EGU General Assembly 2024

Session HS7.4
Future hydroclimatic scenarios
in a changing world

EGU24-12363

Exploring Hurst-Kolmogorov Dynamics: Unraveling the (temporal) link between Flood Insurance Claims and Streamflow Extremes in the contiguous USA

[1]Georgios T. Manolis, [1*]Konstantinos Papoulakos, [1]Theano Iliopoulou, [1]Panayiotis Dimitriadis, [2]Dimosthenis Tsaknias and [1]Demetris Koutsoyiannis

[1]Department of Water Resources, Faculty of Civil Engineering, National Technical University of Athens, Heroon Polytechneiou 5, GR-157 80 Zografou, Greece

[2] Independent researcher, Greece

* Corresponding author. E-mail address: papoulakoskon@gmail.com

Based on a submitted paper that is under review in the *Natural Hazards* scientific journal.
Preprint: https://www.researchsquare.com/article/rs-4184407

CC BY

Abstract

This research investigates the intricate relationship between flood insurance claims and streamflow extremes in the contiguous USA, challenging the conventional belief of independence and non-catastrophic nature of insurable flood losses.

Focusing on the Hurst-Kolmogorov dynamics, which emphasizes the temporal dependence of extreme flood events, we explore the implications of these dynamics on flood insurance practices and streamflow extremes.

By analyzing the US-CAMELS dataset, we investigate the clustering mechanisms' impact on return intervals, event duration, and severity of the over-threshold events, which are treated as proxies for collective risk.

Furthermore, stochastic approaches are developed to explore the correlation between properties of extreme events and recently published FEMA National Flood Insurance Program claims records in an exploratory analysis.

This study aims to contribute valuable insights into the temporal aspects of streamflow extremes, considering the dependencies identified by the Hurst-Kolmogorov dynamics and providing essential information for enhancing the accuracy of flood insurance and reinsurance practices

Highlights

A stochastic investigation of the **clustering dynamics** in streamflow **extremes** is performed using the US-CAMELS dataset.

Clustering of streamflow extremes is associated to **aggregate flood insurance claims** from the new **FEMA database** under the collective risk concept.

Links of a streamflow-based proxy to actual aggregate flood insurance claims are found to be **spatially variable**, reflecting different types of **flood-generating mechanisms** in the USA.

Keywords: Flood insurance claims, Hurst-Kolmogorov dynamics, Clustering, Monte-Carlo simulation, Collective risk assessment, Streamflow extremes, FEMA

US-CAMELS dataset

This analysis is applied on the US-CAMELS dataset, which comprises of **671 daily streamflow time series** from catchments in the contiguous United States (CONUS) that are **minimally impacted by human activities** (Newman et al., 2014).

From this dataset, **360 streamflow time series** with the maximum temporal overlap (namely, 35 years from 1980 to 2014) and less than 10% of missing values **were selected**. Figure 1 shows the study area and stream gauge locations for the full dataset including the finally selected 360 stream gauge locations.

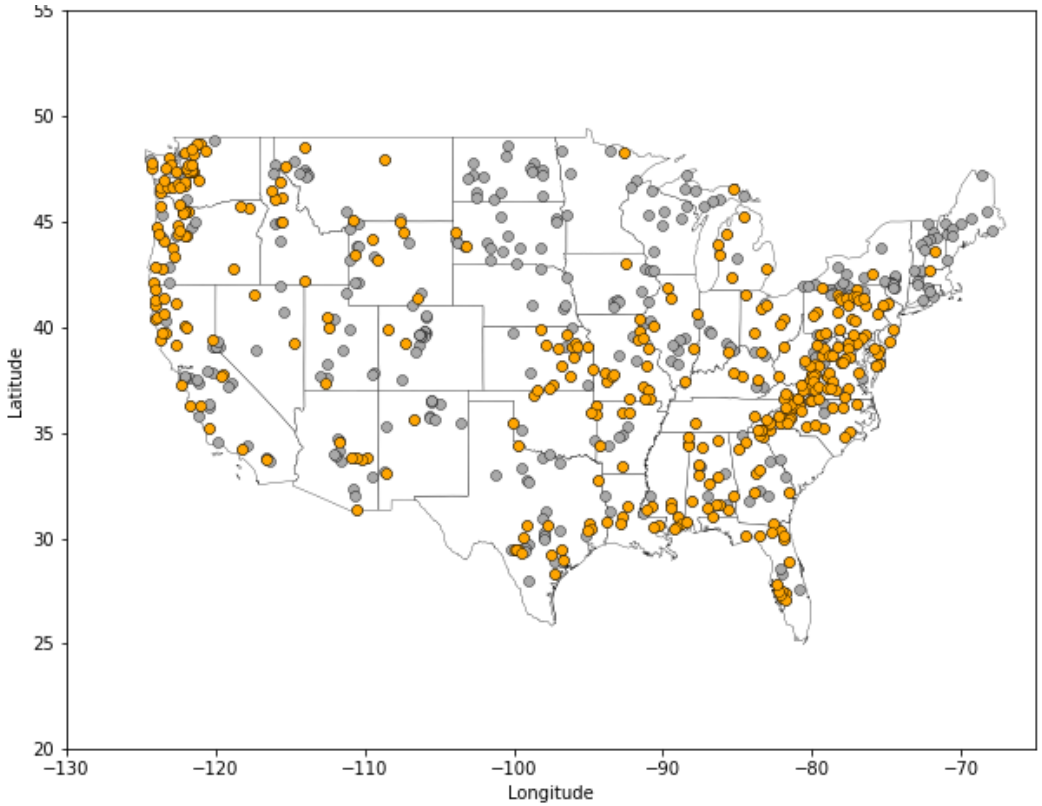


Fig. 1 The 671 US-CAMELS stream gauge locations. The selected 360 US-CAMELS stream gauge locations are colored orange.

FEMA's NFIP claims records dataset

EGU General Assembly 2024

Federal Emergency Management Agency (FEMA) published in **2019** the National Flood Insurance Program (NFIP) data, including more than **two million claims records** dating back to 1970 and more than **47 million policy records** for transactions (FEMA, 2019).

It is evident that this is a **giant contribution** for supporting scientists and policy-makers on their research on **how** the National Flood Insurance Program (NFIP) works, **where** flood damage occurs, and **what** the costs are.

Methodology: Extreme value analysis (EVA) distributions

EGU General Assembly 2024

EVA is widely used and applied as a tool to analyze and study statistics on sample values that **deviate extremely** from the mean of the full sample, in order to develop a deeper understanding of the sample and **precise modeling strategies**. It generates significant applications across many scientific fields such as **hydrology**, **insurance** and **finance** and can be also used to predict the occurrence of rare events, such as **extreme flooding**, large **insurance losses**, crashing of the **stock market** and many others (Reis and Thomas, 2007).

In this manner, **generalized extreme value distribution (GEV)** and **generalized Pareto distribution (GPD)** are introduced as a tool for the statistical analysis of maxima or minima and of exceedances over a given threshold.

Methodology: Threshold selection

Threshold selection is a challenge in insurance and especially in flood insurance practices (Robinson and Botzen, 2020). The threshold should be chosen such that **all losses above the threshold are "extreme losses"** in the sense of the underlying extreme value analysis.

On one hand, we want to choose a **high threshold** in order to investigate the behavior of the (really) extreme events. On the other hand, for the estimation of the parameters in the distribution of the extreme losses, we need **many observations** above the threshold to create a solid statistical foundation for our conclusions, based on a long sequence of values.

Methodology: Threshold selection

In order to characterize the **dynamics of extreme streamflow values**, this study performed a POT analysis using **four** different percentage **thresholds**, i.e., 90%, 95%, 98%, and 99%.

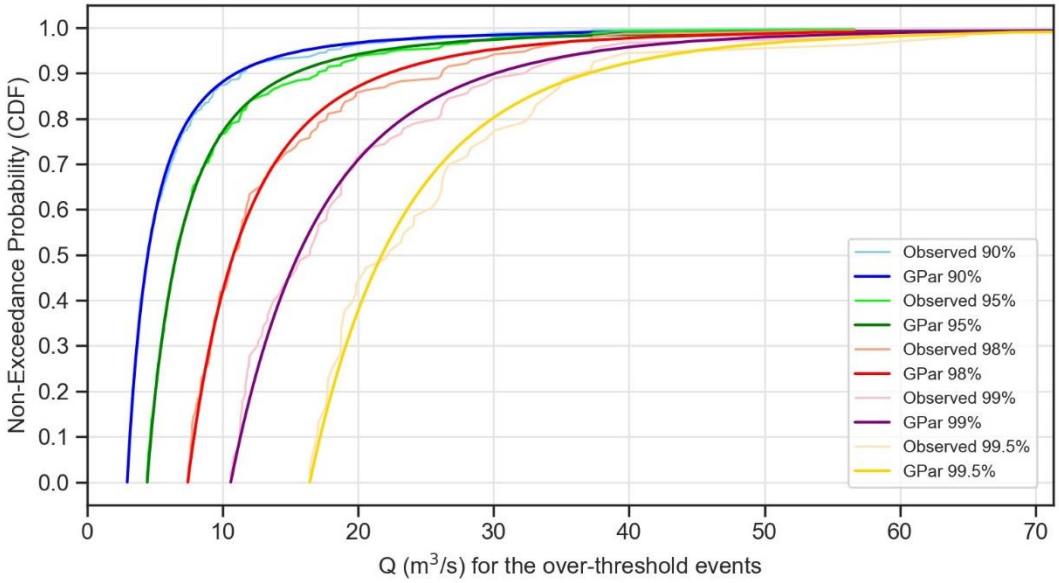


Fig. 2 Diagram that shows the impact of threshold selection on Non-Exceedance Probability (CDF) of streamflow of the over-threshold events regarding the observed streamflow records as well as the ones that were developed by the process of fitting these observed data with the generalized Pareto distribution. Gauge ID: 01552500.

Methodology: Collective risk model in insurance

The distribution of **total claim amounts**, considering the insurance company's portfolio as a collective that produces a random number N of claims in a certain time period, can be described by the **collective risk model** (Kaas et al., 2008).

Collective risk S_x is defined as

$$S_x = X_1 + X_2 + \cdots + X_N, \tag{1}$$

where X_i is the i^{th} claim amount during a certain time period, e.g. a year. Apparently $S_x = 0$ if $N = 0$.

Methodology: Collective risk model in flood insurance

EGU General Assembly 2024

Similarly, regarding **flood insurance** practices and in case of an extreme flood event, the collective risk *S* is the **total claim amount**, considering again the portfolio of (re)insured properties as a collective that produces a random number *N* of claims in a certain time period of **one year** in our case.

Methodology: Collective risk model in flood insurance

Denoting the records y_t of a time series, a **proxy of temporal collective risk *S*** is defined by Serinaldi and Kilsby (2016) as

$$S = \sum_{j=1}^{N} Y_j \tag{2}$$

where Y_j is the jth claim amount proxy (over-threshold flow fluctuation severity). Again, the total claim amounts $S = 0$ if $N = 0$. The definition of collective risk regarding flood insurance practices is a proxy of the actual collective risk, as it involves **hydrological series** and not actual claim amounts. In this study, regarding the aforementioned proxy of temporal collective risk, we use the term **Proxy Aggregated Losses *S***.

Methodology: Sequence of independent variables

In order to characterize the dependence and the clustering mechanisms, it is important to quantify how the time series differs from a sequence of independent variables.

A widely used method to create a sequence of independent variables is to **shuffle** (randomize) the series in order to get a new series which has the **same marginal distribution** but **no correlation**; the quantification of the distance between the independent and the observed variables is performed by **comparing specific characteristics**, i.e. the annual Proxy Aggregated Losses, the duration of the peak-over-threshold events and the occurrence frequency of return periods in the original time series and in the shuffled one.

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.

Methodology: The Hurst – Kolmogorov dynamics

The exhibited **persistence** in many **natural processes**, including streamflow and rainfall dynamics, is known as the Hurst phenomenon or **Hurst-Kolmogorov (HK) dynamics** and is quantified by the Hurst coefficient *H*.

In order to calculate the Hurst coefficient *H* and detect the potential **long-term dependence** (or else persistence, clustering) of a process, the most accurate method is by formulating the ***Climacogram*** (Koutsoyiannis, 2010), which has been shown to outperform estimators based on the autocovariance and power-spectrum (Dimitriadis and Koutsoyiannis, 2015).

Methodology: Generalized-HK (GHK) process

In some cases, such as in this study, fitting of straight line in the Climacogram derived from the observed data cannot capture the full variance behavior of the process at the whole range of scales. Thus, the generalized-HK (GHK) model is applied.

The **generalized-HK (GHK) model** is applied, which exhibits also an HK behavior in large scales but has **more flexibility** in **smaller scales** (Dimitriadis and Koutsoyiannis, 2018; note that a more advanced scheme has been introduced that can preserve any number of moments; Koutsoyiannis and Dimitriadis, 2021).

The ***Climacogram*** of the **GHK model** is the following, where the Hurst coefficient H is bounded between zero and one inclusive, q is positive, while λ and q have dimensions [x2] and [T], respectively:

$$\gamma(k) = \frac{\lambda}{(1 + k/q)^{2-2H}} \quad (3)$$

Methodology: Symmetric – Moving Average (SMA) method

In this study, the symmetric moving average (SMA) scheme (Koutsoyiannis 2000; 2016) is also applied in order to develop and evaluate potential modeling strategies.

EGU General Assembly 2024

SMA is a general algorithm for producing synthetic time series of a physical quantity by **preserving its dependence structure**.

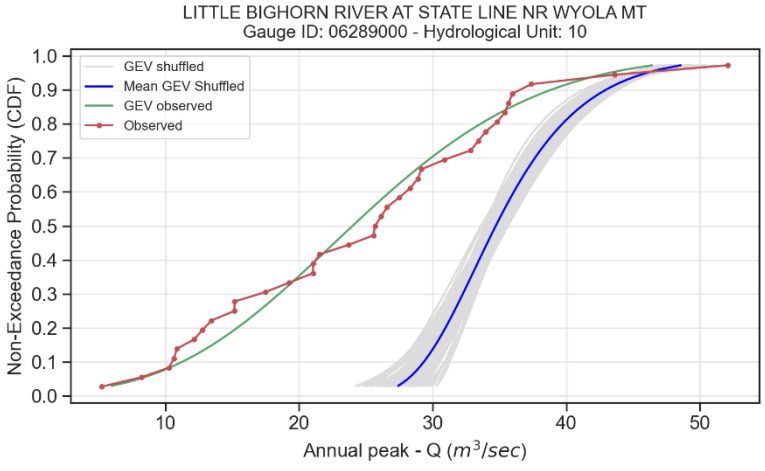
In particular, SMA generation scheme for approximating the **marginal probability function** can replicate a natural process by exactly preserving a selected number of central moments, with four found to be sufficient for **various distributions** commonly applied in geophysical processes (Dimitriadis and Koutsoyiannis, 2018).

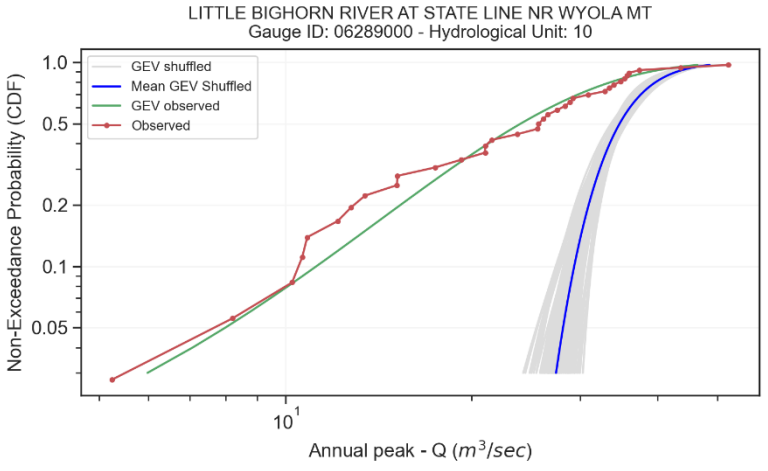
The algorithm to produce time series with the SMA scheme required the first four **central moments**, the ***H*** coefficient of each physical quantity (average, maximum and minimum) as well as the **length** of the time series.

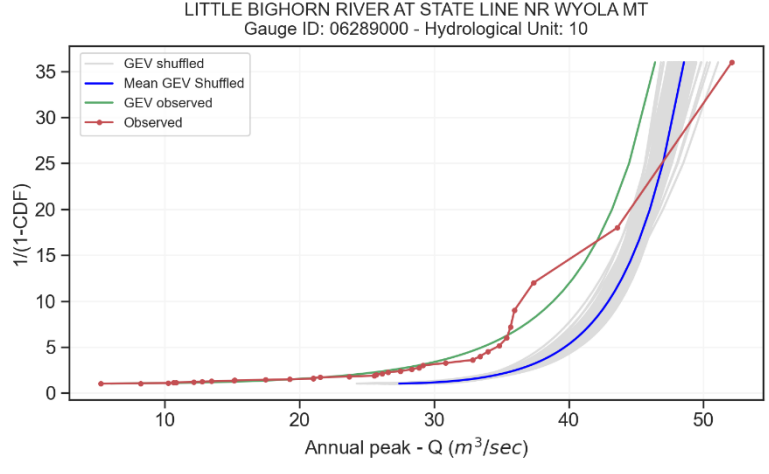
Results: Impact of clustering mechanisms on GEV distribution modelling

EGU General Assembly 2024

Fig. 3 Annual peak Non-Exceedance Probability (CDF) diagrams related with GEV simulations in linear and logarithmic scale, and the return period (1/(1-CDF)) scale (Gauge ID: 11528700).



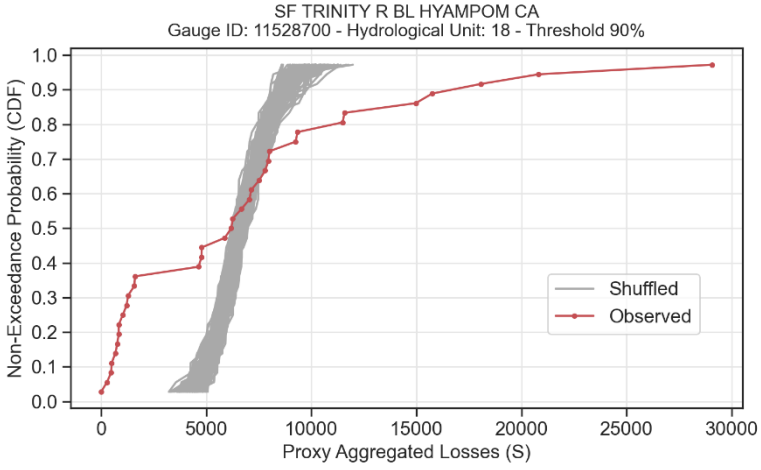


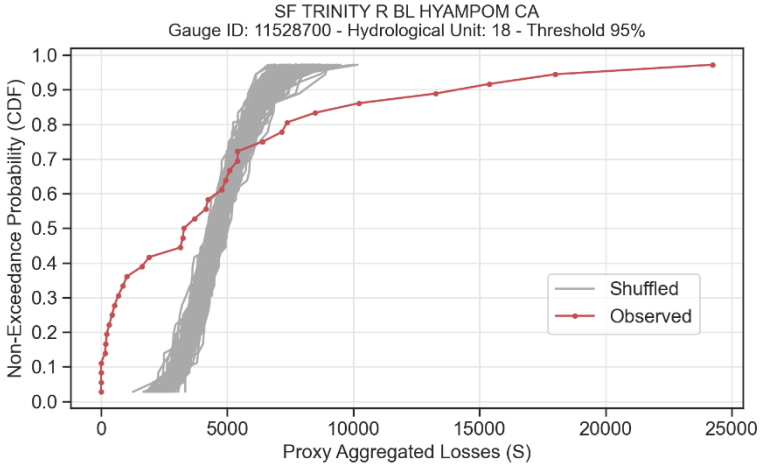


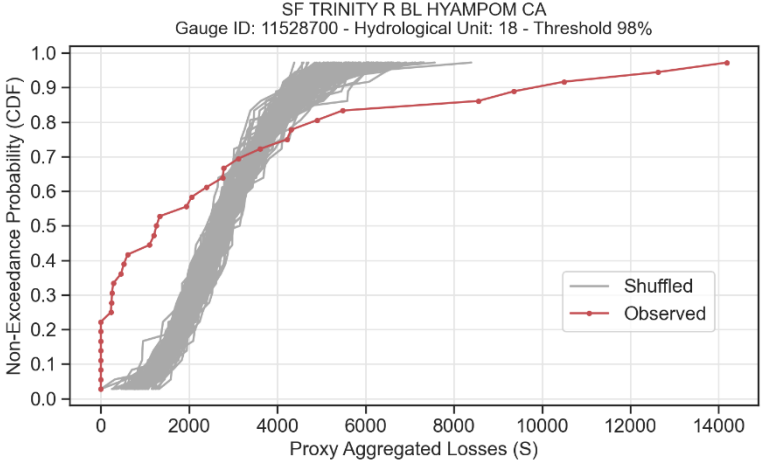
Results: Impact of clustering mechanisms on streamflow-based Proxy Aggregated Losses

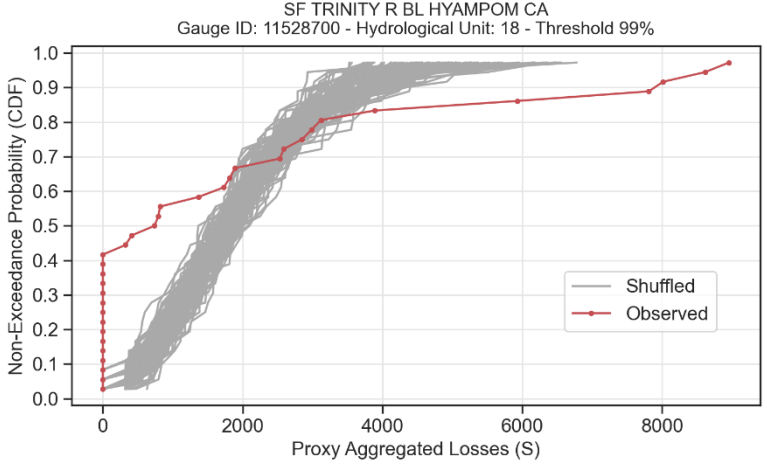
EGU General Assembly 2024

Fig. 4 Proxy Aggregated Losses's Non-Exceedance Probability (CDF) diagrams in linear scale (Gauge location ID: 11528700).





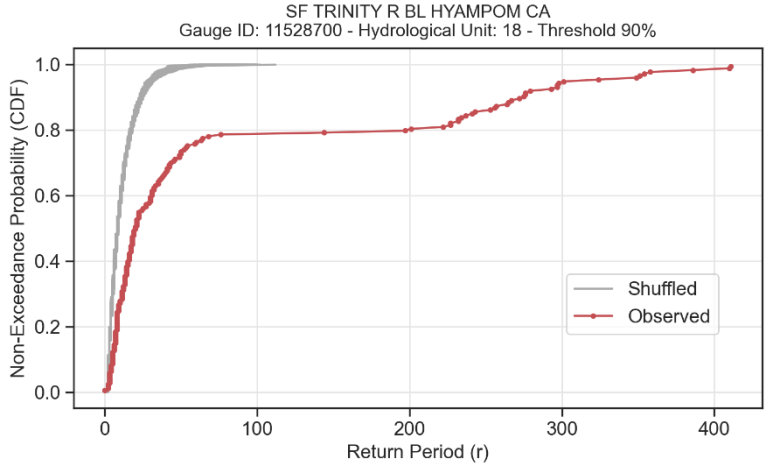


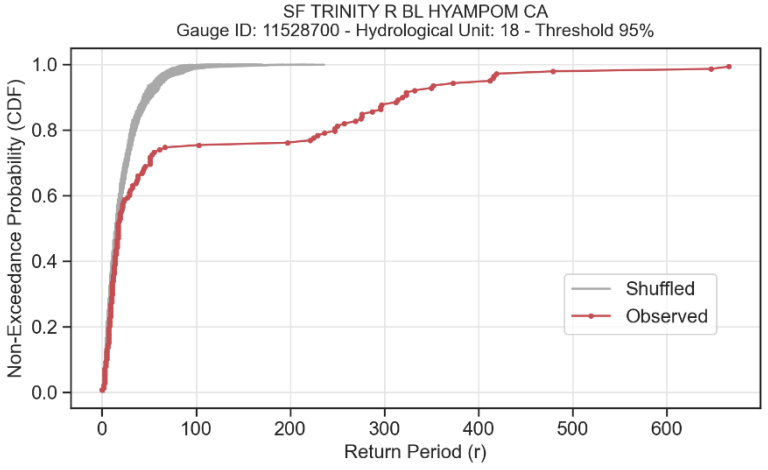


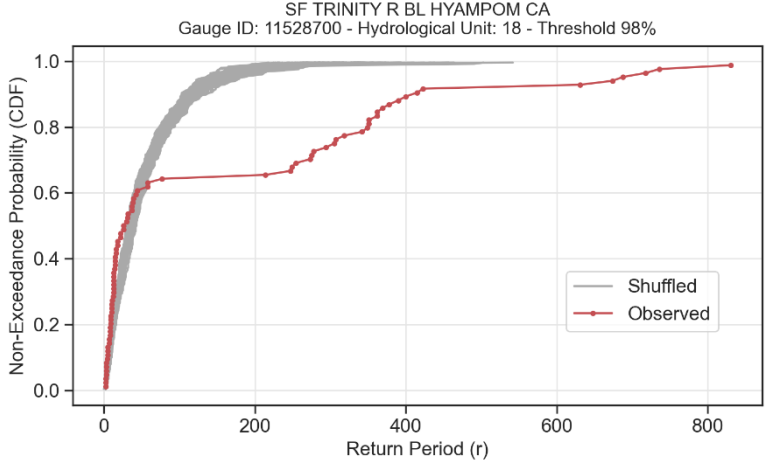
Results:
Impact of clustering mechanisms on return periods

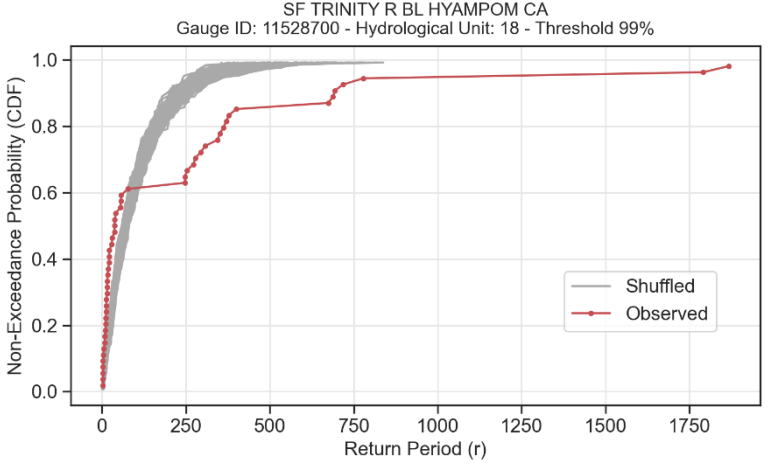
EGU General Assembly 2024

Fig. 5 Return period's Non-Exceedance Probability (CDF) diagrams in linear scale (Gauge location ID: 11528700).





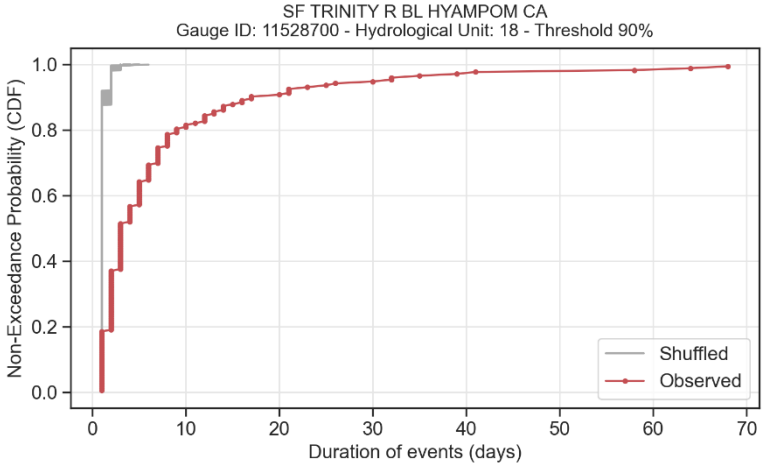


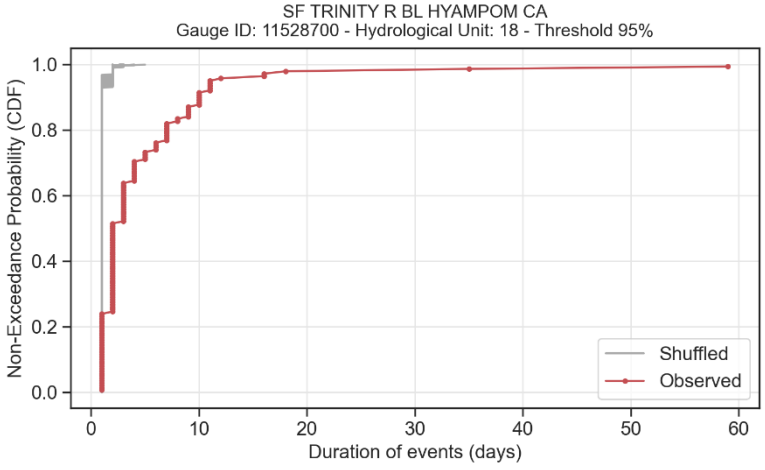


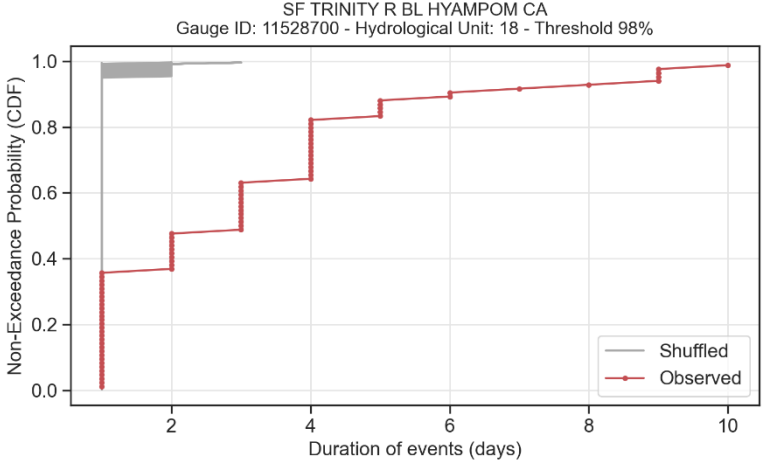
Results:
Impact of clustering mechanisms on the duration of the over-threshold events

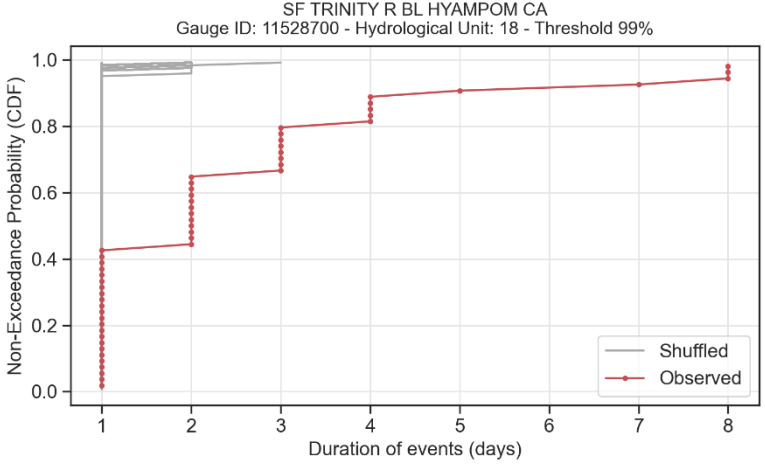
EGU General Assembly 2024

Fig. 6 Events' duration Non-Exceedance Probability (CDF) diagrams in linear scale (Gauge location ID: 11528700).









Reproducing observed clustering using HK dynamics and Monte Carlo simulations

Generalized-HK (GHK) model

EGU General Assembly 2024

Based on the **mean *Climacogram*** of the **GHK process** regarding the 360 empirical streamflow time series of the US-CAMELS dataset, a **persistent behavior** was indicated with parameters H = 0.63 and q = 6.94 days (Figures 7-8).

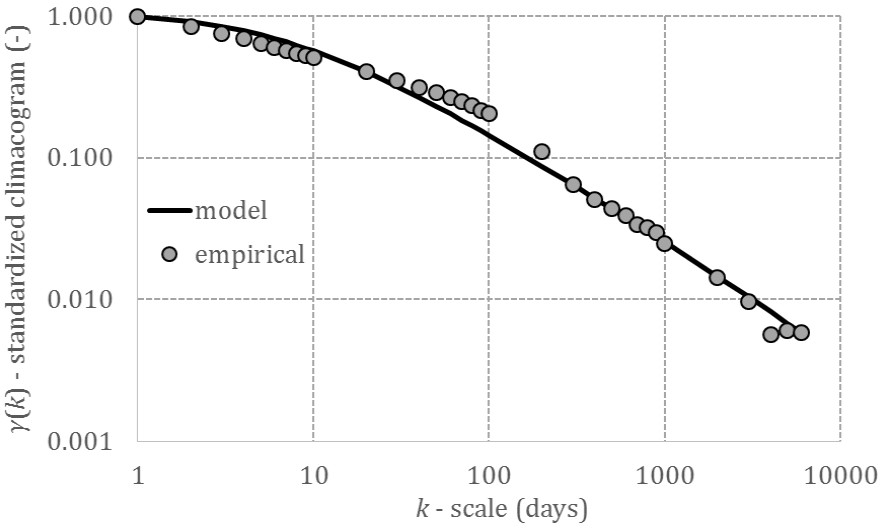


Fig. 7 The mean Climacogram of the 360 selected gauge locations of the US-CAMELS dataset.

Reproducing observed clustering using HK dynamics and Monte Carlo simulations

Generalized-HK (GHK) model

The effect of this **dependence structure** is tracked on the behaviors of POT flows at the annual scale and the estimation of the **Proxy Aggregated Losses**. The behavior of daily streamflow in our dataset is found to be **consistent with HK dynamics** (Dimitriadis et al., 2021) characterized by moderate *H* parameters (in the range 0.6-0.7), through Monte Carlo simulations.

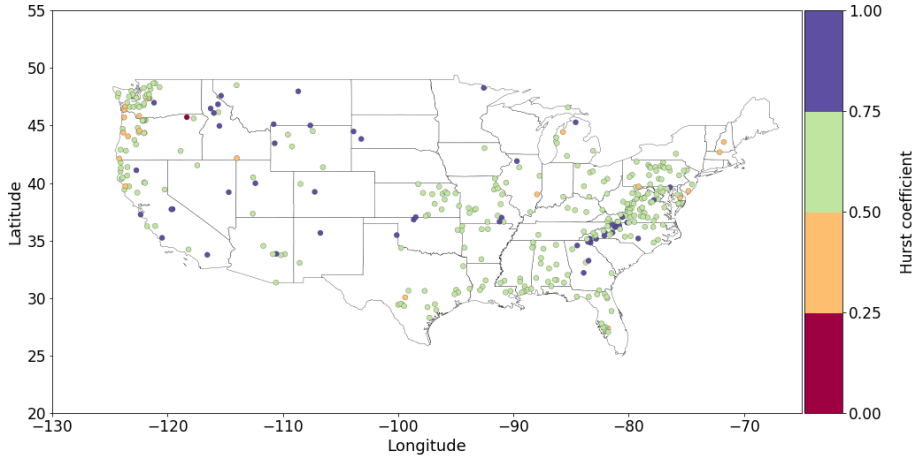


Fig. 8 Hurst coefficient *H* of each one of the 360 selected gauge locations of the US-CAMELS dataset.

Reproducing observed clustering using HK dynamics and Monte Carlo simulations

SMA-GHK model

In order to develop the stochastic simulation of a series with generalized long-range dependence, the **SMA-GHK model** is applied by preserving explicitly the first four central moments of the sample series.

The algorithm to produce synthetic time series from the data of the **observed** (empirical) one with the **SMA scheme**, created by P. Dimitriadis (2018), required as **input** the following: mean (*Sm*), variance (*Sv*), skewness and kurtosis coefficients (*Ss* and *Sk*), Hurst parameter of the GHK model (*H*), scale parameter (*q*), length of synthetic series (*N*).

Reproducing observed clustering using HK dynamics and Monte Carlo simulations

Generalized-HK (GHK) model

EGU General Assembly 2024

The *Climacogram* (Figure 9) was formulated and the **SMA-GHK modelling** simulations were developed regarding the USGS 07071500 gauge located at Eleven Point river near Bardley, State of Missouri, USA, with parameters H **= 0.81** and q **= 1.00 days**.

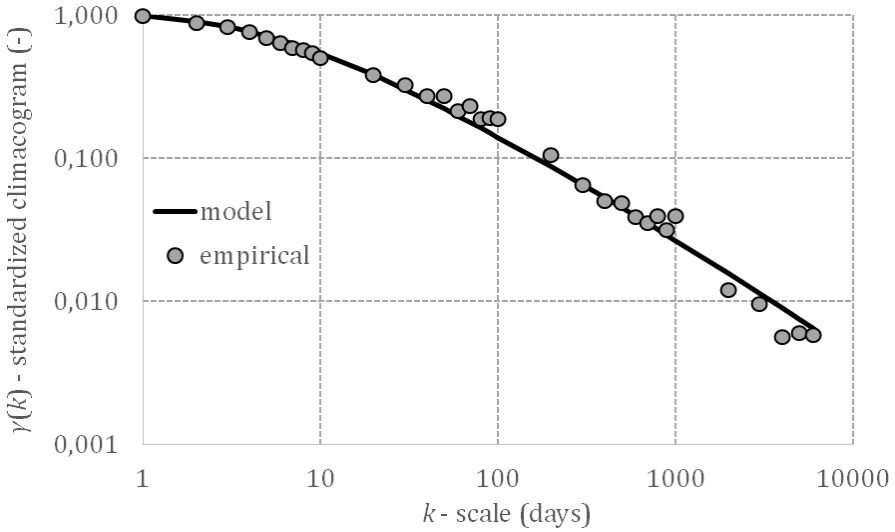


Fig. 9 The Climacogram of the gauge location with ID: 07071500 (H = 0.81, q = 1.00 days).

Reproducing observed clustering using HK dynamics and Monte Carlo simulations

SMA-GHK model

EGU General Assembly 2024

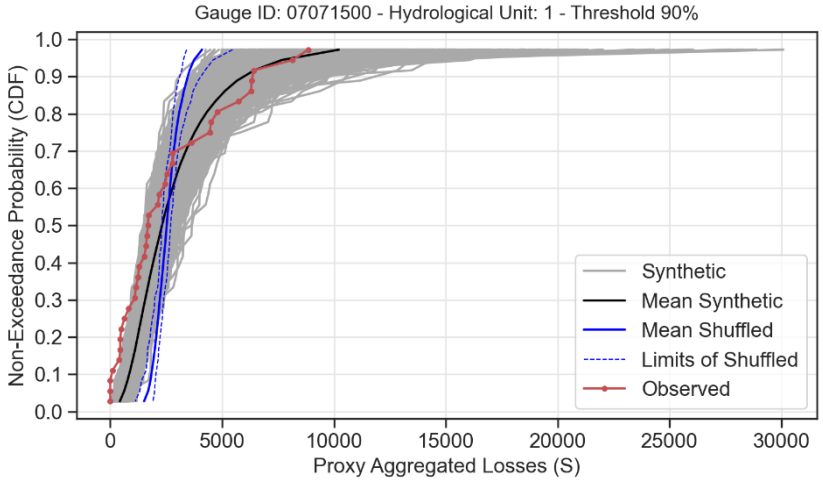
Subsequently, **1000 synthetic time series** through **Monte Carlo** simulations of the USGS 07071500 gauge were developed; the diagrams of the Non-Exceedance Probability (CDF) of Proxy Aggregated Losses for the four thresholds are extracted (Figure 10).

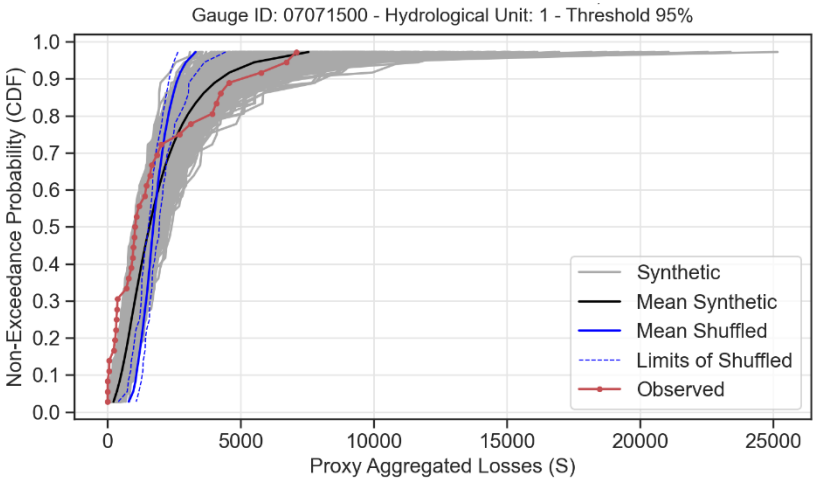
The **Non-Exceedance Probability (CDF)** curve of the observed Proxy Aggregated Losses (Figure 10) is **contained** in the Monte Carlo **prediction limits** by the GHK model, preserving the HK dynamics and the 4 four moments. In contrary, **shuffled** (randomized) curves have a **different** behavior, especially in the **tails** of the distribution.

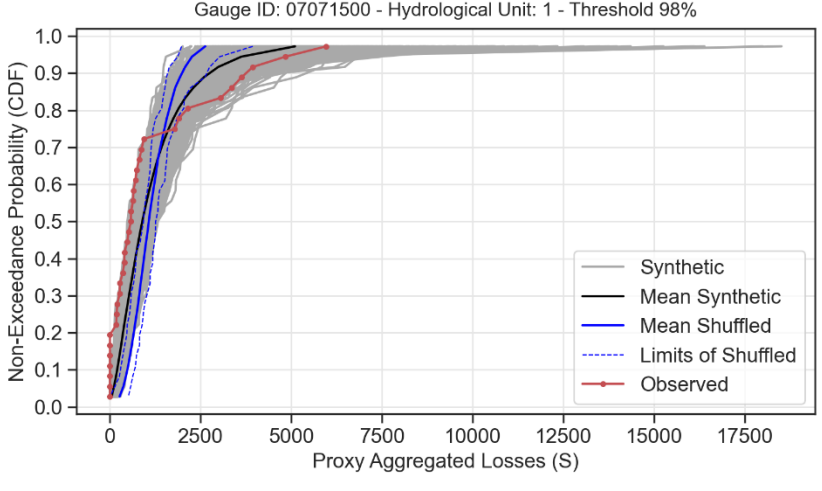
Reproducing observed clustering using HK dynamics and Monte Carlo simulations

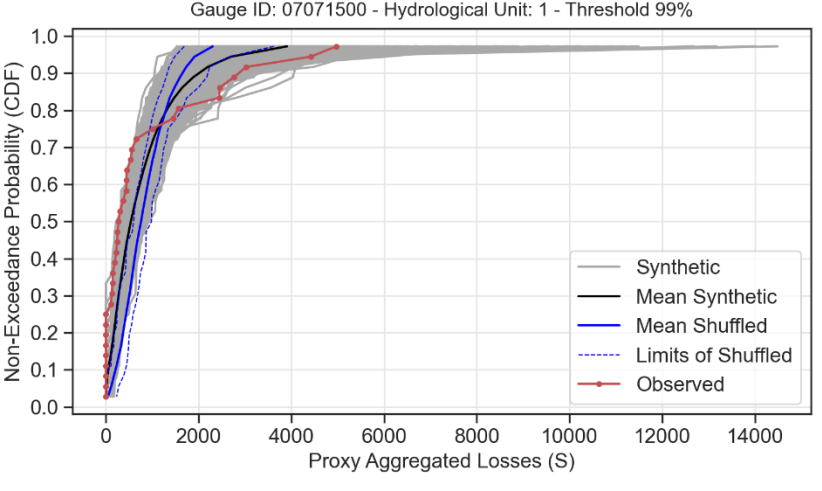
Fig. 10 Proxy Aggregated Losses's Non-Exceedance Probability (CDF) diagrams of observed, shuffled and synthetic time series (SMA-GHK model, Gauge location ID: 07071500).

EGU General Assembly 2024









Is the streamflow-based Proxy Aggregated Losses informative for the dynamics of collective risk deriving from actual flood claims records?

Is the streamflow-based Proxy Aggregated Losses informative for the dynamics of collective risk deriving from actual flood claims records?

EGU General Assembly 2024

The **annual Proxy Aggregated Losses** of the 360 selected gauge locations of the **US-CAMELS dataset** has already been computed, considering the 4 selected thresholds (90%, 95%, 98% and 99%) for the years 1980-2014.

Moreover, the published **FEMA claims records** offers us the opportunity to **investigate the validity** of the developed method on a spatial basis by assessing the **correlation** between these **claims records** with the **results of our study** on streamflow POT events.

The FEMA claims records were **distributed spatially** on the 21 Hydrological Units and the 50 States. In this respect, the **Spearman** correlation coefficient is evaluated between the **annual Proxy Aggregated Losses** for each one of the gauge locations and the **aggregated claims records** of the Hydrological Unit and the State that a specific gauge location belongs to.

Is the streamflow-based Proxy Aggregated Losses informative for the dynamics of collective risk deriving from actual flood claims records?

EGU General Assembly 2024

The **cumulative distribution functions** (Figure 11) and the **boxplots** (Figure 12) of the aforementioned Spearman correlation coefficients for the **360 gauges** and for all the selected thresholds follow, showing that, in general, considering the aggregated claims of **States** tends to **underestimate** the correlation coefficient in contrast to the aggregated claims of the **Hydrological Units**.

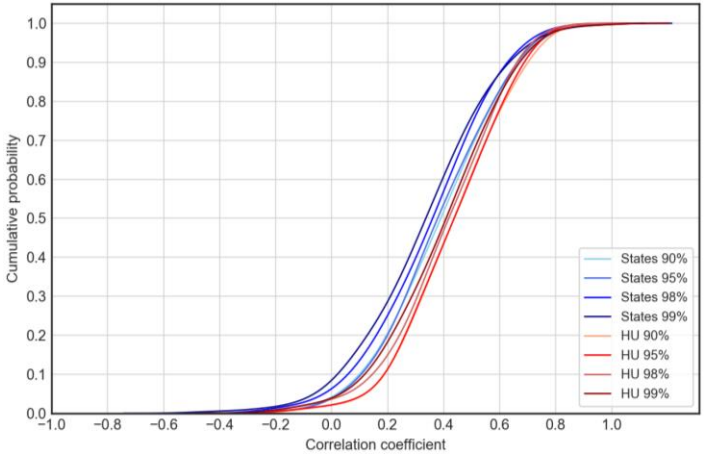


Fig. 11 Distribution functions of the Spearman correlation coefficient between the annual Proxy Aggregated Losses of the 360 gauges and the States/Hydrological Units claims records that a specific gauge location belongs to.

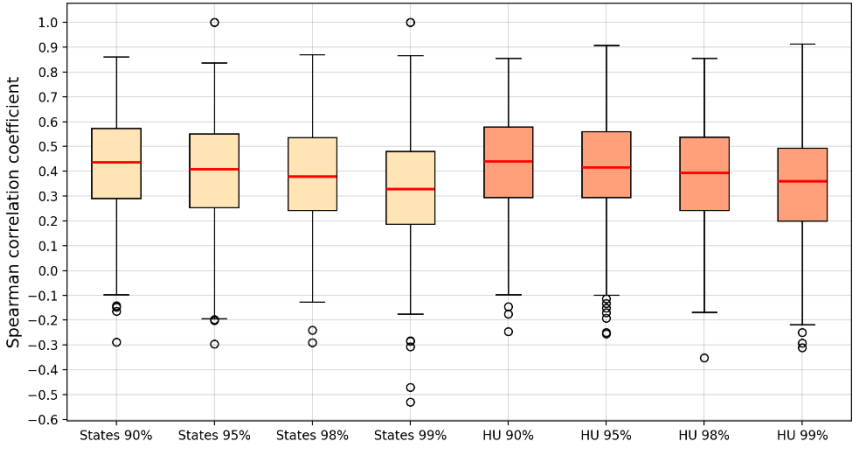


Fig. 12 Boxplot of the Spearman correlation coefficient between the annual Proxy Aggregated Losses of the 360 gauges and the States/Hydrological Units claims records that a specific gauge location belongs to.

Is the streamflow-based Proxy Aggregated Losses informative for the dynamics of collective risk deriving from actual flood claims records?

EGU General Assembly 2024

The **annual Proxy Aggregated Losses** of the 360 selected gauge locations of the **US-CAMELS dataset** has already been computed, considering the 4 selected thresholds (90%, 95%, 98% and 99%) for the years 1980-2014.

Moreover, the published **FEMA claims records** offers us the opportunity to **investigate the validity** of the developed method on a spatial basis by assessing the **correlation** between these **claims records** with the **results of our study** on streamflow POT events.

The FEMA claims records were **distributed spatially** on the 21 Hydrological Units and the 50 States. In this respect, the **Spearman** correlation coefficient is evaluated between the **annual Proxy Aggregated Losses** for each one of the gauge locations and the **aggregated claims records** of the Hydrological Unit and the State that a specific gauge location belongs to.

Is the streamflow-based Proxy Aggregated Losses informative for the dynamics of collective risk deriving from actual flood claims records?

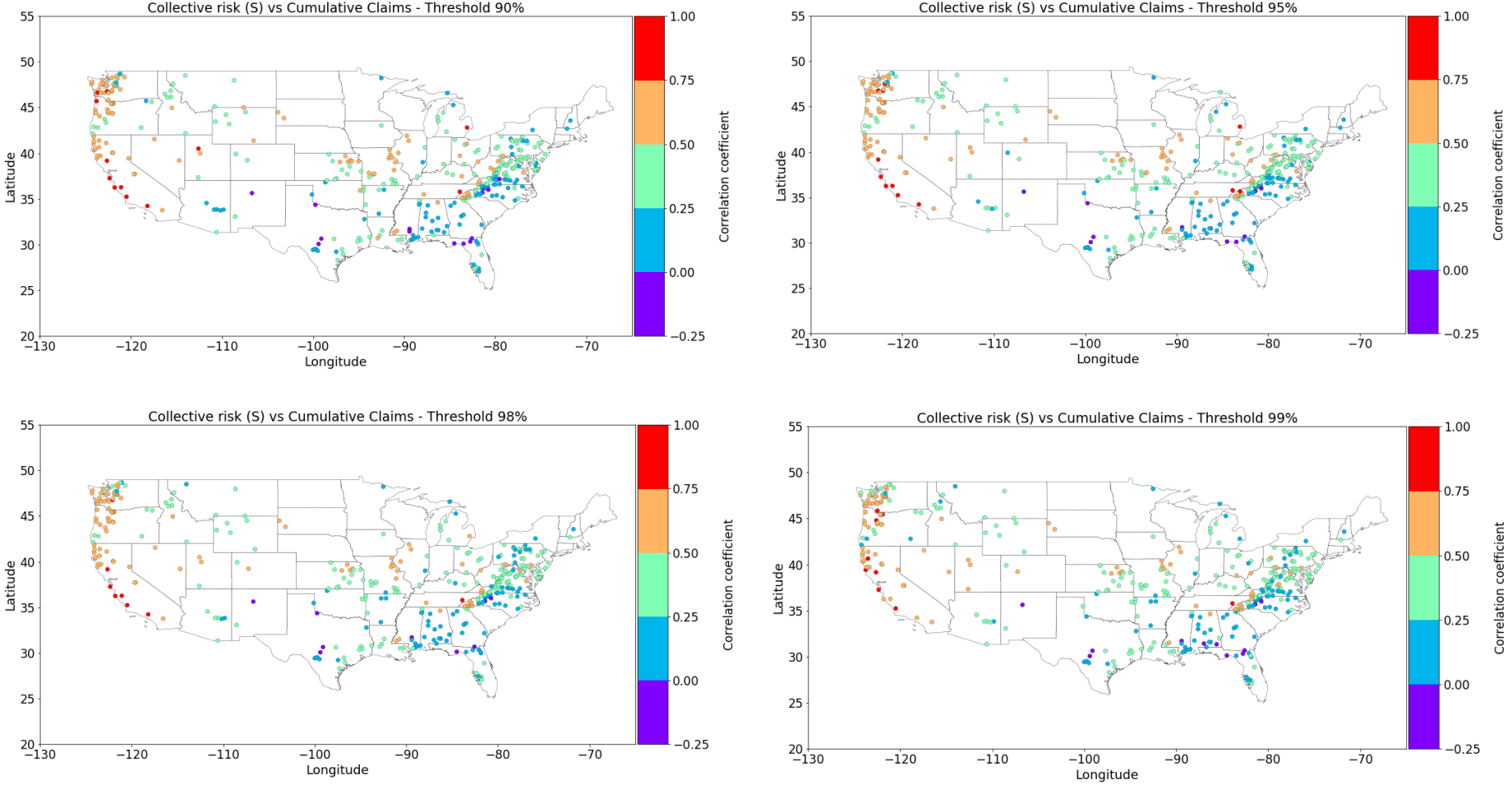
EGU General Assembly 2024

Subsequently, the **USA maps** that show the **Spearman correlation** coefficient between the Proxy Aggregated Losses and the Hydrological Units' claims records for all the gauge locations and the selected thresholds follow (Figure 13), highlighting the **spatial distribution** of the **correlation coefficient** and indicating the **areas** the latter is higher or lower.

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

Fig. 13 Spearman correlation coefficient for each one of the selected 360 gauges between annual Proxy Aggregated Losses and the claims records of the Hydrological Units that a specific gauge location belongs to for four different thresholds, 1st (top) row 90%, 2nd 95%, 3rd 98% and 4th (bottom) row 99%.

EGU General Assembly 2024



Is the streamflow-based Proxy Aggregated Losses informative for the dynamics of collective risk deriving from actual flood claims records?

EGU General Assembly 2024

Box plot of Spearman Correlation coefficient between Proxy Aggregated Losses and the Hydrological Unit's aggregated claims

Data from Paint Rock River (ID: 03574500) show that the **dominant clustering mechanisms** introduce **significant correlation**.

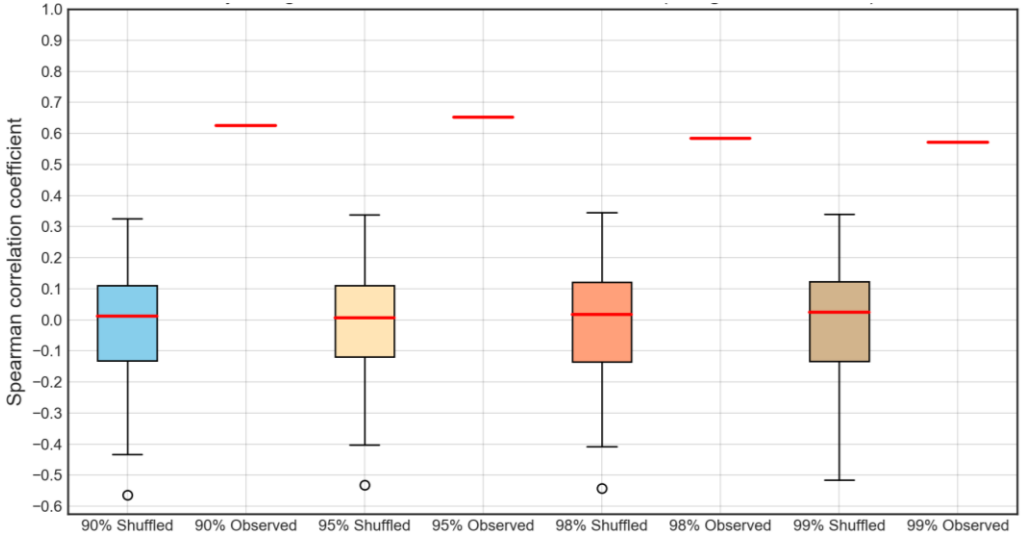


Fig. 14 Box plot of Spearman correlation coefficient between Proxy Aggregated Losses and the Hydrological Unit's aggregated claims that the gauge location (ID: 03574500) belongs to, for all thresholds.

Conclusions

Conclusions on clustering mechanisms

EGU General Assembly 2024

Regarding the impacts of clustering mechanisms on streamflow extremes, the **probabilistic properties** of a streamflow-based proxy for aggregated losses, return periods, and the duration of the over-threshold events from the **US-CAMELS dataset**, were investigated for **four different thresholds**.

Results show that for the **clustering indices**, the divergence between the properties of the observed and the shuffled (randomized, considered as independent) time series is **pronounced** in many gauges.

The latter suggests a deviation from the **independence** assumption and a clear indication for the existence of **clustering in streamflow extremes** which is further quantified through a **stochastic investigation** based on the Hurst-Kolmogorov dynamics.

Conclusions on HK dynamics

EGU General Assembly 2024

Based on the mean ***Climacogram*** and the **GHK process** regarding the 360 empirical streamflow time series of the US-CAMELS dataset, the Hurst parameter was estimated 0.63, which indicates a **persistent** behavior.

Empirical findings regarding the properties of observed streamflow timeseries were also reproduced through Monte Carlo simulations based on the **GHK** and **SMA-GHK model**, preserving the **HK dynamics** and the four moments.

The **Monte Carlo prediction limits** captured the **observed patterns**, whereas in contrary, shuffled (randomized) curves showed a different behavior, especially in the **tails** of the distribution.

Conclusions on

the association with the FEMA's NFIP actual claims records

EGU General Assembly 2024

The association between the streamflow-based **Proxy Aggregated Losses** used herein and the **FEMA's NFIP actual claims records** is **validated** by computing the Spearman correlation coefficient between the two.

A clear **spatial pattern** emerges from this investigation, showing that higher values of the correlation emerge in West Coast, in contrast to the ones in East Coast, which are significantly lower.

As the Proxy Aggregated Losses refer to **fluvial (river) flooding**, these results suggest that this type of flooding is **dominant** in West Coast. In contrast, it is revealed that flooding events that provoke insurance claims in the East Coast exhibit a **different and more complicated pattern**.

Furthermore, the association of the **streamflow-based** Proxy Aggregated Losses to **actual** number of claims records of the Hydrological Unit that the gauge location belongs to, was further **validated** by comparing results from the **observed** to the **shuffled** (independent) time series, which showed no significant correlation.

General Conclusion

Overall, the apparent existence of clustering mechanisms in **streamflow extremes** is shown to be associated to clustering in related **insurance claims** in the USA, yet with spatially variable patterns reflecting different flood generating mechanisms.

Disregarding such **clustering dynamics** may lead to inaccurate **risk assessment** processes and significant **financial** impacts for the insurance and reinsurance sectors, in case of unpredictably large values of aggregate claim amounts **stressing their reserves**.

References

Dimitriadis, P., and Koutsoyiannis, D. (2015). Climacogram versus autocovariance and power spectrum in stochastic modelling for Markovian and Hurst–Kolmogorov processes. Stochastic Environmental Research & Risk Assessment, 29 (6), 1649–1669.

Dimitriadis, P., and Koutsoyiannis, D. (2018). Stochastic synthesis approximating any process dependence and distribution. Stochastic Environmental Research & Risk Assessment, 32 (6), 1493-1515.

Dimitriadis, P., Koutsoyiannis, D., Iliopoulou, T., and Papanicolaou, P. (2021). A Global-Scale Investigation of Stochastic Similarities in Marginal Distribution and Dependence Structure of Key Hydrological-Cycle Processes. Hydrology, 8(2), 59.

Ezer, T., and Atkinson, L.P. (2014). Accelerated flooding along the U.S. East Coast: On the impact of sea-level rise, tides, storms, the Gulf Stream, and the North Atlantic Oscillations. Earth's Future, 2 (8), 362-382.

FEMA (2019). FEMA publishes NFIP claims and policy data. [Online] Available at: https://www.fema.gov/openfema-data-page/fima-nfip-redacted-claims-v2. Last Data Refresh: 12-12-2023 [Accessed: December 2023]

Kaas, R., Goovaerts, M., Dhaene, J., and Denuit, M. (2008). Modern Actuarial Risk Theory Using R. Springer.

Koutsoyiannis, D. (2010). A random walk on water. Hydrology and Earth System Sciences, 14, 585-601.

Koutsoyiannis, D., and P. Dimitriadis, (2021). Towards generic simulation for demanding stochastic processes. Sci, 3, 34, doi:10.3390/sci3030034.

Newman, A., Sampson, K., Clark, M. P., Bock, A., Viger, R. J., and Blodgett, D. (2014). A large-sample watershed-scale hydrometeorological dataset for the contiguous USA. Boulder, CO: UCAR/NCAR.

Reiss, R.D., and Thomas, M. (2007). Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields. Birkhäuser.

Robinson, P.J., and Botzen W.J.W. (2020). Flood insurance demand and probability weighting: The influences of regret, worry, locus of control and the threshold of concern heuristic. Water Resources and Economics, 30, 100144.

Serinaldi, F. and Kilsby, C.G. (2016). Understanding Persistence to Avoid Underestimation of Collective Flood Risk. Water, 8 (4), 152