EGU General Assembly 2020

Session HS3.6
Spatio-temporal and/or (geo) statistical analysis of hydrological events, floods, extremes, and related hazards

Department of Water Resources & Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Greece

#shareEGU20

Konstantinos Papoulakos
Theano Iliopoulou
Panayiotis Dimitriadis
Dimosthenis Tsaknias
Demetris Koutsoyiannis

Investigating the impacts of clustering of floods on insurance practices; a spatiotemporal analysis in the USA

EGU2020-8667

Insurance companies should consider clustering mechanisms in their practices to avoid underestimation of the exceedance probability of collective risk.

(Goulianou et. Al, EGU2019-5981)

Introduction

EGU General Assembly 2020

Recent research has revealed the significance of **Hurst-Kolmogorov dynamics** (Koutsoyiannis, 2011), which is characterized by strong correlation and high uncertainty in large scales (Dimitriadis and Koutsoyiannis, 2015), in flood risk assessment as for example in inundated flood duration (Dimitriadis and Koutsoyiannis, 2020).

However, classic risk estimation for flood insurance practices is formulated under the **assumption of independence** between the frequency and the severity of extreme flood events, which is unlikely to be tenable in real-world hydrometeorological processes exhibiting **long range dependence** (Iliopoulou and Koutsoyiannis, 2019).

Furthermore, insurable flood losses are considered as **ideally independent** and **non-catastrophic** due to the widely spread perception of limited risk regarding catastrophically large flood losses.

As the **accurate risk assessment** is a fundamental process on flood insurance and reinsurance practices, this study investigates the effects of **lack of fulfillment** of these assumptions, paving the way for a deeper understanding of the underlying clustering mechanisms of stream flow extremes.

For this purpose, we present a **spatiotemporal** analysis of the daily flow series from the US-CAMELS dataset (Newman et al., 2014), comprising the impacts of **clustering** mechanisms on return intervals, duration and severity of the over-threshold events which are treated as proxies for **collective risk**.

Moreover, an exploratory analysis is introduced regarding the **stochastic aspects** of the correlation between the properties of the extreme events and the **actual claim records** of the FEMA National Flood Insurance Program which are recently published.

Dataset

We used and processed the US-CAMELS dataset (Newman et al., 2014), which comprises 671 daily stream flow time series across the major basins and hydrological units in USA.

From this dataset, 360 stream flow time series with the maximum temporal overlap (namely, 35 years from 1980 to 2014) and less than 10% of missing values were selected.

Fig. 1. The 671 US-CAMELS stream gauge locations.

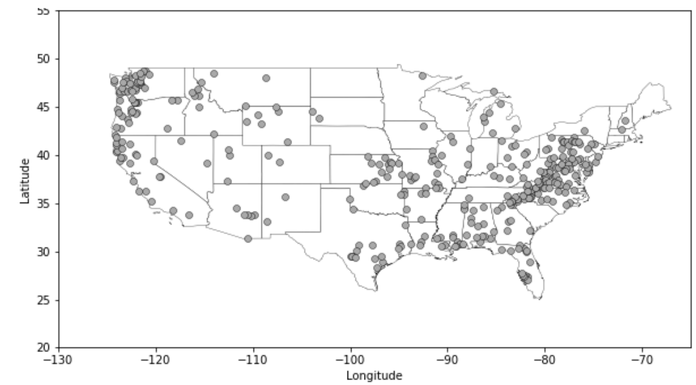


Fig. 2. The selected 360 US-CAMELS stream gauge locations.

Methodology

Collective Risk

Collective risk S is the total claim amount regarding a portfolio of (re)insured properties that produces a random number N of claims in a certain time period. Following Serinaldi and Kilsby (2016), we use POT flows as a proxy for collective risk estimation, defined as:

$$\boldsymbol{S} = \sum_{j=1}^{N} \boldsymbol{Y}_j$$

where Y_j is the jth claim proxy (over-threshold flow fluctuation severity), N is the number of exceedances, and the total claims S=0 if N=0.

Assessing the Collective Risk S is a typical problem faced in insurance sector.

Peak Over Threshold Method

Peak Over Threshold method (POT) has become one of the most preferable extreme value approaches in insurance.

The threshold should be chosen such that all losses above the threshold could be considered as extreme losses, in the sense of the underlying extreme value analysis.

To characterize the dynamics of extreme stream flow values, we selected four different percentage thresholds (90%, 95%, 98%, and 99%).

The behavior of US-CAMELS stream flows is found to be consistent with HK dynamics characterized by moderate *H* parameters (in range of 0.6-0.7), through Monte Carlo simulation.

(Manolis et. al, EGU2020-9357)

Exploratory Analysis

Impacts on Collective Risk S, return intervals and duration of extreme events

EGU General Assembly 2020

Introduction

In order to characterize the clustering mechanisms of a time series, it is of great importance to quantify the divergence between the observed time series and a sequence of independent variables. One of the methods which have been proposed is to shuffle the elements of the observed time series in order to get a new series which has the same dimensional distribution but no correlation (Serinaldi and Kilsby, 2016).

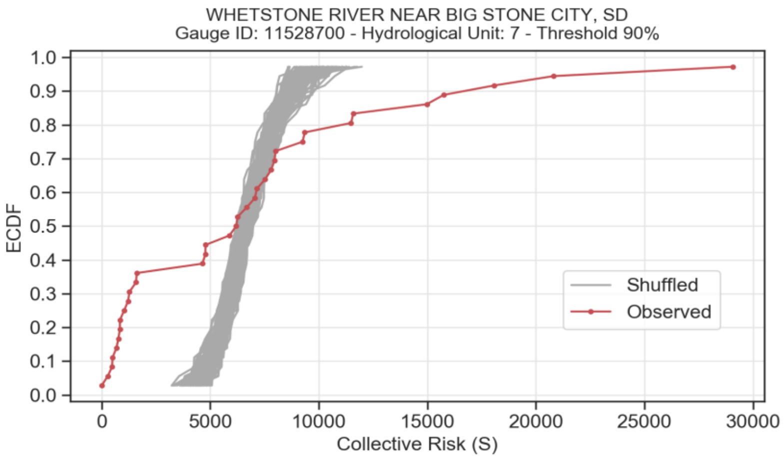
Hence, in order to assess the clustering of extremes of the 360 observed time series, 100 new shuffled time series were reproduced for each one of the 360 original time series.

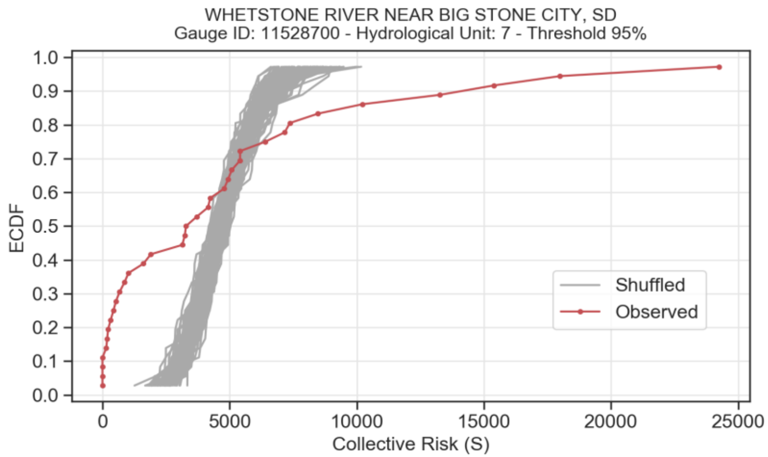
Impacts of clustering mechanisms

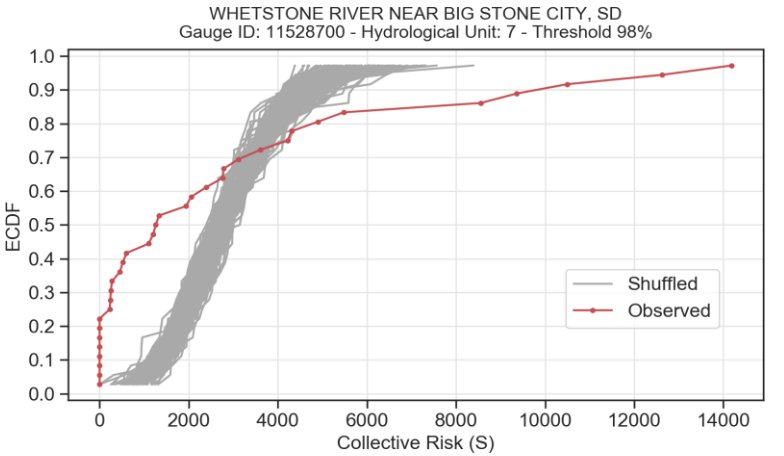
For each one of the 360 gauge locations and for each one of the four selected thresholds (90%, 95%, 98%, 99%), the annual Collective Risk *S*, return intervals and the duration of the over-threshold events were calculated for the observed as well as the shuffled time series. The results from this process are quite impressive as, in many stream gauge locations, the divergence between the observed and the shuffled on the diagram of the Empirical Cumulative Distribution Function (ECDF) is noticeable.

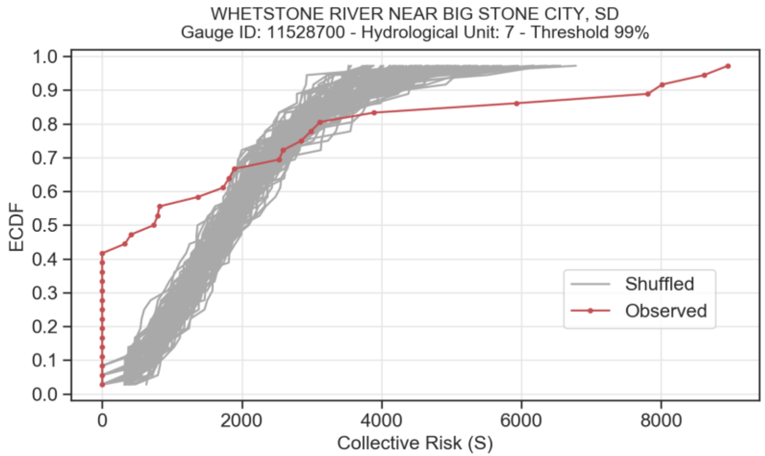
The results regarding the gauge location at the Whetstone River near Big Stone City in South Dakota follow.

Results - ECDF diagram of Collective Risk *S*

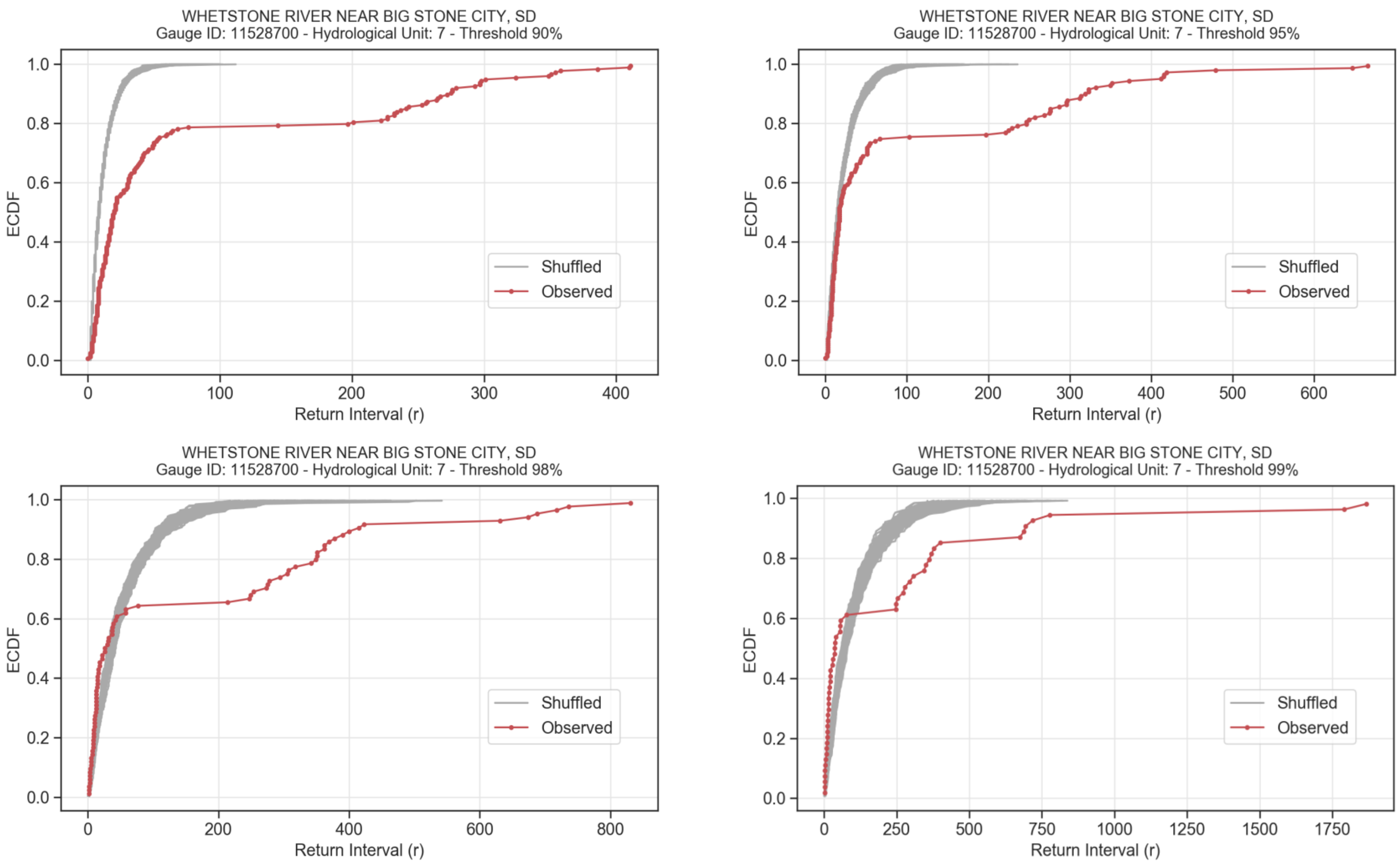




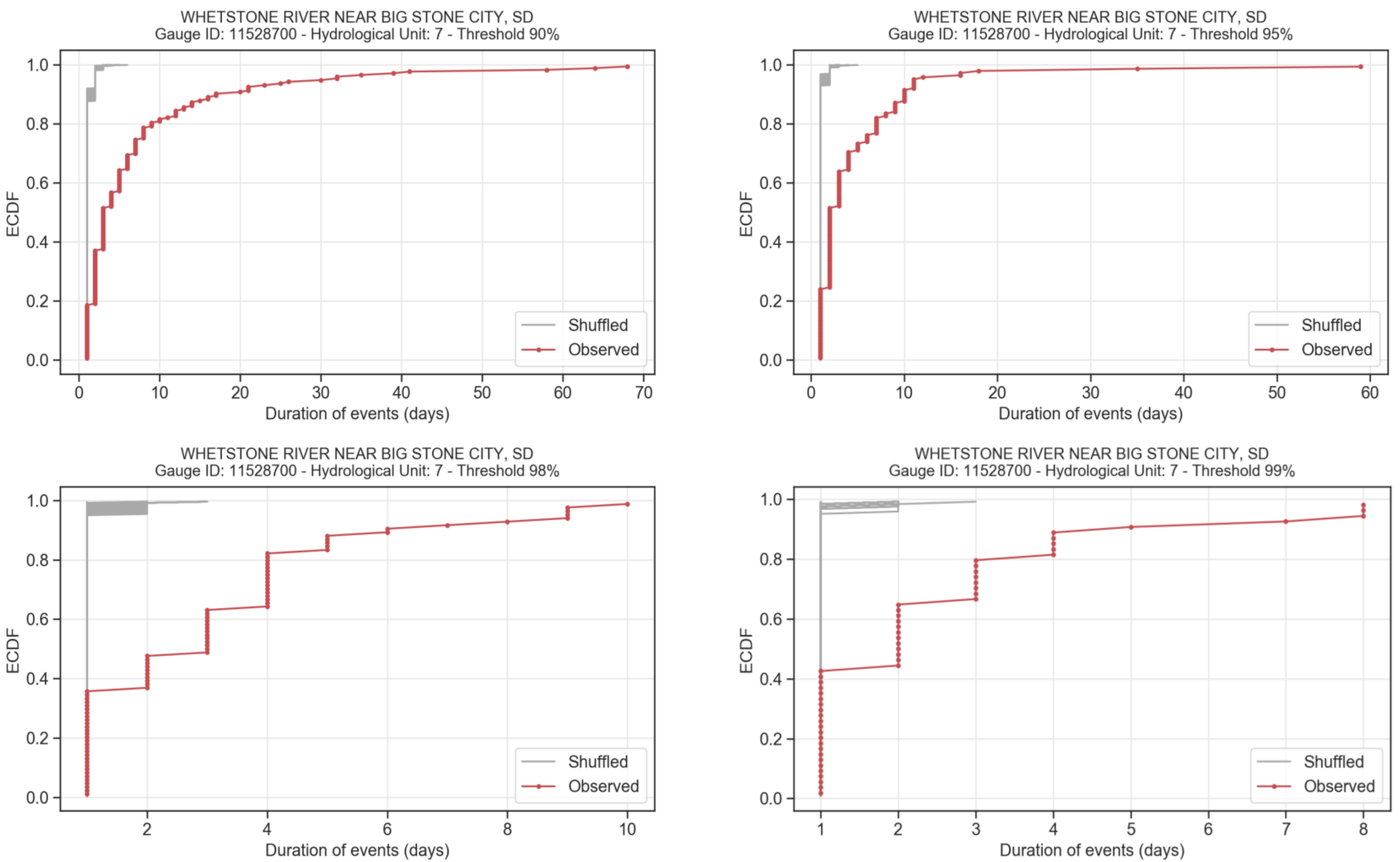




Results - ECDF diagram of Return Intervals



Results - ECDF diagram of Duration of events



The divergence between the observed and the shuffled on the ECDF diagrams of Collective Risk *S*, return intervals and duration of extreme events is evident in many gauge locations.

The ECDF curve of the observed collective risk proxy is contained in the Monte Carlo prediction limits by the GHK model (Koutsoyiannis 2000; 2016), preserving the HK dynamics and the four moments.

(Manolis et. al, EGU2020-9357)

Exploratory Analysis

Spearman correlation coefficient between Average Y_j and Number of Events N

EGU General Assembly 2020

Introduction

A common assumption in insurance is the independence between *Average* Y_j, the average flow which exceeds the selected threshold and is defined as:

$$\frac{1}{N} = \sum Y_j = \frac{S}{N} = Average\ Y_j$$

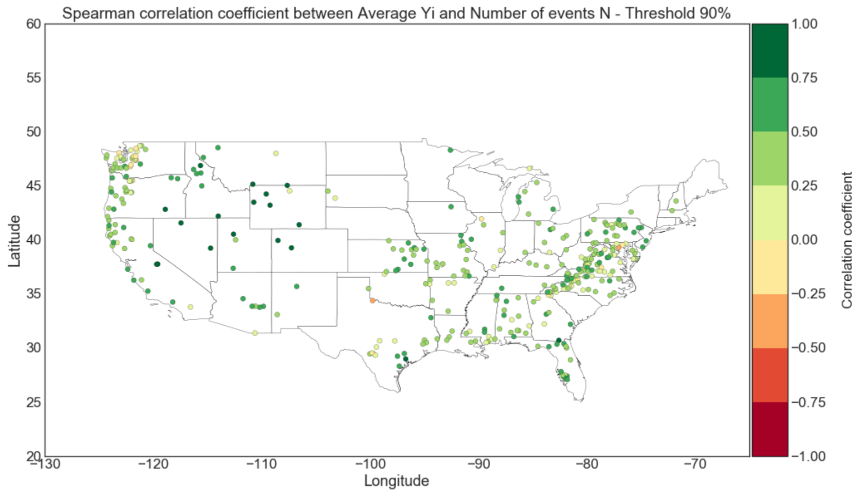
and the number of such exceedances *N* over 365-day time windows (Serinaldi and Kilsby, 2016). Insurance companies' concern about this correlation factor is noteworthy, as they investigate the dependence between the annual number of extreme events and the provoked amount of claims on a specific region.

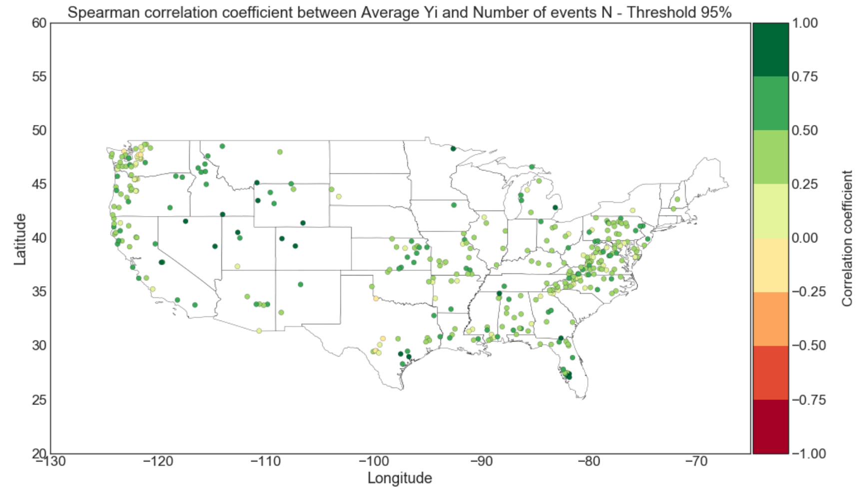
A spatial analysis

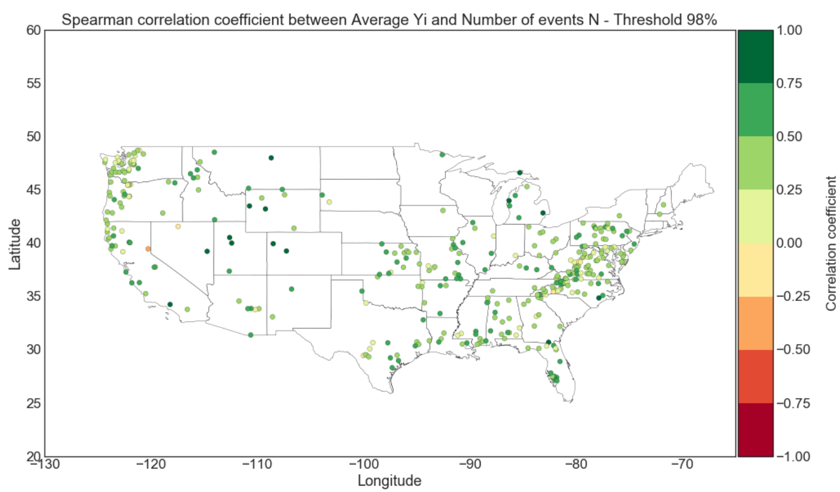
Instinctively, someone would expect that years that are more active in terms of number of events *N* tend to exhibit extreme events also in terms of *Average* Y_j magnitude. Nevertheless, this study shows that this approach cannot be universally applied.

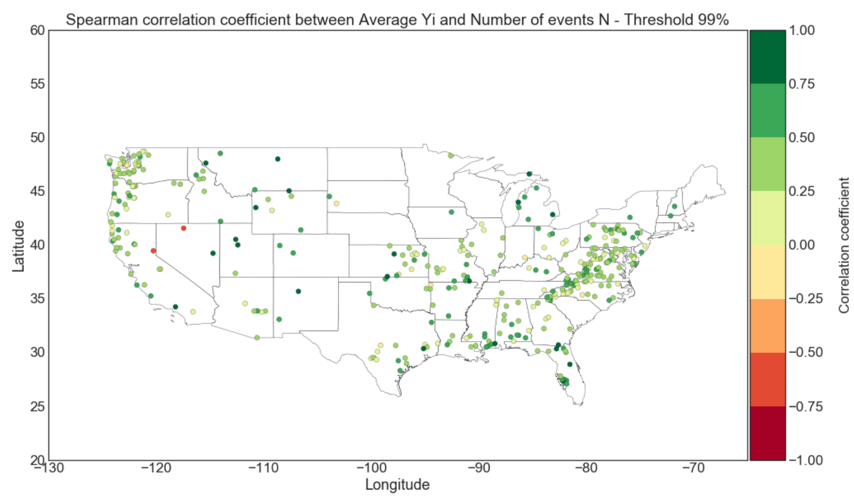
The following figures present the previously mentioned Spearman correlation coefficient of the stream gauge locations for the selected thresholds. This depiction offers a spatial categorization of areas with high correlation coefficient, in contrast with the ones where the correlation coefficient is noticeably lower.

Results - Spearman correlation coefficient of *Average* Y_j and N









This spatial analysis highlights the regions that are subjected to the occurrence of clusters of multiple threshold exceedances of high intensity.

Exploratory Analysis

Clustering mechanisms on the Spearman correlation coefficient between Average Y_j and Number of Events N

EGU General Assembly 2020

Introduction

For each one of the 360 stream gauge locations and for all the selected thresholds, the Spearman correlation coefficient between the *Average* Y_j and the Number of Events N was calculated for the observed as well as the shuffled time series, in order to evaluate the clustering mechanisms on this correlation parameter.

The selected gauge locations which follow are the Suwannee River in GA (ID: 02314500), the Arroyo Seco NR in CA (ID: 11098000), the SF Trinity River in CA (ID: 11528700) and the Cache Creek in WY (ID: 13018300).

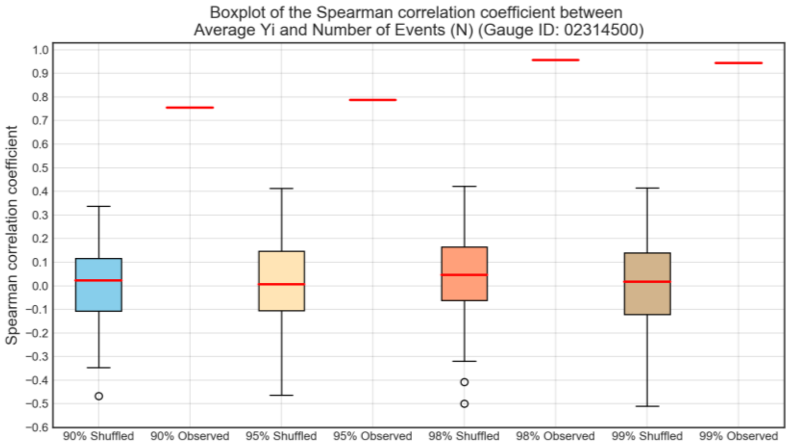
A box-plot depiction

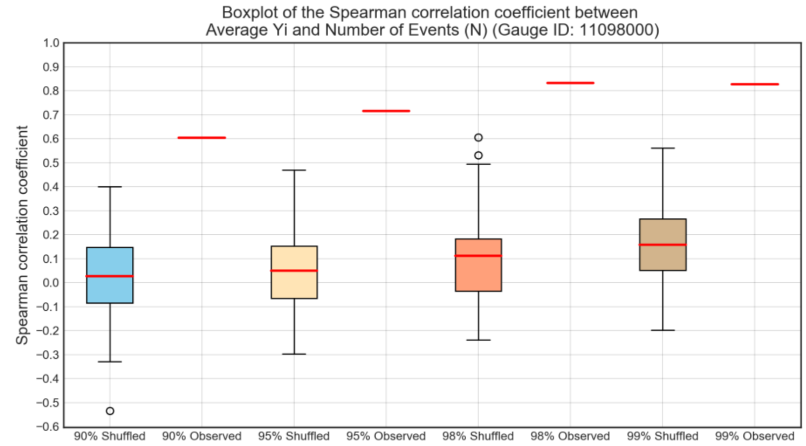
The results of this investigation are quite impressive as the divergence between the correlation coefficient of the observed data and the shuffled ones (considered as independent) is evident.

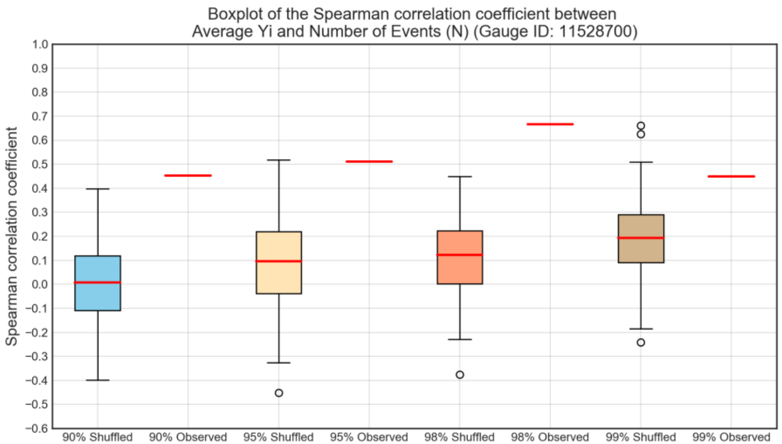
The assumption of independence, which is represented by the shuffled data, could potentially lead policy-makers to inaccurate conclusions.

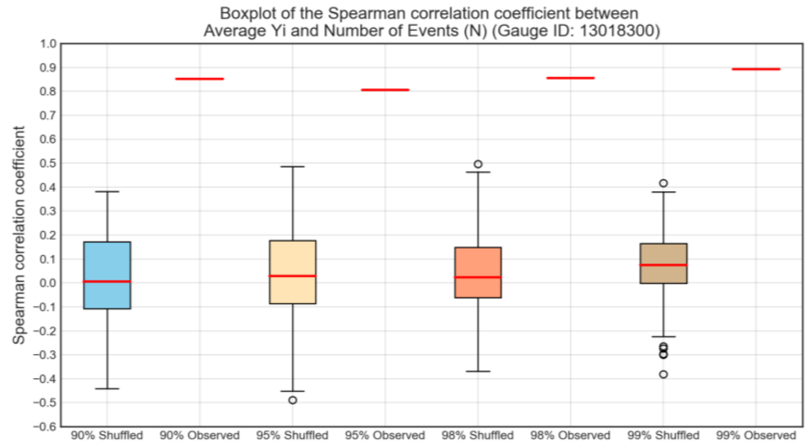
The underestimation of the correlation coefficient between N and the *Average* Y_j, could provoke significant financial impacts.

Results - Clustering mechanisms on *Average Y_j* and *N* Spearman correlation coefficient









The box-plots show that the interplay between short range dependence (SRD) and long range dependence (LRD) introduce significant correlation between the number of events N and the *Average* Y_j.

Exploratory Analysis

Investigating the correlation between the FEMA Claims Records and Collective Risk S

EGU General Assembly 2020

Introduction

The recently published FEMA NFIP Claims Records (FEMA, 2019) offer us the opportunity to investigate the validity of the method we developed on a spatial basis by correlating these records with the results of our study.

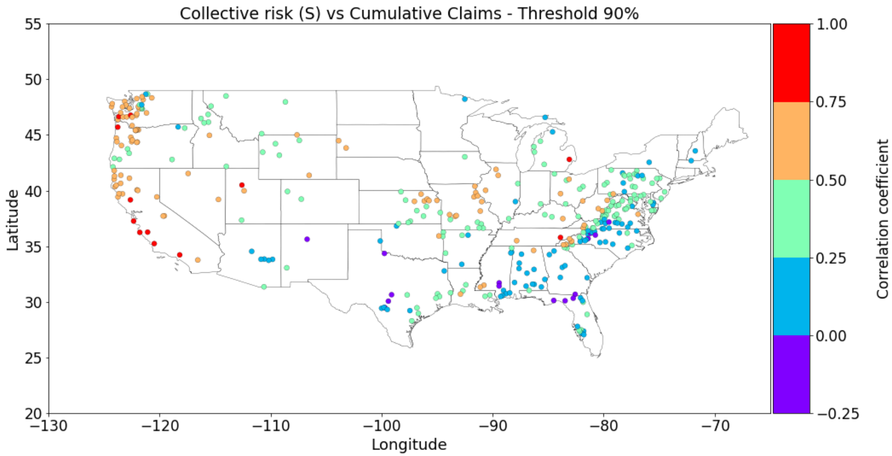
We spatially distributed the FEMA NFIP Claims Records on the 21 Hydrological Units (HU). Subsequently, the Spearman correlation coefficient was evaluated between the annual Collective Risk *S* for the selected gauge locations and the cumulative claims of the Hydrological Unit that each one of these gauge locations belongs to.

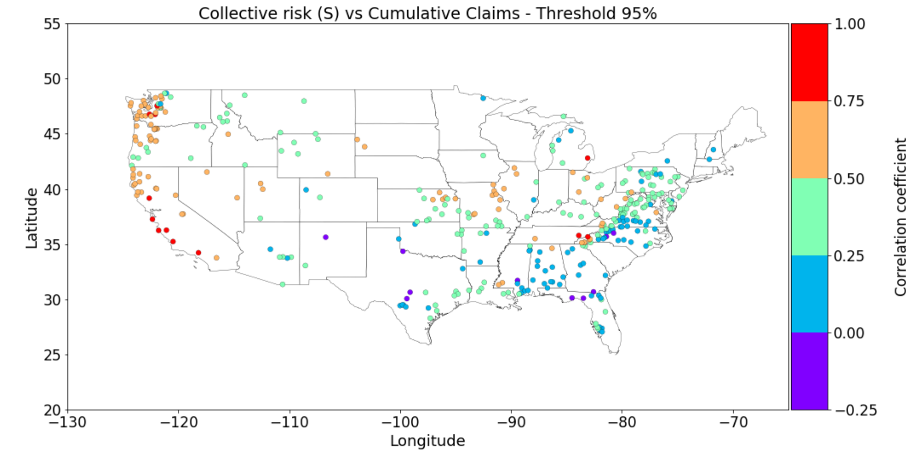
A spatial analysis

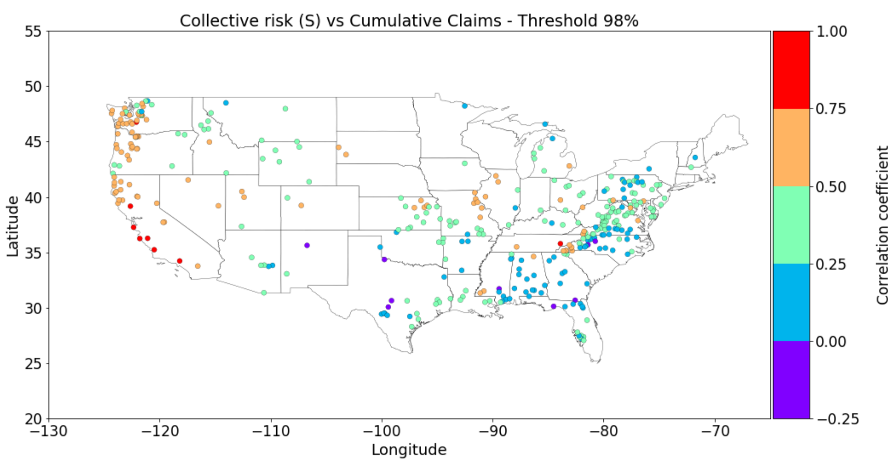
The following USA maps highlight the Spearman correlation coefficient between the Collective Risk *S* and the Hydrological Units' claims records for the selected thresholds and gauge locations, indicating the areas where this parameter is higher or lower.

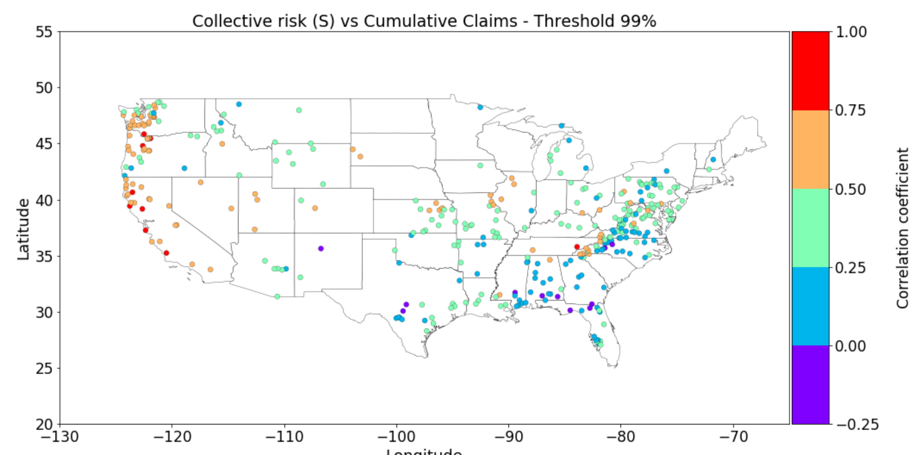
A spatial pattern is evident, showing that higher values of the Spearman correlation coefficient emerge in West Coast, in contrast with the ones in the East Coast, which are significantly lower.

Results – Spearman correlation coefficient between Collective Risk *S* and HU's FEMA NFIP Claims Records









As Collective Risk *S* refers to river flooding, these results show that this type of flooding is dominant in West Coast.

In contrast, it is revealed that the source of flooding events that provoke claims in East Coast present a different and more complicated pattern, mainly due to the significant vulnerability of these areas to hurricane hits and storm surge phenomena.

Exploratory Analysis

Clustering mechanisms on the Spearman correlation coefficient between the FEMA NFIP Claims Records and Collective Risk S

EGU General Assembly 2020

Introduction

For each one of the 360 gauge locations and for all the selected thresholds, the Spearman correlation coefficient between the Collective Risk *S* and the FEMA NFIP Claims Records of the particular Hydrological Unit that each gauge location belongs to was calculated for the observed as well as the shuffled time series, in order to evaluate once again the effect of the clustering mechanisms on this correlation parameter.

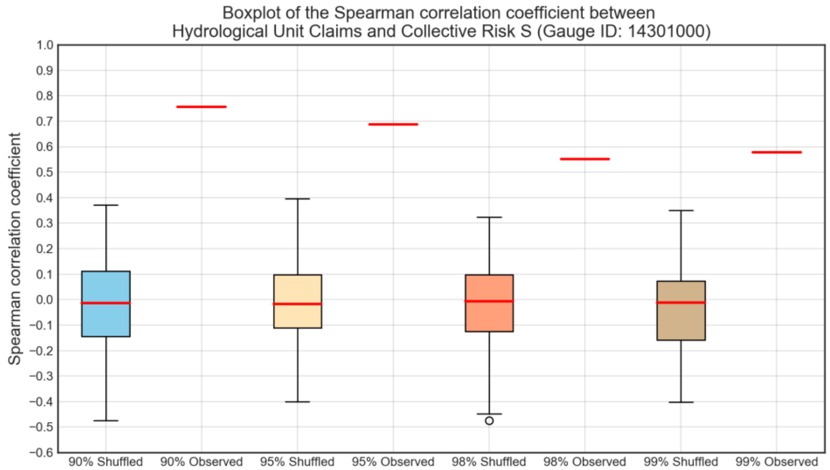
Box plots that follow show that SRD and LRD introduce significant correlation.

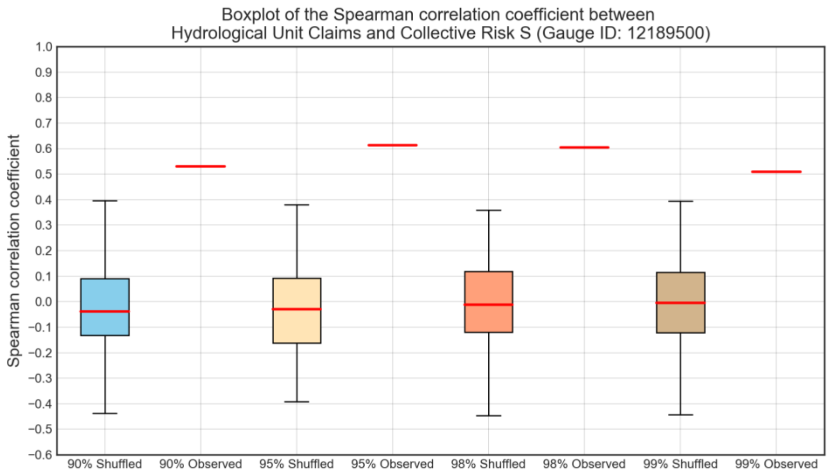
A box-plot depiction

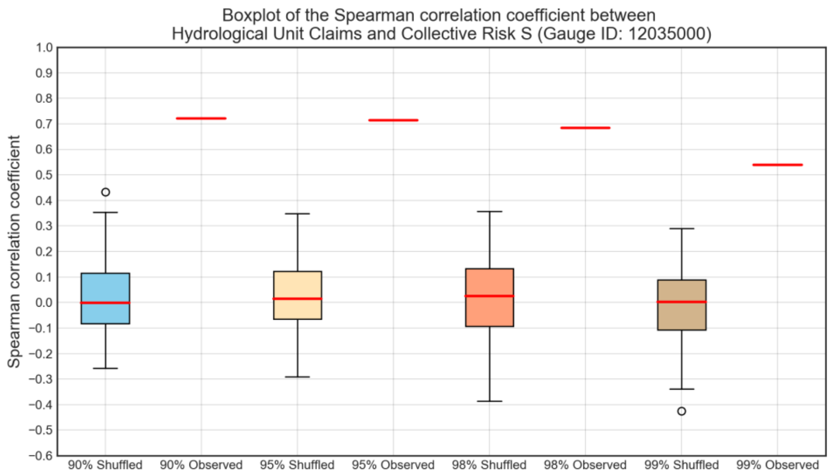
The stream gauge locations presented in the following box-plots are:

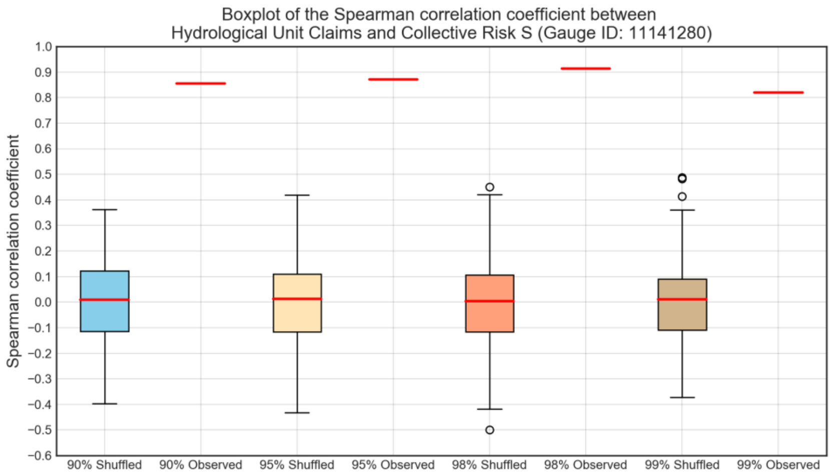
- Nehalem River in OR (ID: 14301000)
- Sauk River Near Sauk in WA (ID: 12189500)
- Satsop River in WA (ID: 12035000)
- Lopez C NR in CA (ID: 11141280)
- Big Bureau Creek in IL (ID: 05556500)
- Paint Rock River in AL(ID: 03574500)
- Shade River in OH (ID: 03159540)
- Piscataway Creek in VA (ID: 01669000)

Results - Clustering mechanisms on Spearman correlation coeff. between Collective Risk *S* and HU's FEMA NFIP Claims Records

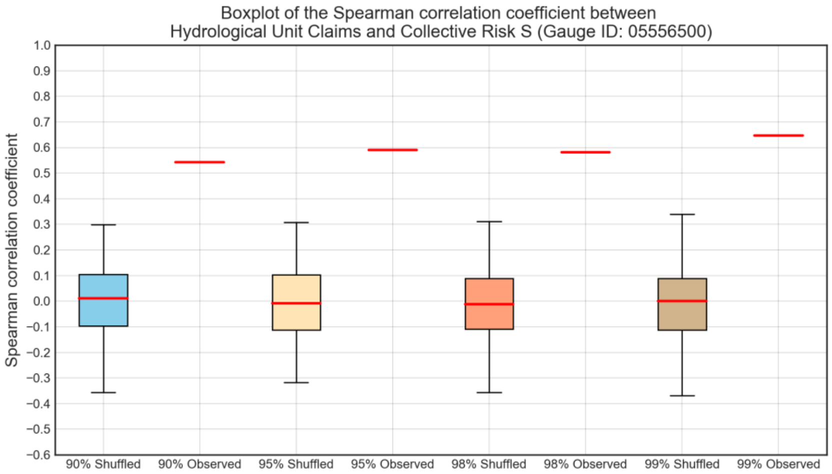


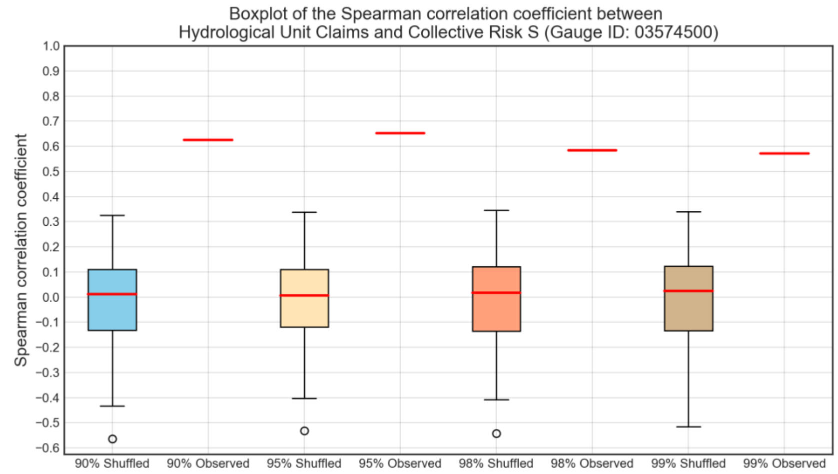


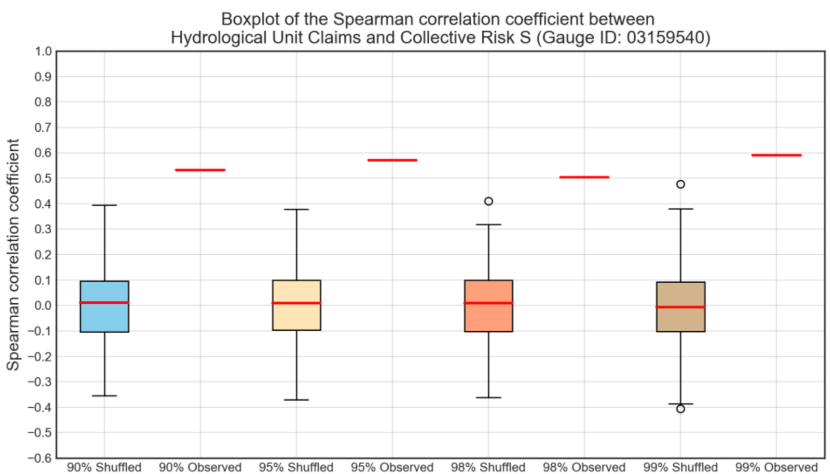


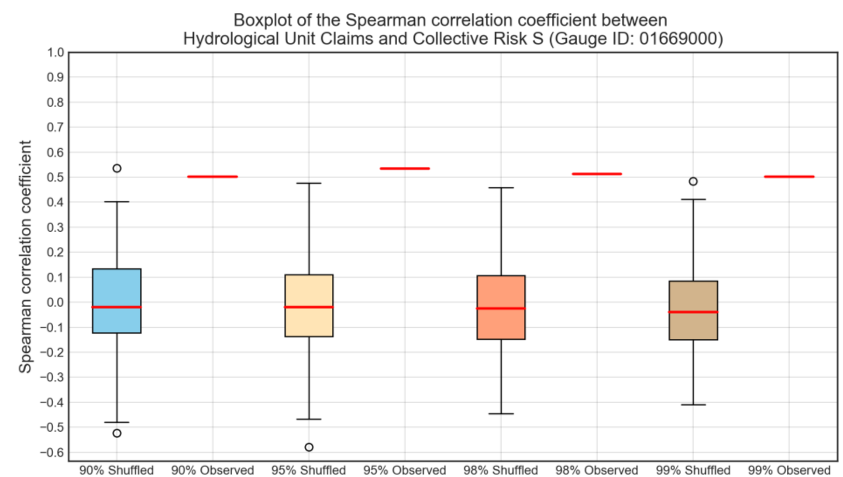


Results - Clustering mechanisms on Spearman correlation coeff. between Collective Risk *S* and HU's FEMA NFIP Claims Records



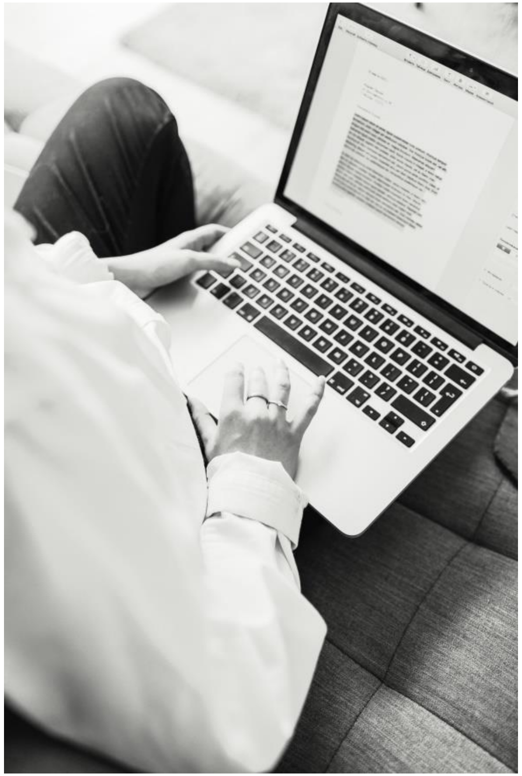






EGU General Assembly 2020

Results show that the shuffled data (independent) are significantly less accurate in comparison with the observed ones, as they tend to underestimate the Spearman correlation coefficient between the Collective Risk *S* and the FEMA NFIP Claims Records.



Learn More

Further information about the project are obtainable at our research team's site https://www.itia.ntua.gr/en/

References

- Dimitriadis, P., and Koutsoyiannis, D., Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst–Kolmogorov processes, *Stochastic Environmental Research & Risk Assessment*, 29 (6), 1649–1669, doi:10.1007/s00477-015-1023-7, 2015.
- Dimitriadis, P., and Koutsoyiannis, D., The mode of the climacogram estimator for a Gaussian Hurst-Kolmogorov process, *Journal of Hydroinformatics*, 22(1), 160–169, 2020.
- FEMA, Federal Insurance & Mitigation Administration National Flood Insurance Program (FIMA NFIP) Redacted Claims Dataset, *https://www.fema.gov/media-library/assets/documents/180374*, 2019.
- Goulianou, T., Papoulakos, K., Iliopoulou, T., Dimitriadis, P. and Koutsoyiannis, D., Stochastic characteristics of flood impacts for agricultural insurance practices, *European Geosciences Union General Assembly 2019*, Vienna, EGU2019-5891.
- Iliopoulou, T., and Koutsoyiannis, D., Revealing hidden persistence in maximum rainfall records, *Hydrological Sciences Journal*, 64 (14), 1673–1689, doi:10.1080/02626667.2019.1657578, 2019.
- Koutsoyiannis, D., A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, *Water Resources Research*, 36 (6), 1519–1533, 2000.
- Koutsoyiannis, D., Hurst-Kolmogorov Dynamics and Uncertainty, *Journal of the American Water Resources Association*, 47(3), pp.481-495, 2011.
- Koutsoyiannis, D., Generic and parsimonious stochastic modelling for hydrology and beyond, *Hydrological Sciences Journal*, 61 (2), 225–244, 2016.
- Manolis, G.N., Papoulakos, K., Iliopoulou, T., Dimitriadis, P., Tsaknias, D., and Koutsoyiannis, D., Clustering mechanisms of flood occurrence; modelling and relevance to insurance practices, *European Geosciences Union General Assembly 2020*, Vienna, EGU2020-9357.
- Newman, A., Sampson, K., Clark, M.P., Bock, A., Viger, R.J., and Blodgett, D., A large-sample watershed-scale hydrometeorological dataset for the contiguous USA, *Boulder, CO: UCAR/NCAR*, 2014.
- Serinaldi, F., and Kilsby, C.G., Understanding Persistence to Avoid Underestimation of Collective Flood Risk, *Water,* 8, 152, 2016.