A data driven approach for the temporal classification of heavy rainfall using Self-Organizing Maps

Konstantinos Vantas, Epaminondas Sidiropoulos and Marios Vafeiadis

Faculty of Engineering, Aristotle University of Thessaloniki

Objectives

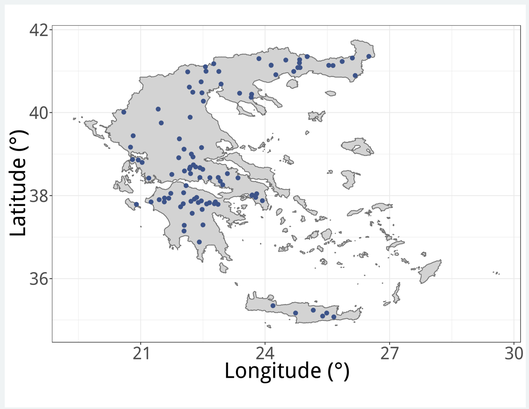
- The identification of heavy rainfall temporal patterns using raw precipitation data.
- Examine if there is seasonality in the occurrence of the different clusters.
- Test the hypothesis of randomness in the structure of standardized heavy rainfall profiles.

Introduction

Knowledge about the temporal distribution of rainfall is essential in current methods of water resources management. Independent rainstorm events can be identified using a Poisson process hypothesis [1] and standardized using their cumulative height and their duration. A method for the classification of standardized rainfall profiles can be found in Huff [2], in which the quartile where the maximum intensity occurs was used. Here, a different, data-driven approach is proposed.

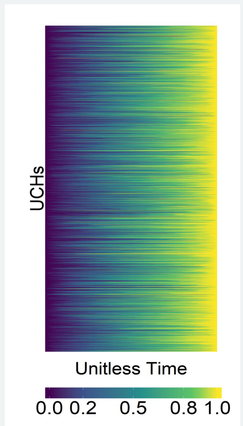
Materials

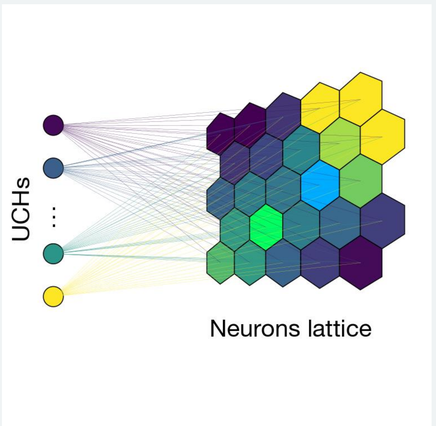
Precipitation data came from 101 stations from Greece [3] with an average length of 23 years per station and 30 minutes time step. 13,211 rain-storms classified as heavy were used in the analysis (height > 12.7 mm, duration > 3hr).

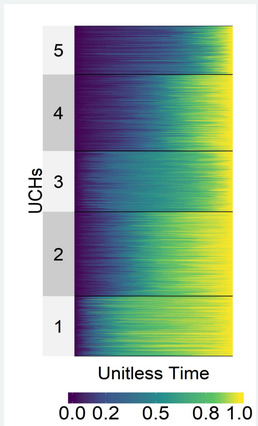


Clustering workflow

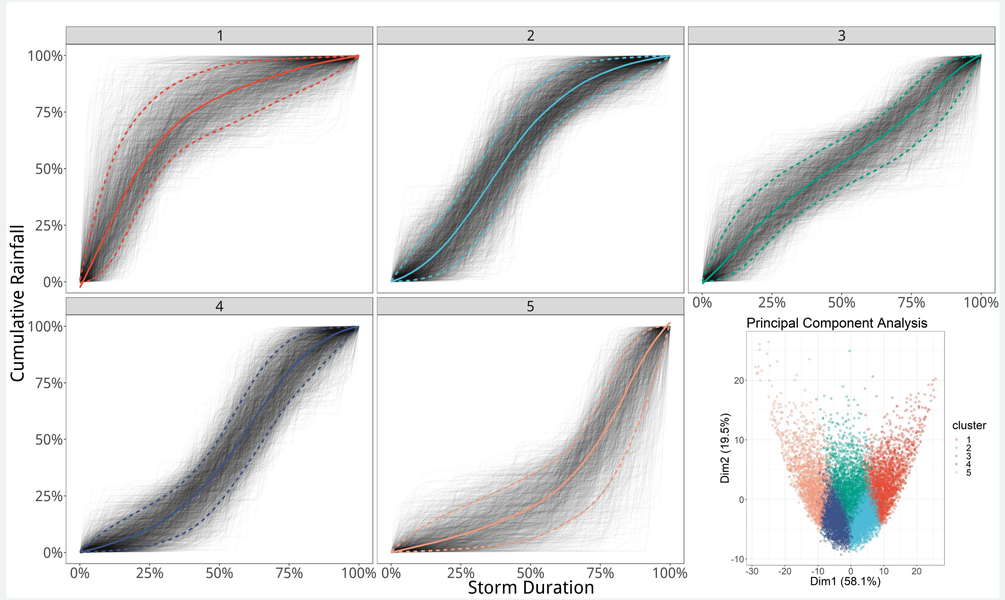
The heavy rainstorms were standardized in the form of Unitless Cumulative Hyetographs (UCHs) and Self Organizing Maps [4] were used, with different neurons lattices, to create clusters of data. Clusters' centers were tested if they come from different distributions utilizing the two-sample Kolmogorov–Smirnov test.



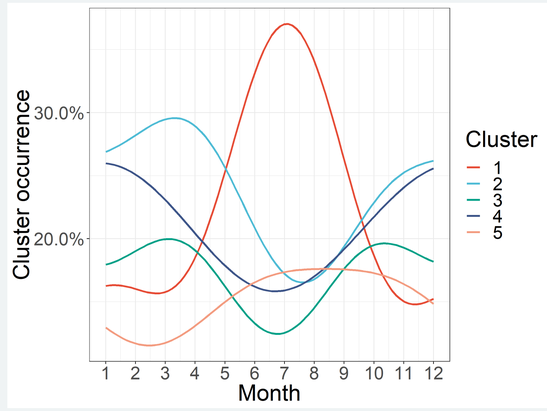




Visualizations of clusters



Seasonal patterns of clusters



Statistical tests

- The hypothesis of random data in the structure of UCHs was rejected using the Hopkins index [5] for clustering tendency.
- The hypothesis that the centers of the clusters were drawn from the same distribution was rejected, using the two-sample Kolmogorov-Smirnov test [6].

Conclusions

1. A limited number of temporal rainfall patterns emerged, in terms of seasonality and different characteristics.
2. The classification of the rainstorm events can be made in an unsupervised manner.
3. The hypothesis that UCHs contain random data was rejected, so there is physical meaning in the categorization of rainstorms

References

[1] Restrepo-Posada, P. J. & Eagleson, P. S. Identification of independent rainstorms. Journal of Hydrology 55, 303–319 (1982).
[2] Huff, F. A. Time distribution of rainfall in heavy storms. Water resources research 3, 1007–1019 (1967).
[3] Vantas, K. hydroscoper: R interface to the Greek National Data Bank for Hydrological and Meteorological Information. Journal of Open Source Software 3, 625 (2018).
[4] Kohonen, T. Self-organizing maps. 30, (Springer Science & Business Media, 2012).
[5] Banerjee, A. & Dave, R. N. Validating clusters using the Hopkins statistic. in 2004 IEEE International Conference on Fuzzy Systems.1, 149–153 (IEEE, 2004).
[6] Conover, W. J. Practical nonparametric statistics. (John Wiley & Sons, 1980).