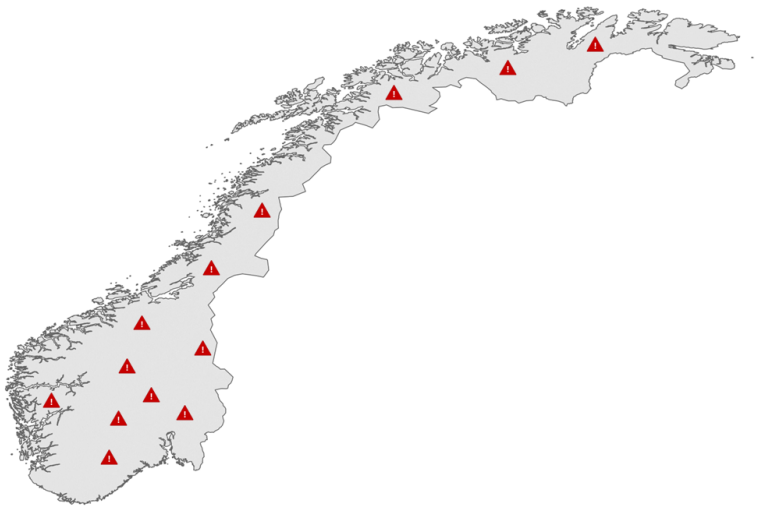
Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Note

Widespread Risk of Extreme Precipitation and Flooding



Note no SAMBA/09/24

Authors Henrik Nordtorp
Thea Roksvåg
Thordis Thorarinsdottir

Date 2nd August 2024

Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

The authors

Henrik Nordtorp is Summer Intern, Thea Roksvåg is Senior Research Scientist and Thordis Thorarinsdottir is Chief Research Scientist II at the Norwegian Computing Center.

Norwegian Computing Center

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in statistical-mathematical modelling, geomodelling, earth observation, machine learning, image analysis, artificial intelligence, language technology, digital security, digital inclusion and digital transformation. The clients include a broad range of industrial, commercial and public service organisations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway, the European Research Council and key customers.

Norsk Regnesentral
Norwegian Computing Center
Postboks 114, Blindern
NO-0314 Oslo, Norway

Besøksadresse
Office address
Gaustadalléen 23
NO-0373 Oslo, Norway

Telefon · telephone
(+47) 22 85 25 00
Telefaks · telefax
(+47) 22 69 76 60

Internett · internet
www.nr.no
E-post · e-mail
nr@nr.no

Title	**Widespread Risk of Extreme Precipitation and Flooding**
Authors	**Henrik Nordtorp , Thea Roksvåg , Thordis Thorarinsdottir**
Date	2nd August 2024
Publication number	SAMBA/09/24

Abstract

Developing a thorough understanding of the prevalence of natural disasters such as floods and extreme precipitation events is of high importance for the society. In this report, it is attempted to investigate and quantify the widespread risk of flood and extreme precipitation events, using statistical methods on historical flood and precipitation data from southern Norway. The use of descriptive methods and clustering revealed patterns in the geographical spread of extreme events, in-line with current process understandings from hydrology and meteorology.

Keywords	extreme precipitation, flooding, risk, spatial clustering, widespread risk
Target group	General
Availability	Open
Project	ClimDesign
Project number	220941
Research field	Statistics, statistical climatology
Number of pages	24

Contents

1 Introduction

Water-related natural disasters like floods and extreme rainfall have a potential to cause a large extent of damage. Traditional risk assessments in Norway associated with extreme precipitation and flooding focus on location-specific estimation of annual exceedance probabilities[1][2] or short- to medium-term warnings at a regional level[3]. In this study, we apply methods developed in Brunner et al. (2020) to bridge the gap between these two approaches by assessing the widespread risk of extreme streamflow discharge or precipitation from observations. A procedure for linking extreme events across different geographical locations is implemented, and the resulting data are used to deduce connections in the prevalence of extreme event occurrences across locations through the use of descriptive statistical methods.

The remainder of the report is organised as follows. The data set used in the study are presented in the next Section 2. The methods, including data pre-procesing and analysis approaches, are presented in Section 3 and the results are summarized in Section 4. A discussion in given in Section 5 and information on data and software availability is given in the last Section 6.

This study was conducted as a part of the project *Climate adjusted design value for extreme precipitation and flooding* (ClimDesign) funded by the Research Council of Norway (project number 302457).

2 Data

A data set consisting of discharge measurements over time for a large number of nationwide measurement stations was supplied by the Norwegian Water Resources and Energy Directorate (NVE). All discharge measurements are registered at a specific station, denoted by an ID-number, on a specific date. The measurements are registered daily, and the unit used for discharge is m^3/s (how many cubic meters of water that passes the stream gauge per second). Table 1 presents a compressed example row from the data set and the locations of the measurement stations are shown on a map in Figure 1.

Table 1. Example row from the discharge data set. Additional columns: *Drainage basin key, year, month, day, region area, main number, degree of regulation, polygon index, polygon area*

Station ID	Discharge (m^3/s)	Date	Total Area	Mean UTM-X	Mean UTM-Y	⋯
200001	0.125540	1990-01-01	37.8500	336001.453	6775081	⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋮

1. https://klimaservicesenter.no/ivf. Accessed 1.8.2024.
2. https://publikasjoner.nve.no/rapport/2022/rapport2022_33.pdf. Accessed 1.8.2024.
3. https://www.varsom.no/. Accessed 1.8.2024.

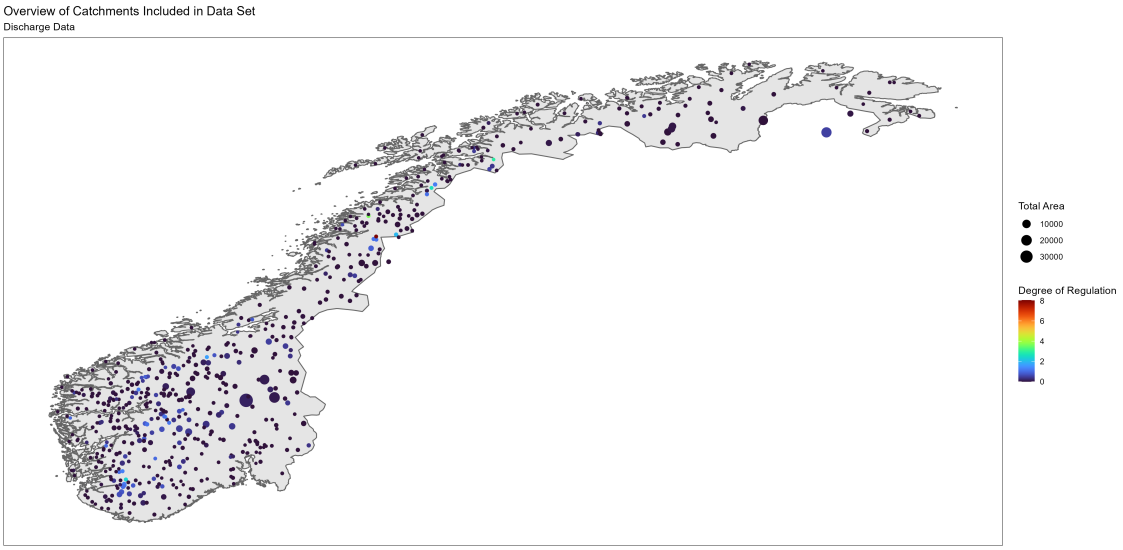


Figure 1. Overview of stream gauging stations in the data set supplied by NVE. Points are sized by catchment area, and coloured by the degree of regulation.

By the pre-processing steps described in Section 3, subsetted, regional data sets for further analysis are created. The data set for eastern Norway, specifically, consists of data from 38 stream gauging stations over the course of 38 years (streamflow data from 1985 up to and including 2022), with only an insignificant number of missing measurements (days with no streamflow values for certain stations).

Equivalently, a raw data set consisting of precipitation measurements over time for a number of measurement stations in eastern Norway was supplied by the Norwegian Meteorological Institute (MET). All precipitation measurements are registered at a specific weather station, also denoted by an ID-number, at a specific time. The time points are separated hourly, and the unit used for discharge is *mm* precipitation. The structure of the precipitation data set is shown in Table 2 and the locations of the measurement stations are shown on the map in Figure 2.

Table 2. Example row from precipitation data set.

Station ID	Timepoint	Precipitation (mm)	Mean UTM-Y	Mean UTM-X
27120	2022-09-30 19:00:00	0.4	6596828	238278
⋮	⋮	⋮	⋮	⋮

By the pre-processing steps described in Section 3, a subsetted data set for further analysis was created. This data set consists of precipitation data from 23 weather stations, mainly located along the inner parts of the Oslo fjord. A large proportion of the stations are located in the cities of Oslo and Drammen. The time series of measurements are not continuous, as data are only kept for the expanded summer months (May - September), and certain time points are removed during the pre-processing, but all locations have

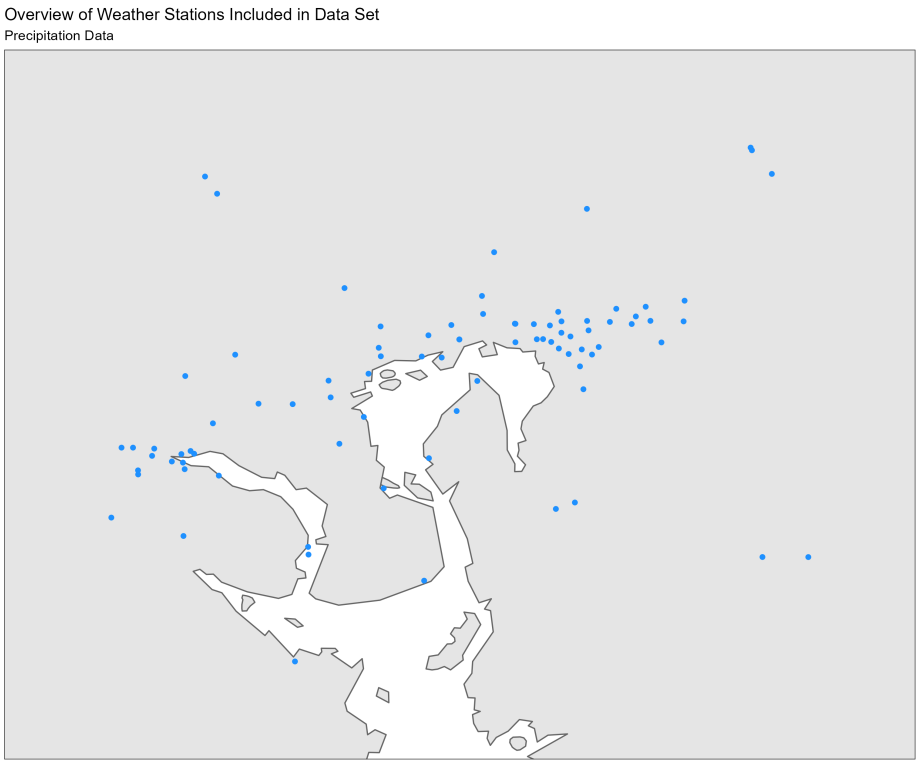


Figure 2. Overview of rainfall stations in the data set supplied by MET.

data for the same time points. A total of 26207 hourly precipitation measurements are present in the final data set for each station.

3 Methods

3.1 Pre-processing

Discharge and precipitation measurements were inconsistently sampled across the different locations, with several gaps in the data (missing data) over their respective time series. In order to allow for widespread risk analysis, data must be overlapping, i.e., consist of consistent measurements over timepoints which have no missing values for all locations. Two iterative methods were implemented for dealing with this, one for discharge data and one for precipitation data. The two methods are similar in principle, but have some important differences.

Initially, the cleaned discharge data were split into five separate data sets corresponding to geographical regions in Norway (east, south, west, mid and north), to allow for region-specific analysis. Identifying dates with missing streamflow values were achieved by first creating a Boolean data frame representing for which dates there exist streamflow values: For each location and every date in the range of the data set, the combination is denoted by a 1 if a streamflow value is present (streamflow > 0), and a 0 if a streamflow is absent. Note that only streamflow values above 0 are considered to be non-missing; thus, if some location have registered a streamflow of 0 (for example in the case of extreme drought),

this will not be included in the data set. This Boolean data frame is used to subset the original data frame. In the first step, all measurements recorded before a certain year, in this case 1985, are discarded. Then, through an iterative process, measurements from certain locations are removed if they contain fewer measurements in a given year than a certain threshold, here set to 364 days. In practice, this ensures that only streamflow series that contain no missing data over the year specified in the first step are kept. The method guarantees that the data are approximately gap free, while still being flexible if certain days (e.g. an extra day due to leap years) are missing.

Identifying time points with missing precipitation measurements was also achieved by first creating a Boolean data frame representing for which time points precipitation values are present in the data: For each location and every time point in the range of the data set, the combination was denoted by a 1 if a measurement was present (precipitation $>= 0$), and 0 otherwise. This Boolean data frame is used to subset the original data set, but in a slightly different manner than that of the discharge data. For each location in the data set, gap sizes are quantified with a separate variable. The data are ordered by time, so when periods of missing data increases, so does the value of the gap size variable. If a location has gap sizes exceeding a certain threshold, in this case 744 hours, the location is removed in its entirety. The choice of this threshold reflects the number of hours in a "long month", i.e. a month with 31 days; In practice, stations with missing values for an entire month are removed. In the next step, specific time points are removed across all stations; If at least one station has a gap of size greater than a certain threshold, in this case 48 hours, these time points are removed. This implies that gaps up to 48 hours are kept (including parts of larger gaps, up to 48 hours from the last registry). The justification of having a threshold like this is that small periods of missing data can likely be safely imputed. Therefore, all remaining missing data after this process are imputed to have a precipitation value of 0 mm.

3.2 Extreme Event Identification

The suitably processed daily streamflow and hourly precipitation time series are used to identify extreme events. An extreme event is defined as a time point where the streamflow or precipitation exceeds a certain threshold. This process consist of two consecutive steps: (1) identifying extreme events at individual locations (stream gauges or weather stations) and (2) determining dates or time points of extreme event occurrences over all locations. For discharge measurements, location-specific peak-over-threshold events are identified as dates where the streamflow exceeds the μth percentile of the daily streamflow over the entire time series for each station. This ensures that extreme event thresholds are relative to catchment sizes, since extremes are identified as streamflow much higher than normal. For precipitation, location-specific peak-over-threshold events are identified as time points where the precipitation exceeds the μth percentile of the hourly precipitation, calculated over the entire time series for all stations, excluding measurements where no precipitation were registered. Thus, the extreme event thresholds are not calculated relative to location. For discharge and precipitation analyses, extreme events where identified using a varying μ of 0.9, 0.95, 0.975 and 0.99 and 0.95, 0.975, 0.99 and 0.999,

respectively.

In a second step, a data set consisting of the dates of extreme event occurrences across all locations is compiled, separately for flood events and extreme precipitation events. These events, henceforth referred to as main events, are defined as events occurring at minimum one location within a certain time window. If there are multiple locations affected by an event in the same time window, the time point where most locations were affected is chosen. The time window is set to 7 days for flood analysis, and 48 hours for precipitation analysis. These main events, which are represented as dates for floods and hourly time points for extreme precipitation, can be viewed as independent events affecting at least one station. In order to gain insight into sensible choices of the time window, the autocorrelation between discharge measurements for a number of stations is calculated.

The set of main events is used to construct a binary event matrix which specifies for each location, which events affected this location. The entries of this matrix are defined by assigning 1 to locations that showed a peak-over-threshold event within a certain time window relative to the time point of the main event, and 0 otherwise. In other words, a location is assigned a 1 if it was affected by the given main event (with a "wiggle room" corresponding to the window size), and 0 if it was not. The time window is set to $\pm$ 2 days for the flood analysis, and $\pm$ 12 hours for the precipitation analysis. The motivation for using a time window like this, and not the main event time points directly, is to compensate for hydrological and meteorological factors such as different catchment sizes and different movement rates of weather systems. Thus, this matrix represents the simultaneity of extreme events across different locations. Mathematically, the binary event matrix can be represented as

$$E \in \mathbb{Z}_2^{m \times n}$$

where m is the number of locations and n is the number of main events. Each entry is assigned the following way:

$$E_{ij} = \begin{cases} 1 \text{ if location } i \text{ was affected by event } j \\ 0 \text{ if location } i \text{ was not affected by event } j \end{cases}$$

3.3 Widespread Hazard Estimation

In order to quantify the prevalence ("widespreadness") of extreme events, the event matrices are used directly (Brunner et al., 2020). Since each column represents an independent event, and the matrix is binary, calculating the column means result in numbers representing the proportion of locations affected by the given extreme events. This is further used to calculate the probability that at least $r\%$ of the locations experience an extreme event at the same time, by dividing the number of events with a column mean greater than a chosen r with the total number of events. This is performed per event matrix, that is, for precipitation and discharge separately. Specifically, we define

$$\text{Widespread Hazard} = \frac{\text{Number of events where at least } r\% \text{ of the locations were affected}}{\text{Number of main events where at least 1 location is affected}}$$

This is calculated for the following values of r: $0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$.

3.4 Cluster Analysis

In order to explore and analyse patterns in the prevalence of extreme events, naive k-means clustering were applied to the event matrices (e.g. Steinley, 2006). The idea is to partition the n number of locations into k clusters, based on similarities and differences in occurences of the main events. Using the k-means algorithm requires manually declaring the number of clusters, k. To gain insight into the optimal number of clusters, the cluster gap statistic was used, which is a goodness of clustering measure. The idea is to compute, for each number of clusters to test (k), the within-cluster sum of squares, defined as $\log(W(k))$, and $E^*[\log(W(k))]$, which is the within-cluster sum of squares on simulated data defined via bootstrapping. More specifically, this latter sum of squares term is based on simulating from a reference distribution determined by the ranges of the original matrix (data) after (1) centering and (2) rotation by singular value decomposition. The gap statistic is then computed as $E^*[\log(W(k))] - \log(W(k))$, and the optimal number of clusters are set according to the *firstSEmax* criterion. For this analysis, the *cluster* package in R was used (Maechler et al., 2022), with 10 initial random cluster centers, a maximum cluster cap of 10 and 50 bootstrap iterations for simulating.

All k-means runs on the different matrix compendiums were mainly performed with the optimal number of clusters, as determined by the gap statistic, with some alterations, especially when the algorithm recommended a low optimal number of clusters compared to the number of locations. In short, the clustering algorithm works as follows: (1) For each predefined cluster number k, a random cluster mean (centroid) is set. (2) All observations (i.e. locations) are assigned to the cluster with the shortest squared Euclidean distance between the observation and the centroid. (3) Centroids are recalculated with respect to the observations. Step 2-3 are repeated until the algorithm converges, which occurs when no assignments change cluster. Clustering was performed on all event matrices (i.e., for all extreme event thresholds for all original data sets). Since centroids initially are assigned randomly, the k-means algorithm tend to produce slightly different results each run. In order to make the results reproducible, a fixed seed was set.

3.5 Season-Specific Analysis

While the clustering analysis summarizes prevalence of the same extreme events geographically, it does not directly reveal information about when during a given year extreme events are likely to occur. For this, circular bar graphs called rose diagrams (Brasseur, 2005) summarising the monthly number of extreme events are used. This analysis is performed for all complete event matrices (showing the seasonal prevalence of extreme events across the entire data sets per extreme threshold) and for all locations clustered together individually, i.e. all members of each cluster (showing the seasonal prevalence for locations that are exposed to the same events historically).

4 Results

Here, we present the results of applying the methods described in Section 3 above to daily streamflow data from 38 stream gauging stations in eastern Norway over the time period 1985-2022 and hourly precipitation data from 23 meteorological stations in a region around Oslo. The precipitation data only cover the summer months May-September and each series contains 26207 measurements.

4.1 Event Matrices

In short, the methods described above mainly identify occurrences of extreme streamflow and rainfall events, and simultaneously link such events between different locations in the data set. This forms the foundation for investigating the geographical prevalence, and makes it feasible to assess the widespread hazard, of these events. This information is formatted as event matrices, which connect events across the different locations using binary representation. Since analyses where done independently for streamflow and precipitation measurements, as well as for different extreme value thresholds, this resulted in multiple matrices. Their properties are summarized in Table 3. The matrix dimension is such that the number of rows equals the number of locations considered and the number of columns equals the number of events identified for the given threshold. Note that the matrix for precipitation with extreme threshold at the 0.999 quantile contains only 22 locations.

Table 3. Overview of event matrices created for streamflow and precipitation data using different thresholds for determining extreme discharge and precipitation.

Type	Extreme event threshold (quantile probability)	Matrix dimension
Streamflow	0.90	38×905
Streamflow	0.95	38×669
Streamflow	0.975	38×491
Streamflow	0.99	38×324
Precipitation	0.95	23×167
Precipitation	0.975	23×135
Precipitation	0.99	23×93
Precipitation	0.999	22×20

4.2 Flooding

4.2.1 Widespread Flood Hazard

For each of the four event matrices describing the occurrences of flood events, a metric for widespread hazard is computed for different choices of r. Seeing that r represents the minimal proportion of locations affected (e.g. "at least 60% of the stations are affected"), it essentially defines a threshold for the prevalance of an event: The greater r is, the more locations are affected for an event to be included in the calculations. Thus the resulting proportion represents how often such an event are likely to occur. These calculations are summarized in Table 4.

Table 4. Hazard estimation of widespread flooding in eastern Norway using event matrices for four different extreme thresholds.

	Extreme threshold	Min. proportion of locations affected	Number of events	Number of total events	Widespread hazard
1	0.90	0.20	480	905	0.53
2	0.90	0.30	347	905	0.38
3	0.90	0.40	248	905	0.27
4	0.90	0.50	190	905	0.21
5	0.90	0.60	133	905	0.15
6	0.90	0.70	80	905	0.09
7	0.90	0.80	37	905	0.04
8	0.90	0.90	12	905	0.01
9	0.95	0.20	277	669	0.41
10	0.95	0.30	187	669	0.28
11	0.95	0.40	129	669	0.19
12	0.95	0.50	92	669	0.14
13	0.95	0.60	54	669	0.08
14	0.95	0.70	25	669	0.04
15	0.95	0.80	11	669	0.02
16	0.95	0.90	2	669	0.00
17	0.975	0.20	163	491	0.33
18	0.975	0.30	102	491	0.21
19	0.975	0.40	56	491	0.11
20	0.975	0.50	44	491	0.09
21	0.975	0.60	21	491	0.04
22	0.975	0.70	12	491	0.02
23	0.975	0.80	4	491	0.01
24	0.975	0.90	2	491	0.00
25	0.99	0.20	77	324	0.24
26	0.99	0.30	42	324	0.13
27	0.99	0.40	25	324	0.08
28	0.99	0.50	15	324	0.05
29	0.99	0.60	5	324	0.02
30	0.99	0.70	4	324	0.01
31	0.99	0.80	1	324	0.00
32	0.99	0.90	0	324	0.00

The computed proportions range between 0.0 and 0.53. The lowest proportions appear for high values of r, and the trend is that the proportion generally decreases with increasing r-threshold. This translates to there being fewer events as the criterion for prevalence increases. For example, where the r-threshold is 0.9 and the proportion is 0, there are no events where at least 90% of all locations in the data set were affected at the same

time. Conversely, in the first row of the table, the proportion is 0.53 and the r-threshold is 0.20, indicating that in this instance over 50% of the events affected more than 20% of the locations. These calculations vary between the different matrices, and therefore depend on the extreme threshold (μ). For example, in the first row, the extreme threshold is 0.9. Generally, the proportions also decrease with increasing extreme threshold.

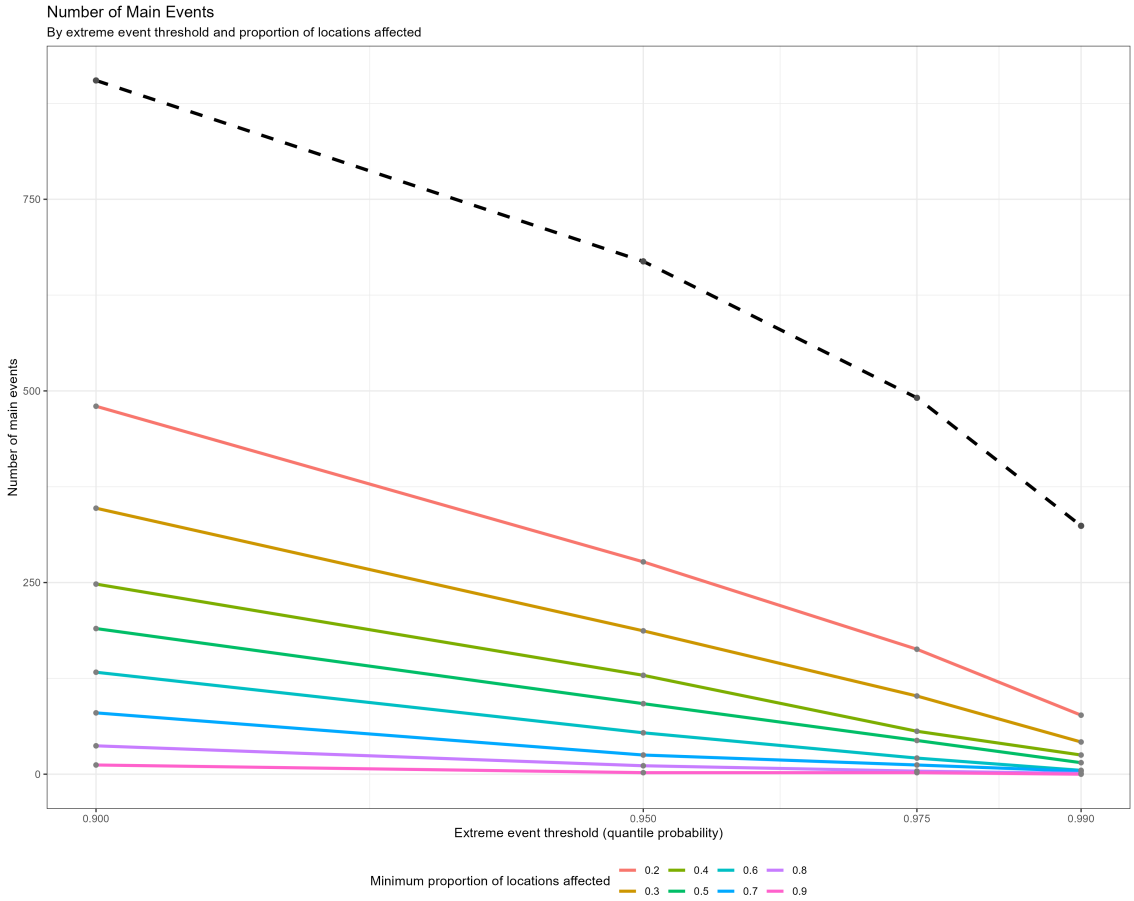


Figure 3. Visual representation of Table 4. Each point corresponds to a certain extreme event threshold (x-axis) and a number of events (y-axis): The black, dashed line corresponds to the total number of events (i.e. the number of events where at least one station is affected), and the coloured, solid lines correspond to the number of events for different choices of the r-threshold.

Figure 3 visualizes the information present in Table 4 as a line diagram where the number of main events for which at least $r\%$ of the locations are affected is described as a function of the extreme event threshold and r. As in the table, this event number decreases with both increasing extreme threshold percentile and increasing r: This can be seen by the scattered lines on the left of the graph which steadily move closer to each other towards the right. The hazard proportions computed in the last column of Table 4 can be found by dividing the y-value of any point with the y-value of the corresponding point (same x-coordinate) on the dashed line.

4.2.2 Spatial Clustering of Floods

The cluster analysis uses the information in the event matrices to describe similarities in flood events across locations. Each cluster represents a group of locations in which

extreme events tend to appear simultaneously. As such, each cluster can be viewed as a individual unit where floods are likely to occur at the same time. In all instances, the number of such units (i.e. clusters) are solely determined by the cluster gap statistic described in Section 3.4.

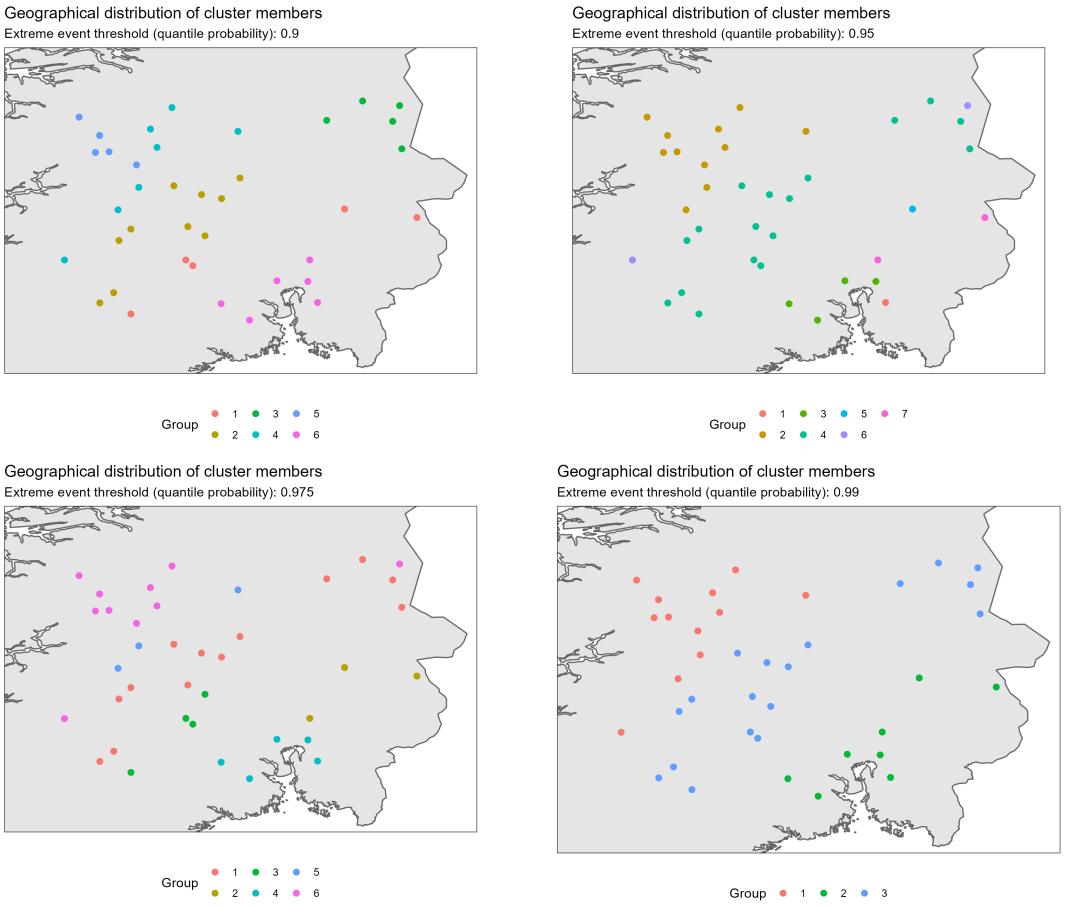


Figure 4. Geographical position of all stream gauge locations, coloured by their cluster membership. Each plots corresponds to clustering performed on one of the four flood event matrices, and therefore they differ based on the choice of extreme event threshold.

Figure 4 visualizes the main result of the cluster analysis, by displaying the cluster members geographically. Generally, there are obvious trends in the plots, with clear spatial groupings of locations belonging to the same clusters. Both these trends and the total number of clusters vary between the four analyses. Note that the numbering of the clusters has no practical meaning, other than for reference. For the event matrix with $\mu = 0.90$, six clear distinct clusters can be observed. Perhaps the most widespread cluster, cluster 1, forms a band of locations stretching along the southernmost parts of the Telemark, Buskerud and Hedmark landscapes. Cluster 2 seems to group most of the locations along the great valleys of Telemark, Buskerud and Oppland, also taking the form of a band crossing different waterways. The other clusters are more spatially cluttered, with cluster 3 grouping locations along the Swedish border in central Telemark, and cluster 4 and 5 grouping locations in the mountainous regions of central Norway. Notably, separation of cluster 4 and 5 seem to correspond well with the east-west mountain plateaus in

central Norway. The final cluster groups the locations along the Oslofjord, also containing locations belonging to separate waterways.

For the event matrix with $\mu = 0.95$, seven clusters can be observed. Clusters 1 and 5 contains only one location, just south-east of Oslo and north-east of Oslo, respectively. Cluster 2 seem to group together more of the mountainous locations in central Norway compared to the corresponding clusters for the matrix with $\mu = 0.90$. Cluster 3 consist of three locations north and south-west of Oslo. The most widespread cluster is cluster 4, which forms a long and wide band of locations from Telemark to the Swedish border. Cluster 6 stands out by containing two stations at two completely different geographical locations, and lastly cluster 7 contains two locations stretching from Oslo east toward the Swedish border.

For the event matrix with $\mu = 0.975$, six clusters can be observed, although these do differ from the clusters resulting from the matrix with $\mu = 0.90$. Cluster 1 forms a band of locations stretching across the east-west direction of the entire region, similar to cluster 4 for the $\mu = 0.95$ clustering, although it contains fewer locations. Cluster 2 groups three locations north-east of Oslo, which previously have belonged to different groups. Cluster 4 contains mostly the same stations as cluster 6 for the $\mu = 0.90$ clustering, centering around the Oslo fjord. Cluster 5 groups three locations along the mountainous region, and cluster 6 groups together most of the stations in the mountainous region, as before.

For the $\mu = 0.99$ clustering, three clear groups are identified. These take the shape of bands stretching from north-east toward south-west, in three north-south layers. Notably, the clustering is clearly not constant between the different matrices. Many of the main grouping patterns are conserved across the different clustering runs, but some locations tend to switch groupings.

4.2.3 Seasonality of Floods

Following the cluster analysis, the monthly distribution of events reveals when flood events occur during the year, both for all locations overall, and for locations grouped together by the clustering. This allows for identifying seasonal geographical differences between occurrences of flood events. The following description focuses on an analysis of the event matrix with $\mu = 0.90$. Corresponding figures for the other matrices can be found in *Appendix A* (available as supplementary material, see Section 6).

Overall, Figure 5 shows that floods are most present during the late spring and early summer months. The total number of flood events decreases throughout the autumn, with very few events happening during winter. However, there are large differences when it comes to the individual groups determined by the clustering. It seems that locations in or near mountainous regions (clusters 2 and 4 especially) tend to experience floods later in the year than those further away from mountains (cluster 1 and 3). More specifically, the first two groups often experience floods during the summer, while the latter groups often experience floods during the spring. An exception is cluster 5, with locations close to the mountains of central Norway which experience flood events at approximately the same time as those more lowland. Cluster 6 stands out by experiencing most flood events in

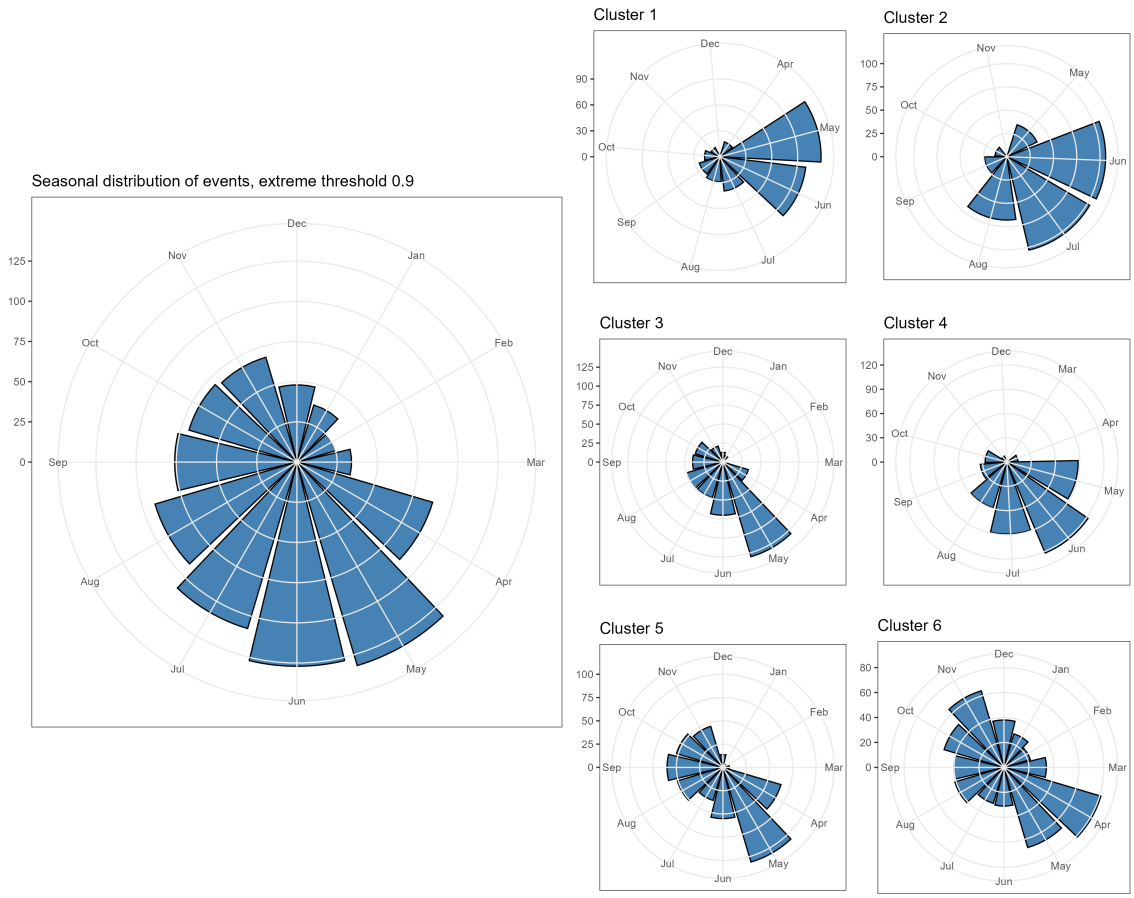


Figure 5. Monthly distribution of flood events for the event matrix with $\mu = 0.9$. The leftmost plot corresponds to the entire event matrix, while each of the six smaller plots correspond to locations in the clusters specified in Figure 4.

the early spring and in the autumn, and almost none during the late spring and summer.

4.3 Extreme precipitation

4.3.1 Widespread Precipitation Hazard

For each of the four event matrices describing the occurences of extreme rainfall events, a metric for widespread hazard is computed for different choices of r, the same way as for flood events. r still represents the minimal proportion of locations affected, and thus defines a threshold for how widespread the events are. The proportion calculations represent how often widespread events defined by r are likely to occur. For precipitation analyses, these calculation are summarised in Table 5.

The computed proportions range between 0.0 and 0.56. As for the flood estimates, the lowest proportions appear for high values of r, and the trend is that the proportion generally decreases with increasing r-thresholds. i.e. there are a lower number of events as the criterion for the proportion of stations affected increases. In the precipitation case, there are fewer extreme events overall and for the two highest thresholds $\mu = 0.99$ and $\mu = 0.999$ many events affect less than 20% of the stations. The highest observed proportion of 0.56 is obtained for an r-threshold of 0.20 and $\mu = 0.95$. In this instance 56% of

Table 5. Hazard estimation of widespread precipitation using the four event matrices for the Oslofjord and Drammensfjord region.

	Extreme threshold	Min. proportion of locations affected	Number of events	Number of total events	Widespread hazard
1	0.95	0.20	93	167	0.56
2	0.95	0.30	75	167	0.45
3	0.95	0.40	54	167	0.32
4	0.95	0.50	46	167	0.28
5	0.95	0.60	40	167	0.24
6	0.95	0.70	30	167	0.18
7	0.95	0.80	21	167	0.13
8	0.95	0.90	14	167	0.08
9	0.975	0.20	61	135	0.45
10	0.975	0.30	50	135	0.37
11	0.975	0.40	36	135	0.27
12	0.975	0.50	29	135	0.21
13	0.975	0.60	18	135	0.13
14	0.975	0.70	12	135	0.09
15	0.975	0.80	10	135	0.07
16	0.975	0.90	5	135	0.04
17	0.99	0.20	36	93	0.39
18	0.99	0.30	25	93	0.27
19	0.99	0.40	13	93	0.14
20	0.99	0.50	10	93	0.11
21	0.99	0.60	4	93	0.04
22	0.99	0.70	0	93	0.00
23	0.99	0.80	0	93	0.00
24	0.99	0.90	0	93	0.00
25	0.999	0.20	2	20	0.10
26	0.999	0.30	1	20	0.05
27	0.999	0.40	0	20	0.00
28	0.999	0.50	0	20	0.00
29	0.999	0.60	0	20	0.00
30	0.999	0.70	0	20	0.00
31	0.999	0.80	0	20	0.00
32	0.999	0.90	0	20	0.00

the events affected more than 20% of the locations. As for the flood estimates, the results vary with μ. Generally, the proportions decrease with increasing extreme threshold μ, indicating that the more extreme events are more localized.

Figure 6 visualizes the information present in Table 5, as a line diagram in which the

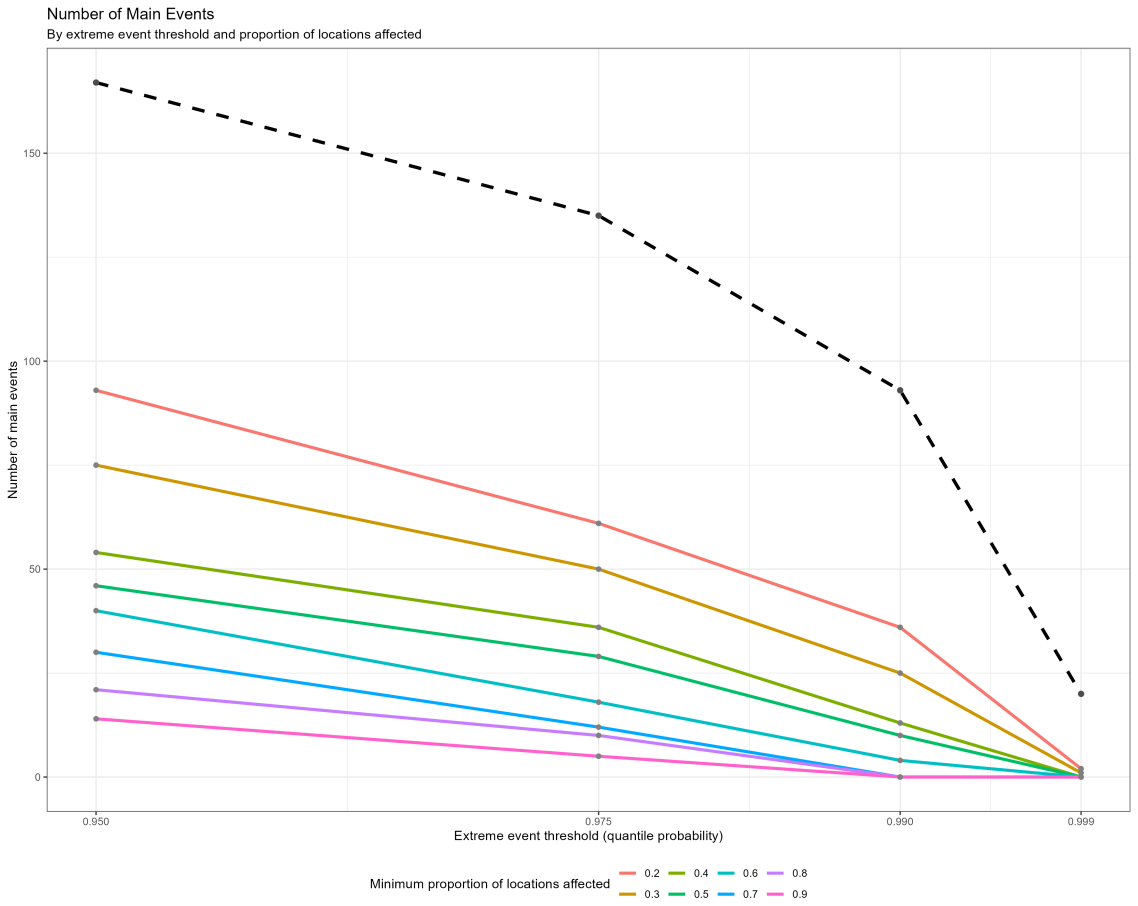


Figure 6. Visual representation of Table 5. Each point corresponds to a certain extreme event threshold (x-axis) and a number of events (y-axis): The black, dashed line corresponds to the total number of events (i.e. the number of events where at least 1 stations is affected), and the coloured, solid lines corresponds to the number of events for different choices of the r-threshold.

number of main events where at least $100r\%$ of the locations are affected is a function of the extreme event threshold and the chosen r. As in the table, this event number decreases with both increasing extreme threshold percentile and increasing r. This can be observed by the scattered lines on the left of the graph which steadily move closer to each other towards the right of the figure. The hazard proportions computed in the last column of Table 5 can be calculated by dividing the y-value of any point with the y-value of the corresponding point along the x-axis on the dashed line.

4.3.2 Spatial clustering of extreme precipitation

As for flood analysis, the cluster analysis on the precipitation matrices compresses the extreme event information to identify similarities in events across locations. Each cluster represents a group of locations in which extreme events tend to appear simultaneously. As such, each cluster can be viewed as an individual unit where the same events tend to occur. In all instances, the number of such units (i.e. clusters) are solely determined by the cluster gap statistic.

Figure 7 visualizes the main result of the clustering, by displaying the geographical dis-

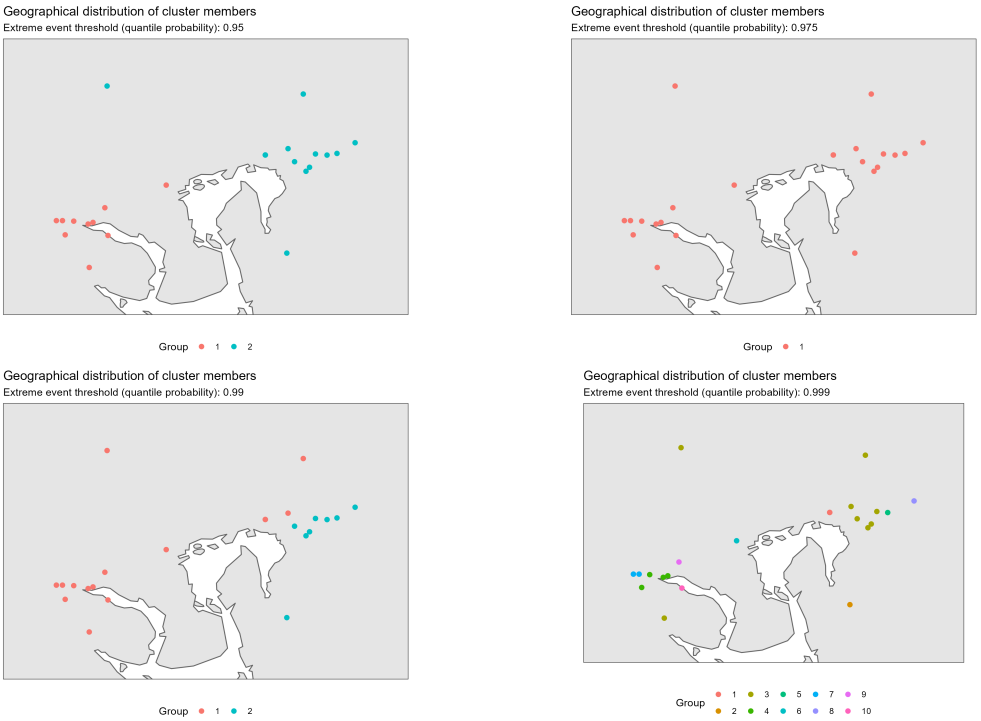


Figure 7. Geographical position of all locations, coloured by their cluster membership. Each figure corresponds to clustering performed on one of the extreme precipitation event matrices, and therefore they differ based on the choice of extreme event threshold μ.

tribution of cluster members. Clear trends can be observed in the figures, with distinct spatial groupings of locations belonging to the same clusters. These trends, as well as the number of clusters as determined by the cluster gap statistic, vary between the four analyses. For the event matrix with $\mu = 0.95$, locations are grouped into two clusters. The spatial separation between these two groups can be viewed as an imaginary border crossing through the plot (approximately at Nesodden) from north-west to south-east. This results in a west-Oslofjord cluster and an east-Oslofjord cluster, approximately. For the event matrix with $\mu = 0.975$, only one cluster is present, containing all stations.

For $\mu = 0.99$, the locations are again grouped into two clusters, but not with the same composition as for $\mu = 0.95$. Now, the separation can be viewed by drawing an imaginary border through the plot (approximately at Nesodden) from north-east to south-east. This results in a western group consisting of all locations from Oslo to the Drammen region, and an eastern group consisting of locations east of Oslo.

For the event matrix with $\mu = 0.999$, 10 clusters are created. Here, one of the stations east of Oslo is omitted from the figure, since no extreme events at this probability level was identified at this station. Further, the clustering is here based on only 20 events in total, with only two events affecting at least 20% of the stations (see Table 5). Seven of the clusters contain single locations (clusters 1, 2, 5, 6, 8, 9, 10), while the three remaining units still group together two or more locations. Clusters 4 and 7 groups together three and two locations, respectively, close to Drammen. More specifically, the members of cluster 4 seems to be located more close to the outlet of the Drammen river than cluster 7.

Lastly, cluster 3 consists of 8 locations scattered geographically, ranging from the eastern parts of Oslo to the southern parts of Drammen.

As for the flood analysis, we find that the clustering is notably not conserved across the different matrices, even though some main patterns can be observed across all clustering runs.

4.3.3 Temporal distribution of extreme precipitation

Following the clustering, the weekly distribution of events reveals when extreme rainfall events occur during the expanded summer months, both for all locations overall, and for locations grouped together by clustering separately. This allows for identifying weekly geographical differences between occurrences of extreme precipitation events. The following description focuses on analysis of the 0.99 event matrix. Corresponding figures for the other matrices are located in *Appendix B* (available as supplementary material, see Section 6).

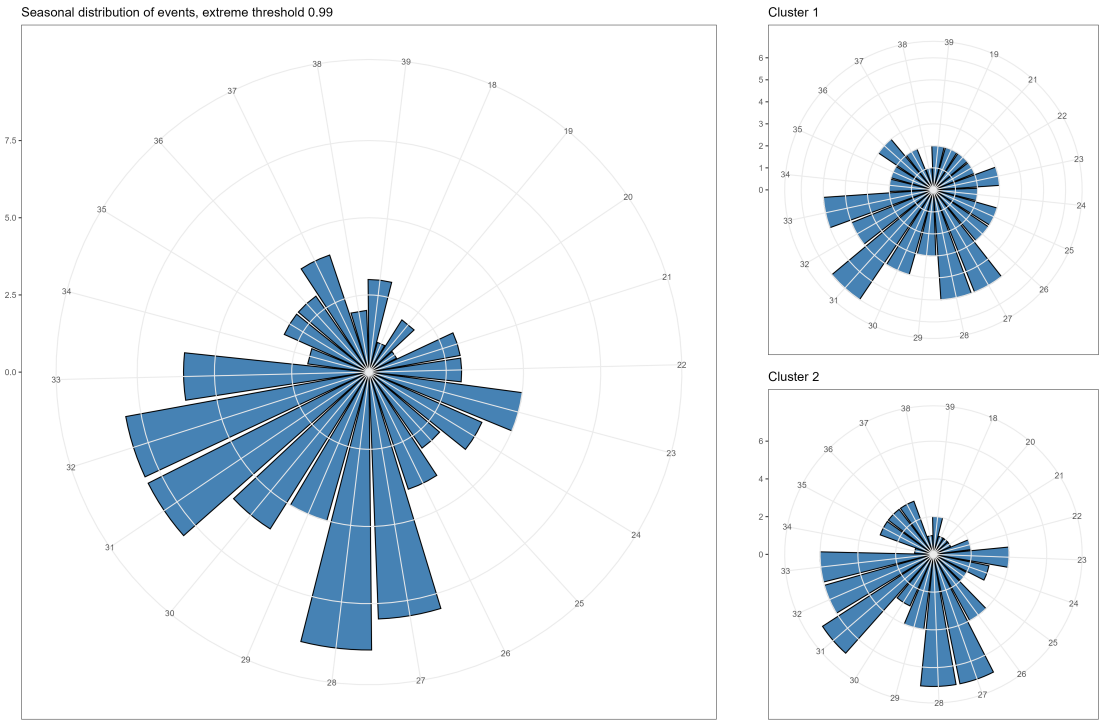


Figure 8. Weekly distribution of flood events for the event matrix with $\mu = 0.99$. The left plot corresponds to the entire matrix overall, while the two smaller right plots correspond to locations in the clusters specified in Figure 7.

Figure 8 shows that most extreme rainfall events occur during mid-summer, more specifically in July and August, with a lower frequency late July and early August. The number of events decreases continuously as the time interval progresses to the ranges of the data (i.e. the start of May and the end of September). However, there are events occurring in all weeks. This pattern is also present when investigating the individual clusters, and there does not seem to be any significant differences in the weekly distribution between the two clusters.

5 Discussion

5.1 Methodology

The procedure for developing the event matrices proved to produce a realistic picture of historical events: The most widespread events identified agreed well with historical records in media and literature. Seeing that the most serious and prevalent events were reflected well in the resulting matrices was a confirmation of the soundness of the analysis performed.

However, the information contained in these matrices is heavily dependent on the choice of parameters for thresholds during the matrix development pipeline. Firstly, there are many ways to define an extreme event, and using percentiles has both advantages and disadvantages. Percentiles efficiently identify peak time points of high precipitation and discharge mathematically, but fail to capture the real-life nuances of what an extreme actually is: It depends on a variety of factors such as topography and infrastructure. For practical purposes, an extreme is often determined by the consequences of the event, not the event itself. This is the main reason for performing the analysis for different extreme thresholds: The higher the threshold, the more likely it is that the events captured actually reflect serious incidents.

The next problematic step is to separate local events into larger main events which cover greater areas. Firstly, the length of extreme events varies. A solution to this is to not only have a threshold for when the event starts, but also when the event ends. However, this complicates connecting events and locations. Secondly, the use of fixed time point thresholds for separating events (such as the 7-day threshold for discharge) proved to be problematic due to varying length of floods. Using a fixed threshold runs the risk of including the same event in the matrices multiple times, which may affect the downstream analyses. How extreme events spread in space also depends on a number of hydrological and meteorological factors. The final time point threshold used to connect local extremes with the main events tries to compensate for such factors (for example different lags based on catchment sizes), but these are also constant. Generally, using such constants does not necessarily generalise the actual conditions well across all locations, which may lead to biased results. To cope with this, thresholds implemented in this study were mainly based on literature and theory.

Another task that proved challenging was pre-processing the raw data to create data sets with time-overlapping measurements suitable for analysis. The methods implemented completed the task successfully, but were not optimised in any way. Some of the subsetting steps were based off visual inspection of the data. This implies that the information gain in the data are probably not maximized, and some good-quality data may have been removed from the data set prior to the analysis.

5.2 Quantifying Widespread Risk

The hazard estimates (proportions) aims to quantify the risk of prevalence of the extreme events. It is questionable that the widespread hazard proportion decreases as the

extreme threshold increases, as it is expected that the most extreme events would affect larger areas, in particular for flooding. However, seeing that the number of total events decreases, it is possible that this might be explained by extreme local events. Also, many flood events probably won't surpass the highest extreme thresholds, which leads to a decreasing number of events present in the matrices. It is important to keep in mind that events present in the matrices for higher thresholds are often also present in the matrices for smaller thresholds (for example, an event present in the 0.99 matrix is almost certainly also present in the 0.9 matrix). It is logical that the number of events decreases when the prevalence criterion (r) increases, since this specifies the minimum proportion of locations affected. In general, by increasing these values, the data captured by the matrices becomes more focused on the most serious events, by eliminating the less extreme occurrences, both in terms of extreme values (i.e. precipitation and discharge values), but also the "widespreadness". The relatively large amount of low proportions in the last columns of Table 4 and 5 indicate that extreme events seldom affect entire country regions (like eastern Norway) overall.

5.3 Geographical and Seasonal Prevalence of Extreme Events

Seeing that even though the unsupervised k-means clustering algorithm only takes the event matrices as input, i.e. binary data without any information about the location of the measurement stations, the locations are clustered very well geographically. This indicates that the method captures real-life patterns in the occurrences of extreme events, which may be explained by geographical, hydrological and meteorological factors.

Not surprisingly, locations in the same catchments are likely to experience the same flood events, of course with different intensity and with a time lag as water progresses downstream. However, the clustering analysis displays groupings crossing the borders of different catchments, indicating that other factors than solely catchment belonging must be the reason for the trends captured. As noted, many clusters seem to take the shape of bands. These bands often seem to correspond well with the different "stages" in multiple catchments, from their highest points toward their outlets. One plausible theory is that these stages experience the same flood events because of influence from the same meteorological phenomena, for example weather systems moving across the different catchments, or the amounts of snow in the areas.

Another possibility is that cluster members are identified as having floods during the same period solely because they experience floods at the same timepoints: Floods start at high altitude in the catchments, and progresses at similar rates toward the outlets. Most likely, a combination of both these theories can explain these bands. Also, it is not guaranteed that a flood progresses through an entire catchment. Either way, it seems that topography plays a vital role in the prevalence of floods. The melting of snow during the spring and early summer months plays a vital role in the development of floods. It is therefore highly likely that locations at similar altitudes experience flood events at the same times, as both the amount of snow and the rate of the melting are similar. This is supported by the seasonal prevalence analysis: Mountainous locations tend to experience floods later than locations lowland, because of temperatures being lower during the

spring at higher altitudes.

Another interesting aspect is the varying number of clusters, as well as differences in cluster compositions, with some locations switching clusters. One theory is that the different matrices capture different types of flood events. Not all flood events progress through entire catchments, but it is expected that the most extreme events affect larger areas. This might explain the reduced number of clusters for the event matrix with $\mu = 0.99$. It seems logical that the cluster would contain more locations with increasing prevalence, but this is not supported by the information available in Table 4: It seems that the widespread risk decreases when the extreme threshold increases. As theorized earlier, this could be caused by local extremes becoming more visible as the matrices shrink.

Another possibility is that there simply are too few events present in the matrices compared to the number of locations for the clustering algorithm to optimally separate them into smaller clusters. Generally, what caused the flood may explain some of the differences in cluster composition: At high altitudes, floods are likely to be heavily influenced by snow melting, but toward the outlets, extreme precipitation could also cause large floods. Differences between events based on this might explain why some clusters seem to be unstable. This may also explain the many floods occurring during the autumn for cluster 6.

For extreme precipitation, locations lying geographically close to each other seem to experience events at the same time. This is logical, indicating that most events are affected by the same weather system moving gradually over a larger area. This is further supported by the formation of only 1-2 clusters for three matrices. The events in these matrices are likely due to weather conditions affecting the entire city regions of Drammen and Oslo. However, the drastic change to 10 clusters may reveal a different type of extreme precipitation. Here, the clusters are more local, in a sense that they are delimited to smaller areas. This clustering is performed based on $\mu = 0.999$, which encompasses only the most extreme events. Remembering that these events are based on hourly precipitation, these events are probably extreme torrential rain. Such rainfall does not last very long, and such the effect are delimited to only a small area. As for flooding, the differences in cluster composition are most likely due to events becoming uncovered (more visible) as the extreme thresholds increase and the matrices shrinking.

The temporal distributions of extreme rainfall show that even though some time periods are more exposed to extreme rainfall, there are no systematic differences between groups present, as observed for the flood analysis. This indicates that torrential rain are, in a sense, random.

5.4 Further Analyses and Concluding Remarks

This study has barely touched the surface of the possibilities for investigating the widespread risk of extreme precipitation and flooding. It would be interesting to extend the analysis to a network analysis based on the methods in Brunner and Gilleland (2021). The idea here is to connect locations together by deducting a regulatory network based on the occurrences of extreme events. Developing the procedures and methods for identi-

fying and separating extreme events is also useful. For example, it is possible to decide on specific start- and end-points of an event using methods by Cunnane (1979) .

While this analysis has mainly focused on generalising and summarising the prevalence for larger regions, but is also interesting to investigate widespread risk on a more local level, by looking at smaller regions, e.g., different catchments. It would also be possible to implement more of the analyses proposed by Brunner et al. (2020). Finally, it could be interesting to try to link the precipitation data with the discharge data, in order to infer knowledge about the causality between precipitation and flooding. This would require having data for the same locations, or eventually define regions that encompass both stream gauging stations and rainfall stations.

This study has shed light on the widespread risk of extreme streamflow and extreme precipitation in eastern Norway. The resulting material is promising, and deduces systematic connections between different locations, in-line with meterological and hydrological knowledge. However, further analyses are necessary to infer more robust knowledge.

6 Data and Code Availability

All data and code are available through the *https://github.com/hnordto/WidespreadExtremeWeather* repository on GitHub. *Appendices A* and *B* can also be found here. Pre-computed event matrices can be found in the folder *mat*. All functions for reproducing the methods used in this study can be found in the *R* folder, while experimental scripts are located in the *dev* folder. Lastly, all data sets are found in the *data* folder.

References

Brasseur, L. (2005). Florence Nightingale's visual rhetoric in the rose diagrams. *Technical Communication Quarterly*, 14(2):161–182. 10

Brunner, M. I. and Gilleland, E. (2021). Complex high- and low-flow networks differ in their spatial correlation characteristics, drivers, and changes. *Water Resources Research*, 57(9):e2021WR030049. 23

Brunner, M. I., Papalexiou, S., Clark, M. P., and Gilleland, E. (2020). How probable is widespread flooding in the united states? *Water Resources Research*, 56(10):e2020WR028096. 5, 9, 24

Cunnane, C. (1979). A note on the poisson assumption in partial duration series models. *Water Resources Research*, 15(2):489–494. 24

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2022). *cluster: Cluster Analysis Basics and Extensions*. R package version 2.1.4. Available from: `https://CRAN.R-project.org/package=cluster`. 10

Steinley, D. (2006). K-means clustering: a half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, 59(1):1–34. 10