T. M. and Thomas, J. A. (2005). Differential entropy. In *Elements of Information Theory*, chapter 8, pages 243–259. John Wiley & Sons, Ltd.
- Jehn, F. U., Bestian, K., Breuer, L., Kraft, P., and Houska, T. (2020). Using hydrological and climatic catchment clusters to explore drivers of catchment behavior. *Hydrology and Earth System Sciences*, 24(3):1081–1100.
- Loritz, R., Gupta, H., Jackisch, C., Westhoff, M., Kleidon, A., Ehret, U., and Zehe, E. (2018). On the dynamic nature of hydrological similarity. *Hydrology and Earth System Sciences*, 22(7):3663–3684.
- Pérez Ciria, T. and Chiogna, G. (2020). Intra-catchment comparison and classification of long-term streamflow variability in the alps using wavelet analysis. *Journal of Hydrology*, 587:124927.
- Rajsekhar, D., Mishra, A. K., and Singh, V. P. (2013). Regionalization of drought characteristics using an entropy approach. *Journal of Hydrologic Engineering*, 18(7):870–887.
- Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, 85:461–464.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423.
- Singh, S. K., McMillan, H., Bárdossy, A., and Fateh, C. (2016). Nonparametric catchment clustering using the data depth function. *Hydrological Sciences Journal*, 61(15):2649–2667.
- Singh, V. P. (1997). The use of entropy in hydrology and water resources. *Hydrological Processes*, 11(6):587–626.
- Viglione, A., Parajka, J., Rogger, M., Salinas, J. L., Laaha, G., Sivapalan, M., and Blöschl, G. (2013). Comparative assessment of predictions in ungauged basins - Part 3: Runoff signatures in Austria. *Hydrology and Earth System Sciences*, 17(6):2263–2279.

## Appendix:

### More details about Shannon's Information Theory (1)

Key measure of information theory is **Shannon's entropy** measures the amount of information content in a generic signal.

The entropy from a variable  $X$  (e.g. streamflow time-series), which has a set of discrete probabilities,  $p_1, \dots, p_n$ , related to  $i$ th class interval, is defined as:

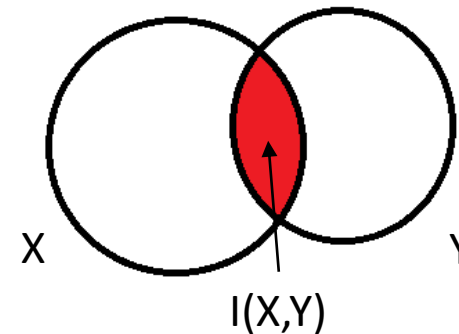
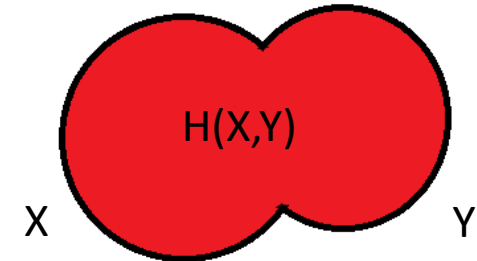
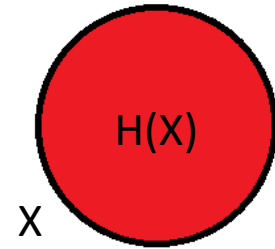
$$H(X) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i)$$

In the multivariate case, **joint entropy** represents the amount of information embedded in two (or more) variables  $X$  and  $Y$ :

$$H(X, Y) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 p(x_i, y_j)$$

It is possible to measure the portion of the information content shared between the two variables  $X$  and  $Y$ . It is named **mutual information** (Cover and Thomas, 2005):

$$I(X, Y) = H(X) + H(Y) - H(X, Y)$$



[Click here to go back to the presentation](#)

## Appendix:

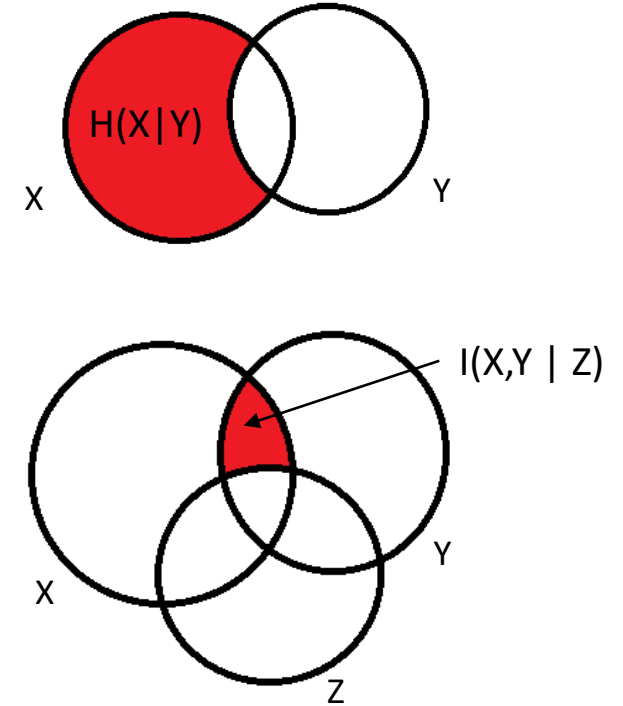
### More details about Shannon's Information Theory (2)

The **conditional entropy**, i.e. the additional information provided by X knowing Y is defined as:

$$H(X|Y) = H(X) - I(X, Y)$$

Similarly, **conditional mutual information** (given a third variable Z) can be also defined as:

$$I(X, Y|Z) = H(X|Z) + H(Y|Z) - H(X, Y|Z)$$



Such quantities are all symmetric. Schreiber (2000) developed a method for accounting for asymmetric **information transfer**. The quantity of our interest, called **transfer entropy** is written as a particular case of conditional mutual information:

$$T_{X \rightarrow Y}(lx, ly) = I(Y_t, X_{t-1}, \dots, X_{t-lx} | Y_{t-1}, \dots, Y_{t-ly})$$

where  $lx$  and  $ly$  are time lags

[Click here to go back to the presentation](#)