Multivariate Flood Frequency Analysis in Large River Basins Considering Tributary Impacts and Flood Types

S. Fischer\(^1\) and A. H. Schumann\(^1\)

\(^1\)Institute of Engineering Hydrology and Water Resources Management, Ruhr-University Bochum, Bochum, Germany

Abstract  In contrast to the basic assumption of a homogeneous population underlying common approaches to flood frequency analysis, flood events often arise from different runoff-generating processes. In many large river basins, the diversity of these processes within tributary basins and the superposition of their flood waves increase the complexity of statistical flood modeling. Under these circumstances, the allocation of the most effective flood protection measures requires a spatially explicit analysis of flood-generating processes and the determination of the probability of downstream flood scenarios. For large basins, flood scenarios are often derived using individual historical floods along with model-based simulations. We, instead, performed hydrograph-based flood-type classification and volume-based runoff analyses for the Upper Danube River to estimate the contributions of subbasins to floods at downstream locations. Using this information, we generated long synthetic samples of peak-volume-pairs to apply a multivariate statistical flood-frequency model that yields a conditional probability of a flood peak given the peaks in tributary stations. The results show that only certain combinations of flood types may result in extreme peaks downstream of confluences. They also highlight the need to distinguish runoff-generation mechanisms for the larger floods from ones that drive smaller, more frequent events. Through an example with the Rhine River, we demonstrate how the statistical model can be generalized for complex river networks featuring several tributary confluences. Finally, design floods for different scenarios of flood-type combinations and assigned probabilities are derived, an approach that can be used to possible climate impacts to flood frequency.

1. Introduction

In the last three decades, several large rivers in Europe have been affected by extreme flood events that have led to damages valued at billions of Euros. In Germany, two extreme flood events affecting the lower parts of the Rhine River in 1993 and 1995 were the starting points for a series of extreme floods in different parts of Germany. In 2002, the Elbe River had one of the largest floods in recorded history (Becker & Grünewald, 2003). In 2013, both the Elbe and Danube Rivers were affected by an extreme flood. Though all these extreme events were similar in their relative magnitudes, there were major differences in how they were generated. In the Rhine Basin, both extreme floods took place in winter. The 1993 flood was mainly caused by a long-duration rainfall event, but the flood in 1995 was heavily impacted by snowmelt. Meanwhile, the 2002 and 2013 events in the Elbe and Danube Basins occurred in the summer but differed with regard to their antecedent moisture conditions and the relative impact of convective precipitation. Such differences in flood events based on meteorological or hydrological processes can be considered through the application of flood typologies, which classify flood events based on meteorological or hydrological processes or the shape of hydrographs (e.g., Merz & Blöschl, 2003; Nied et al., 2014; Tarasova et al., 2019). Most of them are applied to micro- (\(<1\) km\(^2\) area of the watershed; Becker, 1995) and mesoscale (1–10,000 km\(^2\)) catchments. Such classifications have been included in flood statistics to better capture the different statistical behavior of flood types by using mixed population approaches and have become more widely used in recent years (Barth et al., 2017; Fischer, 2018; Hirschboeck, 1987; Murphy, 2001) and to estimate type-specific design hydrographs (Brunner et al., 2017). While flood typologies can be applied for mesoscale catchments, it is more difficult to classify events for large river basins where a multitude of hydrometeorological processes contribute to runoff generation (e.g., rain, rain-on-snow, and snowmelt in the Rhine Basin). A possibility to consider such mixtures of the flood-generating process is proposed by Sikorska-Senoner and Seibert (2020). In large basins, flood-generating processes are much more complex, too, and additional factors in these runoff-generating processes have to be considered, including the superposition of flood waves at confluences (Marsalek et al., 2012).
If a flood peak in the main river is superimposed by a simultaneous peak from a tributary, the magnitude of the flood peak may be increased significantly downstream. Conversely, a temporal offset of both flood waves may result in a substantial increase in volume but only a moderate increase of the peak downstream (Blöschl et al., 2013; Skublics et al., 2016).

Recently, numerous studies have investigated the concurrence of flood events in tributaries of large river basins in Germany/Central Europe, including one affected by anthropogenic modifications. Skublics et al. (2014) analyzed the temporal shift of a flood wave in the Danube between Neu-Ulm and Donauwöhr caused by river training, especially natural retention. Other studies in Germany have also demonstrated that human impacts can cause unfavorable superpositions leading to higher peaks (Vorogushyn & Merz, 2013). Geertsema et al. (2018) demonstrated the impact of the tributaries on the Meuse River to be crucial for flood generation. While all these studies have agreed on the importance of flood superposition for large river basins, the extent of a superposition may change for different flood events (Guse et al., 2020).

With regards to the flood frequency analysis in large river basins, we have to consider that the same flood peak in the main river may result from different combinations of runoff from tributaries. As the exceedance probability of the peak is the main characteristic used in flood risk assessments, the heterogeneity of interactions of tributaries complicates the specification of the probability of flood scenarios. For superposition, the shape of the hydrograph plays an especially crucial role. Events with flashy flood hydrographs have a smaller probability of superposition of peaks than long-duration flood events with large volumes. Hence, for large river basins, the joint consideration of a hydrograph-based flood typology and the flood superposition is beneficial, as the complex network structure and the dependence between concurrent flood events of the tributary and the main river requires multivariate statistical models. Copula models can be used to describe such multivariate structures (e.g., Brunner et al., 2017; Chen et al., 2012; Huang et al., 2018; Kao & Chang, 2012; Peng et al., 2017; Wang et al., 2009). However, in these copula approaches, the flood type is often ignored. Of these studies, only Brunner et al. (2017) develop copulas for specific flood types, who develop type-specific design hydrographs. However, only point-wise estimations are proposed in their study. We extend this idea by taking into account the spatial dependencies and flood superposition for the whole basin under consideration such that also the flood types and hydrographs of upstream catchments are considered.

In this work, a novel approach for design flood scenario estimation at river confluences is proposed. We aim to assess the probability of the simultaneous occurrence of flood events in the tributaries and the main river at the confluence gauge for different flood types. Two different flood typologies are applied: one typology for the characterization of hydrological processes in each tributary and one typology for the classification of the impact of the tributaries on a flood event in the main river. Together, the two typologies form the basis of a statistical simulation procedure, which generates simultaneous peak-volume pairs of flood events at the main river and the tributaries based on their dependence structure and tributary flood types. The generated samples are then used to assess the probability of different scenarios of flood events at confluences. Multivariate statistical models using Vine copulas (Joe et al., 2010) are applied to capture the dendritic structure of the river network. For example, this method can estimate the scenario-probability of a 100-years flood at the confluence gauge given a heavy rainfall flood at the tributary of a given return period and a snowmelt flood at the main river of a given return period. This scenario can be compared with the probability that the same flood peak at the confluence would result from two rainfall-driven floods or combinations of other flood types. To the best of our knowledge, such a combination of two different typologies, one focusing on the runoff generation and one on flood superposition, in a multivariate modeling framework is unprecedented in the literature. It allows the novel construction of flood scenarios that differ in the hydrograph shape due to different tributary flood types and their superposition. The proposed framework extends classical multivariate approaches by adding information on the flood-generation processes.

The developed methodology is demonstrated for the confluence of the Danube River with its Inn River tributary. Flood scenarios with computable probabilities are generated based on different combinations of flood types in the upper Danube and the Inn Basin. We used observed time series to specify the probabilities of peak flood types in tributaries and employ copulas to characterize the probabilities of their superposition. The type-wise construction of the method allows us to understand which drivers have the largest impact on the flood generation downstream. We aim to improve the understanding of the interplay of different flood-generation processes in a large basin. The type-specific estimation of hydrographs in the
flood scenarios allows for a better impact assessment of planned flood retention measures, which could be located in different tributaries or at different locations of the main river, as their performance depends on the relationship between storage capacities and type-specific flood volumes. The superposition of flood waves is also highly relevant for understanding backwater processes and sediment transport. Moreover, we are able to specify impacts of changes of frequencies of several flood types and changing boundary conditions distinguished by their drivers on flood statistics. Therefore, the proposed framework can be the basis for non-stationary modeling by taking into account the changes of individual drivers due to climate change, for example, by modeling the flood-type frequency in a time-variant manner. The framework can be applied to ungauged basins as well.

Since large river basins are considered, two important aspects must be addressed: the variety of catchment characteristics and the heterogeneity of the temporal and spatial distribution of precipitation fields. Especially for large basins that cover more than one climate zone, both impacts can cause a variety of flood generating processes, leading to a mixture of processes downstream and more complex flood-generation scenarios as catchment size increases. These aspects have an impact on the modeling structure. We will present possible extensions to the model that can consider these aspects and will discuss their impact on the results.

More specifically, the following research questions are addressed:

1. How can the impact of a tributary on a flood event at a confluence gauge be characterized?
2. How can flood type and the superposition of floods be included in statistical simulations?
3. What is the most likely combination of flood types from a tributary and main river that leads to an extreme flood downstream?
4. How can the statistical model be generalized to other large river basins with higher network complexity?

2. Data

Two large river basins in Middle Europe are considered in our study: the Danube and Rhine Basins. While a part of the first basin serves as a demonstration case for the methodology due to its rather simple structure (one main river, one tributary), the second basin will be used to demonstrate the suitability of an extended version of our model for more complex river basins. The Danube River is a partially natural river with sectional embankments.

For the Danube basin, we consider the confluence of the tributary Inn River with the main river Danube, close to the German-Austrian border. The database consists of records from three gauges: Passau-Inling (Inn), Hofkirchen (Danube), and Achleiten (Danube), with the Hofkirchen gauge being located upstream and the Achleiten gauge being located downstream of the river confluence point (Figure 1a). The influence of the Alps is much more pronounced for the Inn River Basin than for the upper Danube Basin. Despite the much smaller catchment of the alpine Inn River (55% of the upper Danube Basin at Hofkirchen, Table 1), its proportion of the mean runoff in Achleiten is 52%.

The Rhine River originates in the Swiss Alps, from where it flows through Germany to the Netherlands (Figure 1b). Due to its complex structure, with several large tributaries such as the Moselle and Main Rivers, the formation of flood events is a complex process, changing seasonally as one travels downstream. While in the summer months, the flow regime is dominated by snowmelt and precipitation originating in the Alps, in winter, the precipitation from the uplands (Central Rhine and Moselle) becomes the main driver. The further downstream, the more important becomes the impact of the tributaries, and the distribution of discharge becomes less seasonal (Engel, 1996). The downstream region, especially the Moselle River, has had a major influence on disastrous floods, especially the 1993 flood in Cologne. This is the reason why this study placed special emphasis on this tributary. We examined the impact of the Moselle River at the Andernach/Rhine gauge, which is located just downstream of the confluence of the Moselle River and the Rhine River (Figure 1b). As a reference gauge for the Moselle tributary, the Cochem gauge was chosen. This gauge is located close to the outlet of the Moselle River. To specify the discharge regime of the Rhine River upstream of this confluence, the Kaub gauge was selected. This gauge is situated downstream of the outlet of the tributary Main River into the Rhine. Another reference point is the Worms gauge, which is located between the inlets of the Neckar and Main Rivers into the Rhine. To estimate the impact of the alpine part
of the basin, the Maxau gauge was considered. These five gauges represent the main impact points of the Alpine source region and the tributaries. They were selected due to their location and the availability of long discharge records. The Rhine and its tributaries are affected by several anthropogenic impacts which modified the hydraulic conditions over decades. The most recent one was the installation of barrages for hydropower production in the upper reach upstream from Maxau gauge. It resulted in a change of flood superpositions from waves originating in the upper Rhine with the flood waves from the Neckar River and an increase of the resulting peaks downstream.

Additionally, the Moselle River is widely canalized and dammed since the 1960s. In the Main River, 34 barrages of a total length of 384 km exist, from its confluence with the Rhine River upstream to the mouth of the Regnitz near Bamberg.

The main characteristics for all considered catchments are listed in Table 1.

| Gauge   | River | Catchment size (km²) | Mean elevation (m a.s.l.) | Mean slope of the basin (%) | Mean discharge (m³/s) | Mean annual precipitation (mm) |
|---------|-------|----------------------|---------------------------|-----------------------------|-----------------------|-------------------------------|
| Hofkirchen | Danube | 47,518               | 627                       | 1.77                        | 638                   | 875                           |
| Passau-Ingling | Inn | 26,040               | 1,257                     | 7.67                        | 741                   | 1,142                         |
| Achleiten | Danube | 76,653               | 836                       | 3.76                        | 1,420                 | 963                           |
| Maxau    | Rhine  | 50,196               | 852                       | 5.15                        | 1,136                 | 1,250                         |
| Worms    | Rhine  | 68,827               | 726                       | 4.12                        | 1,045                 | 1,410                         |
| Kaub     | Rhine  | 103,729              | 594                       | 3.16                        | 939                   | 1,650                         |
| Cochem   | Moselle | 27,088              | 343                       | 1.25                        | 865                   | 313                           |
| Andernach | Rhine | 139,549              | 527                       | 2.69                        | 913                   | 2,030                         |

Table 1 Characteristics of the Catchments
Daily series of discharges were obtained from the Global Runoff Data Center (GRDC). To separate flood events, we used a variance threshold equal to the mean of all 3-days variances of daily discharges plus 25% of the standard deviation of all 3-days variances (Fischer et al., 2021). This way, significant increases in the daily discharge values were identified and the corresponding direct runoff was separated for each flood event such that the baseflow components at the beginning and end of it were approximately identical. Flood peaks and flood volumes were derived for each event. Only floods in the joint observation period of all gauges (from 1936 to 2013 for the Danube Basin, respectively, 1931 to 2013 for the Rhine Basin) were considered. Both small and large floods were included in the data set to obtain a large data sample for flood type classification, while small variations in discharge were excluded by the separation algorithm of Fischer et al. (2021) assuming that they are natural variability of flow. For the Danube Basin, we obtained in average 5.7 events per year with a mean duration of 9 days and, for the Rhine Basin, 5.4 events per year with a mean duration of 15 days.

3. Typologies of Flood Events by the Example of the Danube Basin

3.1. Volume-Based Flood Typology to Consider the Spatial Heterogeneity of Runoff Between Tributaries

First, a volume-based flood typology was applied to differentiate flood events by the impact of tributaries on the flood volume in the main river, henceforth referred to as volume-based flood typology. It states whether the majority of the flood volume originates from one part of the river basin only or is a mixture from several (or all) tributaries and the main river. In the first step, for each flood event at the gauge downstream of the confluence, the relation between the direct volume of the corresponding flood event of each tributary and the direct volume at the gauge downstream of the confluence, henceforth the “confluence gauge”, was calculated and divided by the relation between the catchment sizes:

\[
V_{rel,i}(j) = \frac{V_i(j)}{V_C(j)} \cdot \frac{A_{E_{C,i}}}{A_{E_i}}, \quad i = 1, \ldots, n_T, \quad j = 1, \ldots, n
\]

with \(V_i(j)\) being the direct volume of the flood event \(j\) at gauge \(i\), \(A_{E_i}\) being the catchment area for the catchment belonging to gauge \(i\) and \(C\) denoting the confluence gauge. \(n_T\) is the number of tributaries and \(n\) the number of flood events. The direct volume (and later on the direct peak) is here defined as flood event volume (peak) without the inclusion of baseflow.

\(V_{rel,i}(j)\), after normalization to \([0,1]\) by min-max-scaling, were then clustered with the k-MMeans clustering algorithm (Lloyd, 1982) into \(T = 2^n + 1\) volume-based flood types \(T_1, \ldots, T_{2^n+1}\), one for each tributary, the main river, and mixture types, for which the proportion of two or more tributaries is similar. The application of the k-Means clustering to categorize combinations of flood types avoids subjective mixture thresholds and makes the approach applicable to all basins. The number of volume-based flood types can be reduced by hydrological knowledge on the basin under consideration, for example, by identifying improbable combinations for the mixture types. For the Danube basin, we consider the three volume-based flood types “Inn,” “Danube” and “Mixtype.” Also, if a flood event at the confluence gauge was not identified at both gauges upstream because it was below the threshold in the flood event separation procedure at one site, this event is classified as belonging to the other site’s flood-volume type. For example, a flood event in Achleiten without a corresponding flood in Hofkirchen is assigned as the volume-based type “Inn”. If a downstream event would occur without a flood event above the threshold at either of the upstream gauges, this event would be excluded. However, this case did not occur in our case study.

The consideration of volume-based flood types is relevant to determine the differences in the peak-volume relationships at the different gauges upstream of the confluence and their impact on flood peaks downstream. To demonstrate this, the 450 analyzed events were classified by their peaks at Achleiten. For each peak type, the proportion of the three volume-based types was estimated (Figure 2). The ratio of the volume-based type “Inn” increases with larger flood peaks whereas flood waves with relatively small peaks mainly belong to the volume-based type “Danube.” This differentiation becomes relevant for the specification of flood scenarios, as discussed below.
The volume-based flood types for the 20 events with the largest flood peaks at the Achleiten confluence gauge are given in Table S1.

### 3.2. Flood Typologies Considering Runoff—Generation Processes and Hydrograph Shapes

In the second step, the hydrograph-based flood typology proposed by Fischer et al. (2019) was applied to obtain five different event types for each tributary gauge. First of all, the impact of snowmelt was estimated for all flood events in autumn, winter, and spring. To estimate the amount of snowmelt, continuous simulations of the snow coverage and the snow water equivalent for each elevation zone with the HBV-light model (Seibert & Vis, 2012) were applied, using long time series of precipitation and temperature data from the German Weather Service. Model performance of the HBV-light was high with NSE values between 0.778 and 0.820 (Danube), respectively, 0.802 and 0.889 (Rhine). Moreover, the goodness of the simulated snowmelt data was checked by the runoff coefficient, which was all below one, indicating a reasonable amount of snowmelt for each flood event. An event was categorized as “snow-impacted” if more than 20% of the total runoff-generating water originated from snowmelt. The threshold of 20% was already used by Fischer et al. (2019) and is a common value for flood typologies (Kampf & Lefsky, 2016). The snowmelt-impacted flood events were clustered using a k-Means algorithm by their amounts of snowmelt and flood-inducing rain and their runoff coefficient into two types:

1. **S1**: rain-on-snow events, where large amounts of rainfall occur
2. **S2**: snowmelt events, where the amount of rain is negligible.

Next, the rainfall-induced flood events (with a proportion of less than 20% of snowmelt) were characterized by their flood timescale, that is, the relation between direct volume and direct peak (Gaál et al., 2012), each of which have distinctly shaped hydrographs. By assuming a linear relationship between direct flood peak and the volume of the direct runoff, the flood events can be classified into a pre-defined number of flood types, in this case, three (R1, R2, R3). The classification is performed by an optimization of the overall coefficient of determination of all three linear regressions through origin between direct peak and direct volume (Figure S2, Fischer et al., 2019). The following types of rain-induced floods were identified:

1. **R1**: floods that originate from relatively short rain events with durations of less than one day to 5 days (areal precipitation for the total catchment upstream of the Achleiten gauge), which are characterized by flashy flood waves with high peaks and small volumes.
2. **R2**: floods that result from longer individual rain events with a duration of 5–10 days with flood volumes that are significantly larger than for type R1, even if the peak discharge is similar in height.
3. **R3**: long-duration flood waves (>10 days), often with several peaks and high volumes.

A list of all flood events at the outlet gauge was obtained with the corresponding floods at tributary and upstream gauges. The event types were estimated for all three gauges. In the Danube basin, the tributary and the main river were equally important for the generation of an event type downstream of the confluence (Text S1). However, their relative impacts differed across flood types (Table 2). While R1 flood events in Achleiten were mostly caused by R1 events in both the tributary and the main river, an R2 flood event most likely occurred due to an R1 event in the Inn tributary and an R2 flood event at Hofkirchen.

---

**Table 2**

| Type Achleiten/Danube | Type Passau/Inn | Type Hofkirchen/Danube | Relative frequency |
|-----------------------|-----------------|------------------------|-------------------|
| R1                    | R1              | R1                     | **0.705**         |
| R1                    | R1              | R2                     | 0.016             |
| R1                    | R1              | R3                     | 0.016             |
| R1                    | R1              | S1                     | 0.016             |
| R1                    | R2              | R1                     | **0.213**         |
| R1                    | R2              | R2                     | 0.016             |
| R1                    | R2              | S1                     | 0.016             |
| R2                    | R1              | R1                     | **0.359**         |
| R2                    | R1              | R2                     | 0.051             |
| R2                    | R2              | R1                     | **0.487**         |
| R2                    | R2              | R2                     | 0.103             |
| R3                    | R2              | R1                     | 0.500             |
| R3                    | R2              | R2                     | 0.500             |
Comparing the flood types upstream of the confluence with the resulting type at Achleiten gauge, it became obvious how the combination of two R1 flood events leading to an R1 flood event downstream (R1|R1R1) was the most frequent combination, but in several cases, the temporal offset of R1-floods from Inn and the upper Danube resulted in an R2 flood at Achleiten. Different combinations of flood types tended to be more prevalent for peaks of different magnitudes. For example, the proportion of (R1|R1R1)-combinations was significantly greater for the largest 10% of flood peaks compared to all other combinations. In this case, large flood events in Achleiten mostly occurred due to a simultaneous flood event in both the tributary and main river meaning that there were storms with precipitation distributed over spatially extended areas.

4. Statistical Methodology for Multivariate Flood Frequency Analysis (MFFA)

In this section, the basic statistical methodology applied to the Danube basin and the general case of one downstream confluence gauge and several tributaries is described. This methodology will make generating peak-volume pairs possible. Moreover, return periods for different scenarios can be estimated with the introduced Vine-copulas. Later in the Discussion, we describe a generalization of the methodology for more complex river basins like the Rhine basin.

We will now introduce basic notation. Assume that we have a confluence gauge C with \( n \) flood events and \( n_t \) tributary gauges \( 1, \ldots, n_t \). For the sake of simplicity, in the following description, the upstream gauge on the main river is termed as a tributary as well. In this manner, general formula for an arbitrary number of tributaries can be proposed. For the most common case with a downstream gauge at the main river, a tributary gauge, and a confluence gauge, \( n_T = 2 \). Below, general formulas for the statistical simulation and the flood frequency analysis at \( n_T \) gauges are given.

4.1. Statistical Simulation of Synthetic Multivariate Samples of Flood Events

The procedure for the statistical simulation of synthetic samples of flood events is developed to generate arbitrarily large samples of peak-volume pairs for the confluence gauge that have the same bivariate distribution as observed events. Moreover, the respective peak-volume pairs of the tributaries are generated. In these samples, the distribution of all peak-volume pairs must be equal to the observed distribution of all floods, and the same is also true for each event type, that is, R1, R2, R3, S1, and S2, as well as each volume-based type. This requires the consideration of the dependence structure between the tributaries and the confluence gauge. The simulation procedure, which generates a sample of \( N \) pairs of peak and volume, consists of the following seven steps. An overview of the procedure is given in Figure 3.

1. Define a vector of volumes for the tributaries \( V_j(1), \ldots, V_{n_T}(j) \) corresponding to the event \( j, j = 1, \ldots, n \), at the confluence gauge. Result: List of flood event volumes of the confluence gauge with the respective volume-based flood type and corresponding flood volumes of the tributaries. Only those events at the confluence gauge are included for which there had been flood events at at least one of the tributary gauges.

2. For each volume-based type at the confluence gauge: A theoretical probability distribution is fit to the time series of all flood volumes for each tributary separately. Each of the volumes derived in step 1 is transformed on a [0,1] scale using the cumulative distribution function. Then, a copula is fit to the pairs of the transformed volumes. Result: Dependence structure between the volumes of the tributaries for each volume-based type.

3. For each volume-based type at the confluence gauge: A sample of random pairs is generated using the copula derived in step 2. Choose the sample size according to the proportion of the respective volume-based type, \( N_T^j = N \cdot p_T^j, i = 1, \ldots, n_T \). Transform these pairs back to \( \mathbb{R}^2 \) by using the respective marginal distribution quantile function for each tributary identified in Step 2. Result: Simulated pairs of volumes for the tributaries for the respective volume-based type.

4. The volume of each event is calculated for the confluence gauge based on the sum of the volumes of the tributaries for each pair of volumes. Result: Simulated volumes for the confluence gauge and corresponding volumes for the tributaries with volume-based type.

[Optional: Bias-correction for each volume-based types if a systematic bias can be observed. For this, the median error between the observed and simulated quantiles of the volumes, \( q_{\text{lab}} \) and \( q_{\text{sim}} \), can be used.]
1. Specification of event volume pairs
\[
\left( Vol_i(j), \ldots, Vol_{n_f}(j) \right| Vol_c(j) \right)_{j=1, \ldots, n}
\]

2. Selection of copula model (type-specific)
\[
\forall t = T_1, \ldots, T_{n_t+1} : \quad C_t \left( F_t \left( Vol_{1,t} ; \theta_{1,t} \right), \ldots, F_t \left( Vol_{n_f,t} ; \theta_{n_f,t} \right) \right)
\]

3. Simulation of upstream flood volumes (type-specific)
\[
\forall t = T_1, \ldots, T_{n_t+1} : \quad \left( x_{t,j}(j), \ldots, x_{n_f,j}(j) \right| C_t \right)_{j=1, \ldots, N_t}
\]
\[
\text{simVol}_{t,j}(j), \ldots, \text{simVol}_{n_f,j}(j) \right)_{j=1, \ldots, N_t} = \left( F_t^{-1}(x_{t,j}(j); \theta_{t,j}), \ldots, F_t^{-1}(x_{n_f,j}(j); \theta_{n_f,j}) \right)_{j=1, \ldots, N_t}
\]

4. Simulation of downstream flood volumes (type-specific)
\[
\forall t = T_1, \ldots, T_{n_t+1} : \quad \left( \text{simVol}_{t+1,j}(j) \right)_{j=1, \ldots, N_t} = \left( \sum_{i=1}^{n_f} \text{simVol}_{t,i}(j) \right)_{j=1, \ldots, N_t}
\]

Optional: Bias correction
\[
\forall t = T_1, \ldots, T_{n_t+1} : \quad \text{simVol} = \frac{\text{simVol}}{\text{Bias}}, \quad \text{with Bias = median} \left( \frac{q_{vol_1}}{q_{vol_2}} \right)
\]

5. Estimation of flood type probability (diff. quantile ranges)
For \( i=0,1,0,2,\ldots,1 \), \( \forall t = T_1, \ldots, T_{n_t+1}, \forall k=1, \ldots, n_r, r=R1, R2, R3, S1, S2 \)
\[
q_{i,t} = q_{i,r}(\text{obsVol}_{t,k}), \quad q_{i,r} = q_{i,r}(\text{obsVol}_{t,k})
\]
\[
P_{i,t,r} = P\left( \text{Type(\text{obsVol}_{t,k})} = r \mid q_{i,t} \leq \text{obsVol}_{t,k} < q_{i+1,t} \right)
\]
\[
q_{i,r,t} = q_{i,r}(\text{simVol}_{t,k}), \quad q_{i+1,r,t} = q_{i+1,r}(\text{simVol}_{t,k})
\]
\[
P\left( \text{Type(\text{simVol}_{t,k})} = r \mid q_{i,t} \leq \text{simVol}_{t,k} < q_{i+1,t} \right) = p_{i,t,r}
\]

6. Assignment of flood type probability to simulated events
For \( i=0,1,0,2,\ldots,1 \), \( \forall t = T_1, \ldots, T_{n_t+1}, r=R1, R2, R3, S1, S2 \)
\[
q_{i,t} = q_{i,r}(\text{obsVol}_{t,k}), \quad q_{i,r} = q_{i,r}(\text{obsVol}_{t,k})
\]
\[
P_{i,t,r} = P\left( \text{Type(\text{obsVol}_{t,k})} = r \mid \text{Type(\text{obsVol}_{t,k})}, q_{i,t} \leq \text{obsVol}_{t,k} < q_{i+1,t} \right)
\]
\[
q_{i,r,t} = q_{i,r}(\text{simVol}_{t,k}), \quad q_{i+1,r,t} = q_{i+1,r}(\text{simVol}_{t,k})
\]
\[
P\left( \text{Type(\text{simVol}_{t,k})} = r \mid \text{Type(\text{simVol}_{t,k})}, q_{i,t} \leq \text{simVol}_{t,k} < q_{i+1,t} \right) = p_{i,t,r}
\]

7. Estimation of flood peak by peak-volume relationship
For \( r=R1, R2, R3, S1, S2 \)
\[
\text{Peak}_{t+1,j}(j), j=1, \ldots, N_t = (\text{relTQ} \cdot \text{simVol}_{t+1,j}(j))_{j=1, \ldots, N_t}
\]

Figure 3. Schematic overview of the statistical simulation.
1. For each 10%-quantile range of the observed volumes and each volume-based type at the confluence gauge: The probability of occurrence of each event type, R1, R2, R3, S1, and S2, for each tributary, is estimated, and this probability is transferred to the simulated volumes with the respective quantile ranges and volume-based type of the simulated volume pairs. The subdivision of the magnitudes of the volumes is required since the flood type frequency changes with increasing volumes. Result: Event types for the tributaries for the simulated volumes.

2. For each 10%-quantile range of the observed volumes and each volume-based type at the confluence gauge: The probability of occurrence of each event type, R1, R2, R3, S1, and S2, at the confluence gauge, is estimated based on the event types of the tributaries, and this probability is transmitted to the volumes with the respective quantile ranges and volume-based type of the simulated volume pairs. Result: Event types for the confluence gauge for the simulated volumes based on the flood type of the tributaries.

3. For each event type R1, R2, R3, S1, and S2: For each simulated volume at the confluence gauge, the corresponding peak is estimated by linear regression through the origin between the observed peak and volume as described by Fischer et al. (2016), see also Figure S2. The same method can be applied to the simulated volumes of the tributaries. Result: Simulated pairs of peak and volume with corresponding event type and volume-based type for the confluence gauge, as well as the respective flood events for the tributaries.

In step 2, the marginal distribution was selected according to the Anderson-Darling test and the AIC criterion. The AIC criterion was also used to obtain the best-fitting copula model for the previously selected marginal distributions. Moreover, a goodness-of-fit test was applied based on the information matrix equality proposed by White (1982). A p-value below 0.05 implied a rejection of the hypothesis of a copula model fit. This way, both marginal distributions and dependence models, were tested for their goodness-of-fit. This is especially important for right-skewed data, such as flood peaks and volumes, as a proper fit can lead to serious estimation bias if the wrong model is applied in either margins or copula (Fantazzini, 2009). Bias-correction can be helpful in the case where the assumption that the volume at the downstream gauge is equal to the sum of the volumes of the tributaries is not fully met. This could be the case when the gauges of the tributaries and the downstream gauge are not located close together and the runoff between the tributary gauges and the confluence gauge is high. The linear regression in step 7 is straightforward since the flood types, at least the rainfall-induced floods, were classified according to this criterion, meaning that a linear relationship was already assumed for the flood typology. For snowmelt-induced floods or catchments with non-uniform impact of snowmelt, the linear relationship may not hold and a non-linear relation can be linearized by transformation (see for example Text S4). The choice of the model can be supported by goodness-of-fit criteria such as RMSE.

To evaluate the errors of the statistical simulation, the normalized trimmed mean squared error (NTMSE, see, e.g., Kawashima & Fujisawa, 2017; Yang et al., 2018) values were computed. Here, the trimmed version of the mean with \( \alpha = 5\% \) trimming degree was used to avoid having the overall error be disproportionately affected by one incorrectly estimated extreme quantile. NTMSE is given by

\[
NTMSE = \frac{1}{n(1 - 2\alpha)} \sum_{i=\lceil n\alpha \rceil}^{n-\lceil n\alpha \rceil} \left( \frac{q_{\text{sim},i} - q_{\text{obs},i}}{q_{\text{obs},i}} \right)^2,
\]

where \( \alpha \) is the trimming degree and \( \lceil n\alpha \rceil \) denotes the floor function applied to \( n\alpha \). Equation 1 can be applied to the observed and simulated quantiles of either flood peaks or volumes. The NTMSE is derived for each event type separately. As a comparison, also the NTMSE without trimming, that is, \( \alpha = 0\% \), can be used.

4.2. Multivariate Flood Type-Specific Return Periods Considering Confluences

For the estimation of return periods conditional upon flood event types and volume-based types and the tributary flooding, the dependence structure between the tributary, the upstream main river and the confluence gauge must be estimated first. The aim is to obtain a return period for a flood of a given type at the confluence gauge given the return periods and the flood types at the tributaries and the main river. This allows a more flexible estimation of return periods that takes into account the flood type in individual rivers along with the confluence of floods. Since a large amount of data is required to achieve reliable estimates
using this multivariate model, the statistical simulation proposed above is used to generate a sufficiently large data sample. Errors from the simulation are small, as shown in the results section, such that error propagation from the simulated samples to the estimated quantiles can be assumed to be small as well. Then, for each combination of flood types, a Vine-copula approach is applied. Vine-copulas are a tool used to simplify multivariate copulas by only considering bivariate pairs in a dendritic structure conducive to river network modeling. Here, a \((n_T + 1)\) D-Vine is used. Below, the methodology will be described for the special case of \(n_T = 2\) because, for this case, easy-to-handle analytical formulas still exist. Theoretically, the principle is applicable for any value of \(n_T\). Vine-copulas have been applied widely in hydrology for the multivariate modeling of, for example, triples of peak, volume, and duration of flood events at a given site, of several flood peaks measured at neighboring gauges, or for drought frequency analysis (Gröbler et al., 2013; Kao & Govindaraju, 2008; Song & Singh, 2010). Consider the \([0,1]\)-transformed peaks of the tributary \(U\), the main river \(V\) and the confluence gauge \(W\), with respective marginal distributions \(F_U\), \(F_V\) and \(F_W\). The density of a three-dimensional Vine-copula is given by

\[
c_{UVW}(u,v,w) = c_{UUV}(F_U(u|v), F_W(w|v))c_{UV}(u,v)c_{VW}(v,w),
\]

where \(F_{IJ}(i|j)\) is the conditional distribution function of \(I\) given \(J\). With this, conditional probabilities can be estimated. Here, the conditional probability of a flood peak at the confluence gauge given the tributary and the main river flood peak is considered, as given by

\[
P(W \leq w | U = u, V = v) = \frac{\frac{\partial^2}{\partial u \partial v} c_{UVW}(u,v,w)}{\frac{\partial^2}{\partial u \partial v} C_{UV}(u,v)}.
\] (2)

The copulas and their respective parameters are estimated with the R-package VC2copula (Nagler, 2019), where the best fitting model is selected by the AIC and tested with an information matrix equality goodness-of-fit test (see Section 4.1).

As input for the conditional probability in Equation 1, quantiles for given return periods and flood types are estimated for each tributary and the main river with the type-based statistics proposed by Fischer (2018). More precisely, for each flood type, the distribution can be derived with a peak-over-threshold approach, such that the annual distribution function describing the probability in terms of one year is given by

\[
F_t(x) = \sum_{l=0}^{\infty} P(l = k)(F_{POT_2}(x))^k,
\]

where \(P(l = k)\) is the probability that the annual number of exceedances of the threshold \(u_t\) is equal to \(k\) and \(F_{POT_2}\) is the distribution of all flood peaks of type \(t\) above the threshold \(u_t\). The latter is modeled by a Generalized Pareto distribution, and the parameters are estimated with L-moments. For the distribution of the number of exceedances per year, the Poisson distribution is chosen (see Fischer & Schumann, 2016). The threshold \(u_t\) is chosen type-specifically as three times the weighted mean discharge of the respective tributary or the main river. The weights for the mean monthly discharges were chosen according to the relative frequency of the flood type in the respective month. This choice was justified in Fischer (2018) and follows German directives, where a threshold of three times the mean discharge is chosen to define a flood event (DWA, 2012).

For each combination of flood types in the tributaries, main river, and confluence gauge, scenarios with different return periods for the peaks are considered. Return periods are varied between \(T = 2, 5, 10, 20, 25, 50, 100, 200, 500,\) and 1,000 years. Importantly, return periods larger than 100 years are subject to high uncertainty since the simulated sample is based on a series of less than a hundred years. However, they will be presented nevertheless to perform an investigation of the right tail to understand which combinations of tributary event types are most likely to result in extreme floods at the confluence gauge. The respective return periods for the confluence gauge are derived by optimization of the probability given in Equation 2 and the application of the annual distribution.
5. Results for the Danube River Basin

The methodology proposed in Section 4 was applied to the Passau-Inging tributary gauge and the Hofkirchen main river gauge, as well as the Achleiten confluence gauge.

5.1. Statistical Simulation of Synthetic Multivariate Samples of Flood Events

In the first step, 434 distinct flood events at the Achleiten confluence gauge were identified, an average of 5.7 per year. The distribution of the volume-based types is rather uniform: 111 events are of volume-based type Inn, 137 events are of type Danube, and 146 events are of Mixtype.

In the second step, the marginal distributions of the direct volumes of the tributary and the main river, corresponding to the 434 events in Achleiten, differentiated by their volume-based types, were fitted. Several distributions, namely, generalized extreme value distribution (GEV), Gumbel, Normal and (logarithmic) Pearson-III distributions, were considered. For all volume-based types and gauges, the GEV distribution yielded the lowest AIC values without any rejections by the Anderson-Darling test.

Third, bivariate copulas for the transformed direct volumes of Passau-Inging and Hofkirchen had to be fitted. Again, this procedure was performed for each volume-based type separately. The results can be found in Figure S12 and Table 3 together with p-values derived from the goodness-of-fit test with null hypothesis $H_0 : H(\theta) + C(\theta) = 0$, meaning that the Fisher Information can be equivalently calculated as minus the expected Hessian matrix $H$ or as the expected outer product of the score function $C$ which indicates a correct model specification. A p-value below 0.05 in this case would imply rejection of this hypothesis.

It is not surprising that the correlation between the volumes of the tributary and the main river is highest for the Mixtype. For this type, both rivers responded similarly to one precipitation event, and the proportions of the flood volumes of the tributaries relative to the volume of Achleiten were similarly large. Interestingly, for the type Inn, the correlation was much smaller than for the type Danube. As expected, floods that mainly occurred in the Inn were often independent of floods in the upper Danube. Partly, this can be related to the higher number of snowmelt floods in the Inn basin.

The copulas were applied to simulate a large sample ($N = 100,000$) of direct runoff volumes (step 4). A bias correction was not required since differences between observed and simulated volume in terms of NTMSE were below 5%. In step 5, the frequencies of event types R1, R2, R3, S1, and S2 in each volume class (all volumes were subdivided into 10 classes by size) were estimated. These empirical frequencies were used to assign the simulated volumes to classes of flood types at Achleiten (step 6). In Figure 4, Q-Q plots are used to compare the observed distribution of the direct volumes with the distribution of the simulated direct volumes for different flood types. It must be considered that the sample of flood events of each type was rather small. Hence, uncertainty due to sample size can occur. To take this into account, samples of the same size as the observed ones were drawn from the simulated sample. This was repeated 1,000 times, and the mean and the 95% confidence bands were estimated and included in the Q-Q plots. Since the most extreme events seldomly occur during the observation period, the right tail shows comparably larger confidence bands. This is the range where statistical estimation is most difficult and results have to be treated with caution. Due to the small sample sizes of flood types S1 and S2 in the observed sample at Achleiten, the results are not given here due to their high uncertainty.

Finally, the type-specific relationships between flood peaks and volumes were used to specify corresponding direct runoff peaks from the direct volumes for each event type. The respective regression coefficients for the regression $Q_i = \alpha_i V_i$, $i = R1, R2, R3$, at Achleiten are $\alpha_1 = 3.334$, $\alpha_2 = 2.146$, and $\alpha_3 = 1.182$. Analogously to the simulated volumes, the simulated peaks were again compared in a Q-Q plot with 95% confidence intervals (Figure 5).
5.2. Application of Multivariate Return Periods to Specify Design Floods at the Outlet Gauging Station

It was shown before that different combinations of event types in the tributaries resulted in floods at Achleiten. Using the terminology “event type Achleiten| corresponding type Passau-Ingling corresponding type Hofkirchen”, the five most frequent combinations were an R1 type flood in Achleiten caused by an R1 type flood in Passau-Ingling and a R1 type flood in Hofkirchen (R1|R1R1), an R1 type flood in Achleiten caused by an R2 type flood in Passau-Ingling and an R1 type flood in Hofkirchen (R1|R2R1), as well as (R2|R1R1), (R2|R2R1), and (R2|R2R2). However, not all combinations were equally likely for the upper tail. The combination responsible for the largest peaks was (R1|R1R1). We will discuss this in more detail below.

Due to these results, in the following discussion, only these five combinations are considered since they are the most relevant for flood statistics. Similar results could be derived for all other combinations. For the most frequent combination (R1|R1R1), which was also responsible for the largest peaks in Achleiten, the respective return periods of a peak conditional on peaks of a given return period in the tributary and the main river are summarized in Table 4. Please note that all models are based on the stationary assumption. There had been few anthropogenic impacts in the last century in the basin and we did not found evidence for non-stationarity in the samples (by application of Mann-Kendall trend tests and Wilcoxon-test) such that we can assume this assumption to be valid. For some combinations of return periods, only small probabilities can be obtained. This is the case for long return periods at the Passau-Ingling gauge but short return periods for the Hofkirchen gauge, and vice versa. Hence, high rainfall-driven peaks in the Inn tributary rarely concur in combination with small peaks in the upper Danube. This implies that a large R1 flood peak

Figure 4. QQ-plots of observed and simulated direct volume at the Achleiten confluence gauge and 95%-confidence bands for the event types (a) R1, (b) R2, and (c) R3.
at Achleiten cannot arise from the upper Danube or Inn alone, at least for the type combination of two R1 flood events. In other words, this means that large R1 floods at Achleiten are generated with large synoptic-scale storms which cover both basins simultaneously. For example, the combination of an R1 event with a 200-years return period at Paussau-Ingling and an R1 event with a 2-years return period at Hofkirchen has a return period of 10 years. However, the overall probability of such an event considering all flood types is smaller than 0.8.

To compare all possible combinations of event types in the tributaries, all combinations of peaks with a given return period in the tributary and the main river that would result in a flood peak with a return period of 100 years at the Achleiten gauge for the different type-combinations were estimated. This takes into account uncertainty in the estimation. Since only discrete combinations of return periods from upstream gauges were considered for computing the return period at Achleiten to reduce computational time, we considered all events with return periods of 95 up to 105 years to be 100-years events. The results are given in Figure 6a. By estimating the empirical probability of each combination, more detailed information on the scenarios can be derived. Exemplary selected results are given in Table 5, where the probability of each scenario, selected combinations of return periods for peaks at Passau-Ingling and Hofkirchen that result in a 100-years flood (±5 years) downstream in Achleiten, and the respective probability for each volume-based flood type are given.
Table 5 demonstrates that pairs of tributary floods with a wide range of return periods can produce 100-years floods at Achleiten. To show this, three combinations of return periods in Passau-Ingling and Hofkirchen were selected, one with large return period in Passau-Ingling and a small return period in Hofkirchen, one with small return period in Passau-Ingling and large return period in Hofkirchen, and one in which the return periods were as equal as possible. For these return periods, the peak was then estimated with the type-specific distribution function (see Section 3), and the volume was subsequently derived with the peak-volume relation. The same procedure was performed for return periods of 20 and 500 years in Achleiten, and the results are given in the online appendix (Tables S7 and S8).

With these results, a selection of design floods is possible. For example, for a 100-years flood in Achleiten, different scenarios and the respective flood waves at the different gauges can be compared. In Figure 6b, the most probable combinations of flood types in the tributary and main river that lead to a 100-years flood in Achleiten are shown. For each of these combinations, the respective flood waves at the confluence gauge were estimated separately for each volume-based type. Yet, not all combinations have a probability of occurrence greater than zero. The flood hydrographs were obtained by selecting the flood events of a given combination and volume-based flood type from all observed events that had the smallest deviation from the estimated peaks and volumes for the 100-years flood given in Table 5 at all gauges. These events were then rescaled to the estimated peaks and volumes by using the relation between the estimated quantile and the observed maximum discharge. In this way, a meaningful shape of the flood hydrograph at the target gauge,

Table 4
Return Periods of a Type R1 Peak at the Achleiten Confluence Gauge Conditional on Type R1 Peaks for Given Return Periods at the Passau-Ingling Tributary Gauge and Hofkirchen Main River Gauge (R1|R1R1)

| T | Passau-Ingling (Inn) | Hofkirchen (Upper Danube) |
|---|----------------------|---------------------------|
|   | 2                    | 2                         |
|   | 5                    | 5                         |
|   | 10                   | 10                        |
|   | 20                   | 20                        |
|   | 25                   | 25                        |
|   | 50                   | 50                        |
|   | 100                  | 100                       |
|   | 200                  | 200                       |
|   | 500                  | 500                       |
|   | 1000                 | 1000                      |

Note. Combinations highlighted in grey indicate a small probability of less than 0.8.

Figure 6. (a) Combination of return periods of peaks in the tributary Inn (Passau-Ingling gauge) and upper Danube (HofKirchen gauge) that cause a peak with return period of 100 years at the Achleiten confluence gauge for different combinations of event types and respective confidence bands. (b) Design floods for different volume-based types and combination of flood types. Numbers indicate the probability of each scenario. For example, R2|R1R1 denotes a flood of type R2 in Achleiten given by a flood of event type R2 in Passau-Ingling and a flood of type R1 in Hofkirchen.
Achleiten, was ensured that mirrored observed flood scenarios. It can serve as a blueprint for estimating the range of hydrographs associated with design floods.

### 6. Discussion of the Results of the Danube Case Study

#### 6.1. Statistical Simulation of Synthetic Multivariate Samples of Flood Events

Overall, the distribution of the simulated peaks and volumes at the confluence gauge concorded well with the observed peaks and volumes according to a visual inspection of the Q-Q plots (Figures 4 and 5). However, distinct differences between the event types could be observed. While the quantiles of the simulated volumes of type R1 fit those of the observed volumes well for the entire spectrum of the distribution, for the R2 type floods, a deviation in the right tail became evident (Figures 4a and 4b). Here, the statistical simulation overestimated the higher quantiles of the distribution. However, the 95% confidence intervals of the estimated quantiles (derived by 1,000 repetitions of bootstrapping), which were used to consider the sampling uncertainty of the small observed samples, still covered the observed quantiles. Therefore, observed values of the right tail of the distribution are still within the range of estimates, and possible deviations can occur due to sampling uncertainty. Still, the right tail estimation remains afflicted with uncertainty. For flood events of type R3, the quantiles of the simulated direct volumes tended to underestimate the quantiles of the observed volumes (Figure 4c). Though, again, the estimates were within the confidence intervals for almost all quantile ranges. Event type R3 had wide confidence intervals because R3 was the type with the smallest sample size for the confluence gauge. In a future study, Monte Carlo simulations will be used to test the model for type-specific biases.

Estimates of the distribution of simulated peaks were better compared to the volumes (Figure 5). Here, only small deviations in the small and medium quantiles were present for all event types. In the upper tail, some deviation still occurred. Yet, the upper tail is the range with the fewest observations; hence, this range always has high uncertainty. However, the overall visual comparison of the flood peaks indicated satisfactory simulation results for the peak-volume pairs at the confluence gauge. Moreover, the distinction into event types worked well. The results for the NTMSE are given in Table 6.

### Table 5

| Flood type | Probability of occurrence | Probability of volume-based type | Return period of peak | Volume Passau (million m³) | Peak Passau (m³/s) | Volume Hofkirchen (million m³) | Peak Hofkirchen (m³/s) |
|------------|--------------------------|-------------------------------|-----------------------|---------------------------|--------------------|-------------------------------|------------------------|
| R1 R1 R1   | 0.231                    | 0.526 0.474 0 0              | 85 15                 | 6,295                     | 1,422              | 2,706                         | 1,323                  |
|            |                          |                               | 25 30                 | 4,795                     | 1,083              | 2,999                         | 1,466                  |
|            |                          |                               | 15 40                 | 4,235                     | 957                | 3,116                         | 1,523                  |
| R1 R2 R1   | 0.231                    | 0 0.947 0.053                 | 40 15                 | 3,731                     | 1,421              | 2,706                         | 1,323                  |
|            |                          |                               | 15 25                 | 3,175                     | 1,209              | 2,924                         | 1,429                  |
|            |                          |                               | 5 50                  | 2,301                     | 876                | 3,208                         | 1,568                  |
| R2 R1 R1   | 0.050                    | 0 0 1 0                       | 85 15                 | 6,295                     | 1,422              | 2,706                         | 1,323                  |
|            |                          |                               | 35 25                 | 5,184                     | 1,171              | 2,924                         | 1,429                  |
|            |                          |                               | 5 50                  | 3,125                     | 706                | 3,208                         | 1,568                  |
| R2 R2 R1   | 0.026                    | 1 0 0 0                       | 10 100                | 2,301                     | 876                | 3,483                         | 1,703                  |
|            |                          |                               | 8 150                 | 2,717                     | 1,035              | 3,640                         | 1,780                  |
|            |                          |                               | 5 180                 | 2,891                     | 110                | 3,710                         | 1,814                  |
| R2 R2 R2   | 0.308                    | 0.129 0.871 0 0              | 75 20                 | 4,012                     | 1,528              | 2,376                         | 2,518                  |
|            |                          |                               | 10 40                 | 2,891                     | 1,101              | 2,663                         | 2,822                  |
|            |                          |                               | 5 85                  | 2,301                     | 876                | 2,874                         | 3,046                  |

Achleiten, was ensured that mirrored observed flood scenarios. It can serve as a blueprint for estimating the range of hydrographs associated with design floods.
For event types R1 and R2, the resulting errors were negligible, with values smaller than 3%. This implies a good simulation of the distribution of the flood peaks for these event types for the confluence, as well as for the tributary gauges. The estimation of flood peaks of event type R3 for the Achleiten confluence gauge showed a higher but still small error of less than 5%. For the two tributary gauges, a calculation of the error for R3-type floods is not meaningful since the observed samples included fewer than 10 events. However, in general, a satisfactory simulation of the distribution of the peaks and volumes can be observed that makes the simulated sample suitable for further analysis requiring large data samples. Indeed, the trimming in the proposed error measure could cause the estimated peak or volume of the largest flood to be trimmed. However, this only has a small effect on the estimated return periods that were estimated later on from these results, since the estimators used to fit the statistical models are robust. Without trimming, the estimated error for large quantiles (<20%) is higher than it is for small quantiles (Table 6).

The results of the proposed framework were compared with classical approaches of flood frequency analysis. First, the proposed framework was applied without any differentiation of the flood peak types and the volume-based types. Therefore, no information on the runoff-generating processes and the flood superposition were included and the peak-volume samples were treated as homogeneous and of the same origin, a classical assumption in flood frequency analysis (e.g., Kao & Chang, 2012). In a second evaluation, the flood peaks at the gauges simply were modeled by fitting a GEV distribution to the samples using L-moments. Again, no distinction into flood types was performed. This second approach is the standard procedure in flood frequency analysis. It does not take into account any spatial dependency in the catchment and consistent estimation for flood peaks of the tributaries and main river cannot be obtained. The performance of both approaches was evaluated again with the NTMSE with a trimming of 5% as well as the NTMSE for the largest 10% flood peaks without trimming. Results are given in Table 6. Q-Q plots of the observed and simulated flood peaks and volumes for the Achleiten confluence gauge with confidence bands, similar to those given in Figures 4 and 5, are given in Figure S8. It has to be noted that the confidence bands were smaller compared to the proposed type-specific framework since the data basis is larger when all flood events are considered jointly. In general, one can conclude that even though the GEV distribution results in a lower NMTSE, the proposed approach incorporates a lot more process-based information and enables the generation of meaningful flood hydrographs while still estimating peak flows reasonably well.

Here, the assumption of a homogenous sample that does not recognize the mixture of different flood types leads to an underestimation of the large flood quantiles since the distribution function is dominated by many small floods of different genesis. In particular, floods of type R1 are underestimated for different return periods of interest. Compared to the classical flood frequency analysis applied to a peaks-over-threshold series, where a distribution function is fit to the full record of flood peaks, the proposed framework resulted in slightly higher errors (Table 6). However, since only flood peaks can be modeled in standard flood frequency analyses and not peak-volume pairs that are spatially consistent between the gauges not to mention information on the flood superposition or the runoff generation, the increase in errors of less than 1% when applying the proposed framework is negligible.

### 6.2. Multivariate Type-Specific Return Periods Considering Confluences

The estimates of flood quantiles and return periods for the Achleiten gauge based on the tributaries emphasize the importance of taking into account river networks as well as flood types in flood frequency analysis. In Table 4, it can be seen for the case (R1R1R1) that the combination of floods with large return periods in both the tributary and main river led to an increased return period at Achleiten due to flood superposition (worst-case scenario). For example, two 100-years flood peaks of type R1 in the tributary and the main river gauge would lead to a 388-years flood in Achleiten. Such an increase in the return periods for confluences has been observed before (Schulte & Schumann, 2015). This reflects the unlikely case of two concurrent extreme flood events in tributaries and the main river. Although this is rarely the case, such an event occurred for the Inn and Danube River in 2013. In this case, a 10-years flood peak in Hofkirchen together with a 40-years flood in Passau, both of the flood types R1, led to a 100-years flood at the Achleiten confluence gauge, where the flood peak, due to superposition, was also
of type R1. This is the most extreme combination that was observed for this part of the Danube, where the return period of the tributary more than doubled. It was not possible to simulate a combination of large peaks in Hofkirchen and small peaks in Passau-Ingling with sufficiently high probability since no such event ever occurred in the observed sample. This indicates that the proposed approach should be applied to data-rich basins rather than data-sparse basins.

An overview of the differences between the event types and their impact on quantile estimation is given in Figure 6a. In general, it is not necessary for a peak with a return period larger than 100 years to occur in the upper Danube for there to be a peak with a return period larger than 100 years at Achleiten. This is important since one could assume that the return period does not decrease much for the relatively small distance between Hofkirchen and Achleiten despite the potential impact. The largest differences between the event types occur for combinations of small return periods (Figure 6a). For the combination (R1|R2R1), the simultaneous occurrence of flood peaks with small return periods are already sufficient to obtain a flood peak with a return period of 100 years in Achleiten (e.g., a 25-years flood at Passau-Ingling, 15-years flood at Hofkirchen); for the combination (R2|R2R2), much higher flood peaks would have to be observed in the tributary to obtain a 100-years flood at the confluence (e.g., a 50-years flood at Passau-Ingling, 25-years flood at Hofkirchen). This example demonstrates how the event type plays an important role in the generation of extreme events at a confluence gauge.

Interestingly, the estimation for event type combination (R2|R2R1) is counterintuitive. For this combination, only combinations of large return periods (>30 years) of the main river can lead to a 100-years flood in Achleiten (Figure 6a). For distinct combinations of event types, extreme events are much more probable. The combinations with the highest potential for large flood peaks given small peaks in the tributaries were (R1|R1R1) and (R1|R2R1). A drawback of all multivariate return period estimations is that all combinations of the marginal values, here the return periods of flood peaks at Hofkirchen and Passau-Ingling gauges, are treated equally. The case in which one of the two return periods is much higher than the other may be statistically valid and a limit case of the copulas, but it is not hydrologically plausible since the subbasins are expected to react similarly, at least for extreme events where mostly global instead of local drivers cause the flood events. To overcome this problem, critical ranges can be estimated (Chebana & Ouarda, 2011; Volpi & Fiori, 2012). These critical ranges include only those combinations that are likely to occur, based on past observations. Moreover, in this context, not every combination of event types is equally probable.

Table 5 shows that the most likely combination for a 100-years flood in Achleiten is R2 events at both Passau-Ingling and Hofkirchen. For this R2-R2-combination, a flood in Achleiten also has the type R2 with a probability of 31% and is of volume-based flood type “Upper Danube” with a probability of 87%. Therefore, almost all floods in Achleiten that originate from this combination are dominated by the flood in the upper Danube Basin. For these R2|R2R2 events, different combinations of return periods in Passau-Ingling and Hofkirchen lead to 100-years floods. It is striking that the relation between the return periods of the tributary and the main river gauge is not symmetric given their similar mean annual runoff. Yet, the flood regimes of both basins differ (the Inn basin being much more impacted by the Alps and hence snowmelt), which is probably the reason for the difference. Two other probable combinations for a 100-years flood in Achleiten are R1|R1R1 and R1|R2R1. Overall, the general relevance of the combinations changes with the return period.

Table 5 underscores the importance of distinguishing flood types tributary impacts for the flood frequency analysis of large river basins. Different combinations of flood events in the main river and the tributary can lead to different magnitudes of floods downstream at the confluence gauge. With the proposed statistical simulations and the multivariate model, the probability of such combinations can be assessed, and distinct scenarios can be evaluated. Such an evaluation cannot be performed empirically, as the number of observed events is insufficient to estimate such probabilities.

The estimated design flood scenarios (Figure 6b) reveal significant differences in the design floods. The most probable scenario (concerning the joint probability of flood types and volume-based type) is given by an R1 flood at the confluence gauge generated by an R2 flood in the Inn tributary and an R1 flood in the Upper Danube, where also the focus point is located. The difference in the hydrograph shapes associated with different flood types becomes visible. The R1 flood events are associated with hydrographs of steep shape
with high peaks but only small volumes. The R2 floods instead are characterized by hydrographs of wider shape with larger volumes compared to flood type R1. This has to be considered in flood protection, where the higher-volume events can reduce the capacity of reservoirs much more. Additionally, long-duration floods can increase the bank erosion of a river.

7. Generalization of the MFFA to More Complex River Basins–A Case Study on the Rhine Basin

7.1. Set-Up of the Extended Statistical Model With an Appropriated Spatial Structure

The case study of the Danube basin is based on a river network with a relatively simple structure and strong differences between flood regimes of the Inn River and the Upper Danube River. An application of the proposed methodology to more complex river basins with several sub-basins requires the extension of the model structure. We first present the general concept of this extension and then demonstrate it on a case study on the Rhine River basin. The starting point for the statistical modeling is the outlet gauge at the main river and the closest confluence of the main river with a significant tributary upstream (confluence level 1). For the simulation procedure at the outlet, a Vine-copula with a dimension equal to the number of tributaries plus one is chosen. If the considered gauges are located close to the confluence, the dimension of the copula can be reduced to the number of tributaries and no Vine-structure is required. In this latter case, the flood volume at the outlet is defined by the sum of the volumes at the gauges upstream of the confluence, as it was illustrated above for the Danube case. After the joint simulation of this confluence, the simulation proceeds upstream successively. For each confluence of tributaries with the main river, a copula has to be fitted, with dimension depending on the runoff volume from the intermediate catchment between the three gauges. It is possible to consider further sub-catchments of a tributary, however, the dimension of the required copula has to be equal to the number of gauges plus one (the gauge downstream of the point of confluence). In Figure 7a, an example of possible structures is given. Two sub-catchments of Tributary 1.2.1 are considered separately as parts of the Tributary 1.2. This subdivision of the catchment of Tributary 1.2.1 into three parts requires a 4-dimensional copula to combine the statistics of the gauges at Tributaries 1.2.1.1 and 1.2.1.2 with the upstream and downstream gauges of Tributary 1.2.1. Obviously, the “curse of dimensionality” restricts the spatial resolution of large river basins. However, it is possible to simplify the model. If for example, no suitable gauge at a tributary is located close to the confluence, the impact of the...
A tributary can be considered with a bivariate copula between two gauges at the main river, one before and one after the confluence (e.g., for Tributary 1.3 in Figure 7a).

What remains unclear is the choice of the complexity of the model. As discussed above, each gauge in the river network could be included in the model. However, it is more efficient to determine the number of sites to include in the model based on the hydrological information each gauge contains. We will illustrate this for the Rhine basin introduced in Section 2. Our starting point of the Rhine basin case study was the confluence of Moselle River (Cochem gauge) and the Rhine River, which is represented by the downstream Andernach confluence gauge. To describe the flood regime of the Rhine River upstream of the confluence, the Kaub gauge was selected. Further upstream, two main tributaries, the Neckar and Main Rivers enter the Rhine River upstream with the Worms gauge located between both confluences. The impact of these tributaries was measured with the volume-based typology, which had to be extended for this multivariate case. Instead of the trivariate classification at the confluence gauge only, a successive classification was performed downstream that can be applied to any general structure of gauges in large river basins. More precisely, the flood volume at the Maxau gauge, the station furthest upstream whose runoff mainly originates in the Alps, was taken as the initial volume. The relative increase (related to the catchment area) of the flood volume between Maxau and Worms results was largely from the Neckar tributary. The next section, representing the increase in volume between Worms and Kaub, accounts for the impact of the Main River, while finally, the last section from the Kaub gauge to our outlet Andernach gauge is strongly affected by the Moselle River. Using this information, we obtained four relative flood volumes that were compared to the total volume at the Andernach gauge. A comparison of these relative proportions determined where the focal point of the flood event was located: in the alpine parts of the basin, the Neckar, the Main, or the Moselle River basins. Aside from the Cochem tributary gauge, only gauges on the main river were considered. Our event analysis had shown a more relevant impact of the Moselle River on rainfall-generated floods. As we focused our analysis on the pronounced role of this tributary, we decided to join these two separable volume-based types “Neckar River” and “Main River” into one volume type “Mixtype.” As a result, we classified three volume-based flood types: “Alpine”, “Moselle” and “Mixtype” and obtained the model structure given in Figure 7b. Hence, the volume-based typology was not only used to characterize the impact of the tributaries but also to determine the model structure.

7.2. Specification of Design Floods for Given Return Periods at the Andernach Outlet Gauging Station

Here, the design floods are also obtained through a step-by-step procedure: first, a return period with the corresponding design flood is estimated for the confluence gauge. Based on this, using the respective confluence-level-1 (the confluence furthest downstream) copula model (Andernach-Kaub-Cochem), return periods for the actual confluence are estimated. Using these return periods from the first confluence, return periods at gauges of the next confluence upstream are estimated, and so on. In principle, the proposed model is extendable to an arbitrary number of tributaries, but computational capacities and uncertainty propagation in the model chain limit the applicability at high spatial resolutions. In principle, a Vine-copula could also have a number of dimensions equal to the number of gauges in the river basin. However, this procedure would not only be computationally costly, but it also would ignore the interdependence between the gauges. In contrast, our hierarchical approach, instead, is easily generalizable and the step-wise estimation considers the interdependence.

The flood typology was applied as described before. For the Andernach gauge in the Rhine basin, an additional flood type, denoted by R0 became necessary. In periods with high baseflow (resulting from wet initial conditions of watersheds), the basin quickly generates runoff from rainfall. This results in a series of flood events with small direct runoff volumes but relatively high peaks of total runoff. Details can be found in Text S4.

To generate synthetic samples of peak-volume pairs for each flood type for the Andernach outlet gauge, the extended procedure given in Figure 7b was applied. While the simulation for the Andernach outlet gauge in combination with the Cochem tributary gauge and Kaub main river gauge was analogous to the Danube example, additional peak-volume pairs had to be considered for upstream gauges, Worms and Maxau. After simulating the peak-volume pairs of Kaub, the peak-volume pairs of Worms were simulated by using the...
results for Kaub. Then, the simulated pairs were used to derive pairs for Maxau. This successive approach keeps the required model simple with a maximum dimension of three and three partial models. Further details are given in Text S2 and Text S4.

The results show that the NTMSE error for the simulation of the distribution of flood peaks at the Andernach outlet gauge is small for flood types R0, R1, and R2 (Table 7). For the snow-impacted flood types S1 and S2, the errors were slightly larger. This difference can be explained by two factors: on the one hand, the data samples for these types are much smaller than those for the rainfall-induced flood types, introducing uncertainty. On the other hand, an extension of the peak-volume relationship to snow-impacted floods is not straightforward. Snow-impacted flood types were not classified according to the linear peak-volume relationship (Figure S4) and hence do not necessarily show this dependency. Non-linearities in this case can be handled by applying linear transformations.

For the first submodel, including Cochem and Kaub gauges, small errors were observed, mostly below 10%. For the second and third submodels, errors increased with the distance to the outlet gauge because errors propagated in the successive procedure and increased with every new simulation step. One must also consider that, for the error estimation, all flood events observed at the single tributary gauges were used and compared with those events that occurred synchronously at Andernach gauge. It became obvious that not all events could be taken into account for multivariate analyses since in some cases peaks at gauges upstream did not produce flows above downstream flood thresholds due to flood wave diffusion.

With regard to the probabilistic specification of flood scenarios, the most frequent combinations of event types for the first confluence upstream of Andernach given Cochem and Kaub gauges were estimated first. This applies to the combinations (R0|R1R1), (R0|R2R1), (R1|R1R1), (R2|R2R1), (R2|R2R2), (S2|S2S2), and (S2|R1S2). For the Rhine Basin, besides heavy- and medium-rainfall floods, snowmelt floods (event type S2) also play an important role and have to be considered separately. This reflects the large impact of the Alps on flood generation, especially in spring. With the extended model, the more complex interplay between the different flood types can be captured. The most critical combination of flood types causing a 100-years flood downstream is the combination of two contemporaneous heavy rainfall floods (event type R1), leading to an R1 flood downstream (Figure S11). In contrast, a combination of snowmelt-floods with higher return periods is required to produce a 100-years flood. A possible reason could be the asynchronous timing of snowmelt-floods, depending on the catchment elevation.

Going upstream, we considered the combinations at the Worms and Maxau gauges. For the flood scenarios, the corresponding most probable flood types and return periods for these upstream gauges were estimated successively. Mostly, the flood types at the three gauges at the main river were consistent for each flood event. In Table S9, these scenarios are summarized, and probabilities for each combination as well as for the volume-based flood types are given.

Some combinations of flood types in the tributary and main river did come close at all to producing a 100-years flood in Andernach. Especially, this applies to combinations resulting in an R0 event at Andernach gauge. This is reasonable since this flood type at the outlet is related to superimposed flood events that occur during a time period with high baseflow. However, for these flood types, the proportion of direct runoff on the peak is small; hence, it is improbable that a combination of high baseflow and such a direct runoff peak will cause a 100-years flood event.

In Figure 8a, design flood scenarios are given with assigned probability. Here, the procedure was extended to also capture the probability of different flood types at the Worms and Maxau gauges upstream. The derivation of the probability of each scenario is illustrated in Figure 8b. The most probable scenario for a flood at Andernach gauge with a return period of 100 years is an R2 flood resulting from a combination of other R2-floods with a focus point in the Moselle tributary. A combination of R1-flood events in the main river and the Moselle tributary is unlikely to produce an R1 event at Andernach since the response time of the Moselle tributary is much shorter than it is at the Kaub gauge upstream on the Rhine River.

### Table 7

| Gauge     | R0  | R1  | R2  | R3  | S1  | S2  |
|-----------|-----|-----|-----|-----|-----|-----|
| Andernach | 1.34| 0.90| 1.7 | −   | 5.7 | 1.8 |
| Cochem    | −   | 1.9 | 1.1 | −   | 15.8| 0.8 |
| Kaub      | −   | 8.6 | 4.1 | 12.8| 9.5 | 8.9 |
| Worms     | −   | 4.3 | 5.6 | 32.2| 19.8| 7.2 |
| Maxau     | −   | 29.2| 8.8 | 50.3| 38.6| 13.4|

*Note. Normalized Trimmed Mean Square Error could not be calculated for R3 due to insufficient data for Andernach and Cochem gauges.*
Figure 8. Flood event scenarios for the most probable combinations that lead to a 100-years flood at Andernach, differentiated by volume-based flood types. (a) Design hydrographs, numbers indicate the probability of occurrence of each scenario (combination of flood types and volume-based type); (b) flow chart showing how to derive the probabilities. The highlighted path is the one with the highest probability. It reads as follows: the most probable scenario for a 100-years flood at the confluence gauge is an R2 flood with two R2 floods in upstream tributaries (probability 0.341). Having an R2 flood in Kaub, the most probable flood type for Worms is R2 (0.941). For an R2 flood in Worms, the most probable type in Maxau is R2 (0.911). For the volume-based type Moselle (probability 0.473), this gives a joint probability of 0.138 (panel a).
This scenario analysis once again emphasizes the importance of differentiating between spatial distributions of flood-inducing precipitation, flood event types, and flood superpositions. With the proposed generalization, river basins of high complexity can be modeled and statistical scenarios can be obtained. Even topographically diverse catchments within the basin can be captured and their impact on the downstream flood can be assessed.

8. Summary and Conclusions

The concurrence of tributary flood events can have a profound impact on the shape of flood hydrographs downstream of its confluence, which, in turn, can affect the peak flows frequently considered in engineering design. The timing of flood wave superpositions and hydrograph shapes both have strong impacts on the flood peak. In our method, we consider these aspects explicitly by taking into account the different flood types at gauges upstream of a confluence and their impact on the flood type downstream in statistical simulations of synthetic multivariate samples. We are able to simulate the distribution of peak-volume pairs at the confluence gauge for each flood type, as well as the corresponding flood types for tributaries. With these large-sample simulations, we could then generate multivariate (peak-volume) flood statistics for different combinations of tributary flood types. Our simulation model showed a high coherence between the quantiles of the peaks of observed and simulated floods, with errors of less than 5% obtained with NTMSE error measure with a trimming degree of 5%. Floods with higher return periods showed larger errors of up to 20% when trimming was not applied. Compared to classical flood frequency analyses with annual maximum series that do not take into account the flood-generating processes and are much less complex, the errors of the proposed model increased by less than 1%. The consideration of spatial dependency and flood-generating processes, therefore, comes at rather low costs.

By using vine copulas, the conditional distribution of a flood event at the confluence gauge given the flood at the upstream main river and the tributaries were estimated for different combinations of flood types. The results revealed that different combinations of tributary flood types yielded remarkably different probabilities for extreme floods. For instance, the largest increase in return period for the confluence gauge in the Danube basin was shown to occur from a combination of two synchronous short, high-peaked floods in the tributaries, which can be expected. Contrary, two snowmelt floods in the tributaries of the Danube river are unlikely to result in an extreme flood downstream. The relevance of the flood types for producing extreme floods depends much on the catchment, especially the reaction time, as was demonstrated in the second case study of the Rhine basin.

With statistical simulation, the probability of distinct scenarios featuring different tributary event types can be estimated. The results can facilitate the operation of flood protection measurements and water management and lead to a better understanding of the flood dynamics at confluence gauges. For the operation of flood protection measures, the most probable scenarios can be chosen according to the current catchment conditions. For example, in early spring, the snowmelt scenarios should be considered, while in the summer months, the focus should be on the most probable heavy-rainfall scenarios. In addition, sediment transport can be assessed with the design floods provided: not only a larger amount of sediment will be deposited in the main river if only the tributary is affected by a flood. The combination of different flood types in a tributary and the main river may result in different sedimentation or erosion processes downstream. The longer the duration of the flood, the more erosion may take place which does not only to more unstable river banks but can also decrease water quality downstream.

An important point that has to be discussed is the uncertainty introduced by the proposed procedure. There are several sources of uncertainty in the method, all with different impacts on the results. First, distribution functions had to be selected for at least five different flood types as well as copulas for the different volume-based flood types. Moreover, empirical frequencies of the flood types and their combinations were estimated based on observed flood events. Therefore, the uncertainty of the mixture ratio also must be considered (see Barth et al. [2019]). The use of pooled flood volumes, that is, without differentiation of the flood type, could limit the goodness-of-fit of the copula models since the events do not originate from the same population. However, we believe that this practice has less of an impact on the flood volumes compared to the flood peaks since the flood volumes are more affected by the superposition considered in the
volume-based flood types and that the uncertainty introduced by a pooled sample is smaller than the uncertainty introduced by a further distinction of the samples into flood types leading to samples with less than 20 observations. Uncertainty was also created by using the simulated samples to derive the design-flood scenarios and by selecting Vine-copulas for the flood statistics. Propagation of uncertainty became apparent in the more complex study of the Rhine river basin, where a larger number of gauges was considered. It was clearly shown in this example that the complexity of the model is tied to the available data basis and that high-dimensional basins (with many gauges and tributaries) can only be considered for data-rich regions.

Since the statistical model is complex and requires many parameters to be estimated, sufficient long hydrologic records at tributary and confluence stations are required (Brunner et al., 2018). The framework was also tested with record lengths of 30 years, which led to higher errors. We recommend record lengths of at least 80 years of observation to obtain reasonable uncertainty in design flood estimates. Simulation experiments could provide further guidance on this in the future. The use of copula models adds additional uncertainty. The proposed Vine-copulas are highly flexible, which is required for characterizing the complex dependencies between the tributaries. Yet, this comes at the cost of more parameters than simple bivariate copulas. It was shown by Brunner and Sikorska-Senoner (2019) that bivariate modeling of peak-volume pairs is affected by uncertainty due to discharge resolution and model choice, which also may apply in our case. By using goodness-of-fit criteria featuring penalties for additional model parameters, the number of parameters was kept as small as possible.

Another point that must be addressed in the future is the scale-dependence of the model. We assume that the model is not directly transferable to mesoscale catchments with pronounced short-term rainfall-runoff dynamics. For example, heavy convective rainfall that is, spatially limited could lead to rapid catchment responses that would be dampened in large catchments. For mesoscale catchments, the flood-generating processes play a more pronounced role than for macroscale catchments, and more information about the spatial and temporal distribution of rainfall must be considered.

Due to the type-specific nature of the design of floods, changes in either the frequency or the magnitude of floods can be associated with changes in the frequency or magnitude of distinct flood types easily. For example, recent climate change studies assume an increase in the frequency of heavy-rainfall floods in Central Europe (Rajczak et al., 2013). This should be mirrored in an increase of the impact of floods of type R1. At the same time, snowfall is expected to decline and hence snowmelt-impacted floods will probably decrease in frequency. The use of flood types enables the identification of these drivers directly. As a result, the proposed model can be extended to non-stationary cases, incorporating the detected changes for distinct flood types along with changes of flood type ratios at sites with mixed distributions. For example, the flood type frequencies could be linearly increasing for flood type R1. By the application to discharge series modeled using projected climate scenarios, such changes could be investigated in greater detail. Again, the type-specific methodology offers the opportunity to directly compare the flood-frequency implications of changes to each flood type in future studies. Changes or trends in the magnitude can be detected with classical stationary tests like Mann-Kendall- or Wilcoxon-tests, while changes in frequency can be detected using variance-change-point-tests as demonstrated in Fischer et al. (2019), who identified an increase in frequency of R1 floods in eastern Germany. In the next step, time-variant distributions could be applied as marginal distributions such that the distribution parameters can vary in time, either by an abrupt change or a monotonic trend. The copula models can be extended to capture non-stationarity as well, by using time-variant copula models (Feng et al., 2020).

The seven-step method proposed here can be applied to an ungauged confluence location as well. To do so, the estimated probabilities of occurrence of the event types simply have to be estimated regionally or based on a nearby located gauge. The same holds for the peak-volume relationship, where regionalization is easily possible, as shown by Fischer et al. (2019).

It was shown that the method is generalizable to complex river networks without major reservoirs. For all basins under consideration, the stationarity assumption was not rejected despite river training, for example, in the upper Rhine basin close to Kaub gauge, where several weirs have been installed in the mid of last century. Yet, in Europe, many reservoirs have been built in the last century. These reservoirs change the flood regimes of rivers. Future studies will concentrate on a detailed analysis of these factors and their impact on
the model performance and how the model can be adapted to accommodate them. For example, the copula models could be extended to non-stationary cases where the different impact of reservoirs on different flood magnitudes is included. Possible generalizations could also capture nonstationary features of reservoir-impacted river reaches. For example, the acceleration of flood waves in tributaries or flood retention measures with impact on the hydrographs can be considered as modifications of the dependence structure.

The proposed procedure can also be seen as an alternative to continuous distributed hydrological simulations. Our statistical modeling techniques require less data compared to these physically-based models, which often require detailed precipitation scenarios and temperature data as main input as well as soil moisture states or routing routines. Soil moisture data are rarely available for long time periods and, if substituted with outputs from other models, inject additional uncertainty into the model parameter and structural uncertainty. In contrast, our approach does not have all these uncertainties. Future work should compare the uncertainty cascades in our proposed statistical approach with the uncertainty in a continuous distributed hydrological simulation.

The proposed statistical model that takes into account flood event types and the spatial dependency of floods within a basin, therefore, offers the opportunity to model the flood-generating processes in a basin in greater detail while only having slightly increased errors compared to classical flood frequency analyses methods. Future changes in climate and impacts on the flood protection measures can be evaluated more easily with this model and give space for many future applications.

Data Availability Statement
Climatic data can be obtained from the German Weather Service (DWD; https://www.dwd.de/EN/climate_environment/cdc) and the E-OLBS data set (https://surfobs.climate.copernicus.eu/dataaccess/access_eobs.php). The manuscript provides all the information needed to replicate the results. R-Codes for flood typology and flood event separation are available in the R-package FloodR (https://github.com/PhilippBuehler).

References
Barth, N. A., Villarini, G., Nayak, M. A., & White, K. (2017). Mixed populations and annual flood frequency estimates in the western United States: The role of atmospheric rivers. Water Resources Research, 53, 257–269. https://doi.org/10.1002/2016WR0191964
Barth, N. A., Villarini, G., & White, K. (2019). Accounting for mixed populations in flood frequency analysis: Bulletin 17C perspective. Journal of Hydrologic Engineering, 24(3), 10–1061. https://doi.org/10.1061/ASCEHE.1943-5584.0001762
Becker, A. (1995). Problems and progress in macroscale hydrological modelling. In R. A. Feddes (Ed.), *Space and time variability and interdependencies in hydrological processes*. International Hydrology Series (pp. 134–143). Cambridge University Press. https://doi.org/10.1017/CBO9780511564475.012
Becker, A., & Grünwald, U. (2003). Flood risk in Central Europe. Science, 300(5622), 1099. https://doi.org/10.1126/science.1083624
Blöschl, G., Nester, T., Komma, J., Parajka, J., & Perdigão, R. A. P. (2013). The June 2013 flood in the Upper Danube Basin, and comparisons with the 2002, 1954 and 1899 floods. Hydrology and Earth System Sciences, 17, 5197–5212. https://doi.org/10.5194/hess-17-5197-2013
Brunner, M. I., Sikorska, A. E., Furrer, R., & Favre, A.-C. (2018). Uncertainty assessment of synthetic design hydrographs for gauged and ungaged catchments. Water Resources Research, 54(3), 1493–1512. https://doi.org/10.1029/2017WR021129
Brunner, M. I., & Sikorska-Senoner, A. E. (2019). Dependence of flood peaks and volumes in modeled discharge time series: Effect of different uncertainty sources. Journal of Hydrology, 572, 620–629. https://doi.org/10.1016/j.jhydrol.2019.03.024
Brunner, M. I., Viviroli, D., Sikorska, A. E., Vannier, O., Favre, A.-C., & Seibert, J. (2017). Flood type specific construction of synthetic design hydrographs. Water Resources Research, 53, 1390–1406. https://doi.org/10.1002/2016WR019535
Chebana, F., & Ouarda, T. (2011). Multivariate quantiles in hydrological frequency analysis. Environmetrics, 22(1), 63–78. https://doi.org/10.1002/env.1027
Chen, L., Singh, V. P., Shenglian, G., Hao, Z., & Li, T. (2012). Flood coincidence risk analysis using multivariate copula functions. *Journal of Hydrologic Engineering, 17*(6), 742–755. https://doi.org/10.1061/(asce)he.1943-5584.0000504
DWA. (2012). *Merkblatt-M552: Ermittlung von Hochwasserwahrscheinlichkeiten* (01477440 edition). DWA.
Engel, H. (1996). The flood events of 1993/1994 and 1995 in the Rhine river basin. Destructive water: Water-caused natural disasters, their abatement and control (Proceedings of the Conference held at Anaheim, California, June 1996). IAHS Publication No. 239, 1997.
Fantazzini, D. (2009). The effects of misspecified marginals and copulas on computing the value at risk: A Monte Carlo study. *Computational Statistics & Data Analysis, 53*(6), 2168–2188. https://doi.org/10.1016/j.csda.2008.02.002
Feng, Y., Shi, P., Qi, S., Mou, S., Chen, C., & Dong, F. (2020). Nonstationary flood coincidence risk analysis using time-varying copula functions. Scientific Reports, 10, 3385. https://doi.org/10.1038/s41598-020-60264-3
Fischer, S. (2018). A seasonal mixed-POT model to estimate high flood quantiles from different event types and seasons. *Journal of Applied Statistics, 45*, 2831–2847. https://doi.org/10.1080/02664763.2018.1441385
Fischer, S., & Schumann, A. (2016). Robust flood statistics: Comparison of peak over threshold approaches based on monthly maxima and TL-moments. Hydrological Sciences Journal, 61(3), 457–470. https://doi.org/10.1080/02626667.2015.1054391
Fischer, S., Schumann, A., & Bühler, P. (2019). Timescale-based flood typing to estimate temporal changes in flood frequencies. *Hydrological Sciences Journal, 64*, 457–470. https://doi.org/10.1080/02664667.2018.1679376
Fischer, S., Schumann, A., & Bühler, P. (2021). A statistics-based automated flood event separation. Journal of Hydrology X, 10, 100070. https://doi.org/10.1016/j.hydrol.2020.100070

Fischer, S., Schumann, A., & Schulte, M. (2016). Characterisation of seasonal flood types according to timescales in mixed probability distributions. Journal of Hydrology, 539, 38–56. https://doi.org/10.1016/j.jhydrol.2016.05.005

Gaál, L., Szolayi, J., Kohnová, S., Parajka, J., Merz, R., Viglione, A., & Blöschl, G. (2012). Flood timescales: Understanding the interplay of climate and catchment processes through comparative hydrology. Water Resources Research, 48, W04511. https://doi.org/10.1029/2011WR011599

Geertsema, T. J., Teuling, A. J., Uijlenhoet, R., Torfs, P. J. F., & Hoitink, A. J. F. (2018). Anatomy of simultaneous flood peaks at a lowland confluence. Hydrology and Earth System Sciences, 22, 5599–5613. https://doi.org/10.5194/hess-22-5599-2018

Gräler, B., van den Berg, M. J., Vanendenbergh, S., Petrozelli, A., Grimaldi, S., De Baets, B., & Verhoest, N. E. C. (2013). Multivariate return periods in hydrology: A critical and practical review focusing on synthetic design hydrograph estimation. Hydrology and Earth System Sciences, 17(4), 1281–1296. https://doi.org/10.5194/hess-17-1281-2013

Guse, B., Merz, B., Wiertzke, L., Ullrich, S., Viglione, A., & Voroguschyn, S. (2020). The role of flood wave superposition in the severity of large floods. Hydrology and Earth System Sciences, 24, 1633–1648. https://doi.org/10.5194/hess-24-1633-2020

Hirschboeck, K. K. (1987). Hydroclimatically-defined mixed distributions in partial duration flood series. In V. P. Singh (Ed.), Hydrologic frequency modeling (pp. 199–212). Springer. https://doi.org/10.1007/978-94-009-3953-0_13

Huang, K., Chen, L., Zhou, J., Zhang, J., & Singh, V. P. (2018). Flood hydrograph coincidence analysis for mainstream and its tributaries. Journal of Hydrology, 565, 341–353. https://doi.org/10.1016/j.jhydrol.2018.08.007

Joe, H., Li, H., & Nikoloulopoulos, A. K. (2010). Tail dependence functions and vine copulas. Journal of Multivariate Analysis, 101(1), 252–270. https://doi.org/10.1016/j.jmva.2009.08.002

Kamp, S. K., & Lefskay, M. A. (2016). Transition of dominant peak flow source from snow melt to rainfall along the Colorado Front Range: Historical patterns, trends, and lessons from the 2013 Colorado Front Range floods. Water Resources Research, 52(1), 407–422. https://doi.org/10.1002/wrcr.v52.110.2015wr017784

Kao, S.-C., & Chang, N.-B. (2012). Copula-based flood frequency analysis at ungauged basin confluences: Nashville, Tennessee. Journal of Hydrologic Engineering, 17(7), 790–799. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000477

Kao, S.-C., & Govindaraju, R. S. (2008). Trivariate statistical analysis of extreme rainfall events via the Plackett family of copulas. Water Resources Research, 44(2), 2008. https://doi.org/10.1029/2007wr006261

Kawashima, T., & Fujisawa, H. (2017). Robust and sparse regression via divergence. Entropy, 19(11), 608. https://doi.org/10.3390/e19110608

Lloyd, S. P. (1982). Least squares quantization in PCM. In IEEE Transactions on Information Theory (Vol. 28, pp. 129–137). https://doi.org/10.1109/tit.1982.1056489

Marsalek, J., Watt, W. E., Zeman, E., & Sieker, F. (2020). Flood Issues in contemporary water management. Nato Science Partnership Sub-series: 2 Springer Netherlands.

Merz, R., & Blöschl, G. (2003). A process typology of regional floods. Water Resources Research, 39(12), 1340. https://doi.org/10.1029/2002wr001952

Murphy, P. J. (2001). Evaluation of mixed-population flood-frequency analysis. Journal of Hydrologic Engineering, 6(1). https://doi.org/10.1061/(asce)1084-0699(2001)6:1(62)

Nagler, T. (2019). VC2copula: Extend the 'copula' package with Families and models from 'VineCopula'. R package version 0.1.0. Retrieved from https://CRAN.R-project.org/package=VC2copula

Nied, M., Pardowitz, T., Nissen, K., Ulbrich, U., Hundehe, Y., & Merz, B. (2014). On the relationship between hydro-meteorological patterns and flood types. Journal of Hydrology, 519, 3249–3262. https://doi.org/10.1016/j.jhydrol.2014.09.089

Peng, Y., Chen, K., Yan, H., & Yu, X. (2017). Improving flood-risk analysis for confluence flooding control downstream using copula Monte Carlo method. Journal of Hydrologic Engineering, 22(8), 04017018. https://doi.org/10.1061/(asce)he.1943-5584.0001526

Rajczak, J., Pall, P., & Schär, C. (2013). Projections of extreme precipitation events in regional climate simulations for Europe and the Alpine Region. Journal of Geophysical Research: Atmospheres, 118, 3610–3626. https://doi.org/10.1002/jgrd.50297

Schulte, M., & Schumann, A. (2015). Extensive spatio-temporal assessment of flood events by application of pair-copulas. Proceedings of the IAHS, 370, 177–181. https://doi.org/10.5194/iahs-370-177-2015

Seibert, J., & Vis, M. J. (2012). Teaching hydrological modeling with a user-friendly catchment-runoff-model software package. Hydrology and Earth System Sciences, 16(9), 3315–3325. https://doi.org/10.5194/hess-16-3315-2012

Sikorska-Senoner, A. E., & Seibert, J. (2020). Flood-type trend analysis for alpine catchments. Hydrological Sciences Journal, 65(8), 1281–1299. https://doi.org/10.1080/02626667.2020.1749761

Skublics, D., Bloeschl, G., & Rutschmann, P. (2016). Effect of river training on flood retention of the Bavarian Danube. Journal of Hydrology and Hydromechanics, 64, 349–356. https://doi.org/10.1515/johh-2016-0035

Skublics, D., Seibert, S. P., & Ehret, U. (2014). Abbildung der Hochwasserretention durch hydrologische und hydrodynamische Modelle unter unterschiedlichen Randbedingungen – Sensitivitätsanalyse am Donauabschnitt zwischen Neu-Ulm und Donauwörth. Hydrology Wasserbewirts, 56, 178–189. https://doi.org/10.5675/HyWa_2014_3_2

Song, S., & Singh, V. P. (2010). Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data. Stochastic Environmental Research and Risk Assessment, 24(3), 425–444. https://doi.org/10.1007/s00477-009-0331-1

Tarasova, L., Merz, R., Kiss, A., Basso, S., Blöschl, G., Merz, B., et al. (2019). Causative classification of river flood events. WIREs Water, 6, e1353. https://doi.org/10.1002/wat2.1353

Voß, E., & Fiori, A. (2012). Design event selection in bivariate hydrological frequency analysis. Hydrological Sciences Journal, 57(8), 1506–1515. https://doi.org/10.1080/02626667.2012.726357

Voroguschyn, S., & Merz, B. (2013). Flood trends along the Rhine: The role of river training. Hydrology and Earth System Sciences, 17(10), 3871–3884. https://doi.org/10.5194/hess-17-3871-2013

Wang, C., Chang, N. B., & Yeh, G. T. (2009). Copula-based flood frequency (COFF) analysis at the confluences of river systems. Hydrological Processes: International Journal, 23(10), 1471–1486. https://doi.org/10.1002/hyp.7273

White, H. (1982). Maximum likelihood estimation of misspecified models. Econometrica, 50, 1–26.

Yang, E., Lozana, A. C., & Aravkin, A. (2018). A general family of trimmed estimators for robust high-dimensional data. Electronic Journal of Statistics, 12, 3519–3553. https://doi.org/10.1214/18-ejs1470