Analysis of Flow Signatures and Catchment Similarity Indices for Catchment Classification in Yesilirmak Basin

Batuhan SOYUGUR, Koray K. YILMAZ

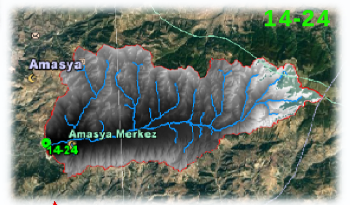
Department of Geological Engineering, Middle East Technical University, Ankara, Turkey

batuhan.soyugur@metu.edu.tr, yilmazk@metu.edu.tr

1) STUDY OBJECTIVES

- Catchment classification schemes aim to identify groups of hydrologically similar catchments to enable a mapping between catchment physical characteristics and hydro-climatic conditions with the catchment functioning.
- This mapping, together with the quantified uncertainties, potentially facilitates improved process understanding, transfer of this understanding to ungauged catchments, model parameter regionalization and hence improve operational applications and watershed management.
- The aim of this study is:
  1. To carefully derive hydrologically relevant similarity metrics from catchment physical (elevation, area, aspect, slope, geology, soils, land use), climatic and hydrologic response characteristics (flow signatures).
  2. To utilize Affinity Propagation clustering algorithm to determine the optimal number of groupings based on individual as well as a combination of these similarity metrics.

2) STUDY AREA



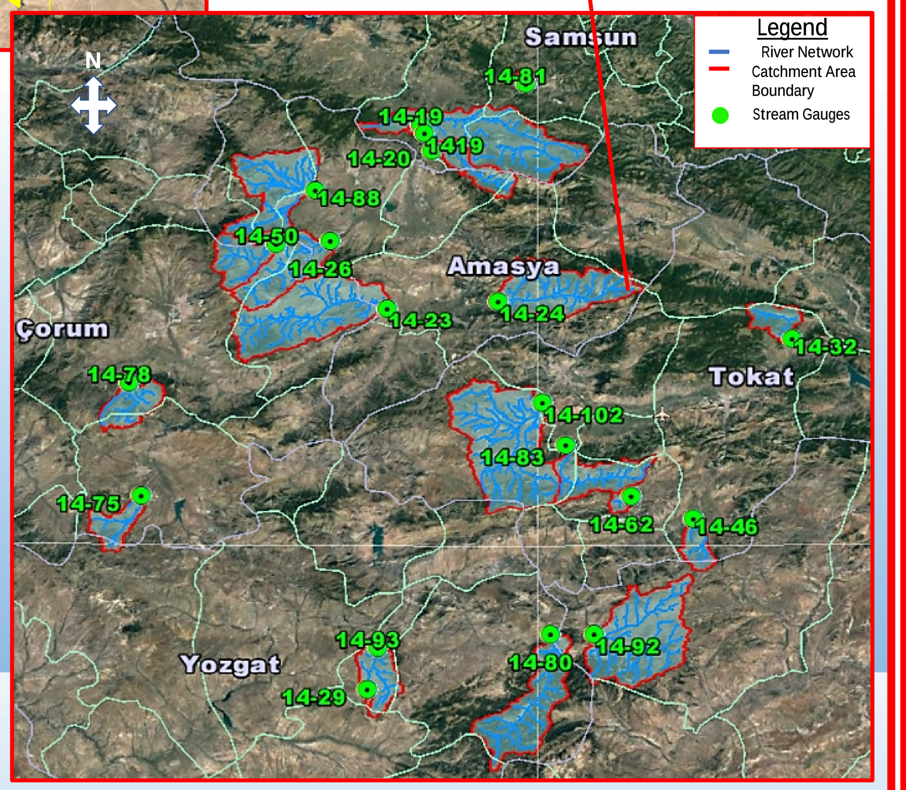


Figure 1 - Study Area

- The Yesilirmak Basin, with an area of 38,730 km², is the third largest river basin in Turkey.
- Yesilirmak Basin lies within 39˚30'-41˚21'N lattitude bands and 34˚40'-39˚48'E longitude bands.
- The study area includes 20 sub-catchments of the Yesilirmak Basin – selected based on data availability.
- Area of the study catchments range from 6,8 km² to 613 km²
- The study time period covers 1975 – 1980 based on data availability (6 years).

3) DATA SETS

- **STREAMFLOW DATASET:**
  - Daily streamflow data is provided by the General Directorate of State Hydraulic Works (DSİ).
- **RAINFALL DATASET:**
  - There are 75 rain gauge data (monthly precipitation) provided from Turkish State Meteorological Service.
- **SOIL MAP:**
  - National Soil Map of Turkey with 1:25000 resolution
- **DIGITAL ELEVATION MODEL:**
  - A digital elevation model having a resolution of 30 m x 30 was used.
- **GEOLOGICAL MAPS:**
  - 1 / 100 000 scale geological maps were provided by General Directorate Of Mineral Research And Exploration (MTA)
- **LAND COVER:**
  - Corine 2000 Land Cover Map

REFERENCES

1. Ali, G., Tetzlaff, D., Soulsby, C., McDonnell, J. J., & Capell, R. (2012). A comparison of similarity indices for catchment classification using a cross-regional dataset. Advances in Water Resources, 40, 11-22.
2. Chiverton, A., Hannaford, J., Holman, I., Corstanje, R., Prudhomme, C., Bloomfield, J., & Hess, T. M. (2015). Which catchment characteristics control the temporal dependence structure of daily river flows?. Hydrological Processes, 29(6), 1353-1369.
3. Cramér, Harald. "Mathematical methods of statistics." (1947): 169-173.
4. Frey BJ, Dueck D. Clustering by passing messages between data points. Science 2007; 315:372-6
5. Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20, 53-65.
6. Sawicz, K. A. (2011). Catchment classification: emğrical analysis of hydrologic similarity based on catchment function in the eastern USA.

4) METHODS

SELECTED VARIABLES

AREA	AREA		
CLIMATE	Annual Total Precipitation		
FLOW INDICES	Q5: High Flow Conditions		Q50: Medium Flow Conditions
	Q85: Low Flow Conditions		Slope of FDC: Catchment Response to Precipitation
PHYSICAL PROPERTIES	TOPOGRAPHIC	Elevation	Slope
		Aspect	
	GEOLOGY	Alluvium (A)	Ophiolites-Melanges (OP-M)
		Volcanic Rocks (VR)	Gypsum (G)
		Plutonic Rocks (PR)	Marble (M)
		Clastics (C)	Metamorphic Rocks (MT)
		Limestone (LST)	Olistoliths (OL)
		Schists (SCH)	
	SOIL	Gray Brown Podzolic Soils (GBPDS)	Reddish Brown Soils (RBS)
		Brown Forest Soils (BFS)	Alluvial Soils (AS)
		Brown Forest Soils Without Lime (BFSWL)	Hydromorphic Soils (HS)
		Chestnut (maron) Color Soils (CCS)	Colluvial Soils (CS)
		Reddish Maron Soils (RMS)	Settlement Area (SA)
		Brown Soils (BS)	Bare Rocks and Rubble Areas (BRRA)
		Brown Soils Without Lime (BSWL)	Missed Data (MD)
		Dam - Small Dam (DS)	
	CORINE LAND COVER	Mineral Extraction Sites (MES)	Water Bodies (WB)
		Vineyards (VY)	Discontinuous Built-Up Areas with Family Houses with Gardens (DVAFHG)
		Pastures (PS)	Arable Land With Large Fields (ALWLF)
		Land Principally Occupied by Agriculture (LPOA)	Permanently Irrigated Arable Land (PIAL)
		Broad-Leaved Forest (BLF)	Orchards (OR)
		Coniferous Forest (CF)	Complex Cultivation Patterns without Scattered Houses (CCPWSH)
		Mixed Forests (MF)	Complex Cultivation Patterns With Scattered Houses (CCPSH)
		Natural Grassland (NG)	Bare Rocks (BR)
		Transitional Woodland Shrub (TWS)	Beaches, Dunes and Sand Plains (BDSP)
		Sparsely Vegetated Areas (SVA)	

- All the variables were normalized (Z-score) prior to the analysis to match the scale.
- Flow duration curves were prepared for all gauging stations and high flows (Q_5), medium flows (Q_{50}), and low flows (Q_{85}) were calculated for the study time period (1975-1980).

CLUSTERING ALGORITHM

- ***AFFINITY PROPAGATION*** (Frey and Dueck, 2007)
  - Three advantages of the algorithm:
    - The algorithm does not require an initial parameterization (i.e. exemplars, # of clusters)
    - An exemplar is identified which is the most representative of each cluster
    - Highly effective algorithm with lower clustering error compared to majority of the existing methods

ASSOCIATION BETWEEN CLUSTERS OF DIFFERENT VARIABLES

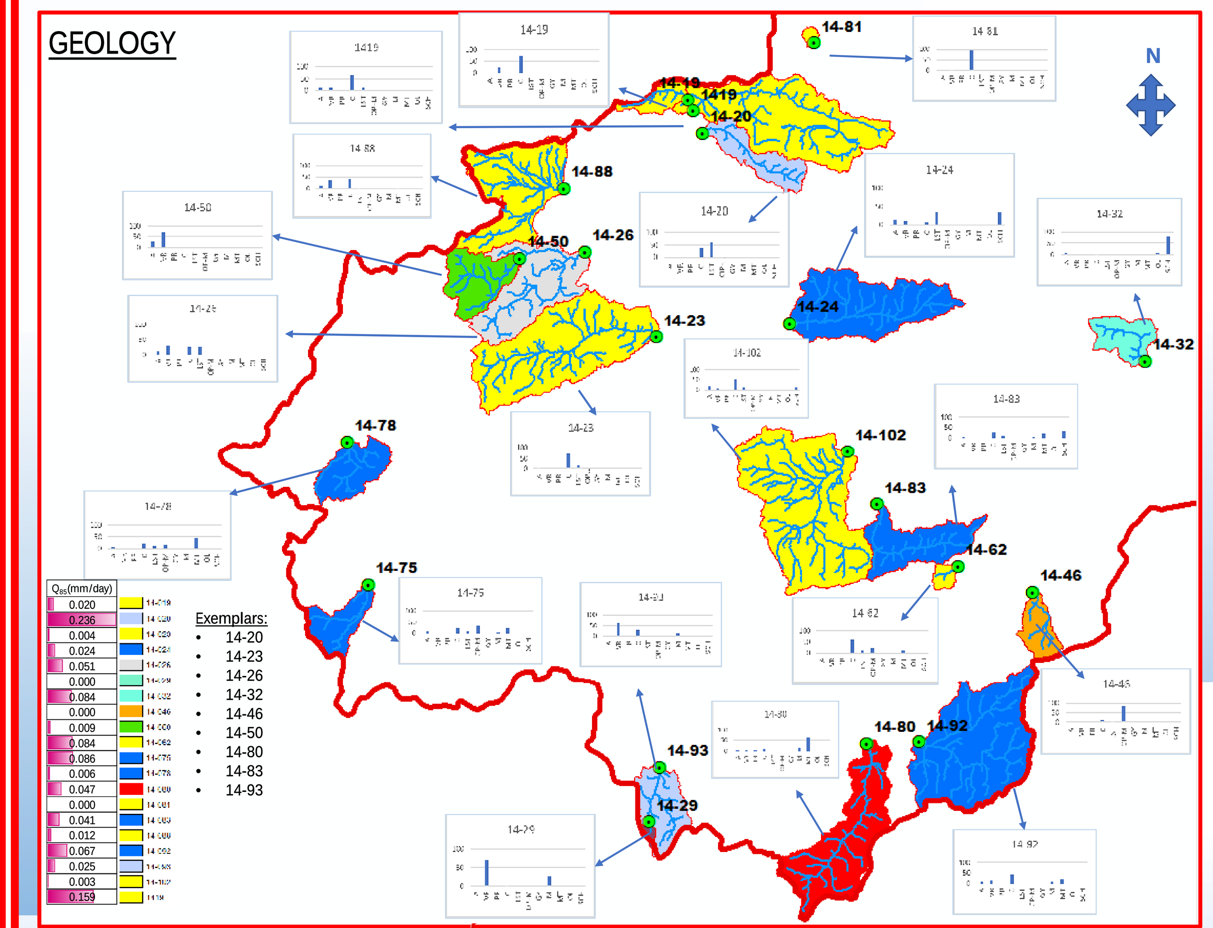
- ***CRAMER'S V COEFFICIENT***
  - Cramer's V Coefficient is a way of estimating the extent of relationship between two variables, in our case clusters based on different variables.Cramer's V Coefficient varies between 0 and 1.
  - Groups which have a larger value for Cramer's V can be considered to have a strong relationship between the variables, with a smaller value for V indicating a weaker relationship.

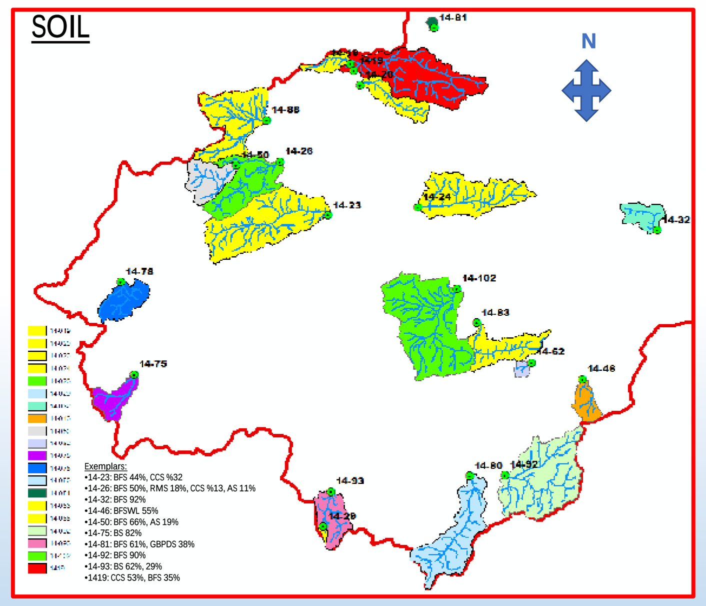
$$V = \sqrt{\frac{\varphi^2}{\min(k-1, r-1)}} = \sqrt{\frac{\chi^2/n}{\min(k-1, r-1)}}$$

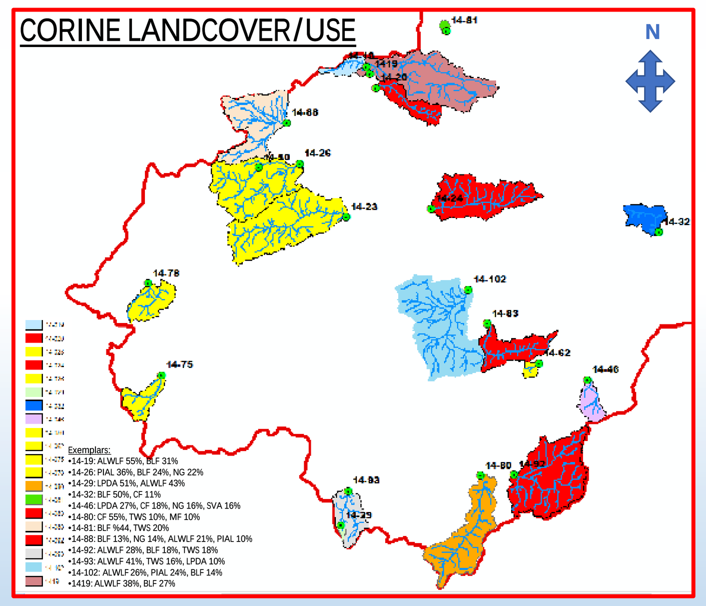
χ^2 = Pearson's chi-squared test
n = Grand total of observations
k= Number of columns
r= Number of rows

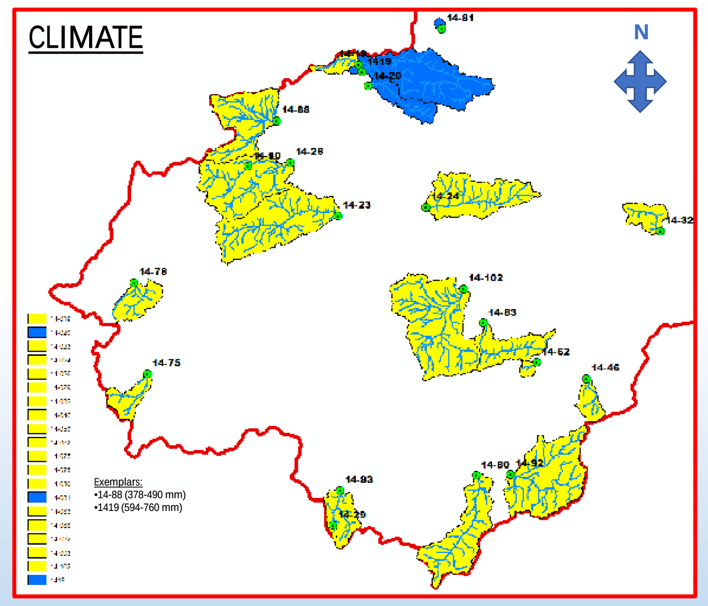
5) RESULTS

CLUSTER MAPS





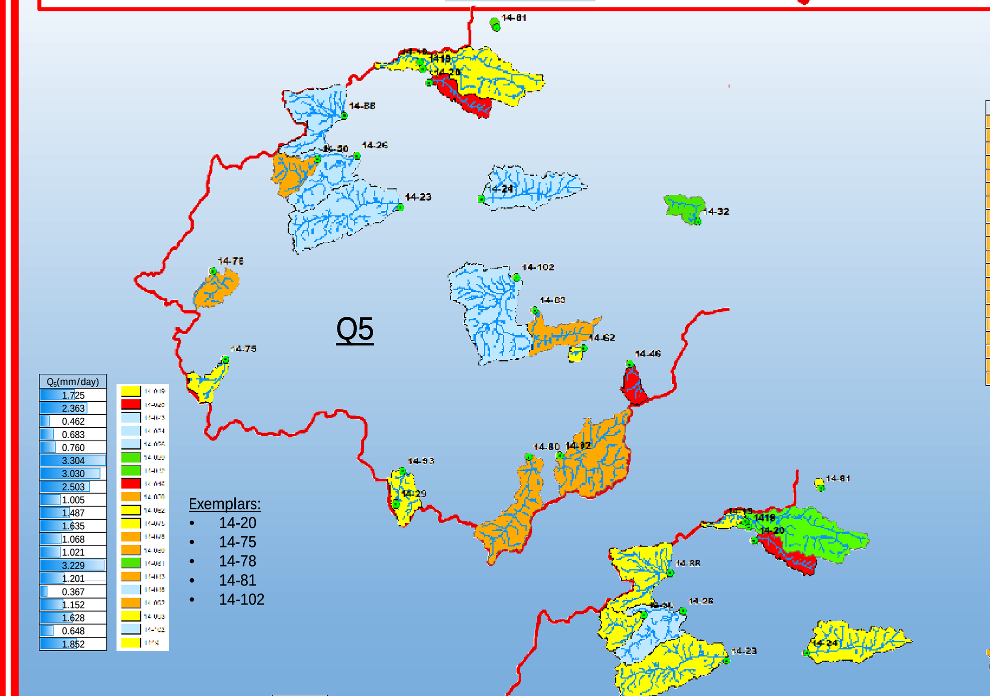


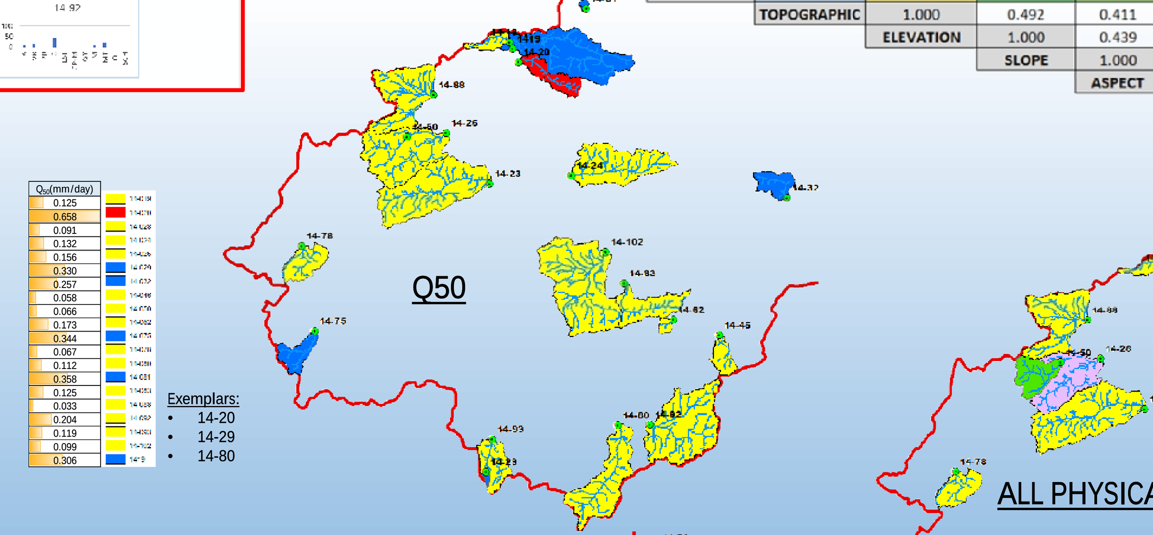


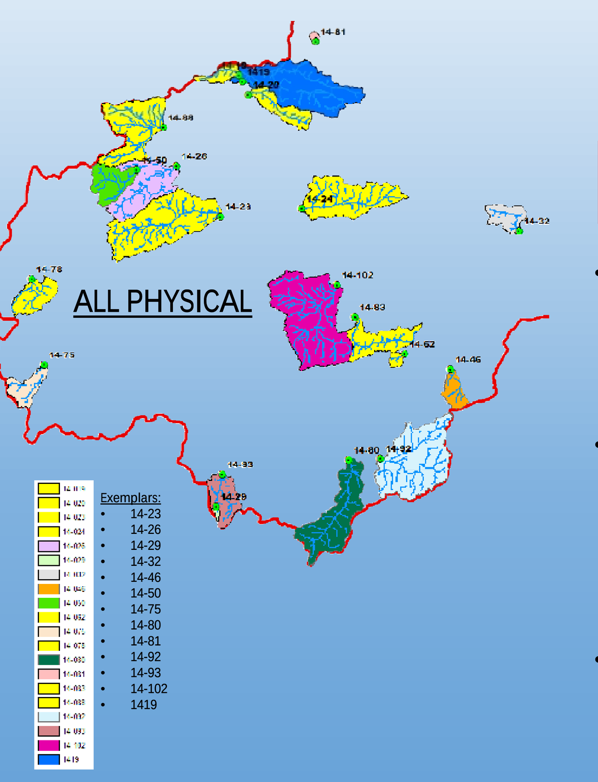
ASSOCIATION OF VARIABLES

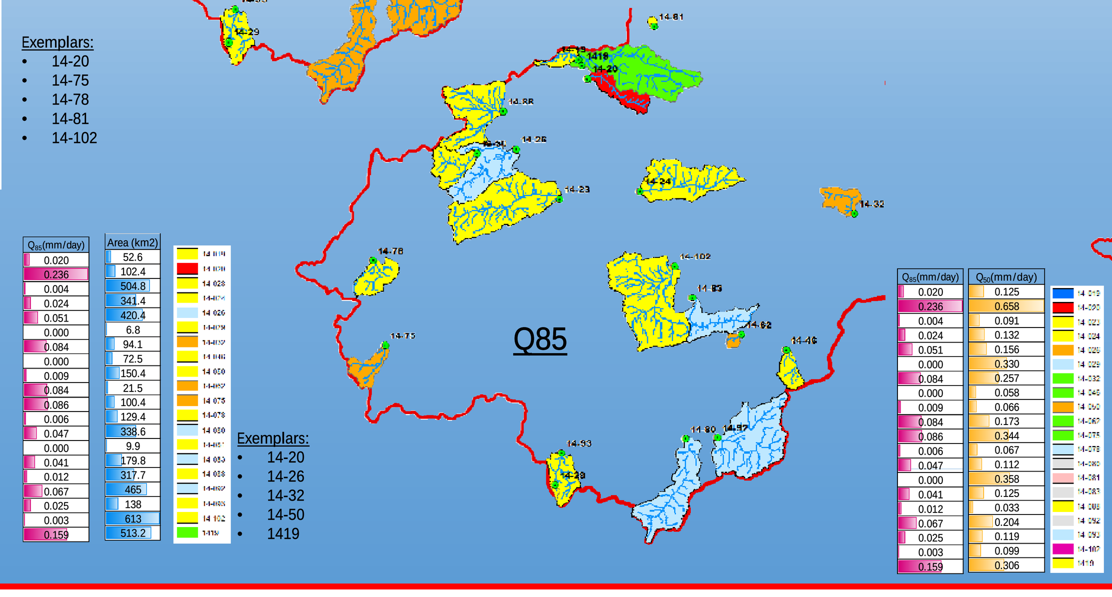
CRAMER'S V COEFFICIENT (Flow Variables: Q5, Q50, Q85, SFDC)

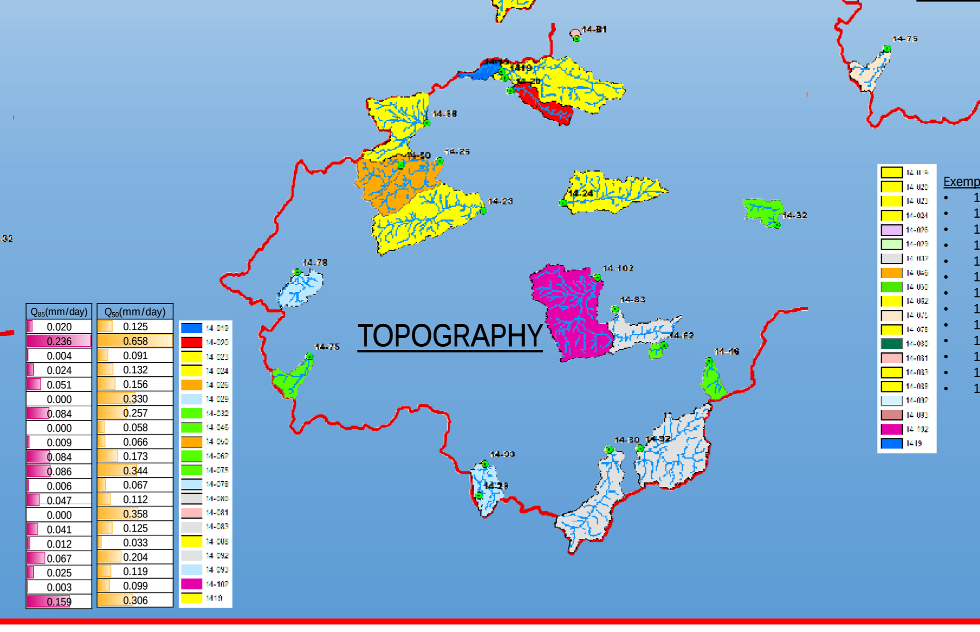
	AREA	CLIMATE	ALLFLOW	TOPOGRAPHIC	ELEVATION	SLOPE	ASPECT	GEOLOGY	SOIL	CORINE	ALLPHYSICAL	Q5	Q50	Q85	SFDC	ALLWITHAREA	ALLWITHOUTAREA
AREA	1.000	0.057	0.471	0.870	0.741	0.408	0.082	0.627	0.734	0.717	0.781	0.764	0.323	0.560	0.618	0.692	0.692
CLIMATE		1.000	0.610	0.737	0.262	0.642	0.183	0.663	0.815	0.840	0.810	0.478	0.728	0.802	0.388	1.000	1.000
ALLFLOW			1.000	0.699	0.382	0.481	0.333	0.905	0.735	0.856	0.830	0.749	0.758	0.615	0.719	0.830	0.830
TOPOGRAPHIC				1.000	0.492	0.411	0.333	0.732	0.709	0.766	0.724	0.699	0.785	0.834	0.650	0.770	0.812
ELEVATION					1.000	0.439	0.428	0.835	0.766	0.796	0.807	0.523	0.316	0.426	0.495	0.818	0.793
SLOPE						1.000	0.246	0.734	0.759	0.785	0.778	0.465	0.527	0.487	0.452	0.725	0.722
ASPECT							1.000	0.816	0.743	0.912	0.772	0.446	0.255	0.578	0.343	0.772	0.809
GEOLOGY								1.000	0.756	0.789	0.716	0.742	0.794	0.716	0.713	0.789	0.766
SOIL									1.000	0.788	0.841	0.710	0.618	0.711	0.777	0.792	0.765
CORINE										1.000	0.788	0.787	0.697	0.732	0.819	0.762	0.762
ALLPHYSICAL											1.000	0.804	0.729	0.782	0.822	0.846	0.891
Q5												1.000	0.755	0.543	0.490	0.885	0.837
Q50													1.000	0.829	0.330	1.000	1.000
Q85														1.000	0.543	0.916	0.916
SFDC															1.000	0.793	0.756
ALLWITHAREA																1.000	0.965
ALLWITHOUTAREA																	1.000











6) DISCUSSIONS

- Clustering based on FDC-based flow variables were associated (high Cramer's V) with clustering based individually on: TOPOGRAPHY, GEOLOGY, SOILS, LANDCOVER & CLIMATE. An interplay between the latter exist.
- Strong association was found between clusters based on flow variables and clusters based on all other variables combined (ALLWITHAREA). Note that Area does not have strong influence on clusters (compare ALLWITHAREA & ALLWITHOUTAREA).
- For example, catchment 14-20, characterized by high flow variables (more specifically, Q85) is dominated by limestone geology and wet climate.
- In future work, we will include 1) more climate parameters, 2) more detailed soil map and alternative clustering algorithms such as hierarchical clustering and two-step clustering.