Spatiotemporal clustering of flash floods in a changing climate (China, 1950-2015)

Nan Wang[1,2], Luigi Lombardo[3], Marj Tonini[4], Weiming Cheng[1,2,5,6,*], Liang Guo[7,8], Junnan Xiong[1,9]

Abstract

The persistence over space and time of flash flood disasters – flash floods that have caused either economical or life losses, or both – is a diagnostic measure of areas subjected to hydrological risk. The concept of persistence can be assessed via clustering analyses, performed here to analyze the national inventory of flash flood disasters in China that occurred in the period 1950-2015. Specifically, we investigated the spatiotemporal pattern distribution of the flash flood disasters and their clustering behavior by using both global and local methods: the first, based on the Ripley's K-function, and the second on Scan Statistics. As a result, we could visualize patterns of aggregated events, estimate the cluster duration, and make assumptions about their evolution over time, also with respect to the precipitation trend. Due to the large spatial (the whole Chinese territory) and temporal (66 years) scale of the dataset, we were able to capture whether certain clusters gather in specific locations and times, but also whether their magnitude tends to increase or decrease. Overall, the eastern regions in China are much more subjected to flash flood disasters compared to the rest of the country. Detected clusters revealed that these phenomena predominantly occur between July and October, a period coinciding with the wet season in China. The number of detected clusters increases with time, but the associated duration drastically decreases in the recent period. This may indicate a change towards triggering mechanisms which are typical of short-duration extreme rainfall events. Finally, being flash flood disasters directly linked to precipitation and their extreme realization, we indirectly assessed whether the magnitude of the trigger itself has also varied through space and time, enabling considerations in the context of climatic changes.

[1]State Key Laboratory of Resources and Environmental Information Systems, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, China

[2]University of Chinese Academy of Sciences, Beijing, 100049, China

[3]University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC), PO Box 217, Enschede, AE 7500, Netherlands

[4]Institute of Earth Surface Dynamics, Faculty of Geosciences and Environment, University of Lausanne, CH-1015 Lausanne, Switzerland

[5]Jiangsu Center for Collaborative Innovation in Geographic Information Resource Development and Application, Nanjing, 210023, China

[6]Collaborative Innovation Center of South China Sea Studies, Nanjing, 210093, China

[7]Research Center on Flood and Drought Disaster Reduction of the MWR, Beijing, 100038, China

[8]State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100038, China

[9]School of Civil Engineering and Architecture, Southwest Petroleum University, Chengdu, 610500, China

Keywords: Spatiotemporal clustering; Flash flood disasters; Extreme rainfall; China

1 Introduction

Flash floods are among the most destructive surface processes around the world, especially in mountainous areas (Au, 1998; Borga et al., 2011; Gomez and Kavzoglu, 2005; Jonkman, 2005). They are mainly initiated by rapid and intense rainfall, often discharged in few hours (e.g., Borga et al., 2007; Bout et al., 2018; He et al., 2018; Lóczy et al., 2012), and by complex interactions of the climatic conditions with topography and hydrology (e.g., Hatheway et al., 2005). Because of the very rapid raise in water levels caused by flash floods, it is challenging to take timely and effective actions to contain the associated damage. Flash flood disasters are essentially flash floods that have caused losses either in terms of human lives or economy, or both (Gaume et al., 2009; Jonkman and Kelman, 2005; Kelman and Spence, 2004). In China, approximately 70% of the total area is covered by mountains and hills, which exposes a substantial surface of the national territory to flash flood disasters' risk (Liu et al., 2018). Additionally, the more frequent extreme precipitation associated with climate change has increased the number of flash flood disasters in recent decades (Sampson et al., 2015).

The susceptibility to hydro-geomorphological processes is commonly assessed by considering only the spatial distribution of observed events (Cama et al., 2015, 2017; Santangelo et al., 2012; Zaharia et al., 2017). However, this is purely a convenient assumption from the modeling perspective. Recently, a growing amount of evidence indicates that these events tend to aggregate in space conditioned by the temporal variability, attesting for an interaction between space and time on event frequency and distribution (Gariano and Guzzetti, 2016; Kouli et al., 2010; Zhang and Cong, 2014; Fuchs et al., 2015; Merz et al., 2016; Tonini and Cama, 2019). In other words, when an event occurs at a specific location, a temporary increase in the probability that other events will cluster at nearby locations should be accounted for. This increase in probability can be captured through clustering analyses and various examples already exist in literature where this has been done at different spatial and temporal scales and via different analytical approaches. Notably, this type of application spans in many areas of natural hazards and have become mainstream in case of seismicity (e.g., Fischer and Horálek, 2003; Georgoulas et al., 2013; Varga et al., 2012; Woodward et al., 2018; Yang et al., 2019), joint sets and their orientation in rock outcrops (e.g., Tokhmechi et al., 2011; Zhan et al., 2017), groundwater monitoring (Chambers et al., 2015), wildfires (e.g., Orozco et al., 2012; Costafreda-Aumedes et al., 2016; Fuentes-Santos et al., 2013; Tonini et al., 2017), and landslides (e.g., Lombardo et al., 2018, 2019; Tonini and Cama, 2019). In the specific case of flooding, Zhao et al. (2014) used the projection pursuit theory to cluster spatial data and to build a dynamic risk assessment model for flood disasters. Moreover, Renard (2017) detected flood vulnerability accounting for clustering effects in key areas with high flood risk. Pappadà et al. (2018) also investigated the flood risks in a given region and identified clusters where the floods show a similar behavior with respect to multivariate criteria. Gu et al. (2016a,b) indicated the floods in Tarim River basin showed evident inter-annual clustering pattern. Another example can be found in Merz et al. (2016) where the authors analyzed the inter-annual and intra-annual flood clustering in Germany. All these examples confirm a substantial scientific interest in recent years dedicated to investigate the clustering behaviors of flash floods and the associated risk; and, more generally, to concurrently analyze their spatial and temporal persistence. However, despite the scientific efforts, detecting flash flood patterns at long temporal scale is still scarce in literature, mainly because of technical limitations. In fact, limited information and records are available in digital form reporting locations and dates of flash floods (and flash flood disasters), especially over long periods. Nevertheless, very recent advances in data collection and sharing techniques are gradually filling this gap, and an increasing number of databases are being published and made available to the scientific community with the records of historical and hydro-geomorphological disasters at the global, continental, or regional scale over long periods (Gourley et al., 2013; Haigh et al., 2017; Vennari et al., 2016; Liu et al., 2018; Archer et al., 2019; de Bruijn et al., 2019; Nowicki Jessee et al., 2020; Wood et al., 2020). Among these, Chinese historical inventories of flash flood disasters are a precious source of information allowing to investigate their spatiotemporal pattern distribution and evolution. Furthermore, this information can be related with the geomorphological settings of the area and the climatic/meteorological conditions to detect triggering factors, highlight the more vulnerable areas, and to prevent and forecast their effects in the future.

Typically, flash flood disasters (as many other hydro-geomorphological disasters) can be considered as a stochastic point processes (Stoyan, 2006) acting in both spatial and temporal dimensions (e.g., Lombardo et al., 2020). Point patterns can be analyzed in terms of their random distribution, dispersion and clustering behaviour (Merz et al., 2016; Tonini and Cama, 2019). Several methods can be implemented to deal with stochastic properties. Some classic models, such as Moran's I (Moran, 1950), Ripley's K-function (Ripley, 1977), fractal dimension (Lovejoy et al., 1986), and Allan factor (Allan, 1966), have been used to detect clustering behaviour in space and in time. Representative models for local clustering analysis (i.e. allowing to detect clusters and their specific location) include Geographical Analysis Machine (GAM, Openshaw et al., 1987), Turnbull's Cluster Evaluation Permutation Procedure (CEPP, Turnbull et al., 1990), Scan Statistics (Kulldorff, 1997), and DBSCAN (Ester et al., 1996). For flash floods, which are triggered by storms, the temporal dependency among persistent events is mainly driven by climatic and meteorological conditions. However, global cluster indicators only take into consideration one dimension, disregarding the interaction between space and time. In this sense, spatiotemporal Scan Statistics is a good tool to detect clusters since it allows to identify statistically significant excess of observations thanks to a moving cylindrical window that scans all locations both in space and time (Kulldorff et al., 1998).

Therefore, it is especially useful to investigate large spatiotemporal inventories of hydroand geo-morphological processes, such as flash floods. Indeed, the detection of clusters originated by events closer both in space and in time can be more informative that the simply investigation of their purely temporal and purely spatial pattern distribution. For example, understanding the duration of the spatiotemporal clusters of flash floods is key tool to investigate their dynamic and to highlight more vulnerable area and frame period.

In light of this, the main objective of the present research is to explore the pattern distribution of flash flood disasters which have caused life and/or economic losses in China over a 66-years period (daily data from 1950 to 2015). Firstly, the Ripley's K-function was applied to explore the deviation of flash flood disasters from a random process. Results allow to assess at which spatial and temporal scales events are clustered. Then, a local cluster indicator, namely Scan Statistics, was implemented to map statistically significant spatiotemporal clusters. To the best of our knowledge, this study represents the first attempt of investigating the spatiotemporal cluster behaviour of flash flood disasters affecting a huge area, such as the entire Chinese territory. Moreover, the volume of the data that we analyzed represents an additional challenge allowing to provide useful insights on flood dynamics over a large spatiotemporal domain and enabling considerations in the context of climatic changes. To this end, we finally compared the dynamic of the clusters, detected from the early to the recent period, with the extreme rainfall evolution, computed each 10-years, which is assumed as a local climatic proxy factors.

2 Material and methods

2.1 Data description

2.1.1 Study area

China lies between latitudes 18° and 54° N, and longitudes 73° and 135° E. With an area of about 9.6 million square kilometers, it is the world's third-largest country. The landscape varies significantly across this vast area, ranging from the Gobi and Taklamakan deserts in the north to the subtropical forests in the wetter south. The eastern plains and southern coasts are the location of most of China's agricultural land and settlements. The southern areas consist of hilly and mountainous terrain. The west and north of the country are dominated by sunken basins (such as the Gobi and the Taklamakan desert), towering massifs and rolling plateaus, including part of the highest tableland on earth, the Tibetan Plateau. Based on its topography, China can be divided into six homogeneous geomorphological macro-regions (Wang et al., 2020): eastern plain, southeastern hills, southwestern mountains, north-central plateaus, northwestern basins and Tibetan Plateau. Mountains (33% of the territory), plateaus (26%) and hills (10%) account together for nearly 70% of the entire surface.

In recent years, the precipitation intensity shows an increasing trends over China (Zhang and Cong, 2014). Influenced by the East Asian summer monsoon and the geomorphologic settings, the climatic condition across the whole country varies considerably (Wu et al., 2019). In general, the wet season in China lasts from May to September (Song et al., 2011b). In the Eastern area, the annual rainfall decreases from south to north with an average annual precipitation that ranges from 250 to 750 mm (Zhang et al., 2007). In the west and central part of North China, due to its far distance away from ocean, the climate tends to be more arid and the landscape transitions to large deserts. The Tibetan plateau is characterized by wet and humid summers with cool and dry winters. More than 60–90% of the annual total precipitation falls between June and September (Xu et al., 2008).

2.1.2 Flash flood disaster inventory

The dataset used in this study has been collated and made accessible for the present research as part of a national effort carried out by the Chinese Institute of Water Resources and Hydropower Research (Liu et al., 2018). It reports flash flood occurrences in China since 1950 until 2015 together with available information, namely longitude and latitude, date, fatalities and economic losses. Due to the lack of specific terminology and/or detailed descriptions of the disaster process in the database, the data does not differentiate the initial mechanism, be it water floods or debris floods/flows (e.g., Fernández and Lutz, 2010; Gartner et al., 2014). The only common information is that for each specific case, a large amount of overland flows, mixed with an unspecified solid fraction, rapidly flooded a given area with disastrous effects (e.g., Pierson et al., 1987; Chang et al., 2011).

To better understand the spatiotemporal dynamics of flash floods and associated disasters, as well as the relationship with the triggering factors, the date of occurrence is of vital importance. Therefore, for consistency reasons, we considered only the records whose metadata contained a full temporal description (year-month-day) resulting in a subset of 32,473 flash flood disasters (accounting for 68% of the entire dataset) precisely located in space and time (Figure 1).

2.2 Methodological overview

2.2.1 Spatiotemporal K-function

The Ripley's K-function ($K_{(s)}$) is largely applied in environmental studies to analyse the pattern distribution of spatial point processes and to detect deviation from spatial randomness. $K_{(s)}$ allows to determine if a set of mapped punctual events show a random, dispersed or cluster distribution pattern over increasing distance values (Ripley, 1977). It is computed as the ratio between the expected number of events falling at a distance r from an arbitrary event and the average number of points per unit area, corresponding to the intensity of the spatial point process (λ). In the same way, it is possible to define the temporal K-function ($K_{(t)}$) allowing to asses for the randomness of events in time. The spatiotemporal K-function ($K_{(s,t)}$) is a generalization of the univariate Ripley's K-function which allows to test for the independence between two variables, space (s) and time (t). Therefore, the $K_{(s,t)}$ is a suitable

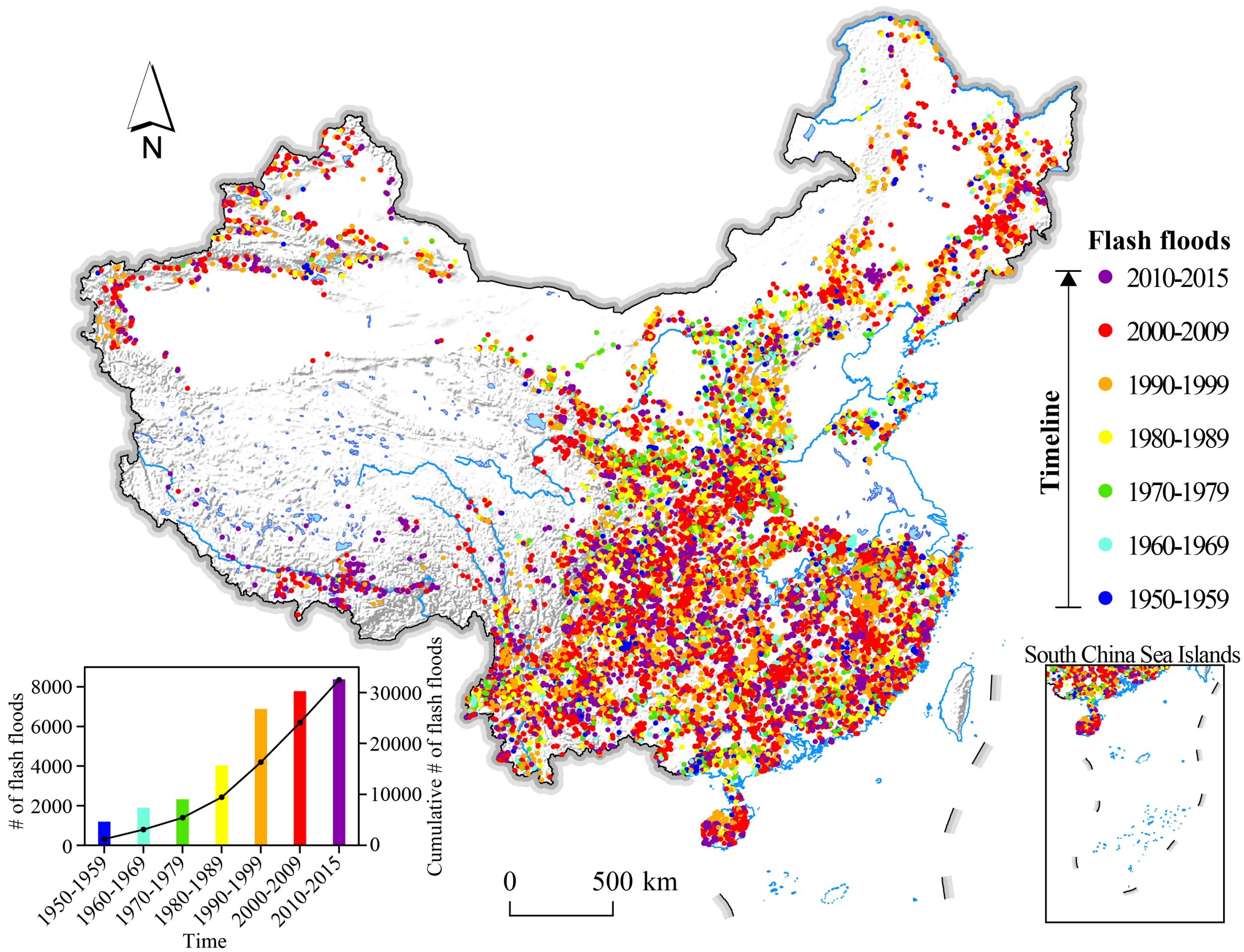

Figure 1: Distribution of flash flood disasters and background setting of China. Dashed lines correspond to the geo-political boundary of China.

tool to investigate the clustering behaviour of a set of events occurred in a given area at a given time. For a point process X with intensity λ, according to equation 1, it is defined as the number of expected further events (E) occurring within a distance r and time t from an arbitrary event u, where a define the contouring circle.

$$K_{(s,t)} = 1/\lambda \times E[n(X \cap a(u,r,t)u)|u \in X] \tag{1}$$

To illustrate the interaction between space and time, it can be useful to evaluate the value $D_{(s,t)}$, defining the difference between the spatiotemporal K-function and the product of the purely spatial and the purely temporal K-function (see equation 2).

$$D_{(s,t)} = K_{(s,t)} - K_{(s)} \times K_{(t)} \tag{2}$$

If space and time are independent variables, this value equals to zero. Otherwise, positive values of $D_{(s,t)}$ indicates the interaction among events in space and in time. In other words, events closer in space are more likely to occur in a closer time. On the contrary, the negative values means a dispersed pattern.

In this study, spatiotemporal K-function analyses were performed with the package "Spatial and Space-Time Point Pattern Analysis" (splancs, Rowlingson and Diggle, 2017) in R (R Core Team, 2019).

2.2.2 Spatiotemporal scan statistics

Scan statistic was originally developed by Naus (1965a,b) to detect cluster in a one-dimensional point process. Subsequently Kulldorff (1997) extended this approach to multi-dimensional point process, introducing the use of scanning windows. The procedure was implemented into a free software, SaTScan$^{\text{TM}}$ (satscan.org) which can handle a purely spatial, purely temporal or spatiotemporal datasets and includes different probability models depending on the nature of the data and the scope of the research (e.g. for prospective or retrospective cluster detection). In the purely spatial case, the aim of scan statistics is the early detection of clusters, allowing to map them and to assess their statistical significance. Moving windows scan the region increasing their radius up to a fixed limit (R_{max}) and count the number of events falling inside and outside the area. The probability that a window contains more observations than expected is assessed via the likelihood ratio, by comparing with the background population. Then, the null hypothesis of randomness is tested via Monte Carlo simulations, based on repeated random sampling. The spatiotemporal scan statistic use cylinders instead of circular windows, where the height of the cylinder account for the temporal dimension.

In order to deal with flash floods, the retrospective spatiotemporal permutation scan statistics (STPSS, Kulldorff et al., 2005) seems to be the most adequate model. Indeed, for environmental processes, the definition of the background population at risk needed for the statistical significance assessment of the detected clusters is quite problematic. STPSS assesses the expected number of cases using only the observed cases by permutation, supposing that each event has the same probability for all the times. Computationally, if C is the total number of observed cases and c_{zd} the number of cases observed in a specific zone z and a day d, the expected number of cases per zone and day (μ_{zd}) is equal to:

$$\mu_{zd} = \frac{1}{C}\left(\sum_{z} c_{zd}\right)\left(\sum_{d} c_{zd}\right) \tag{3}$$

It follows that, for a spatiotemporal cylinder A, the expected number of cases μ_A can be estimated as the sum of each μ_{zd} inside the cylinder A:

$$\mu_A = \sum_{z\,,d \in A} \mu_{zd} \tag{4}$$

If C_A is the number of observed cases in A, considered as Poisson-distributed with mean μ_A, the Poisson generalized likelihood ratio (GLR) can be computed as:

$$GLR = \left(\frac{c_A}{\mu_A}\right)^{c_A}\left(\frac{C - c_A}{C - \mu_A}\right)^{C - c_A} \tag{5}$$

This ratio is calculated and maximized for every possible scanning cylinder. The cylinder with the highest GLR-value is the most likely cluster, that is, the cluster least likely to be due to chance, while the following are secondary clusters. Then, Monte Carlo simulations are performed and the statistical significance (p-value) of the detected clusters can be assigned by comparing the rank (R) of GLR from the real data set with the GLR from the simulated one. Thus, the p-value can be estimated by dividing R by the number, plus one, of performed simulations.

3 Results

3.1 Deviation from a random process

In the present study, the spatiotemporal K-function was used to assess the global cluster behavior of flash flood disasters generated by the interaction between these two variables. To this end, the perspective 3D-plot of D(s,t) represents a useful visual tool allowing to estimate the distribution pattern of events along the spatial and the temporal dimensions. In more details, positive values attest for a cluster distribution, while values close to zero indicate a random pattern, with no interaction between space and time. In our case, the 3D-plot (Figure 2) shows that at any distance, from hundred to thousands meters, and from few years to decades, flash flood disasters display a cluster behaviour, which is more pronounced at increasing distance-values. In addition, the spatiotemporal K-function was computed considering individually the southeastern and the northwestern area in China, given that the first corresponds to the rainiest zone, highly affected by flash floods, while the second is predominantly desert. It results that (Figure 3) in the southeastern China (panel a) clusters arise at a shorter spatial distance and closer in time than in the northwestern China (panel b). As regards the temporal dimension, the two areas show a similar cluster behaviour, with a strong attraction among events up to 10-years, and than lasting in time with a more relaxed clustering behaviour.

To summarize, the spatiotemporal K-function reveals a deviation of flash flood disasters and associated spatiotemporal pattern distribution from a random process at specific scales, measured and quantified both in space, as distances-values, and in time, as yearly periods. These values can provide a useful indication to set up the parameters for further clustering algorithms, acting at local scale such as, for example, the spatiotemporal scan statistics.

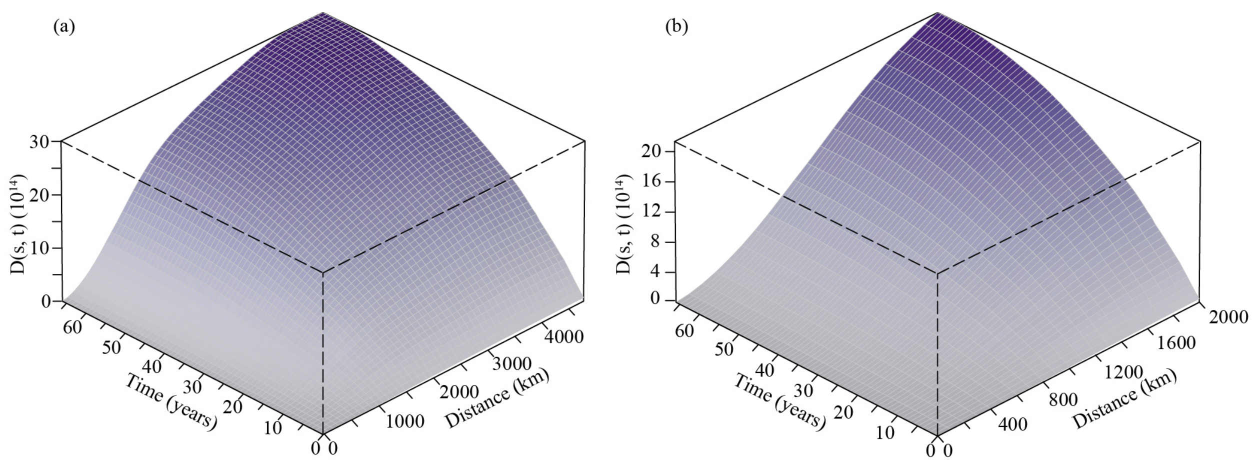

Figure 2: Perspective 3D-plot of flash flood disasters in China during 1950-2015 (panel a) with a zoom up to 2000 km (panel b).

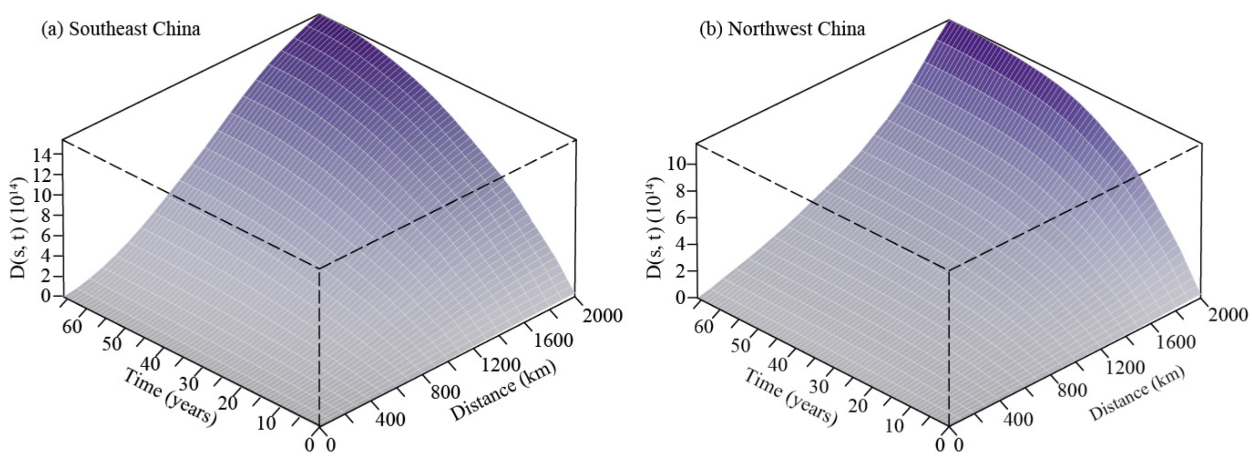

Figure 3: Perspective 3D-plot of flash flood disasters in the southeast (panel a) and northwest China (panel b).

3.2 Spatiotemporal cluster detection

3.2.1 Cluster parametrization and their spatial distribution

Scan statistics was performed to detect spatiotemporal clusters of flash flood disasters. The size and the duration of the detected clusters are influenced by the input parameters of the scanning windows, namely the maximum radius (R_{max}), the maximum temporal duration (T_{max}), and the time aggregation (T_{agg}). Indeed, values of R_{max} exceeding the 50% of the total area or, for T_{max}, the 50% of the entire study period, can result in an exceptionally low rate outside the scanning window rather than detecting an exceptionally high rate inside. T_{agg} is used to adjust the aggregation of the data over time and allows adjusting for cyclic temporal trends: for example, a time aggregation of one year automatically adjusts for the seasonal variability, while the contrary happen with monthly aggregations. Moreover, both spatial and temporal aggregations can highly reduce the computer processing time. Different values for R_{max} were tested for the southeast and northwest areas in China, as suggested by looking at the respective perspective 3D-plot. Nevertheless, the performed analyses indicated that the effect onto the detected clusters were negligible and that finally the distribution of spatiotemporal clusters of flash flood disasters in the country can be analysed as a whole. A set of possible combinations of R_{max} and T_{max} was tested, while T_{agg} was initially fixed to one year.

More specifically, since results of the K-function revealed that flash floods disasters in the study area and investigated period are globally clustered even at short distances, we chose R_{max} of 100, 200 and 300km, and T_{max} equals to 1, 3 and 5 years. The choice for R_{max} is corroborated by Zhang et al. (2010) who report measurements constantly less than 500 km for the radius of typical convective storms in the Chinese mainland, which can trigger flash floods. Results of STPSS for each of the nine combinations of these parameters are shown in Figure 4.

Table 1: Number of detected spatiotemporal clusters of flash flood disasters in China during 1950-2015 using different parameters, as indicated.

R_{max} (km)	T_{max} (year)		
	1	3	5
100	131	128	130
200	85	77	75
300	58	54	53

As shown in Table 1, the largest variation in the number of detected clusters is mainly associated with R_{max} rather than with T_{max}; as R_{max} increases, the number of detected clusters decrease. Indeed, large R_{max} values affect the detection of clusters acting at a fine

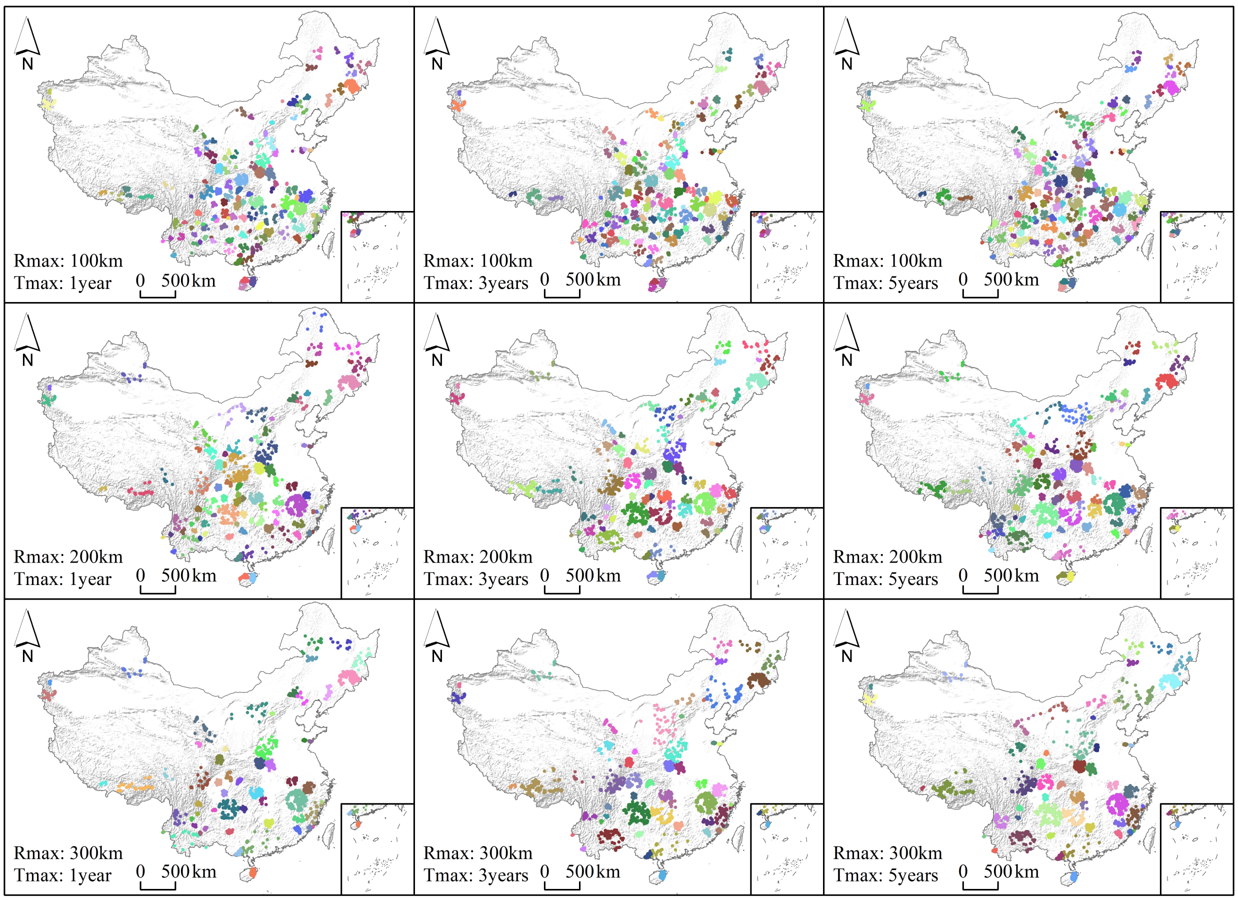

Figure 4: Significant (p<0.005) spatiotemporal clusters of flash flood disasters in China during 1950-2015.

scale, by merging small cluster close each others into big ones or eventually by neglecting very small flash floods aggregations. Conversely, very large clusters, acting at a coarse spatial scale, are detected for any value of R_{max}, as can be geographically visualised in the south-easternmost sector of China (Figure 4). Changes of T_{max} have almost no effect on the number of detected clusters since, even allowing for a maximum duration of 5 years, almost all the clusters do not exceed the duration of one year.

To confirm this finding, we computed the temporal duration of the first ten clusters of flash flood disasters detected by applying a T_{max} equals to 3 years and for the three models, defined by using values of R_{max} equal to 100, 200 and 300 km (Table 2). Results confirm that clusters duration, expresses as start and end date, never exceed one year. The most significant cluster (ranked as ID=1) is the same for any model and dated to 1975. Secondary clusters (just from the second to the tenth) are almost the same using R_{max} of 200 or 300 km, while, reducing the radius at $100km$, their size and ranking can change, due to the merging of small clusters into bigger ones. Finally, it is worth noting that the top-ten clusters are well distributed over the entire study period, with the oldest one detected between 1963 and 1969 and the latest in 2010.

Table 2: Temporal duration of the first 10 clusters of flash flood disasters detected via three different models (left: $R_{max} = 100km$; center: $R_{max} = 200km$; right: $R_{max} = 300km$).

ID	Radius	Start date	End date	ID	Radius	Start date	End date	ID	Radius	Start date	End date
1	81.04	1975/1/1	1975/12/31	1	81.04	1975/1/1	1975/12/31	1	81.04	1975/1/1	1975/12/31
2	64.51	2010/1/1	2010/12/31	2	146.06	1998/1/1	1998/12/31	2	146.06	1998/1/1	1998/12/31
3	60.73	2006/1/1	2006/12/31	3	64.51	2010/1/1	2010/12/31	3	64.51	2010/1/1	2010/12/31
4	72.76	2010/1/1	2010/12/31	4	60.73	2006/1/1	2006/12/31	4	60.73	2006/1/1	2006/12/31
5	94.42	1998/1/1	1998/12/31	5	72.76	2010/1/1	2010/12/31	5	72.76	2010/1/1	2010/12/31
6	73.13	1969/1/1	1969/12/31	6	73.13	1969/1/1	1969/12/31	6	73.13	1969/1/1	1969/12/31
7	56.67	1963/1/1	1963/12/31	7	176.96	1982/1/1	1982/12/31	7	176.96	1982/1/1	1982/12/31
8	49.51	1996/1/1	1996/12/31	8	70.57	1984/1/1	1984/12/31	8	70.57	1984/1/1	1984/12/31
9	70.57	1984/1/1	1984/12/31	9	129.06	1996/1/1	1996/12/31	9	157.14	2010/1/1	2010/12/31
10	35.27	1987/1/1	1987/12/31	10	157.14	2010/1/1	2010/12/31	10	56.18	1960/1/1	1960/12/31

We opted to carry out additional analyses using a T_{agg} of three months (hereafter referred as *monthly model*). Results are shown in Figure 5 where information on the spatial distribution of the detected clusters is shown. Overall, the clusters chiefly appear along the main river systems in China, namely the Yangtze, the Yellow, the Pearl and the Yarlung Zangbo Rivers. In addition, some clusters stand out on high mountains, such as the Qinling-Daba and the Changbai Mountains.

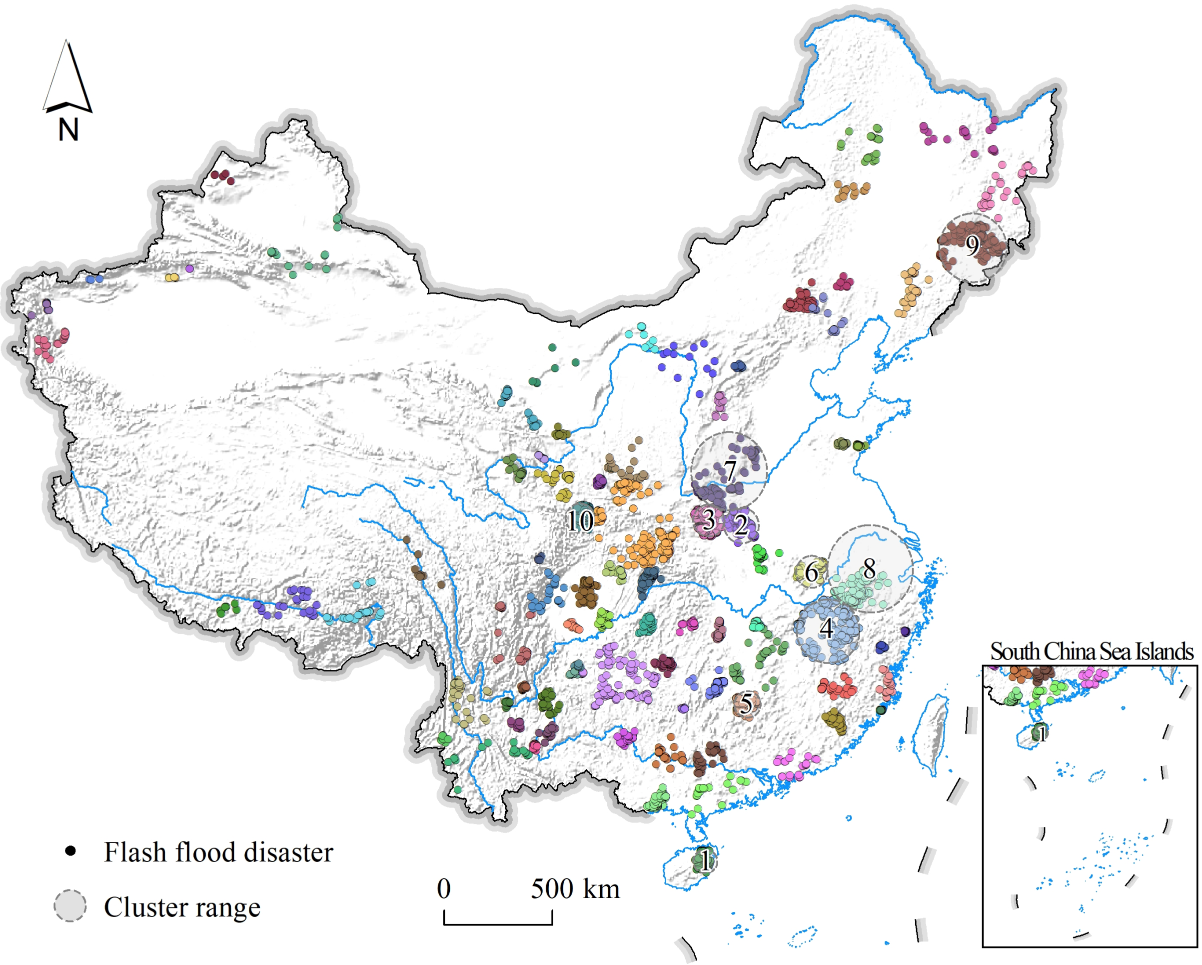

Figure 5: Significant (p<0.005) spatiotemporal clusters of flash flood disasters occurring in China during 1950-2015 ($R_{max} = 200km$, $T_{max} = 3years$, $T_{agg} = 3months$).

3.2.2 Temporal characterization of detected clusters

It emerges that clusters detected by using the different parameters for the scanning cylinders overlap both in space and in time. Therefore, in the present analysis, seeking at investigating in more details the years of occurrence and the temporal duration of detected clusters of flash floods, only results from the model with $R_{max} = 200km$, $T_{agg} = 1year$ and $T_{max} = 3year$ are presented. Considering all the statistically significant clusters, they emerged during almost the each year of the investigated period, but are more frequent starting from 1980. The relative small number of clusters detected between 1950 and 1980 may imply that the data acquisition and report in the Chinese database of hydro-morphological disasters was not fully operational at the time. Conversely, from 1980 to present days the Chinese database has evolved into a mature and detailed geographic information system. Another factors explaining this distribution can be the more frequent extreme precipitations observed in the recent decades, which can have increased the frequency of flash flood disasters in this last period. The precipitation regime can also explain the variation in the duration of the detected clusters. Indeed some clusters have a temporal extent up to three years, which could results from a persistent precipitation pattern over a delimited prone area.

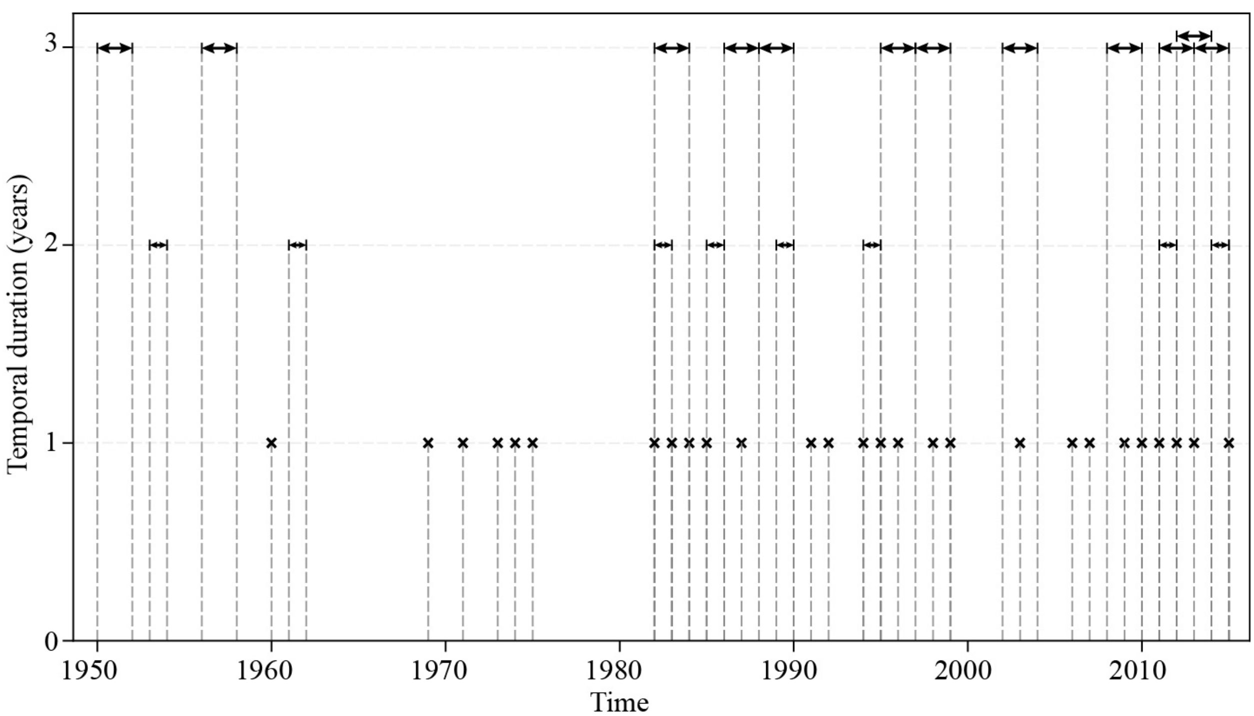

Figure 6: Temporal duration of flash flood disaster clusters detected via the *yearly model* ($R_{max} = 200km$, $T_{agg} = 1year$, $T_{max} = 3years$).

To highlight the influence of the seasonal variability in cluster detection, we carried out additional analyses using a T_{agg} of three months (hereafter referred as *monthly model*). As shown on 5, the detected clusters are spread along the main river systems in China, namely the Yangtze, the Yellow, the Pearl and the Yarlung Zangbo Rivers. In addition, some clusters stand out on high mountains such as the Qinling-Daba and the Changbai Mountains.

Table 3: Temporal duration of the first 10 clusters of flash flood disasters detected via the *monthly model* ($R_{max} = 200km$, $T_{max} = 1year$, $T_{agg} = 3months$)).

ID	Radius	Start date	End date
1	54.88	2010/10/1	2010/12/31
2	81.04	1975/4/1	1975/9/30
3	72.76	2010/7/1	2010/9/30
4	146.06	1998/4/1	1998/9/30
5	60.73	2006/7/1	2006/9/30
6	73.13	1969/4/1	1969/9/30
7	178.05	1982/7/1	1982/9/30
8	199.88	1996/4/1	1996/6/30
9	157.14	2010/7/1	2010/9/30
10	67.05	1984/4/1	1984/9/30

Forcing the model parameterization to aggregate over three months allows to investigate potential seasonal effects. Indeed, even if the maximum temporal duration is still of one year, looking at the ten most significant clusters detected under the *monthly model* (Table 3), it results that all of them have a duration of three (six clusters) or six (four clusters) months. Notably, almost every cluster (nine clusters) encompass the period from July to September, with an earlier start date (in April) for the ones which have a longer duration.

To visualize the seasonality trend, we summarized these result using a cyclic representation (Figure 7). The majority of the cluster have a 3-months duration, concentrated in the period between July and October. Furthermore, clusters of 6-months temporal duration are most likely to occur from January to July or from April to October. As for clusters with 9-months temporal duration, these mostly cover the period of July-August-September, irrespective of the starting month. Ultimately, as noticed for the *yearly model*, also in the *monthly model* much more clusters were detected in the recent period. Overall, the vast majority of flash flood disasters clusters happened between July and October, a period coinciding with the wet season in China.

3.2.3 Clusters pattern evolution at decade-scale

The previous analyses allowed to detect yearly and seasonal clusters. However, environmental changes usually act on a longer time-span. To better investigate this factor, we considered a temporal subdivision of the dataset into six subsets, each one lasting ten years (starting from 1956). Each subset was analysed separately, using the following parameter for the scanning widows: $R_{max} = 200km$, $T_{max} = 1year$, while no temporal aggregation was applied. This allowed to precisely evaluate the temporal duration of each cluster, given as number of days

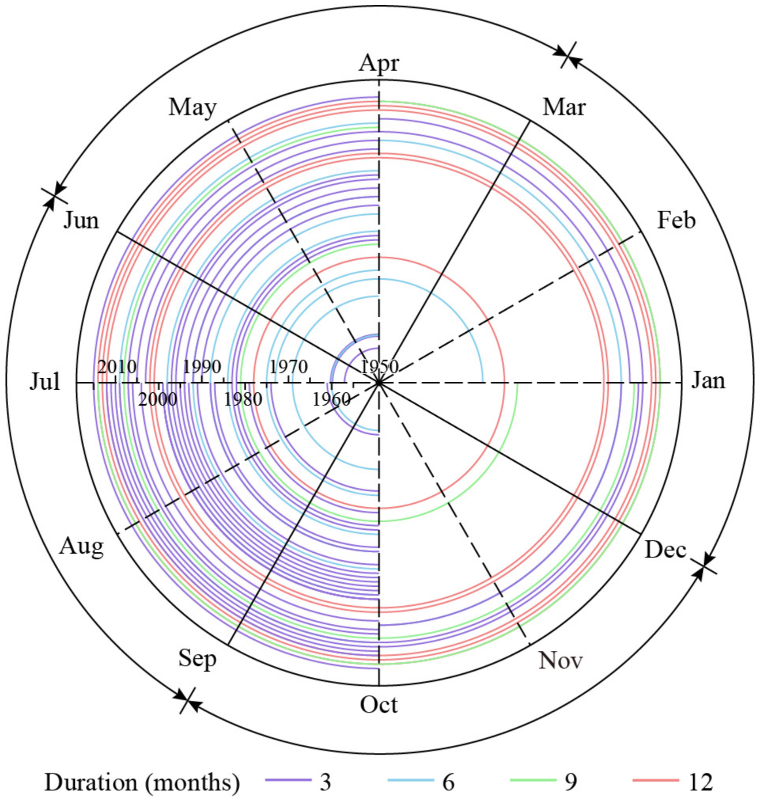

Figure 7: Seasonal effect of flash flood disaster clusters detected via the *monthly model* ($R_{max} = 200km$, $T_{max} = 1year$, $T_{agg} = 3months$).

between the earliest and the latest flash flood single event within a cluster. As shown in Figure 8, the number of detected clusters increases from the early to recent periods. These are compared with the rainfall distribution, derived from the daily rainfall data provided by the China Meteorological Administration (http://data.cma.cn/). In the present study, only the weather stations (a total of 699 rain gauges) with complete data for the period 1955-2015 were considered. From these, we computed the extreme precipitation as follows. Out of the rainfall records available per weather station, we initially extracted 5% of the time series corresponding to the rainfall values greater than the 95^{th} percentile (Karl and Easterling, 1999; Klein Tank and Können, 2003). Then we cumulated these values per station over decadal time periods corresponding to 1956-1965, 1966-1975, 1976-1985, 1986-1995, 1996-2005 and 2006-2015. From these cumulated extreme rainfall values per station and per decade, we computed the mean over the ten year time span and then interpolated over the whole spatial domain under consideration via a Ordinary Kriging. The data was regionalized on a $2km \times 2km$ lattice. The procedure returned six maps of the mean extreme events per decade over the Chinese territory. It results that flash floods detected clusters are mainly located in the southeastern most humid regions in every period. However, in the last two decades, clusters appear also in the northwestern arid regions. Even if the rainfall distribution, averaged over each decades, does not allow to discover clear changes along the subsequent periods, these newly detected clusters can be due to the intensification of the extreme rainfall events occurring in the area in recent periods. This assumption is confirmed by the statistics on clusters duration (Figure 9). From the boxplot summarizing the descriptive statistics it is evident that the median values of clusters duration tends to slightly decrease from 46 days (1956-1965), to 17 days (1986-1995), to stabilise at a value around 20 days in the two last decades. At the same time, the overall duration, measured as difference between the maximum and the minimum value, is higher in the late periods (140 days in 1956-1965, and 93 and 74 days respectively in the two following decades) than in the early periods (about 65 days for the last two decades). This is even more evident looking at the inter quantile ranges, which decrease with time. To resume, from these analyses, it results that the number of detected clusters globally increase in time, but their duration drastically decreases in the recent period.

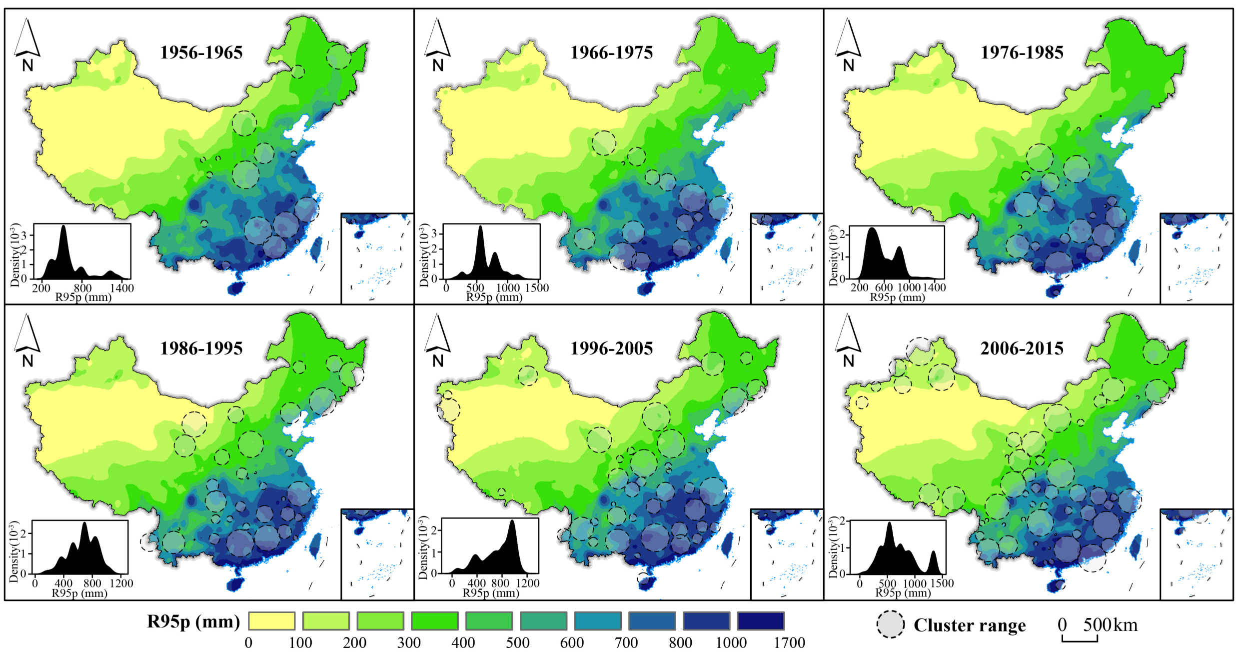

Figure 8: Significant ($p < 0.005$) spatiotemporal clusters of flash flood disasters in China every ten years. The size of the circles indicates the spatial coverage of the flash flood clusters we detected.

Spatiotemporal clusters of flash flood disasters detected in China by decades were further assembled in a unique image. To this end, the centroid of each cluster (with reference to Figure 8) was extracted and intersected with the catchment boundaries. Then, we computed the total number of clusters per catchment (Figure 10, panel (a)) as well as the average interval of time at which two consecutive clusters have been detected in the same catchment (Figure 10, panel (b)). It results that the catchments mainly affected by clusters of flash floods along several decades are mainly located in the southeast sector and essentially in the coastal mountains and that, on average, most of the cluster occur within an interval of 10-20

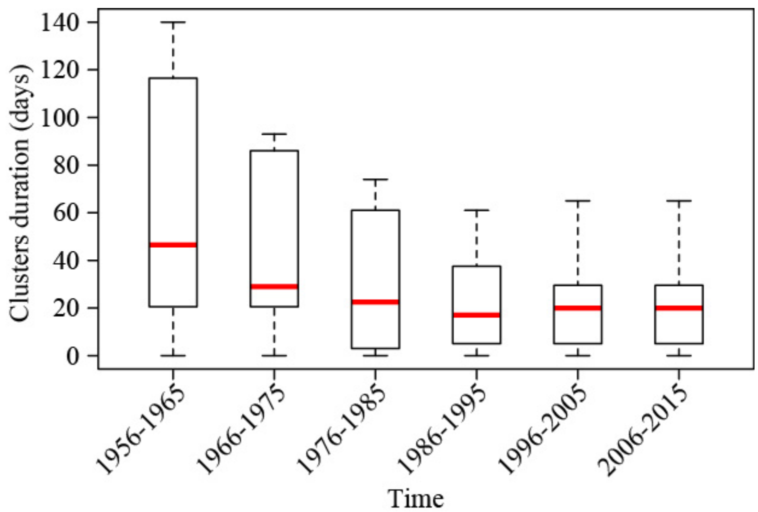

Figure 9: Boxplots summarizing the descriptive statistics of the duration of clusters reported on Figure 8.

years.

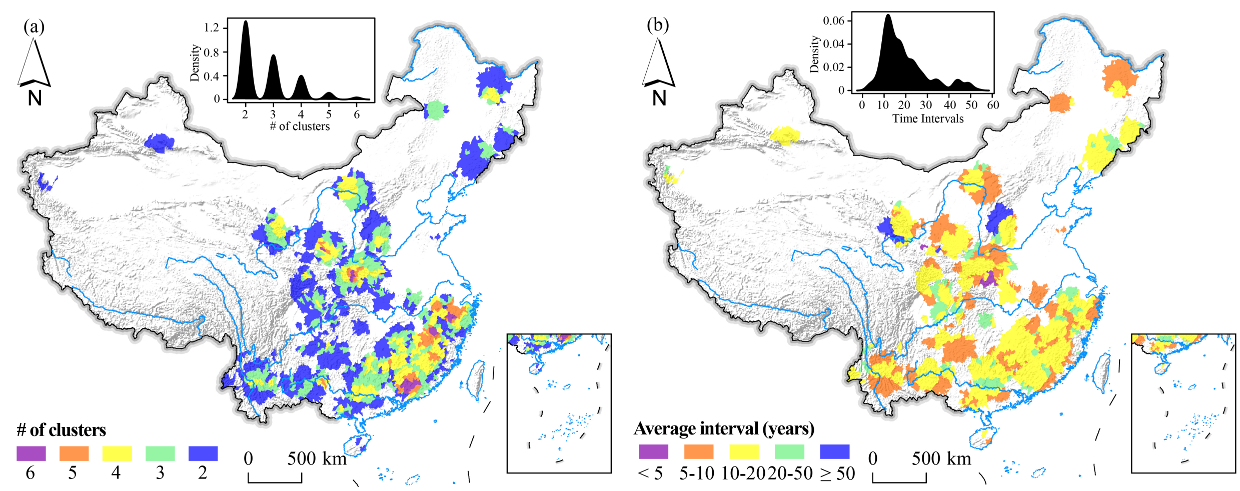

Figure 10: Number of time a cluster has been detected by catchment and by decade (a). Average time-interval between two clusters detected over the same catchment by decade (b)

4 Discussions

The present study aims at exploring the spatiotemporal clustering characteristics, in terms of spatial location and temporal duration, of flash flood disasters in China. For this purpose, we analyzed the official historical inventory, which covers a several decades from 1950 to 2015. Results are interpreted with a particular regard to the extreme rainfalls distribution, being these two processes highly related (Wei et al., 2018). Actually the spatiotemporal pattern distribution of flash floods can also been induced by the geomorphological setting of the area and by anthropogenic pressures, such as land use and land cover changes (Yang and Tian, 2009). However, in the present study we are considering both the spatial and the temporal dimension with the aim of detecting clusters occurring as a consequence of the interaction between these two variables. Therefore these clusters are likely to be related with dynamic factors such as rainfalls, which is the only triggering factors that covers and varies across the same spatiotemporal domain as the clusters themselves. Thus, our results are interpreted and discussed on the basis of this hypothesis.

The spatiotemporal K-function firstly computed reveals a deviation of flash flood disasters from a random process at specific scales, measured and quantified both in space, as distances-values, and in time, as yearly periods. Nevertheless this indicator can not provide the location at which clusters appears, or their duration. To this end, the spatiotemporal permutation Scan Statistics was then performed. Results allowed to identify statistically significant clusters together with the start and end date of their occurrence, and to detect areas and periods more susceptible to flash flood disasters. We opted for a set of possible combinations for the maximum spatial and temporal extension of the scanning windows, while dates were aggregated both at yearly and at seasonal scale (i.e. over three months). Among the dozens or even hundreds of clusters detected by the different models, the top ten most significant clusters resulting from the yearly model were analysed in detail. These appears to be almost the same for any increasing value of R_{max}, even if their size and ranking can change. This is a consequence of the fact that small clusters detected when using an R_{max} of $100km$ can merge into bigger cluster when R_{max} increases at 200 and 300 km. As for the occurrence time, these top-ten clusters are well distributed over the entire study period, with the earliest one dated to 1963 and the latest to 2010. Results of the monthly model show that the top ten most significant clusters have a duration of three (six clusters) or six (four clusters) months. Notably, almost every cluster encompasses the period from July to September, coinciding with the wet season in China, with an earlier start date (in April) for the clusters that have a longer duration. The same behaviour can be observed for the subsequent secondary clusters detected under the monthly model which, in addition, reveals an increasing number on cluster detected in the recent period.

Overall, clusters are chiefly located along the main river systems in China (the Yangtze, the Yellow, the Pearl and the Yarlung Zangbo Rivers). In addition, some clusters stand out on high mountains such as the Qinling-Daba and the Changbai Mountains.

Finally, to monitor the cluster pattern evolution, data were grouped and analysed by decades. As for the previous analyses, detected clusters are mainly located in the southeastern most humid regions in every period. However, in the last two decades, clusters appear also in the northwestern arid regions. These newly detected clusters can be due to the intensification the extreme rainfall events occurring in the area in recent periods, as a consequence of climate changes (Song et al., 2011a). This important fact is confirmed by checking the descriptive statistics of the duration of clusters: globally, the number of detected clusters increases in time, but the duration drastically decreases in recent periods, indicating a possible activation induced by short-duration extreme rainfall events. Another factor that can induce flash floods in China are the tropical cyclones (Hu et al., 2018). Indeed, it is well known tropical cyclones induce torrential rains which are a major trigger of catastrophic flood hazards in many coastal regions around the world (Rappaport, 2000; Dare et al., 2012; Zhang et al., 2019). A recent study by Lai et al. (2020) show that slow-moving tropical cyclones, characterized by lower translation speed, occurred more frequently after 1990 in the Pearl River Delta in southern China. In addition, their findings suggest that these cyclones tends to elevate local rainfall totals and thus impose greater flood risks at the regional scale. Essentially clusters results to be outnumbered in the last three decades, but their duration drastically decreases in the recent period, indicating a possible activation induced by short-duration extreme rainfall events.

As concern the spatial distribution of detected clusters, our analyses revealed that the more affected catchments with frequent clusters are mainly located in the southeast sector and essentially in the coastal mountains. China is indicated as one of the hotspot with global flood-exposed coastal population (Van Coppenolle and Temmerman, 2020). Therefore, we can assume these catchments to be exposed at the highest potential risk across the whole Chinese territory also in the short to long term future. In addition, catchments with clusters occurring within a short interval (5 to 10 years) may also pose a relevant threat, especially in the near future.

In the present study spatiotemporal clusters of flash floods were detected chiefly on the basis of two parameters (R_{max} and T_{max}), without featuring terrain attributes, precipitation regimes and anthropogenic pressure. However, these factors may have played and still play a significant role to explain the distribution of flash flood disasters. For instance, the approach we adopted may over-rely on spatial distances to detect clusters. In fact, the natural landscape has mountain belts that can act as orographic barriers to the incoming cloudbursts, effectively limiting the rainfall distribution – hence flash flood occurrences – on one or the other side of a catchment divide (at various scales). As for the temporal scale, due to the large time-span, the detected temporal patterns may reflect more information due to long-term climatic variations rather than specific conditions. For this reason, we are planning to extend our spatiotemporal cluster analyses to more complex models, which can concurrently capture multivariate contributions featuring environmental effects, even at the latent level (Lombardo et al., 2018, 2019).

5 Conclusion

In this work, we explore the national archive of flash flood disasters in China from 1950 to 2015. The term disaster is meant to describe the destructiveness of the flash floods, since each record in this archive has produced economic, life losses, or both.

The clustering procedure highlighted distinct spatial and temporal patterns at different scales. For instance, the statistically significant clusters of flash flood disasters detected in the present study occur in specific area and have a characteristic duration which closely follow the extreme rainfall patterns. The performed analyses allowed us to distinguish seasonal, yearly and even long-term flash flood persisting behaviors. The persistence of disasters is a crucial information because it indicates the risk that a community may undergo in response to a flash flood. Moreover, we studied the cycle of such disasters with particular emphasis on their repeated occurrence per catchment and by decade. As a result, we highlighted that the south-easternmost sector of China is subjected to a much larger number of flash flood clusters compared to the rest of the country. However, in terms of how these clusters are manifested through time with regards to their average re-occurrence time, the catchments in the south-eastern sector suffer from flash floods as frequently as the rest of the central and eastern sectors of the country. This complementary information can be further used in relation to engineering and structural design. In fact, infrastructure is usually built to sustain the damage of an event of certain return time. In our analyses we show that at catchment level, the very same area can be affected by clusters at least two up to six times in the last 60 years, considering a time-unit of ten years. This may suggest locally-tailored structural improvements which may lengthen the life expectancy of specific infrastructure as well as reduce the number of victims.

We would like to stress that, as advanced as it may be, our clustering framework is essentially a descriptive tool. And yet, the amount of information one can draw from a descriptive tool can be extremely valuable. Nowadays, the hazard community's effort is mainly dedicated to predictive modeling of various natures and purposes, thus leaving under-explored or even unexplored some basic concepts and interpretative conclusions that data description and visualization can provide. Long time series of national hazard phenomena are one of these examples where studying variations over space and time can highlight very important environmental dynamics, even in the direction of climate change and its implications.

Author contribution

Nan Wang and Weiming Cheng conceived and designed the experiments; Nan Wang performed the experiments; Luigi Lombardo and Marj Tonini analyzed the results; Nan Wang, Luigi Lombardo, and Marj Tonini wrote the paper; Liang Guo and Junnan Xiong revised the paper.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This work was supported by the China National Flash Flood Disasters Prevention and Control Project. The authors are grateful for financial support from the National Natural Science Foundation of China, grant number No. 41590845, and the China Institute of Water Resources and Hydropower Research (IWHR), grant number No. SHZH-IWHR-57.

References

Allan, D. W. (1966) Statistics of atomic frequency standards. Proceedings of the IEEE **54**(2), 221–230.

Archer, D., O'donnell, G., Lamb, R., Warren, S. and Fowler, H. J. (2019) Historical flash floods in England: New regional chronologies and database. Journal of Flood Risk Management **12**, e12526.

Au, S. (1998) Rain-induced slope instability in Hong Kong. Engineering Geology **51**(1), 1–36.

Borga, M., Anagnostou, E., Blöschl, G. and Creutin, J.-D. (2011) Flash flood forecasting, warning and risk management: the HYDRATE project. Environmental Science & Policy **14**(7), 834–844.

Borga, M., Boscolo, P., Zanon, F. and Sangati, M. (2007) Hydrometeorological analysis of the 29 August 2003 flash flood in the Eastern Italian Alps. Journal of Hydrometeorology **8**(5), 1049–1067.

Bout, B., Lombardo, L., van Westen, C. and Jetten, V. (2018) Integration of two-phase solid fluid equations in a catchment model for flashfloods, debris flows and shallow slope failures. Environmental Modelling & Software **105**, 1–16.

de Bruijn, J. A., de Moel, H., Jongman, B., de Ruiter, M. C., Wagemaker, J. and Aerts, J. C. (2019) A global database of historic and real-time flood events based on social media. Scientific Data **6**(1), 1–12.

Cama, M., Lombardo, L., Conoscenti, C., Agnesi, V. and Rotigliano, E. (2015) Predicting storm-triggered debris flow events: application to the 2009 Ionian Peloritan disaster (Sicily, Italy). Natural Hazards and Earth System Science **15**(8), 1785–1806.

Cama, M., Lombardo, L., Conoscenti, C. and Rotigliano, E. (2017) Improving transferability strategies for debris flow susceptibility assessment: Application to the Saponara and Itala catchments (Messina, Italy). Geomorphology **288**, 52–65.

Chambers, J. E., Meldrum, P. I., Wilkinson, P. B., Ward, W., Jackson, C., Matthews, B., Joel, P., Kuras, O., Bai, L., Uhlemann, S. et al. (2015) Spatial monitoring of groundwater drawdown and rebound associated with quarry dewatering using automated time-lapse electrical resistivity tomography and distribution guided clustering. Engineering Geology **193**, 412–420.

Chang, C.-W., Lin, P.-S. and Tsai, C.-L. (2011) Estimation of sediment volume of debris flow caused by extreme rainfall in Taiwan. Engineering Geology **123**(1-2), 83–90.

Costafreda-Aumedes, S., Comas, C. and Vega-Garcia, C. (2016) Spatio-temporal configurations of human-caused fires in Spain through point patterns. Forests **7**(9), 185.

Dare, R. A., Davidson, N. E. and McBride, J. L. (2012) Tropical cyclone contribution to rainfall over Australia. Monthly Weather Review **140**(11), 3606–3619.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X. et al. (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In Kdd, volume 96, pp. 226–231.

Fernández, D. and Lutz, M. (2010) Urban flood hazard zoning in Tucumán Province, Argentina, using GIS and multicriteria decision analysis. Engineering Geology **111**(1-4), 90–98.

Fischer, T. and Horálek, J. (2003) Space-time distribution of earthquake swarms in the principal focal zone of the NW Bohemia/Vogtland seismoactive region: period 1985–2001. Journal of Geodynamics **35**(1-2), 125–144.

Fuchs, S., Keiler, M. and Zischg, A. (2015) A spatiotemporal multi-hazard exposure assessment based on property data. Natural Hazards and Earth System Sciences **15**(9), 2127–2142.

Fuentes-Santos, I., Marey-Pérez, M. and González-Manteiga, W. (2013) Forest fire spatial pattern analysis in Galicia (NW Spain). Journal of Environmental Management **128**, 30–42.

Gariano, S. L. and Guzzetti, F. (2016) Landslides in a changing climate. Earth-Science Reviews **162**, 227 – 252.

Gartner, J. E., Cannon, S. H. and Santi, P. M. (2014) Empirical models for predicting volumes of sediment deposited by debris flows and sediment-laden floods in the transverse ranges of southern California. Engineering Geology **176**, 45–56.

Gaume, E., Bain, V., Bernardara, P., Newinger, O., Barbuc, M., Bateman, A., Blaškovičová, L., Blöschl, G., Borga, M., Dumitrescu, A. et al. (2009) A compilation of data on European flash floods. Journal of Hydrology **367**(1-2), 70–78.

Georgoulas, G., Konstantaras, A., Katsifarakis, E., Stylios, C. D., Maravelakis, E. and Vachtsevanos, G. J. (2013) "Seismic-mass" density-based algorithm for spatio-temporal clustering. Expert Systems with Applications **40**(10), 4183–4189.

Gomez, H. and Kavzoglu, T. (2005) Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. Engineering Geology **78**(1-2), 11–27.

Gourley, J. J., Hong, Y., Flamig, Z. L., Arthur, A., Clark, R., Calianno, M., Ruin, I., Ortel, T., Wieczorek, M. E., Kirstetter, P.-E. et al. (2013) A unified flash flood database across the United States. Bulletin of the American Meteorological Society **94**(6), 799–805.

Gu, X., Zhang, Q., Singh, V. P., Chen, X. and Liu, L. (2016a) Nonstationarity in the occurrence rate of floods in the Tarim River basin, China, and related impacts of climate indices. Global and Planetary Change **142**, 1–13.

Gu, X., Zhang, Q., Singh, V. P., Chen, Y. D. and Shi, P. (2016b) Temporal clustering of floods and impacts of climate indices in the Tarim River basin, China. Global and Planetary Change **147**, 12–24.

Haigh, I. D., Ozsoy, O., Wadey, M. P., Nicholls, R. J., Gallop, S. L., Wahl, T. and Brown, J. M. (2017) An improved database of coastal flooding in the United Kingdom from 1915 to 2016. Scientific Data **4**, 170100.

Hatheway, A. W., Kanaori, Y., Cheema, T., Griffiths, J. and Promma, K. (2005) 10th annual report on the international status of engineering geology—year 2004–2005; encompassing hydrogeology, environmental geology and the applied geosciences. Engineering Geology **81**(2), 99–130.

He, B., Huang, X., Ma, M., Chang, Q., Tu, Y., Li, Q., Zhang, K. and Hong, Y. (2018) Analysis of flash flood disaster characteristics in China from 2011 to 2015. Natural Hazards **90**(1), 407–420.

Hu, P., Zhang, Q., Shi, P., Chen, B. and Fang, J. (2018) Flood-induced mortality across the globe: Spatiotemporal pattern and influencing factors. Science of the Total Environment **643**, 171–182.

Jonkman, S. N. (2005) Global perspectives on loss of human life caused by floods. Natural Hazards **34**(2), 151–175.

Jonkman, S. N. and Kelman, I. (2005) An analysis of the causes and circumstances of flood disaster deaths. Disasters **29**(1), 75–97.

Karl, T. R. and Easterling, D. R. (1999) Climate extremes: selected review and future research directions. Climatic Change **42**(1), 309–325.

Kelman, I. and Spence, R. (2004) An overview of flood actions on buildings. Engineering Geology **73**(3-4), 297–309.

Klein Tank, A. and Können, G. (2003) Trends in indices of daily temperature and precipitation extremes in Europe, 1946–99. Journal of Climate **16**(22), 3665–3680.

Kouli, M., Loupasakis, C., Soupios, P. and Vallianatos, F. (2010) Landslide hazard zonation in high risk areas of Rethymno Prefecture, Crete Island, Greece. Natural Hazards **52**(3), 599–621.

Kulldorff, M. (1997) A spatial scan statistic. Communications in Statistics-Theory and methods **26**(6), 1481–1496.

Kulldorff, M., Athas, W. F., Feurer, E. J., Miller, B. A. and Key, C. R. (1998) Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. American Journal of Public Health **88**(9), 1377–1380.

Kulldorff, M., Heffernan, R., Hartman, J., Assunção, R. and Mostashari, F. (2005) A Space–Time Permutation Scan Statistic for Disease Outbreak Detection. PLoS Med **2**(3), e59.

Lai, Y., Li, J., Gu, X., Chen, Y. D., Kong, D., Gan, T. Y., Liu, M., Li, Q. and Wu, G. (2020) Greater flood risks in response to slowdown of tropical cyclones over the coast of China. Proceedings of the National Academy of Sciences **117**(26), 14751–14755.

Liu, Y., Yang, Z., Huang, Y. and Liu, C. (2018) Spatiotemporal evolution and driving factors of China's flash flood disasters since 1949. Science China Earth Sciences **61**(12), 1804–1817.

Lóczy, D., Czigány, S. and Pirkhoffer, E. (2012) Flash flood hazards. In Studies on water management issues. IntechOpen.

Lombardo, L., Bakka, H., Tanyas, H., van Westen, C., Mai, P. M. and Huser, R. (2019) Geostatistical modeling to capture seismic-shaking patterns from earthquake-induced landslides. Journal of Geophysical Research: Earth Surface **124**(7), 1958–1980.

Lombardo, L., Opitz, T., Ardizzone, F., Guzzetti, F. and Huser, R. (2020) Space-time landslide predictive modelling. Earth-Science Reviews p. 103318.

Lombardo, L., Opitz, T. and Huser, R. (2018) Point process-based modeling of multiple debris flow landslides using INLA: an application to the 2009 Messina disaster. Stochastic Environmental Research and Risk Assessment **32**(7), 2179–2198.

Lovejoy, S., Schertzer, D. and Ladoy, P. (1986) Fractal characterization of inhomogeneous geophysical measuring networks. Nature **319**(6048), 43–44.

Merz, B., Nguyen, V. D. and Vorogushyn, S. (2016) Temporal clustering of floods in Germany: Do flood-rich and flood-poor periods exist? Journal of Hydrology **541**, 824–838.

Moran, P. A. (1950) Notes on continuous stochastic phenomena. Biometrika **37**(1/2), 17–23.

Naus, J. L. (1965a) Clustering of random points in two dimensions. Biometrika **52**(1-2), 263–266.

Naus, J. L. (1965b) The distribution of the size of the maximum cluster of points on a line. Journal of the American Statistical Association **60**(310), 532–538.

Nowicki Jessee, M., Hamburger, M., Ferrara, M., McLean, A. and FitzGerald, C. (2020) A global dataset and model of earthquake-induced landslide fatalities. Landslides pp. 1–14.

Openshaw, S., Charlton, M., Wymer, C. and Craft, A. (1987) A mark 1 geographical analysis machine for the automated analysis of point data sets. International Journal of Geographical Information System **1**(4), 335–358.

Orozco, C. V., Tonini, M., Conedera, M. and Kanveski, M. (2012) Cluster recognition in spatial-temporal sequences: the case of forest fires. Geoinformatica **16**(4), 653–673.

Pappadà, R., Durante, F., Salvadori, G. and De Michele, C. (2018) Clustering of concurrent flood risks via Hazard Scenarios. Spatial Statistics **23**, 124–142.

Pierson, T. C., Costa, J. E. and Vancouver, W. (1987) A rheologic classification of subaerial sediment-water flows. Debris flows/avalanches: process, recognition, and mitigation. Reviews in Engineering Geology **7**, 1–12.

R Core Team (2019) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rappaport, E. N. (2000) Loss of life in the united states associated with recent atlantic tropical cyclones. Bulletin of the American Meteorological Society **81**(9), 2065–2074.

Renard, F. (2017) Flood risk management centred on clusters of territorial vulnerability. Geomatics, Natural Hazards and Risk **8**(2), 525–543.

Ripley, B. D. (1977) Modelling spatial patterns. Journal of the Royal Statistical Society: Series B (Methodological) **39**(2), 172–192.

Rowlingson, B. and Diggle, P. (2017) Splancs: spatial and space-time point pattern analysis. R package version 2.01-40.

Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., Alfieri, L. and Freer, J. E. (2015) A high-resolution global flood hazard model. Water Resources Research **51**(9), 7358–7381.

Santangelo, N., Daunis-i Estadella, J., Di Crescenzo, G., Di Donato, V., Faillace, P., Martín-Fernández, J., Romano, P., Santo, A. and Scorpio, V. (2012) Topographic predictors of susceptibility to alluvial fan flooding, Southern Apennines. Earth Surface Processes and Landforms **37**(8), 803–817.

Song, Y., Achberger, C. and Linderholm, H. W. (2011a) Rain-season trends in precipitation and their effect in different climate regions of China during 1961–2008. Environmental Research Letters **6**(3), 034025.

Song, Y., Achberger, C. and Linderholm, H. W. (2011b) Rain-season trends in precipitation and their effect in different climate regions of China during 1961–2008. Environmental Research Letters **6**(3), 034025.

Stoyan, D. (2006) Fundamentals of point process statistics. In Case studies in spatial point process modeling, pp. 3–22. Springer.

Tokhmechi, B., Memarian, H., Moshiri, B., Rasouli, V. and Noubari, H. A. (2011) Investigating the validity of conventional joint set clustering methods. Engineering Geology **118**(3-4), 75–81.

Tonini, M. and Cama, M. (2019) Spatio-temporal pattern distribution of landslides causing damage in Switzerland. Landslides **16**(11), 2103–2113.

Tonini, M., Pereira, M. G., Parente, J. and Orozco, C. V. (2017) Evolution of forest fires in Portugal: from spatio-temporal point events to smoothed density maps. Natural Hazards **85**(3), 1489–1510.

Turnbull, B. W., Iwano, E. J., Burnett, W. S., Howe, H. L. and Clark, L. C. (1990) Monitoring for clusters of disease: application to leukemia incidence in upstate New York. American Journal of Epidemiology **132**(supp1), 136–143.

Van Coppenolle, R. and Temmerman, S. (2020) Identifying global hotspots where coastal wetland conservation can contribute to nature-based mitigation of coastal flood risks. Global and Planetary Change **187**, 103125.

Varga, P., Krumm, F., Riguzzi, F., Doglioni, C., Süle, B., Wang, K. and Panza, G. (2012) Global pattern of earthquakes and seismic energy distributions: Insights for the mechanisms of plate tectonics. Tectonophysics **530**, 80–86.

Vennari, C., Parise, M., Santangelo, N. and Santo, A. (2016) A database on flash flood events in Campania, southern Italy, with an evaluation of their spatial and temporal distribution. Natural Hazards and Earth System Science **16**, 2485–2500.

Wang, N., Cheng, W., Wang, B., Liu, Q. and Zhou, C. (2020) Geomorphological regionalization theory system and division methodology of China. Journal of Geographical Sciences **30**(2), 212–232.

Wei, L., Hu, K.-h. and Hu, X.-d. (2018) Rainfall occurrence and its relation to flood damage in China from 2000 to 2015. Journal of Mountain Science **15**(11), 2492–2504.

Wood, J., Harrison, S., Reinhardt, L. and Taylor, F. (2020) Landslide databases for climate change detection and attribution. Geomorphology **355**, 107061.

Woodward, K., Wesseloo, J. and Potvin, Y. (2018) A spatially focused clustering methodology for mining seismicity. Engineering Geology **232**, 104–113.

Wu, Y., Ji, H., Wen, J., Wu, S.-Y., Xu, M., Tagle, F., He, B., Duan, W. and Li, J. (2019) The characteristics of regional heavy precipitation events over eastern monsoon China during 1960–2013. Global and Planetary Change **172**, 414–427.

Xu, Z., Gong, T. and Li, J. (2008) Decadal trend of climate in the Tibetan Plateau—regional temperature and precipitation. Hydrological Processes: An International Journal **22**(16), 3056–3065.

Yang, J., Cheng, C., Song, C., Shen, S., Zhang, T. and Ning, L. (2019) Spatial-temporal distribution characteristics of global seismic clusters and associated spatial factors. Chinese Geographical Science **29**(4), 614–625.

Yang, Y. and Tian, F. (2009) Abrupt change of runoff and its major driving factors in Haihe River Catchment, China. Journal of Hydrology **374**(3-4), 373–383.

Zaharia, L., Costache, R., Prăvălie, R. and Ioana-Toroimac, G. (2017) Mapping flood and flooding potential indices: a methodological approach to identifying areas susceptible to flood and flooding risk. Case study: the Prahova catchment (Romania). Frontiers of Earth Science **11**(2), 229–247.

Zhan, J., Xu, P., Chen, J., Wang, Q., Zhang, W. and Han, X. (2017) Comprehensive characterization and clustering of orientation data: A case study from the Songta dam site, China. Engineering Geology **225**, 3–18.

Zhang, D. D., Zhang, J., Lee, H. F. and He, Y.-q. (2007) Climate change and war frequency in Eastern China over the last millennium. Human Ecology **35**(4), 403–414.

Zhang, Q., Wei, Q. and Chen, L. (2010) Impact of landfalling tropical cyclones in mainland China. Science China Earth Sciences **53**(10), 1559–1564.

Zhang, X. and Cong, Z. (2014) Trends of precipitation intensity and frequency in hydrological regions of China from 1956 to 2005. Global and Planetary Change **117**, 40–51.

Zhang, Y., Wang, Y., Chen, Y., Liang, F. and Liu, H. (2019) Assessment of future flash flood inundations in coastal regions under climate change scenarios—A case study of Hadahe River basin in northeastern China. Science of the Total Environment **693**, 133550.

Zhao, J., Jin, J., Guo, Q., Liu, L., Chen, Y. and Pan, M. (2014) Dynamic risk assessment model for flood disaster on a projection pursuit cluster and its application. Stochastic Environmental Research and Risk Assessment **28**(8), 2175–2183.