Austrian Journal of Statistics
2025, Volume 54, 124-149.
http://www.ajs.or.at/
doi:10.17713/ajs.v54i3.2101

Exploring Process Heterogeneity in Environmental Statistics: Examples and Methodological Advances

Gregor Laaha
Institute of Statistics
BOKU University
Vienna, Austria

Johannes Laimighofer
Institute of Statistics
BOKU University
Vienna, Austria

Nur Banu Özcelik
Institute of Statistics
BOKU University
Vienna, Austria

Svenja Fischer
Department of Environmental Sciences
Wageningen University & Research
The Netherlands

Abstract

Environmental models typically rely on stationarity assumptions. However, environmental systems are complex, and processes change over states or seasons, leading to often overlooked heterogeneity. This paper explores methods to incorporate process heterogeneity into statistical models to improve their performance. It considers problems from natural hazards and earth system sciences, demonstrating the effects of process heterogeneity and proposing methodological advances through model extensions. The first problem addresses flood frequency analysis, where floods are generated by different processes in catchment and atmosphere. A mixture model combining peak-over-threshold distributions of flood types can handle this heterogeneity, especially regarding tail heaviness, making it relevant for flood design. The second problem involves minimum flow frequency analysis, with heterogeneity from different summer and winter processes. A mixture distribution model for minima and a copula-based estimator can incorporate seasonal distributions and event dependence, showing significant performance gains for extreme events. The third problem examines process heterogeneity in rainfall models. Clustering event characteristics (e.g., duration, intensity) using Gower's distance and a lightning index helps distinguish between convective and stratiform events, showing potential to enhance rainfall generators. The fourth problem deals with parameter variation in temporal models of environmental variables, using daily streamflow series. A tree-based machine learning model shows that prediction performance and model parameters vary with quantile loss optimization, suggesting the need for different or combined models for full time series in the presence of process heterogeneity. The study highlights the importance of considering process heterogeneity in modeling from the outset and encourages a better understanding of statistical assumptions and the enrichment of physical knowledge in environmental statistics.

Keywords: environmental statistics, process-based modeling, flood, drought, hybrid rainfall distribution, extreme events, cluster analysis, statistical learning models, quantile regression.

1. Introduction

Modeling environmental variables is a challenging task. Accurate models are needed in the context of planning and engineering design, where values representing the magnitude of events with specified probabilities or return periods are critical. Models are also needed to answer questions about the temporal evolution of a variable and to make predictions, often for a future point in time, including climate projections and forecasting. In addition, they need to answer questions about the spatial distribution and make predictions at locations without observations. To achieve these goals, a wide range of statistical models is employed. This includes distribution models for frequency analysis of extreme events, time series models for stochastic simulations (such as those used in rainfall generators), regression-type approaches for modeling the temporal evolution or spatial patterns of environmental variables, and geostatistical models for similar problems.

The complexity of environmental systems complicates modeling efforts. Common approaches assume that observations are homogeneous and independent. In statistics, homogeneity is often interpreted as stationarity, implying that observations are realizations of a set of identically distributed random variables. The importance of stationarity has been widely discussed in the context of global change, leading to controversial views. Some researchers argue that "stationarity is dead", meaning that past experience and data are less representative in a changing climate (Milly *et al.* 2008), whereas others assert that "stationarity is immortal" and remains crucial for achieving efficient solutions despite non-stationarity (Montanari and Koutsoyiannis 2014). However, the original definition of homogeneity is broader. Gumbel (1941) defines homogeneity as "subject to a common set of forces", suggesting that both the homogeneity of statistical properties and the generation processes are important. A common problem in modeling environmental systems is that samples may consist of events resulting from distinctly different processes. In such cases, models may need to be adapted to account for process-heterogeneity.

This paper explores various concepts for incorporating process heterogeneity into statistical models to improve their performance. It is motivated by specific problems from natural hazards and earth system sciences, and demonstrates model extensions using mixture models, copula-based estimators, clustering techniques, and machine learning approaches.

The first problem pertains to process heterogeneity in flood hydrology. An introduction to the problem is provided by a discussion of how natural phenomena often violate the statistical assumptions used in modeling. Different types of floods pose a particular challenge to frequency analysis if they follow different distributions. The second section proposes a mixture model that combines peak-over-threshold (POT) distributions of flood types into an overall distribution to calculate annual return periods of events.

The second problem also concerns river flows but focuses on droughts. Here, the heterogeneity relates to different generation processes of summer and winter events, which are easier to separate. However, the independence assumption may be violated. The proposed methods transfer mixture distribution approaches originally defined for maxima to annual minimum series, and a copula-based method is proposed to incorporate dependence of series.

The third problem relates to meteorology and discusses the role of process heterogeneity in rainfall models. We propose clustering based on multiple event characteristics (such as duration and severity) to separate rainfall series into types of events. The resulting classes of rainfall are compared regarding to their event characteristics for plausibility. Finally, we compare the distributions of event types to discuss the value of mixture distribution approaches for rainfall modeling.

The fourth problem addresses parameter variation in temporal models, using daily streamflow as an example. Extremes of environmental variables are often generated by different processes, leading to the expectation that parameters for temporal covariates will vary across the variable's range. This section demonstrates the impact of process heterogeneity on model parametrization using a simple tree-based machine learning model. We explore whether an overall parametrization can provide fair estimates for the entire series or whether alternative parametrizations tailored to specific quantile ranges offer greater advantages.

2. Homogeneity in flood statistics

The first example we would like to introduce has its origin in flood statistics. Flood statistics is one of the major topics in hydrological research. The statistics form the basis for many practical approaches, often related to design flood estimation. Design floods, i.e., the flood quantile for a given probability or return period based on a time series of flood peaks, are used for multiple purposes, e.g., the dimensioning of flood protection structure or the operation of reservoirs (Volpi *et al.* 2024). Statistical theory of flood statistics dates back until the early work of Emil J. Gumbel, who proposed the use of extreme value statistics for the estimation of design floods (Gumbel 1941). By making use of this approach, flood quantiles for return periods that exceed the observation period by many years can be obtained, e.g., for a return period of 100 years, which would be the design event for a dam protecting settlements in Germany and Austria.

Traditional flood statistics usually make use of one of two main concepts: the Annual Maximum Series (AMS), a block maximum approach; and POT approach. Let X_i be a sequence of random variables representing a record of measurements x_i on a hourly or even higher resolution time step, and let Z_i be its maxima, either of blocks of n observations, or peaks over some high threshold u representing extreme events in the AMS and the POT approach (Coles 2001). On this basis, the AMS approach considers the maximum flood peak in each (hydrological) year. Thus, the probability of occurrence p obtained by this approach has a direct relationship to the return period T by $T = 1/p$. The theory of extreme value statistics provides the asymptotic distribution of these block maxima under proper normalization, the Generalized Extreme Value (GEV) distribution (Fisher 1928). The direct relation between T and p makes the AMS a frequently-applied approach in science and practice. However, it has the disadvantage that exactly one event per year is considered, leading to the neglect of other large events in the same year or the forced inclusion of small events in low-flow years. To overcome this issue, the POT approach is applied, where all events above a predefined threshold are considered. Here again, extreme value statistics provide the asymptotic distribution as the Generalized Pareto Distribution (GPD; Balkema and de Haan 1974; Pickands 1975). To obtain annual return periods for this approach, the non-occurrence probability of the flood peak has to be combined with the probability of a given number of events per year, e.g., in a Poisson-Pareto model. Both approaches, AMS and POT, assume independence of the events, which can be ensured by application of independence criteria such as a certain time between consecutive peaks. However, both approaches also assume identically distributed, homogeneous samples, i.e., *subject to a common set of forces* (Gumbel 1941). This is a crucial assumption that heavily impacts the estimation.

Hydrological processes in nature, though, are never homogeneous. Flood events can be caused by multiple and vastly different processes in either atmosphere or the catchment itself, and their interplay defines the flood events, especially the hydrograph shape. Typical examples of floods in Europe are heavy-rainfall floods, where rainfall events of high intensity but mostly only short duration cause steep and flashy flood events, or stratiform-rainfall floods, where long-duration rainfall over several days causes moist soils and a slow but large flood hydrograph. Additionally, floods can be also caused by snowmelt especially in the spring months, where the accumulated snow cover melts and contributes to the runoff (Jiang, Bevacqua, and Zscheischler 2022). Given these different processes and also the related different flood events, one cannot speak of a common set of forces.

It is therefore not surprising that there has been criticism on the assumption of homogeneity in flood statistics for quite some time. For example, already the Bulletin No. 17, which contains flood guidelines for the United States, questions the homogeneity assumption for catchments with different types of floods. Practical consequences, though, were only taken slowly. First attempts mainly focused on splitting the sample based on pure data distinction (Potter 1958) or by making use of a mixture distribution (Singh 1974; Rossi, Fiorentino, and Versace 1984). While the latter approach is promising from a statistical point of view, as it clearly provides distinct samples based on a best fit, the hydrological community was critical as clearly physical knowledge was missing. Alternatively, one could also apply Generalized Additive Models for Location, Scale, and Shape (GAMLSS) with cyclical splines to account for variations throughout the year, though again process knowledge might be missing here and the high number of parameters could introduce comparably high uncertainty. Only later approaches aimed to split samples according to their physical processes (Waylen and Woo 1982; Hirschboek 1988).

A statistical sound and nowadays more and more frequently applied approach is the split of the sample into different event types, the fit of separate distributions to each subsample and the combination of these distributions in a mixture model. The simplest split that still takes into account physical processes would be the split into seasons (e.g., Yan, Xiong, Liu, Hu, and Xu 2017; Veatch and Villarini 2022). This approach does not require process knowledge, nor does it usually reduce the number of events per sample when considering block maxima, as seasonal maxima are available for most catchments. For example, Fischer, Schumann, and Schulte (2016) split the sample of flood events into summer (Z_S) and winter (Z_W) maxima, fit a GEV distribution F_S and F_W to each sample and combine both distributions in a multiplicative mixture model for the whole year F_{mix}. The multiplicative model is chosen here, as one is interested in the maximum value (Todorovic and Rouselle 1971):

$$F_{mix}(z) = P(\max(Z_S, Z_W)) = P(Z_S)P(Z_W) = F_S(z)F_W(z).$$

Summer and winter in this case are defined according to the hydrological calendar year in Europe, i.e., winter starting in November and lasting until the end of April (in some regions until end of May to consider the snowmelt period), while summer starts in May (June) and lasts until the end of October. However, by using a seasonal split it is still not guaranteed that floods with different genesis are split in different samples, as for example in winter snowmelt- as well as rainfall-induced floods could occur. Moreover, it is difficult to define strict seasons in regions where, e.g., snowmelt occurs over a prolonged period in spring and under consideration that seasons are shifting with a changing climate. The above model also assumes independence of winter and summer events, which is not necessarily given with a strict split in seasons. To overcome these issues, the flood samples are often split according to some process-based classification, the so-called flood types. There exist numerous flood classifications with different focus points, e.g., on the hydroclimatic processes or the hydrograph (Tarasova *et al.* 2019; Fischer, Schumann, and Bühler 2019).

These different flood types then can be modeled in a mixture model. Such a model was proposed, e.g., by Fischer and Schumann (2023) or Yan, Xiong, Ruan, Xu, Yan, and Liu (2019). For a total number of J flood types, assume that for each flood type a corresponding sample of flood peaks $Z_1^{(j)}, \dots, Z_{n_J}^{(j)}$, $j = 1, \dots, J$ of sample size n_J associated with that flood type exists. First, each flood type sample is considered separately. For this purpose, a POT-approach is applied with a type-specific threshold u_j. The distribution of the exceedances of this threshold, $Z_i^{(j)} > u_j$, for each flood type j, $j = 1, \dots, J$, is chosen as the GPD, following the consideration above, defined as

$$G_j(z; \theta_j = (\kappa_j, \beta_j), u_j) = 1 - \left(1 + \kappa_j \left(\frac{z - u_j}{\beta_j}\right)\right)^{-\frac{1}{\kappa_j}},$$

for a shape parameter $\kappa_j \neq 0$ and scale parameter $\beta_j > 0$ with support $z \geq u_j$. The choice of the threshold u_j can be motivated by statistical consideration, e.g., using the mean residual life plot (Coles 2001), or by empirical or hydrological considerations. Fischer and Schumann (2023) propose u_j to be chosen type-specifically as thrice the type-weighted mean discharge.

The type-weighted mean discharge is derived from the monthly means of discharges weighted according to the relative frequency of the flood type in the respective months. The POT-approach can be generalized to an annual type-specific distribution $\tilde{G}_j$ for each flood type simply by application of the total probability theorem to obtain

$$\tilde{G}_j(z) = \sum_{k=0}^{\infty} P_j(l = k)(G_j(z; \theta_j, u_j))^k,$$

where $P_j(l = k)$ is the probability that the annual number l of flood peaks of type j above the threshold u_j is equal to k and can be described by the Poisson distribution with parameter λ_j,

$$P_j(l = k) = \frac{\lambda_j^k}{k!} e^{-\lambda_j},$$

as described above for the general POT approach. To obtain the annual joint distribution of all flood types, a mixture model is applied:

$$H(z) = \prod_{j=1}^{J} (G_j(z; \theta_j, u_j)(1 - F_j(u_j; \vartheta_j)) + F_j(u_j; \vartheta_j)).$$

The model is denoted as the Type-based Mixture model of Partial duration Series (TMPS). In this model, the POT distribution G_j is multiplied with the exceedance probability of the threshold $1 - F_j$. F_j is modeled by a GEV distribution based on all flood peaks in the sample of the respective flood type

$$F_j(z; \vartheta_j = (\xi_j, \mu_j, \sigma_j)) = \exp\left(-\left(1 + \xi_j \frac{z - \mu_j}{\sigma_j}\right)^{-\frac{1}{\xi_j}}\right),$$

for $1 + \xi_j(z - \mu_j)/\sigma_j > 0$, where $\xi_j \in \mathbb{R}$ is the shape parameter, $\sigma_j > 0$ is the scale parameter and $\mu_j \in \mathbb{R}$ is the location parameter. Alternatively, F_j can also be modeled by the empirical frequency to reduce the number of parameters in the model or by a Gamma distribution, which has been shown to be suitable for many discharge series.

In addition to the proposed mixture models, other approaches to incorporate inhomogeneity into flood statistics are also worth considering. This includes Bayesian hierarchical models or Generalized Linear Mixture Models (GLMMs). Currently, such models are mainly applied for regionalization purposes, modeling the inhomogeneity in space (Lima, Lall, Troy, and Devineni 2016; Hofrichter, Harum, and Friedl 2016; Sampaio and Costa 2021). A basic assumption in regionalization, i.e., the estimation of design variables such as floods in ungauged catchments by transferring information from gauged catchments, is the homogeneity within the groups of catchments. This approach can be seen as similar to having different flood types which have to be homogeneous each. However, such models usually require huge statistical knowledge and might thus be less easily applicable. Moreover, the type-based mixture model allows for an easy interpretation of the impact of each flood type, and thus also a link to the hydrological processes. This is why the focus in this study is placed on mixture models.

By making use of the type-specific distribution and the joint mixture model, it can be analyzed which flood type contributes most to the extremes of the distribution and thus to potentially heavy tails. For example, in Figure 1 two examples of the TMPS model and the respective type-specific distributions fitted to the flood peaks of two catchments in Europe are given, derived from the open access GRDC data base (The Global Runoff Data Centre 2024) and classified by a hybrid-hydrograph-based flood typology (Fischer *et al.* 2019). The first catchment, Saint-Cyran-du-Jambot/ Indre River in France, had discharge data available for the years 1968 to 2012. In total, 78 flood events were identified, 25 of which being associated with heavy-rainfall floods (R1), 18 to synoptic-rainfall floods (R2), 12 to sequence-of-rainfall

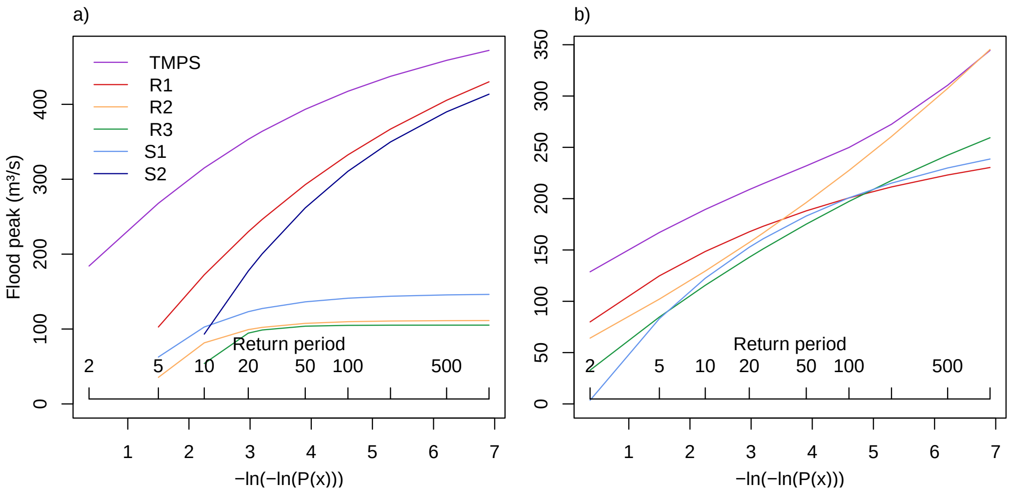

Figure 1: Two examples of the TMPS model fit to the flood peaks at a) Saint-Cyran-du-Jambot/ Indre River, France and b) Tenbury/ Teme River, Great Britain. Flood types are: R1 (heavy-rainfall floods with high convective share), R2 (synoptic rainfall floods), R3 (sequence-of-rainfall floods), S1 (rain-on-snow floods), S2 (snowmelt floods).

floods (R3), 14 to rain-on-snow floods (S1) and 9 to snowmelt-induced floods (S2). The second catchment, Tenbury/ Teme River in Great Britain, had discharge data available for the years 1956 to 2018. In total, 216 flood events were identified, 97 of which being of type R1, 61 of type R2, 30 of type R3, 28 of type S1 and none of type S2. It is clear to see that only certain flood types contribute to the right tail of the distribution, while others might be more relevant in the lower range of the return periods. The use of traditional block maxima approaches would mix these flood types in one sample and underestimate the right tail by considering a kind of leveled mean behavior. Moreover, information on the flood type also implicitly contains information on the flood volume. For example, for the Tenbury catchment one has to consider medium-sized volumes for the most extreme floods, as this is what can be expected for synoptic-rainfall floods. Flood types with larger volume, such as sequence-of-rain floods or snowmelt floods do not contribute much to the extremes.

The combination of flood types and a mixture model has many advantages in flood statistics. First, the improved homogeneity of the samples will lead to less violation of the statistical assumptions. With a meaningful flood classification, the genesis of flood events can be considered such that the events in one sample should have the same origin. Yan *et al.* (2019) showed that the performance of the statistical model increased when flood types were considered. But the model also has practical advantages: it incorporates physical information on the flood events, making the results easier interpretable and allowing for a plausibility check. The missing of physical knowledge in statistical models in flood statistics is a frequent criticism (Klemes 1971). In addition, also the information content of the statistics is increased, as implicitly not only one flood characteristics (the peak) is considered, but also information on the flood volume and the hydrograph shape are available.

However, there are also difficulties and drawbacks that come with this type of model. Compared to traditional flood statistical approaches, the number of parameters in the model is increased by a factor equal to the number of flood types. This can increase the uncertainty in the estimation. In a large Monte Carlo simulation study, Fischer and Schumann (2023) analyzed the cases, where the use of a traditional block maxima model is beneficial compared to the TMPS model in terms of bias and uncertainty. They showed that only in cases where the probability distributions of the different flood type samples are similar in shape and scale parameter and where only few events per year occur, the block maxima model leads to acceptable bias and uncertainty. As soon as one flood type distribution deviates from the remaining ones the block maxima model can no longer be recommended due to large bias. The same holds true in case of a large number of flood events per year. In both cases, application of the block maxima approach will lead to severe underestimation of the flood quartiles. So, in many cases mixture models can reduce the bias compared to traditional block maxima approaches (Fischer and Schumann 2023) Additionally, it was shown that the Root Mean Square Error (RMSE) of the TMPS model, despite its higher number of parameters, decreases faster than that of the block maxima approach for increasing sample sizes. This supports the assumption that inhomogeneity can heavily affect the estimation of flood quantiles and that even a higher uncertainty in the estimation is preferable compared to the bias in the block maxima model. In summary, there is definitely a trade-off between bias and variance. Even for the small sample sizes one usually has in flood statistics, the bias when using mixture models is reduced significantly compared to block maxima approaches. This comes at the cost of higher RMSE and thus variance. The latter can only be reduced by increasing sample sizes, which offers a perspective for the future in which increasing data samples might reduce RMSE.

3. Mixture distribution approaches for low flow

Our second example focuses on the frequency analysis of drought events. Similar to floods, low flows are driven by a complex interplay of processes, but here the separation into process types is more straightforward. In a seasonal climate, two different types of low flows can be distinguished. Summer low flows are triggered by a precipitation deficit that occurs during prolonged periods of dry weather, leading to the drying out of soils and the depletion of groundwater and other stored sources. Such droughts are often associated with heat waves, which increase evaporation and thus water deficits and the severity of events. Summer low flows are therefore more pronounced in lowlands and occur over large spatio-temporal scales. Winter low flows, on the other hand, are caused by long periods of frost when water is stored in snow and ice. These events are most pronounced in higher mountains and occur on smaller spatio-temporal scales.

Clearly, such mixing of processes violates the homogeneity assumption of common statistical models and may lead to inaccurate conclusions. There have been many attempts to deal with process heterogeneity in spatial low-flow models. Early studies focused on regression methods that divided the study area into groups or regions and fitted individual regressions to each. This is known as a regional regression approach (Institute of Hydrology 1980). The spatial separations were established based on soil classification (Gustard and Irving 1994), cluster analysis of several catchment characteristics (Nathan and McMahon 1990) or the seasonal occurrence of low-flow events (Laaha and Blöschl 2006). Surprisingly little efforts have been made for considering process heterogeneity for low-flow frequency analysis. Recently, two different approaches have been developed.

In the first approach, a mixture distribution model for annual minima series has been adopted (Laaha 2023b) and implemented in R (R Core Team 2024). It assumes independent and seasonally separable types of low-flow events. Serial dependency has been found to be weak for annual minima series related to hydrological years starting at the end of the groundwater recharge period and is typically neglected in studies. Exceptions occur in climates prone to multiannual droughts, where block sizes longer than one year can be chosen to reduce serial correlation of events. However, unlike flood frequency analysis, a POT (or rather "pit under threshold") approach was not considered because of the long time scales of droughts and groundwater processes in the catchment, which make low-flow periods in the same season serially dependent.

In the approach, the annual minima series

$$M_n = \min\{X_1, \ldots, X_n\}$$

is regarded as the minima of the annual summer minima $M_{n,S}$ and winter minima $M_{n,W}$:

$$M_n = \min\left\{M_{n,S}, M_{n,W}\right\}.$$

In this notation, X refers to the original variable (daily flow), and M_n to its minima within each block of size n (typically one year). Assuming that summer and winter events are independent of each other, the probability of occurrence of an event of magnitude z is obtained by multiplying its respective non-occurrence probabilities in the summer and winter seasons (Stedinger, Vogel, and Foufoula-Georgiou 1993). This leads to the definition of the mixed probability p_{mix} for minima:

$$p_{mix} = 1 - (1 - p_S)(1 - p_W),$$

where p_S and p_W are the respective probabilities of occurrence in the summer and winter seasons. These can be estimated using an empirical probability estimator based on the rank m of the event z within the (increasingly ordered) sample of size n. The theoretical probability estimator is obtained when inserting the marginal extreme value distributions of the summer and winter series, $F_S(z)$ and $F_W(z)$. The Cumulative Distribution Function (CDF) $F(z)$ of the annual low flow is estimated as

$$F_{mix}(z) = 1 - \{1 - F_S(z)\}\{1 - F_W(z)\}.$$

A Weibull (Extreme Value type III) distribution is used as the probability model as this is the limiting distribution for minima according to the Fisher-Tippett-Gnedenko theorem. The mixed Weibull-model for minima can be written as

$$G_{mix}(z) = 1 - \left\{\exp\left[-\left(\frac{z - \zeta_S}{\beta_S}\right)^{\delta_S}\right]\right\}\left\{\exp\left[-\left(\frac{z - \zeta_W}{\beta_W}\right)^{\delta_W}\right]\right\}.$$

The $\zeta_. > z$, $\beta_. > 0$ and $\delta_. > 0$ are the location, scale and shape parameters of the summer (index S) and winter (index W) marginal distributions.

The properties of the mixed probability estimator are demonstrated for two example stations in Figure 2. The fact that the seasonal distributions differ strongly reflects the effect of process heterogeneity on the frequency analysis. In Figure 2a, the conventional annual distribution deviates from the mixture distribution and is expected to produce biased results. Such a bias would not be evident from the empirical distribution of the annual series, which in both example stations show no signs of irregularity. The biases are most pronounced for highly mixed regimes, which are typically associated with mountain forelands (Figure 2a). They tend to be much smaller for strong summer regimes in the lowlands and winter regimes in the high mountains, with the differences eventually disappearing when the annual series consists only of summer or winter events (Figure 2b).

Due to the long time scale of the events, it is unlikely that summer and winter low flows are completely independent. The second approach extends the mixture distribution model to capture seasonal dependence using a copula estimator (Laaha 2023a). Copulas have been used in streamflow models either for a multivariate event characterization (e.g., volume and duration) or to aggregate streamflow along the river (Fischer and Schumann 2021). The use of copulas to account for seasonal dependence in extreme value statistics constitutes a novel approach. It uses a bivariate probability model to calculate the joint probability that summer low flows S and winter low flows W are less than or equal to magnitudes s and w, respectively:

$$P(S \leq s, W \leq w) = F_{S,W}(s, w).$$

Following the theorem of Sklar (1959), the joint CDF can be written as:

$$F_{S,W}(s, w) = C\left[F_S(s), F_W(w)\right]. \quad (1)$$

The approach requires the choice of a copula model C, with the Gumbel-Hougaard copula being considered an appropriate choice for modeling the mixture of summer and winter low flows (Laaha 2023a). Alternatively, different families can be considered and the most appropriate model is selected using on a copula information criterion (Grønneberg and Hjort 2014).

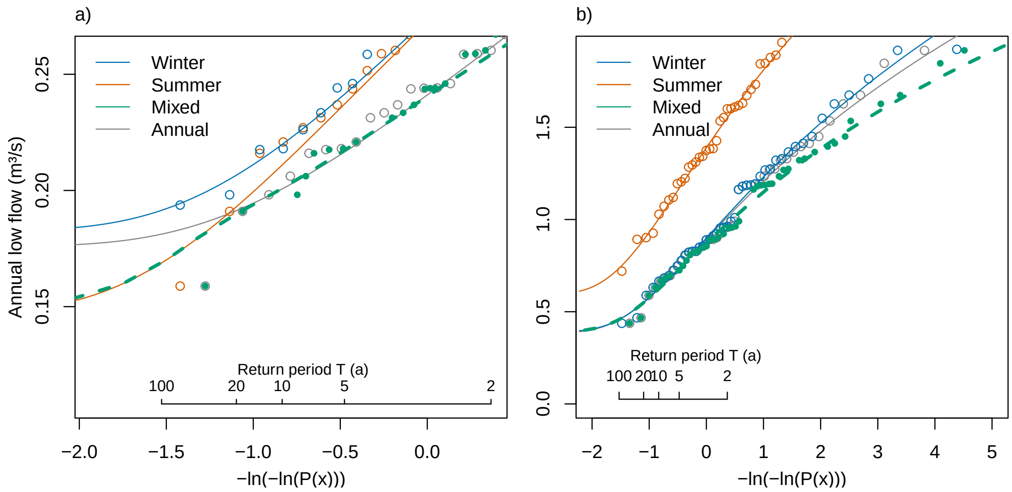

Figure 2: Annual versus seasonal and mixture distribution approaches for low-flow frequency analysis for (a) gauge Weg, river Isen in Germany and (b) gauge St. Peter-Freienstein at river Vordernberger Bach. Empirical distributions (circles) and fitted Weibull distributions (lines) for annual and seasonal (summer and winter) AMS series. "Mixed" refers to annual estimates calculated using the mixed probability estimator (Laaha 2023b).

The probability of event z occurring in any season is finally obtained by the mixed copula estimator

$$F_{mix,C}(z) = F_S(z) + F_W(z) - C[F_S(z), F_W(z)]. \tag{2}$$

As for the mixed probability estimator, a Weibull distribution is used to model the marginal distributions of summer and winter events. In the same way as for the theoretical estimator, we can also define a generalized empirical estimator. It can be formulated in accordance with the theoretical probability estimator in Equation (2) as

$$p_{mix,C}(z) = p_S(z) + p_W(z) - C_n\left[p_S(z), p_W(z)\right],$$

where $p_S(z)$ and $p_W(z)$ denote the empirical probability of z to occur in the summer season and the winter season, respectively. The $C_n(\cdot)$ is the empirical copula, which defines the empirical multivariate distribution in analogy to Equation (1).

The actual value of the mixture distribution approaches has been investigated in two studies. Laaha (2023b) evaluated the performance of the mixed probability estimator using a dataset of long-term (1976–2010) records of daily streamflow at 329 gauging stations in Austria. The study focused on the change in the return period obtained by two different models. With a relative deviation of 21%, 39% and 63% when estimating low flows with return periods of 20%, 50 and 100 years, the differences to the conventional annual distribution estimator were found to be large. For the 100-year event, 41% of the stations show a performance gain (reduction in the absolute relative deviation of the inferior to the superior model) of more than 50% when using the mixed probability estimator. The success of the model is strongest for highly mixed regimes, but was found to be relevant not only for mountain forelands, but for a wide range of catchment conditions. The effects are strongest for high return periods, where the results vary widely across all seasonal regimes.

The relative advantage of the mixed copula estimator over the other two approaches has been demonstrated by Laaha (2023a). The difference between the two mixture models decreases as the return period increases, so that for the most severe events the two models are in agreement. However, for moderate events, the comparison reveals a bias in the mixed probability estimator, which can be corrected using the copula approach. The bias can be related to the stronger serial correlations found for the moderate events and almost no correlation for the most severe events. The mixed copula estimator is a valid generalization of the mixed probability estimator that relaxes the assumption of independent events. Due to its favorable characteristics, it should be used as a new standard for frequency analysis of extreme low-flow events.

The mixed probability estimators have many conceptual advantages. They provide an efficient way to deal with sample heterogeneity, which is considered to be a major obstacle to meeting the assumptions of the annual minimum approach to low-flow frequency analysis. Their approach is also more process-oriented, leading to more realistic models and allowing easier plausibility checks. However, perhaps the greatest advantage of the mixture models is that they provide a consistent framework for summer, winter and annual events. Seasonal characteristics are directly relevant to a number of water management tasks, including environmental flows and hydropower generation. For other problems, annual characteristics, which describe the absolute minimum regardless of the season, are more relevant. The mixture models combine the seasonal distributions into a valid annual distribution, which allows the calculation of consistent return periods for summer, winter and annual events.

4. Clustering rainfall types for stochastic simulations

Our third example discusses process heterogeneity in rainfall modeling. Long-term rainfall records are essential for a variety of applications, where rainfall series serve as input in process-based models. These models are instrumental in predicting runoff for locations or time periods lacking direct runoff observations and in generating runoff scenarios under changing system conditions. Other model applications include natural hazards such as soil erosion or mass movements caused by heavy rainfall, and drought impact on soil moisture or groundwater levels during periods of low rainfall. However, in many instances, observational records of rainfall are too short to yield reliable conclusions. To address this limitation, stochastic rainfall generators have been developed to provide simulated rainfall data.

A variety of methods are employed for rainfall generation. Classical approaches typically focus on point rainfall and consist of two coupled components: a rainfall occurrence model, often represented by a Markov-chain model, and a distribution model for rainfall intensities (Wilks and Wilby 1999). Rainfall intensities are frequently modeled using a Gamma distribution, although this approach tends to underestimate heavy rainfall events (Furrer and Katz 2008; Papalexiou, Koutsoyiannis, and Makropoulos 2013). To better capture the rainfall distribution, especially for sub-daily rainfall, more heavy-tailed distributions such as the Mixed Gamma and Generalized Pareto distributions have been developed (Vrac and Naveau 2007).

However, rainfall is also subject to important process heterogeneity. A frequently used typology distinguishes between convective and stratiform events. Convective events are storms that typically occur in summer when moist air moves upward due to heating and condenses to produce rainfall. They tend to be of short duration and small spatial extent, but typically result in intense rainfall. Stratiform events occur with weather fronts when warm air masses pass over cold air masses. This can lead to longer but less intense events, but the event characteristics differ between cyclone types, with cold fronts leading to a faster transition and warm fronts tending to lead to longer events. Similar processes occur with orographic lifting, where moist air is forced to rise over a mountain range or elevated terrain.

Since rainfall is generated by such different processes, we can expect that rainfall series will be composed of events with different distributions. But the separation is not straightforward. Here, we test an event separation based on features derived from Yevjevich's theory of runs (Yevjevich 1967). It provides a systematic approach to exploit the temporal characteristics of rainfall events for event classification. Originally proposed for droughts, this method has recently been applied to analyze cold spells and electricity production deficits, such as those underlying the 2021 Texas winter blackout (Gruber, Gauster, Laaha, Regner, and Schmidt 2022).

To adapt this method for rainfall events, a threshold is defined above which rainfall is considered significant. Rainfall events are then detected as temporally contiguous wet spells exceeding this threshold. These events are characterized by several features, including peak magnitude, duration, severity (i.e., total rainfall volume), average rainfall intensity (total volume divided by duration), and time to peak as a ratio of event duration. These characteristics are subsequently used for event classification. During a rainfall event, short pauses in rainfall or disturbances can artificially split the event into several interdependent events. To avoid such splitting, pooling criteria are applied, using an inter-event time and volume threshold of 1.27 mm in 6 hours, a criterion commonly used in soil erosion studies (Vásquez *et al.* 2024).

The obtained event characteristics contain complementary information about the temporal dynamics that should be related to the generating processes. They appear useful to separate the rainfall series into types of events. The event information is supplemented by a binary lightning index derived from the European lightning location system (EUCLID), which has a detection efficiency of 96% (Schulz, Diendorfer, Pedeboy, and Poelman 2016) for downward flashes. An event has a lightning index of one when lightning was recorded in a circular neighborhood of 7 km during the event.

Separation of events into types can be informed by unsupervised learning methods such as cluster analysis. The aim is to obtain classes that are homogeneous not only with respect to the rainfall distribution, but also with respect to the multivariate signature of the temporal event characteristics. In the analysis presented here, two methods of cluster analysis were considered.

The first method is model-based clustering. We used the method implemented in the R package **MixAll** (Iovleff 2024), which is suitable for handling large data sets. A Gamma mixture model was adopted, taking into account the following event characteristics: peak magnitude, duration, and severity. Other event characteristics were not included as they were found to be redundant in a preliminary analysis. The lightning index was also not included at this stage. The data were clustered based on the conditional independence assumption. The optimal number of clusters was found by the ICL (Integrated Complete-data Likelihood) criterion (Biernacki, Celeux, and Govaert 2000).

As a second method, the robust Partitioning Around Medoids method (PAM; Kaufmann and Rousseeuw 1990) was tested. Here we considered all event characteristics and additionally the binary lightning index for clustering. Each variable k ($k = 1, \ldots, p$) was entered as log-transformed values, except for the lightning index. Given the mixed data types, we used Gower's distance (Gower 1971) to define the (dis)similarity between events:

$$S_{ij} = \frac{\sum_{k=1}^{p} w_{ijk} s_{ijk}}{\sum_{k=1}^{p} w_{ijk}},$$

where the w_{ijk} are weights (here set to 1) and s_{ijk} is the similarity between the the i-th and j-th event regarding their k-th variable. For continuous event characteristics, the similarity corresponds to the range-normalized Manhattan distance (after logarithmic transformation):

$$s_{ijk} = 1 - \frac{|x_{ik} - x_{jk}|}{R_k}$$

with R_k being the range of the k-th variable. For the binary lightning index, the s_{ijk} take values 0 or 1, with 1 indicating equality of the two events with respect to the lightning index. All variables were given equal weight in the distance metric. Cluster analysis was performed with different numbers of clusters, and the optimal number of clusters was determined by the silhouette plot and the average silhouette width.

The event classification was tested on 27 stations in Austria. Here we report on the station Geboltskirchen, located in the foothills of the Alps in Austria, which represents a typical behavior seen in the data set. The analyses were carried out for sub-daily time series with a 5-minute resolution, covering the summer months (April to October) from 2011 to 2020. A threshold of 1.27 mm was chosen to identify rainfall events, resulting in a total of 869 events encompassing 12,487 5-minute individual rainfall measurements.

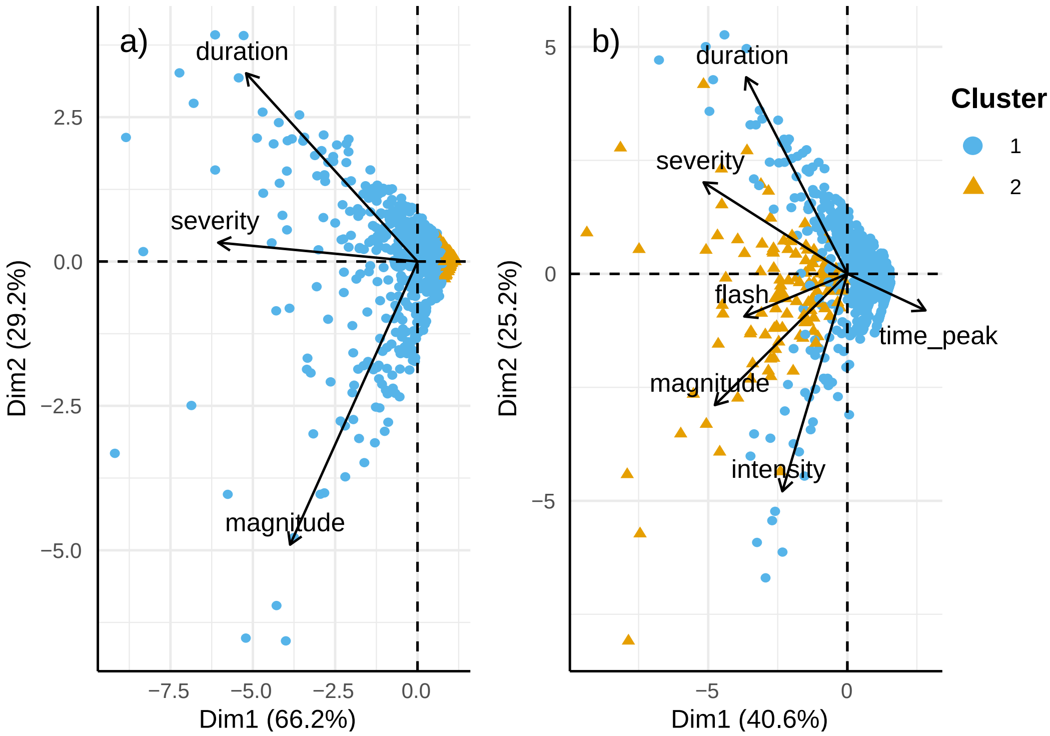

Figure 3: Rainfall event clustering for the Geboltskirchen station. (a) Model-based clustering without lightning. (b) PAM clustering including lightning. Analyses are based on the following event characteristics: peak magnitude (magnitude), duration (duration), total rainfall volume (severity), average rainfall intensity (intensity), relative time to peak as a ratio of duration (time_peak), and the lightning index (flash). The biplots display rainfall events as points and variables as vectors along the first two principal components. The values in parentheses refer to the portion of variance explained by each principal component.

Figure 3 shows the results of rainfall event clustering. For model-based clustering (Figure 3a) two clusters were optimal based on the ICL criterion. However, the clusters show only small differences in the distribution of rainfall intensities (Figure 4a). The allocation to rainfall types is not straightforward, as convective events (lightning) are distributed across both clusters. PAM clustering including the lightning index, on the other hand, produces two clusters with a clear separation between convective and stratiform events (Figure 3b). This can be seen from the lightning index, which shows an exact separation into events with and without lightning. The convective events (Cluster 2) have a higher average intensity, peak magnitude and overall severity. The clusters also have quite different rainfall distributions. These demonstrate the effectiveness of event clustering using a combination of event characteristics derived from the rainfall record and the lightning index representing the connectivity of events.

The analysis shows that event clustering is a straightforward yet effective method to reduce process heterogeneity in rainfall frequency analysis. In our approach, the event characteristics are obtained directly from the rainfall records, without requiring additional weather data, which would necessitate interpretation of weather maps and satellite data. Consequently, these characteristics are easier to obtain than those in other process classifications that require additional steps in the analyses, which is considered a major advantage of our approach. The potential of mixture-model-based clustering can be further enhanced by incorporating lightning data, which will be investigated in a subsequent study.

While this study was limited to an exploration, the methods can be extended to perform more effective classification of rainfall distributions. Finite mixture models, such as those implemented in package **flexmix** (Leisch 2004; Grün and Leisch 2025), can be used for classification tasks conditional on event characteristics. These models appear promising and will be tested in subsequent research. Alternatively, GAMLSS models with cyclical splines could be

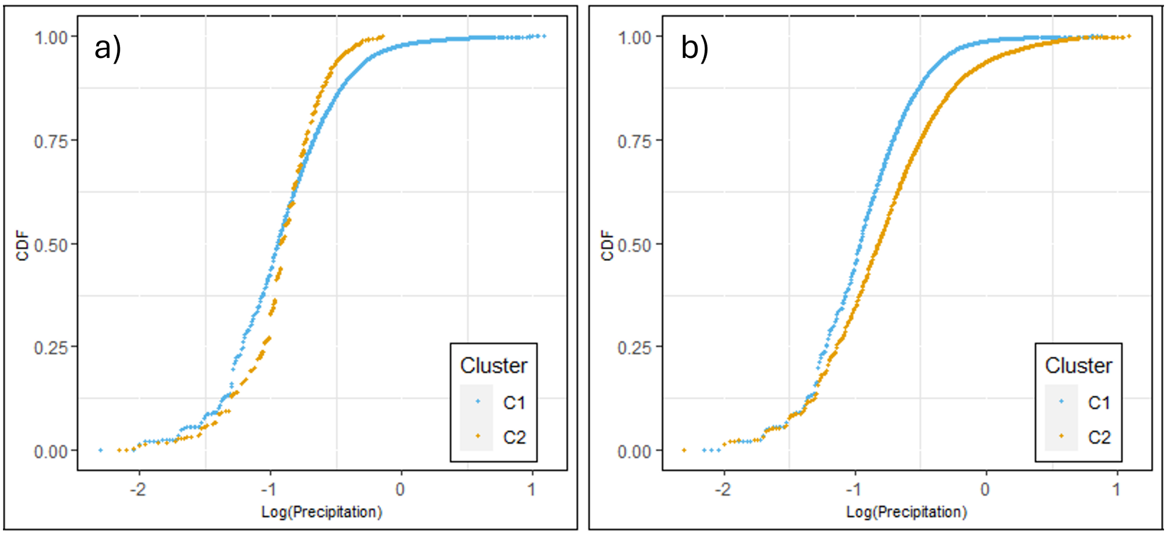

Figure 4: Rainfall distributions for station Geboltskirchen by cluster (a) Model-based clustering without lightning, (b) PAM including lightning.

considered to account for seasonal variations, but this approach would be less process-based than taking the different characteristics of rainfall types into account.

5. Quantile regression for temporal streamflow modeling

Our last example focuses on the prediction of the time series of daily river discharges. This is often referred to as rainfall-runoff modeling because one of the dominant drivers of runoff is precipitation. We can mainly distinguish between physically based models (Guo, Zhang, Zhang, and Wang 2021; Razavi and Coulibaly 2013) and stochastic approaches, on which we will focus here (Solomatine and Ostfeld 2008; Kratzert, Klotz, Herrnegger, Sampson, Hochreiter, and Nearing 2019a; Kratzert, Klotz, Shalev, Klambauer, Hochreiter, and Nearing 2019b; Lees *et al.* 2021). The simplest case for both modeling approaches is to train the model on a subset of the time series and evaluate it on the remaining dataset. This can give an indication of how well the fitted models will predict outside the calibration period or extrapolate if only a few years of data are available. Training these models usually involves optimizing some predefined loss function. Most commonly, some adaptation of a mean squared error called the Nash-Sutcliffe Efficiency (NSE; Nash and Sutcliffe 1970) is used to train the models, which is basically the coefficient of determination:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}, \quad (3)$$

where $\hat{y}_i$ is the predicted and y_i the observed time series of daily discharge with n observations. If modelers want to give more weight to the lower part of the distribution, optimization is done using the logarithm of observations and predictions known as a NSE_{log}-criterion. Although these methods approximate average runoff conditions well, they fail to estimate extreme events (such as floods and droughts). The use of some asymmetric loss functions such as quantile regression (Koenker and Bassett 1978) or expectile regression (Aigner, Amemiya, and Poirier 1976; Newey and Powell 1987) is less common – exceptions include Toth (2016) and Laimighofer, Melcher, and Laaha (2022) for prediction, or Tyralis, Papacharalampous, and Khatami (2023) or Papacharalampous *et al.* (2019) for probabilistic forecasting. However, these options could improve the estimation of extreme events, which are often the main focus of studies. In this study, we aim to demonstrate the impact of the loss function on model performance, variable selection, and variable importance. We hypothesize that there is no single model that is optimal for the full range of discharge or environmental variables in general, which has important implications for building interpretable models that capture processes in a realistic way.

This case study uses 184 stations in Austria with a daily discharge time series from 1982 to 2017, provided through an open source dataset (LamaH-CE) by Klingler, Schulz, and Herrnegger (2021). We train an XGBoost model (Chen and He 2015; Chen and Guestrin 2016; Chen *et al.* 2021) for the sub-period 1982–2000 and evaluate the model on the remaining years. As predictor variables, we used daily mean temperature, evapotranspiration, soil moisture, and precipitation. Rolling sums or averages of these variables were added for one to five days, to mimic the short-term memory of daily discharge. Additionally, accumulations over 15, 30 and 60 days were used as predictors to capture long-term processes. Further, mean annual monthly and daily meteorological quantities are included to characterize seasonality. Finally, the month of the year and the day of the year are added as numeric variables.

Our assessment is carried out in two steps. First, we test the ability of a quantile loss function to approximate the high and low extremes of the time series. This is done by training the model with a quantile error loss of $\tau = 0.1$ for the lower quantile, and $\tau = 0.9$ for the higher quantile, and comparing these to the squared error loss (i.e., Ordinary Least Squares (OLS) or mean regression approach). The response at each station was transformed by the square root, which already gives more weight to the low values for the mean regression approach. The second analysis aims to investigate how different optimization strategies modify the selected variables and their contributions to the predictions. For this purpose, we identify the 25 variables that most significantly contribute to the reduction of the gradient loss function in the models built for the years 1982 to 2000. These top predictors are then used to train a final model on the dataset from 2001 to 2017 at each station. The goal of this final model is to analyze how the predictor contributions change over the range of predictions. This is achieved by computing SHAP (SHapley Additive exPlanations) values (Aas, Jullum, and Løland 2021; Lundberg and Lee 2017). The multivariate dependency structure between the predictor variables is hereby modeled using a conditional forest approach (Hothorn, Hornik, and Zeileis 2006; Hothorn and Zeileis 2015). The code, data and final results can be found in the accompanying Zenodo publication (Laimighofer 2024).

Starting with the most common metric in statistical hydrology, we compare the models using R^2 (Equation 3). Calculated for each station (and the test data period of 2001–2017), we obtain a median over the entire dataset of 0.62 for the mean regression. The median R^2 for the lower quantile regression ($\tau = 0.1$) is 0.04, and 0.43 for the higher quantile regression ($\tau = 0.9$). When models focus on low-flow events, it is often common to report the R^2 of log-transformed streamflow, using $\text{NSE}_{\log}$ as the performance measure. Not surprisingly, the $\text{NSE}_{\log}$ would give a higher median performance score for the mean regression (0.66) and the lower quantile regression (0.20), since the models were optimized by the square-root transformed daily discharge. The higher quantile regression, on the other hand, would receive a lower median performance ($\text{NSE}_{\log}$ of 0.36).

Both performance measures undoubtedly favor the mean regression approach; however, average conditions are often not the major concern in hydrology. Extreme events such as droughts or floods pose a higher risk to society and nature. Hence, we propose to take into account a broader mix of error metrics, focusing on different streamflow quantiles, to give a more holistic picture of the potential of using quantile regression. First, we will consider the Quantile Loss performance (QL_τ) adapted from Koenker and Bassett (1978), which is analogous to a pseudo R^2:

$$QL_\tau = 1 - \frac{\sum_{i=1}^{n} L^{\tau}_{QL,i}(y_i, \hat{y}_i)}{\sum_{i=1}^{n} L^{\tau}_{QL,i}(y_i, q(y_i, \tau))},$$

where $q(y_i, \tau)$ is used as the naive estimate of the model considering the specific quantile of

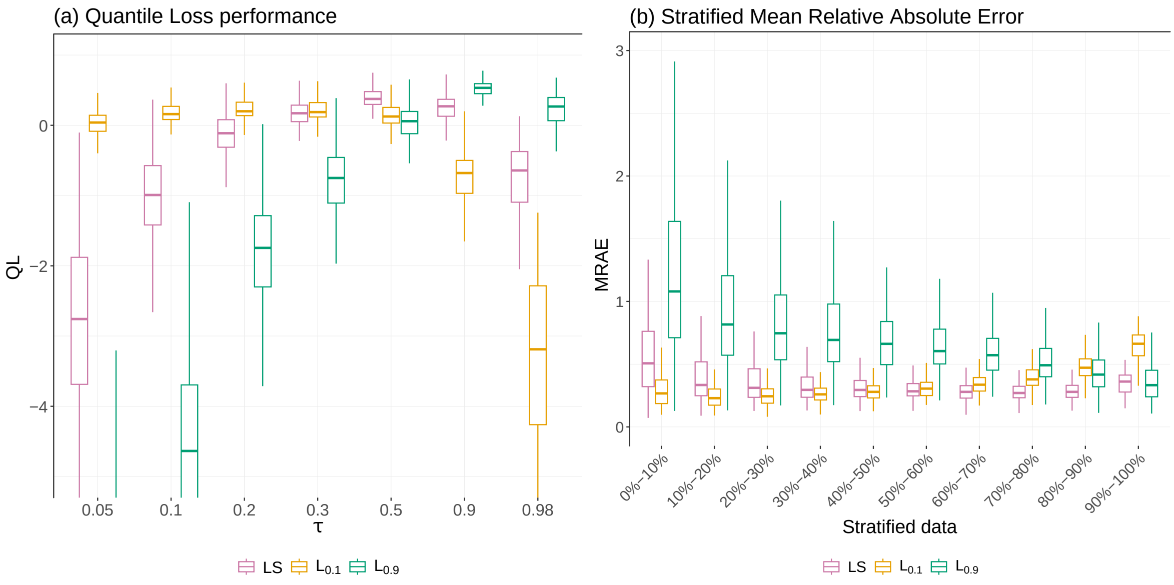

Figure 5: Model performance for lower to upper time series quantiles depending on model optimization using the quantile loss criterion. (a) Quantile Loss performance (QL) over all stations vs. the τ-th streamflow quantile (values < -5 not shown). (b) Mean Relative Absolute Error (MRAE) over all stations stratified into streamflow decile bins.

the time series and L^{τ}_{QL} is the quantile loss function, given by:

$$L^{\tau}_{QL,i} = \begin{cases} \tau \cdot |y_i - \hat{y}_i|, & (y_i - \hat{y}_i) \geq 0, \\ (1-\tau) \cdot |y_i - \hat{y}_i|, & (y_i - \hat{y}_i) < 0, \end{cases}$$

where $\hat{y}_i$ are the predicted quantiles at time step i. A value of 1 indicates that the observations are perfectly fit by the predictions, and a value below 0 indicates that the model predictions are worse than the naive estimate.

As expected, Figure 5a shows that for low streamflow quantiles represented by low τ values, our lower quantile estimator ($L_{0.1}$) provides the best approximation, while for high τ values the upper quantile estimator ($L_{0.9}$) performs best. The mean regression (LS) performs well only in the middle part of the distribution. The results indicate that for low streamflow quantiles ($\tau \leq 0.3$), the lower quantile estimator outperforms the mean regression approach. Additionally, for low τ values, the mean regression approach yields a strongly biased estimate (as indicated by negative QL values), whereas for higher τ values, such strong biases are only observed for the most extreme quantiles (τ of 0.98, representing flood events).

By stratifying the observations into 10 equal parts of the empirical distribution, we can obtain a complementary view of the models' performance. For each of these parts, we calculate the Mean Relative Absolute Error (MRAE). The results are shown in Figure 5b. For the lower 50% of the data, the lower quantile regression approach reveals advantages over the mean regression approach. In contrast, the higher quantile estimate provides only a slightly better estimate of the upper 10% of the data. The use of an expectile regression approach would place even more emphasis on the highest values and thus improve the estimates for the upper extremes.

In the second assessment, we investigate how the predictor contributions change over the full range of data, depending on the regression approach. For this purpose, we consider the 25 most important variables and compute their SHAP values for $\tau = 0.01, 0.02, \ldots, 0.99$ quantiles of each daily discharge time series, where τ are the respective quantiles of the quantile loss. As the baseline estimation (mean or quantile) for each regression approach differs, we estimate the relative contribution for each variable, defined by the contribution divided by the sum of all absolute contributions, and summarize these relative SHAP values for each decile of the

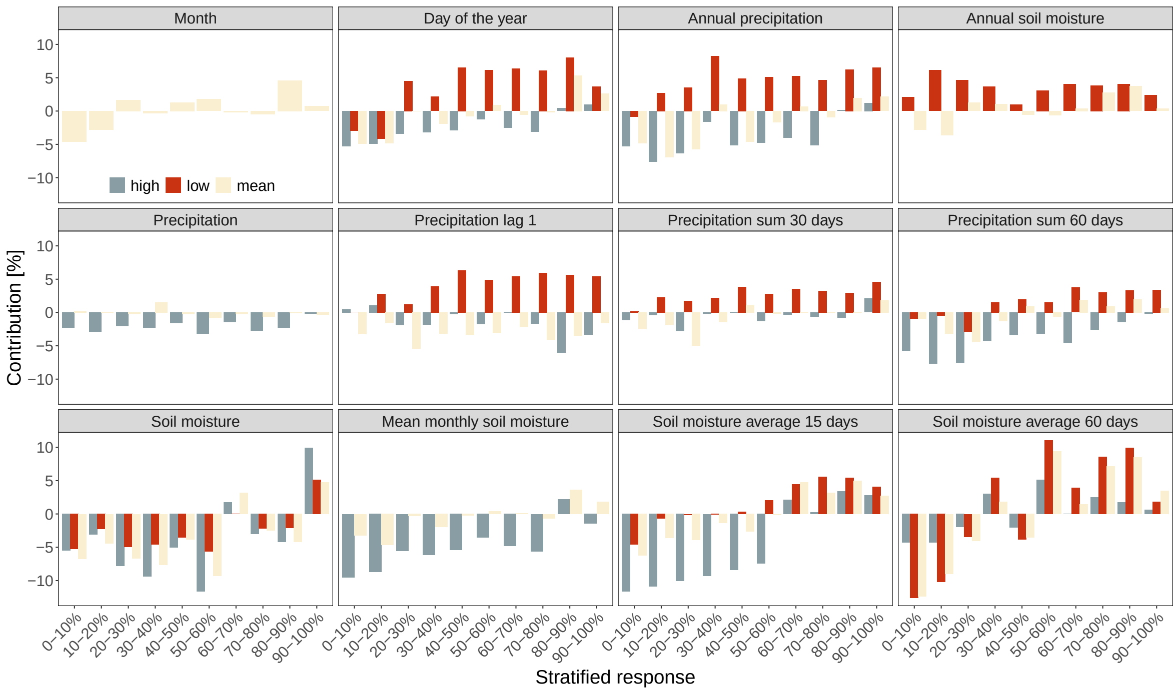

Figure 6: Relative averaged contribution of the main predictors for each decile of the stratified response. Results are shown for the river Kamp for a mean regression approach (mean) and the two quantile regression approaches (high and low). An overview of the estimated contribution of all predictors can be found in Laimighofer (2024).

distribution.

The analysis is carried out for the river Kamp near Zwettl (results are summarized in Figure 6). We expect low flows in the catchment to be triggered by prolonged meteorological droughts, often accompanied by high temperatures and consistently low precipitation. Floods, on the other hand, are expected to be triggered by relatively short but intense precipitation events in combination with high antecedent soil moisture. For low-flow events, we are primarily interested in SHAP contributions that lower the baseline estimate, i.e., have a negative effect on the prediction. Seasonal information, such as the month and the day of the year, decreases the prediction for the mean regression (about 10%), but only by about 3% for the low quantile regression. This suggests that seasonal information can actually provide some adjustment to the mean regression for the low-flow season, while the quantile regression is already adjusted for the low-flow season and requires no further adjustment. The typically low precipitation in summer (lags and accumulated transformations) has a negative contribution for the mean regression, but has no or a positive contribution for low quantile regression. This indicates that the low precipitation is only essential for lowering average streamflow, but not for extreme events below the 0.1 quantile. One main contributor to low streamflow is short-term and accumulated (15–60 days) soil moisture. This is true for the mean regression and the low quantile regression, suggesting that the deficit in soil moisture was found important by both approaches. Regarding high streamflow events, the main predictor for both the high quantile regression and the mean regression is soil moisture, considering current and antecedent conditions with time lags of up to 7 days. Differences between the two regression approaches, however, are difficult to discern.

Overall, the example demonstrated how an asymmetric loss function can provide higher prediction accuracy for the tails of the distribution which correspond to the extremes of the time series. It also illustrated how the contribution of the predictors changes across the range of the variable of interest. The asymmetric loss functions resulted in different parametrizations of the models, all of which were physically interpretable. Consequently, quantile regression approaches offer a pathway to more process-oriented statistical models when combined with a process-oriented preselection of predictors. This is exemplified by the station at the river Kamp, where the variable contributions at low and high quantiles could be linked to key drivers of low flows and floods. However, a limitation of quantile regression in a time series context is the detection of too many extreme events, leading to a high rate of false positives. Therefore, selecting an appropriate τ value is critical to avoid such overdetection.

6. Discussion and conclusion

Nature is complex, and environmental systems are characterized by intricate interactions of processes that can vary significantly between different states of the system, such as between summer and winter, or dry and wet conditions. This process heterogeneity poses a major challenge for all types of models. Although many standard statistical models assume homogeneity, this assumption is often invalid, leading to inaccurate model results. This study explores the implications of process heterogeneity through problems related to natural hazards and earth system sciences. It presents various approaches to account for this heterogeneity.

For example, typical applications of extreme value statistics, such as hydrological extreme events, are influenced by different forces. Thus, different types of maxima or minima distributions are to be expected, and classical approaches to frequency analysis that assume sample homogeneity may be misleading. In our case study on floods, the extreme value distributions differed strongly when POT samples were stratified according to flood types related to snow, rainfall type and antecedent wetness. Other studies have shown that considering flood types can decrease the uncertainty in the estimation of design floods for large return periods compared to traditional methods which do not consider process heterogeneity (Yan *et al.* 2019, 2023). And even though the larger number of parameters in type-based mixture models can increase the uncertainty in the estimation, it was demonstrated in the discussion above that the consideration of flood types in flood statistics leads to a smaller bias as soon as there is one flood type that deviates from the remaining flood types regarding its statistical distribution (Fischer and Schumann 2023). Moreover, the flood types make a different contribution to the lower and upper tail of the overall frequency distribution of floods, leading to a complex shape and heavy tail that may deviate substantially from a GEV of a homogeneous sample. It is well-known that the mixture of distributions with different tail behavior can amplify the overall tail behavior (Merz *et al.* 2022). This can then again violate many standard approaches in statistics, such as bootstrap or ordinary moments (Vogel, Papalexiou, Lamontagne, and Dolan 2024). The use of mixture models allows to consider the origin of this heavy tail behavior and provides a physical explanation for a statistical phenomenon, the latter being often criticized to miss in statistical hydrology (Klemes 1971).

The same holds true for low flows, where the extreme value distributions differed strongly when annual series were calculated for summer and winter seasons as there is a clear differentiation between generating processes in summer and winter. Similar to the mixture of flood distributions, Laaha (2023b) showed for low flows that a frequency distribution can become biased as soon as single events from a different process type are mixed in. The study gives a quantification of the possible errors made when using a classical frequency analysis approach that neglects different process types and fits the extreme value distribution to the heterogeneous sample. Errors were found to be particularly large for extreme events with long return periods, for mixed alpine and lowland regimes, and for climates with cold winters and warm summers. So again, a physical explanation of the observed statistical behavior can be found. A possible drawback of incorporating process heterogeneity by mixing seasonal distributions is the greater dependence between seasonal events compared to annual events, which has resulted in a considerable bias in the upper tail of the distribution at a number of stations. Correlation between annual low flow events is usually weak when series are derived for hydrological years starting in string when groundwater stores are refreshed, and therefore typically neglected in extreme value statistics. A copula-based estimator can additionally incorporate event dependence and has been shown to provide robust estimates not only for extreme events, but also for moderate events. An additional benefit of the mixture distribution approaches is that they allow the calculation of consistent return periods on an annual and seasonal scale, all of which are relevant indices for water management (Van Lanen *et al.* 2016; Laaha *et al.* 2017).

The physical difference between events was not only found on the hydrological scale. It was shown that already in the atmospheric part of the water cycle distinct differences between processes were visible, which was demonstrated by the third example. Similarly to the flood example, the consideration of atmospheric covariates helped to stratify different types of rainfall, providing a direct link to the different types of flooding. Apart from the classification based on rainfall event characteristics proposed here, there were only relatively few attempts to address heterogeneity in rainfall distributions so far. One approach is rainfall clustering based on temporal distributions of rainfall events described by the so-called Huff curves (Dolšak, Bezak, and Šraj 2016; Dunkerley 2022), which appeared useful to explain rainfall variability compared to flood types (Oppel and Fischer 2020) and to soil erosion events (Vásquez *et al.* 2024). The Huff curves contain more precise information on the timing of rainfall, but are standardized by event duration, so the information on event severity is removed. They can thus be seen as complementary information to the event characteristics derived by Yevjevich's threshold approach. As both event descriptors are complementary, their combination should be beneficial for clustering rainfall events. Another approach would be to use other process variables such as the lightning index in our study, or the Showalter index of the potential for convection calculated using temperature and dew point at 850 hPa and temperature at 500 hPa (Showalter 1953). This is currently tested in an ongoing study to perform space-time modeling of rainfall distributions. Alternatively to the classification of the rainfall itself, often a classification based on large-scale atmospheric indices like circulation patterns or weather patterns is performed to understand the different processes in the hydrosphere (Bárdossy and Filiz 2005; Huth *et al.* 2008; Hofstätter, Chimani, Lexer, and Blöschl 2016). Compared to the classification based on rainfall characteristics, such a hydroclimatic classification focuses more on large-scale patterns, making it harder to interpret and to apply for small-scale processes like flash floods, where small-scale processes of rainfall and catchment characteristics play a major role (Tarasova *et al.* 2019). The proposed classification based on rainfall, considering Yevjevich's threshold approach, instead is developed more in the spirit of the flood classification proposed in the first example, where also local dynamics (of hydrographs) are directly classified (Fischer *et al.* 2019). It might therefore be beneficial in small-scale applications like flood-forecasting and rainfall-runoff modeling.

The first three examples have in common that they not only introduce statistical concepts to cope with heterogeneity in environmental data sets, they also increase the physical knowledge that is included in statistics. This provides a better interpretability of the results, an opportunity to validate the findings, and a chance to improve communication of statistical results to practice. Furthermore, consideration of heterogeneity in environmental processes could also lead to improvement in further applications. This also includes a broad range of statistical models. As suggested in the last example, in temporal modeling, adopting a single optimization strategy will only improve the performance of the model in the corresponding part of the distribution. This is true not only for a tree-based model considered in this study, but also for state-of-the-art prediction models for hydrological time series, such as deep-learning architectures like LSTMs (Hochreiter and Schmidhuber 1997). Although, they have been shown to provide an improvement for rainfall-runoff modeling in general (Kratzert *et al.* 2019a,b; Lees *et al.* 2021), and also for extremes (Frame *et al.* 2022), their potential may not be fully exploited, if they are not directly tailored to the research question, using an adequate optimization strategy. However, model optimization is not limited to chasing the best error metric. It should always aim to be process-based, to improve the physical understanding of the processes specific to the phenomenon under consideration, and to advance science in general (Strobl and Leisch 2024). With this concept in mind, our last example specifically showed that not only does model performance suffer from an inappropriate loss function, but the selection and parametrization of variables in the model is also affected. Our study suggests that the parametrization of the model will differ along the range of the target variable, necessitating the use of different parametrizations or a combination of models if model predictions are sought for the full time series in the presence of process heterogeneity.

All examples demonstrate how incorporating process heterogeneity leads to enhanced methods that address a broad range of research areas and are crucial for various applications. This calls for further model developments. Our study showed that mixture distribution approaches provide more comprehensive event descriptions and accurate predictions. Accurate extreme value statistics are essential for water management tasks such as reservoir and dam design, flood protection planning, and managing the residual risk of overtopping. This involves disciplines from meteorology, hydrology, and natural hazards, to risk research, spatial planning, social sciences, and policy. Often, both flood peak and volume are critical, which are marginals of the same event distribution. In this context, an extended TMPS method would be particularly suitable for estimating multivariate design values. For drought management, as another example, seasonal minima are often relevant for tasks like designing water treatment plants or for ecosystem assessment (WMO 2008; DWA 2022). This calls for specific methods, including seasonal models such as the ones presented in this paper, and multivariate model extensions. Future research may focus on compound prediction methods for seasonal and annual characteristics of multiple variables, such as streamflow, temperature, and nutrients. These should be supported by operational tools such as calculation software or regionalized water resources maps to transfer science into practice.

Accurate rainfall distributions through rainfall clustering benefits various applications, such as rainfall generators. Rainfall is essential for agriculture and forestry, but it also poses natural hazards like soil erosion, landslides, debris flows, and rockfalls. Extending our mixture models with hybrid distributions appears beneficial for better covering the entire range of rainfall, from dry spells to extreme events. Real-world applications that will benefit from more process-realistic models include rainfall simulations used for soil erosion and flood design, as well as rainfall forecasting. Studies such as Goodarzi, Banihabib, Roozbahani, and Dietrich (2019) have shown the advantages of advanced statistical models in operational forecasting. These methods can also inform climate impact projections, as demonstrated by Parajka *et al.* (2016) and Laaha *et al.* (2016) in the context of droughts.

Finally, incorporating process heterogeneity enhances not only distribution approaches but also spatio-temporal models. Our research highlights the dangers of ignoring process heterogeneity in temporal models, which can lead to inappropriate parametrizations and biased predictions. Therefore, addressing process heterogeneity in model design – by including appropriate model structure, optimization criteria, and process-oriented variable selection – is crucial. Similar benefits, as found for temporal models, are expected in spatial and spatio-temporal modeling contexts. This is illustrated by Laimighofer and Laaha (2025), who demonstrated the advantage of incorporating deterministic trend and seasonality components with machine learning. Their advanced explainable machine learning method outperformed simpler models, further showcasing the benefits of process-based statistical models. Moreover, our experience from transdisciplinary projects shows that better representation of processes in stochastic models, achieved by considering heterogeneity, also increases acceptance in practice. This is particularly evident in the field of machine learning, where the "black box" nature of some methods is often criticized.

In summary, this study should be seen as an encouragement to better understand statistical assumptions in the applied models and to enrich the physical knowledge included in environmental statistics. Heterogeneity of processes often turns out to be an obstacle in modeling. The approach of including heterogeneity in the statistical models could be seen as an alternative to the development of more complex models that can cope with the variability and heterogeneity in the data, which often suffer from limited data bases. While both perspectives offer important knowledge on the statistical theory, consideration of heterogeneity in combination with a linkage to physical processes could also improve interdisciplinary knowledge transfer. In some instances, even the combination of both approaches could be beneficial, developing more sophisticated models as well as increasing the incorporated statistical and physical knowledge.

Acknowledgments

This research has been supported by the Austrian Climate and Energy Fund under the programs "ACRP13" (grant no. KR20AC0K17974) and "ACRP14" (grant no. C265154). The computational results presented have been achieved in part using the Vienna Scientific Cluster (VSC). We would like to thank the editors and the reviewers for their valuable comments on the manuscript.

References

Aas K, Jullum M, Løland A (2021). "Explaining Individual Predictions when Features Are Dependent: More Accurate Approximations to Shapley Values." *Artificial Intelligence*, **298**, 103502. doi:10.1016/j.artint.2021.103502.

Aigner DJ, Amemiya T, Poirier DJ (1976). "On the Estimation of Production Frontiers: Maximum Likelihood Estimation of the Parameters of a Discontinuous Density Function." *International Economic Review*, **17**(2), 377–396. doi:10.2307/2525708.

Balkema AA, de Haan L (1974). "Residual Life Time at Great Age." *The Annals of Probability*, **2**(5), 792–804. doi:10.1214/aop/1176996548.

Bárdossy A, Filiz F (2005). "Identification of Flood Producing Atmospheric Circulation Patterns." *Journal of Hydrology*, **313**(1–2), 48–57. doi:10.1016/j.jhydrol.2005.02.006.

Biernacki C, Celeux G, Govaert G (2000). "Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(7), 719–725. doi:10.1109/34.865189.

Bulletin No. 17 (1976). *Guidelines for Determining Flood Flow Frequency*. Washington D. C., USA.

Chen T, Guestrin C (2016). "XGBoost: A Scalable Tree Boosting System." In *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pp. 785–794. Association for Computing Machinery, New York, NY, USA. doi:10.1145/2939672.2939785.

Chen T, He T (2015). "Higgs Boson Discovery with Boosted Trees." In G Cowan, C Germain, I Guyon, B Kégl, D Rousseau (eds.), *Proceedings of the NIPS 2014 Workshop on High-Energy Physics and Machine Learning*, volume 42 of *Proceedings of Machine Learning Research*, pp. 69–80. PMLR, Montreal, Canada. URL https://proceedings.mlr.press/v42/chen14.html.

Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, Chen K, Mitchell R, Cano I, Zhou T, Li M, Xie J, Lin M, Geng Y, Li Y (2021). ***xgboost**: Extreme Gradient Boosting*. R package version 1.4.1.1, URL https://CRAN.R-project.org/package=xgboost.

Coles S (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer-Verlag, London. doi:10.1007/978-1-4471-3675-0.

Dolšak D, Bezak N, Šraj M (2016). "Temporal Characteristics of Rainfall Events under Three Climate Types in Slovenia." *Journal of Hydrology*, **541**(Part B), 1395–1405. doi:10.1016/j.jhydrol.2016.08.047.

Dunkerley D (2022). "Huff Quartile Classification of Rainfall Intensity Profiles ('Storm Patterns'): A Modified Approach Employing an Intensity Threshold." *CATENA*, **216**(Part A), 106371. doi:10.1016/j.catena.2022.106371.

DWA (ed.) (2022). *Merkblatt DWA-M 541 Statistische Analyse von Niedrigwasserkenngrößen*. Number 541 in DWA-Regelwerk, Juli 2022 edition. Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall e.V. (DWA), Hennef.

Fischer S, Schumann A, Bühler P (2019). "Timescale-Based Flood Typing to Estimate Temporal Changes in Flood Frequencies." *Hydrological Sciences Journal*, **64**(15), 1867–1892. doi:10.1080/02626667.2019.1679376.

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. doi:10.1016/j.jhydrol.2016.05.005.

Fischer S, Schumann AH (2021). "Multivariate Flood Frequency Analysis in Large River Basins Considering Tributary Impacts and Flood Types." *Water Resources Research*, **57**(8), e2020WR029029. doi:10.1029/2020wr029029.

Fischer S, Schumann AH (2023). *Type-Based Flood Statistics: An Interlink between Stochastic and Deterministic Flood Hydrology*. Springer-Verlag. doi:10.1007/978-3-031-32711-7.

Fisher RA (1928). "Limiting Forms of the Frequency Distribution of the Largest and Smallest Member of a Sample." *Mathematical Proceedings of the Cambridge Philosophical Society*, **24**(2), 180–190. doi:10.1017/s0305004100015681.

Frame JM, Kratzert F, Klotz D, Gauch M, Shalev G, Gilon O, Qualls LM, Gupta HV, Nearing GS (2022). "Deep Learning Rainfall–Runoff Predictions of Extreme Events." *Hydrology and Earth System Sciences*, **26**(13), 3377–3392. doi:10.5194/hess-26-3377-2022.

Furrer EM, Katz RW (2008). "Improving the Simulation of Extreme Precipitation Events by Stochastic Weather Generators." *Water Resources Research*, **44**(12), 2008WR007316. doi:10.1029/2008wr007316.

Goodarzi L, Banihabib ME, Roozbahani A, Dietrich J (2019). "Bayesian Network Model for Flood Forecasting Based on Atmospheric Ensemble Forecasts." *Natural Hazards and Earth System Sciences*, **19**(11), 2513–2524. doi:10.5194/nhess-19-2513-2019.

Gower JC (1971). "A General Coefficient of Similarity and Some of Its Properties." *Biometrics*, **27**(4), 857–871. doi:10.2307/2528823.

Grønneberg S, Hjort NL (2014). "The Copula Information Criteria: The Copula Information Criteria." *Scandinavian Journal of Statistics*, **41**(2), 436–459. doi:10.1111/sjos.12042.

Gruber K, Gauster T, Laaha G, Regner P, Schmidt J (2022). "Profitability and Investment Risk of Texan Power System Winterization." *Nature Energy*, **7**(5), 409–416. doi:10.1038/s41560-022-00994-y.

Grün B, Leisch F (2025). ***flexmix**: Flexible Mixture Modeling*. R package version 2.3-20, URL https://CRAN.R-project.org/package=flexmix.

Gumbel EJ (1941). "The Return Period of Flood Flows." *The Annals of Mathematical Statistics*, **12**(2), 163–190. doi:10.1214/aoms/1177731747.

Guo Y, Zhang Y, Zhang L, Wang Z (2021). "Regionalization of Hydrological Modeling for Predicting Streamflow in Ungauged Catchments: A Comprehensive Review." *Wiley Interdisciplinary Reviews: Water*, **8**(1), e1487. doi:10.1002/wat2.1487.

Gustard A, Irving KM (1994). "Classification of the Low Flow Response of European Soils." In *FRIEND: Flow Regimes from International Experimental and Network Data. Proceedings of an International Conference Held at the Technical University of Brunswick, Germany, 11 to 15 October 1993*, pp. 113–117. IAHS Press, Wallingford.

Hirschboek KK (1988). "Flood Hydroclimatology." In VR Baker, RC Kochel, PC Patton (eds.), *Flood Geomorphology*, pp. 27–49. John Wiley & Sons.

Hochreiter S, Schmidhuber J (1997). "Long Short-Term Memory." *Neural Computation*, **9**(8), 1735–1780. doi:10.1162/neco.1997.9.8.1735.

Hofrichter J, Harum T, Friedl H (2016). "Statistical Modelling of Annual Maxima in Hydrology." *Austrian Journal of Statistics*, **35**(1), 21–30. doi:10.17713/ajs.v35i1.345.

Hofstätter M, Chimani B, Lexer A, Blöschl G (2016). "A New Classification Scheme of European Cyclone Tracks with Relevance to Precipitation." *Water Resources Research*, **52**(9), 7086–7104. doi:10.1002/2016wr019146.

Hothorn T, Hornik K, Zeileis A (2006). "Unbiased Recursive Partitioning: A Conditional Inference Framework." *Journal of Computational and Graphical Statistics*, **15**(3), 651–674. doi:10.1198/106186006x133933.

Hothorn T, Zeileis A (2015). "**partykit**: A Modular Toolkit for Recursive Partytioning in R." *The Journal of Machine Learning Research*, **16**(1), 3905–3909.

Huth R, Beck C, Philipp A, Demuzere M, Ustrnul Z, Cahynová M, Kyselỳ J, Tveito OE (2008). "Classifications of Atmospheric Circulation Patterns: Recent Advances and Applications." *Annals of the New York Academy of Sciences*, **1146**(1), 105–152. doi:10.1196/annals.1446.019.

Institute of Hydrology (1980). "Low Flow Studies Report." *Technical Report 3 volumes*, Institute of Hydrology, Wallingford, UK.

Iovleff S (2024). ***MixAll**: Clustering and Classification Using Model-Based Mixture Models.* R package version 1.5.16, URL https://CRAN.R-project.org/package=MixAll.

Jiang S, Bevacqua E, Zscheischler J (2022). "River Flooding Mechanisms and Their Changes in Europe Revealed by Explainable Machine Learning." *Hydrology and Earth System Sciences*, **26**(24), 6339–6359. doi:10.5194/hess-26-6339-2022.

Kaufmann L, Rousseeuw PJ (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, New York. doi:10.1002/9780470316801.

Klemes V (1971). "Some Problems in Pure and Applied Stochastic Hydrology." In *Proceedings of a Symposium on Statistical Hydrology*. U.S. Department of Agriculture.

Klingler C, Schulz K, Herrnegger M (2021). "LamaH-CE: LArge-SaMple DAta for Hydrology and Environmental Sciences for Central Europe." *Earth System Science Data*, **13**(9), 4529–4565. doi:10.5194/essd-13-4529-2021.

Koenker R, Bassett Jr G (1978). "Regression Quantiles." *Econometrica*, **46**(1), 33–50. doi:10.2307/1913643.

Kratzert F, Klotz D, Herrnegger M, Sampson AK, Hochreiter S, Nearing GS (2019a). "Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning." *Water Resources Research*, **55**(12), 11344–11354. doi:10.1029/2019wr026065.

Kratzert F, Klotz D, Shalev G, Klambauer G, Hochreiter S, Nearing G (2019b). "Towards Learning Universal, Regional, and Local Hydrological Behaviors via Machine Learning Applied to Large-Sample Datasets." *Hydrology and Earth System Sciences*, **23**(12), 5089–5110. doi:10.5194/hess-23-5089-2019.

Laaha G (2023a). "A Mixed Distribution Approach for Low-Flow Frequency Analysis – Part 1: Concept, Performance, and Effect of Seasonality." *Hydrology and Earth System Sciences*, **27**(3), 689–701. doi:10.5194/hess-27-689-2023.

Laaha G (2023b). "A Mixed Distribution Approach for Low-Flow Frequency Analysis – Part 2: Comparative Assessment of a Mixed Probability vs. Copula-Based Dependence." doi:10.5194/hess-27-2019-2023.

Laaha G, Blöschl G (2006). "A Comparison of Low Flow Regionalisation Methods–Catchment Grouping." *Journal of Hydrology*, **323**(1–4), 193–214. doi:10.1016/j.jhydrol.2005.09.001.

Laaha G, Gauster T, Tallaksen LM, Vidal JP, Stahl K, Prudhomme C, Heudorfer B, Vlnas R, Ionita M, Van Lanen HAJ, Adler MJ, Caillouet L, Delus C, Fendekova M, Gailliez S, Hannaford J, Kingston D, Van Loon AF, Mediero L, Osuch M, Romanowicz R, Sauquet E, Stagge JH, Wong WK (2017). "The European 2015 Drought from a Hydrological Perspective." *Hydrology and Earth System Sciences*, **21**(6), 3001–3024. doi:10.5194/hess-21-3001-2017.

Laaha G, Parajka J, Viglione A, Koffler D, Haslinger K, Schöner W, Zehetgruber J, Blöschl G (2016). "A Three-Pillar Approach to Assessing Climate Impacts on Low Flows." *Hydrology and Earth System Sciences*, **20**(9), 3967–3985. doi:10.5194/hess-20-3967-2016.

Laimighofer J (2024). *Code and Data to "Quantile Regression for Temporal Streamflow Modeling"*. doi:10.5281/zenodo.14066026.

Laimighofer J, Laaha G (2025). "Statistical Learning and Topkriging Improve Spatio-Temporal Low-Flow Estimation." *Water Resources Research*. Accepted for publication.

Laimighofer J, Melcher M, Laaha G (2022). "Low-Flow Estimation beyond the Mean – Expectile Loss and Extreme Gradient Boosting for Spatiotemporal Low-Flow Prediction in Austria." *Hydrology and Earth System Sciences*, **26**(17), 4553–4574. doi:10.5194/hess-26-4553-2022.

Lees T, Buechel M, Anderson B, Slater L, Reece S, Coxon G, Dadson SJ (2021). "Benchmarking Data-Driven Rainfall–Runoff Models in Great Britain: A Comparison of Long Short-Term Memory (LSTM)-Based Models with Four Lumped Conceptual Models." *Hydrology and Earth System Sciences*, **25**(10), 5517–5534. doi:10.5194/hess-25-5517-2021.

Leisch F (2004). "FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R." *Journal of Statistical Software*, **11**(8), 1–18. doi:10.18637/jss.v011.i08.

Lima CHR, Lall U, Troy T, Devineni N (2016). "A Hierarchical Bayesian GEV Model for Improving Local and Regional Flood Quantile Estimates." *Journal of Hydrology*, **541**(Part B), 816–823. doi:10.1016/j.jhydrol.2016.07.042.

Lundberg SM, Lee SI (2017). "A Unified Approach to Interpreting Model Predictions." In I Guyon, UV Luxburg, S Bengio, H Wallach, R Fergus, S Vishwanathan, R Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

Merz B, Basso S, Fischer S, Lun D, Blöschl G, Merz R, Guse B, Viglione A, Vorogushyn S, Macdonald E, Wietzke L, Schumann A (2022). "Understanding Heavy Tails of Flood Peak Distributions." *Water Resources Research*, **58**(6), e2021WR030506. doi:10.1029/2021wr030506.

Milly PCD, Betancourt J, Falkenmark M, Hirsch RM, Kundzewicz ZW, Lettenmaier DP, Stouffer RJ (2008). "Stationarity Is Dead: Whither Water Management?" *Science*, **319**(5863), 573–574. doi:10.1126/science.1151915.

Montanari A, Koutsoyiannis D (2014). "Modeling and Mitigating Natural Hazards: Stationarity Is Immortal!" *Water Resources Research*, **50**(12), 9748–9756. doi:10.1002/2014wr016092.

Nash JE, Sutcliffe JV (1970). "River Flow Forecasting through Conceptual Models Part I – A Discussion of Principles." *Journal of Hydrology*, **10**(3), 282–290. doi:10.1016/0022-1694(70)90255-6.

Nathan RJ, McMahon TA (1990). "Identification of Homogeneous Regions for the Purposes of Regionalisation." *Journal of Hydrology*, **121**(1–4), 217–238. doi:10.1016/0022-1694(90)90233-n.

Newey WK, Powell JL (1987). "Asymmetric Least Squares Estimation and Testing." *Econometrica*, **55**(4), 819–847. doi:10.2307/1911031.

Oppel H, Fischer S (2020). "A New Unsupervised Learning Method to Assess Clusters of Temporal Distribution of Rainfall and Their Coherence with Flood Types." *Water Resources Research*, **56**(5), e2019WR026511. doi:10.1029/2019wr026511.

Papacharalampous G, Tyralis H, Langousis A, Jayawardena AW, Sivakumar B, Mamassis N, Montanari A, Koutsoyiannis D (2019). "Probabilistic Hydrological Post-Processing at Scale: Why and How to Apply Machine-Learning Quantile Regression Algorithms." *Water*, **11**(10). doi:10.3390/w11102126.

Papalexiou SM, Koutsoyiannis D, Makropoulos C (2013). "How Extreme Is Extreme? An Assessment of Daily Rainfall Distribution Tails." *Hydrology and Earth System Sciences*, **17**(2), 851–862. doi:10.5194/hess-17-851-2013.

Parajka J, Blaschke AP, Blöschl G, Haslinger K, Hepp G, Laaha G, Schöner W, Trautvetter H, Viglione A, Zessner M (2016). "Uncertainty Contributions to Low-Flow Projections in Austria." *Hydrology and Earth System Sciences*, **20**(5), 2085–2101. doi:10.5194/hess-20-2085-2016.

Pickands III J (1975). "Statistical Inference Using Extreme Order Statistics." *The Annals of Statistics*, **3**(1), 119–131. doi:10.1214/aos/1176343003.

Potter WD (1958). "Upper and Lower Frequency Curves for Peak Rates of Runoff." *Eos, Transactions American Geophysical Union*, **39**(1), 100–105. doi:10.1029/tr039i001p00100.

Razavi T, Coulibaly P (2013). "Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods." *Journal of Hydrologic Engineering*, **18**(8), 958–975. doi:10.1061/(asce)he.1943-5584.0000690.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rossi F, Fiorentino M, Versace P (1984). "Two-Component Extreme Value Distribution for Flood Frequency Analysis." *Water Resources Research*, **20**(7), 847–856. doi:10.1029/wr020i007p00847.

Sampaio J, Costa V (2021). "Bayesian Regional Flood Frequency Analysis with GEV Hierarchical Models under Spatial Dependency Structures." *Hydrological Sciences Journal*, **66**(3), 422–433. doi:10.1080/02626667.2021.1873997.

Schulz W, Diendorfer G, Pedeboy S, Poelman DR (2016). "The European Lightning Location System EUCLID – Part 1: Performance Analysis and Validation." *Natural Hazards and Earth System Sciences*, **16**(2), 595–605. doi:10.5194/nhess-16-595-2016.

Showalter AK (1953). "A Stability Index for Thunderstorm Forecasting." *Bulletin of the American Meteorological Society*, **34**(6), 250–252. doi:10.1175/1520-0477-34.6.250.

Singh K (1974). "A Two-Distribution Method for Fitting Mixed Distributions in Hydrology." In *Proceedings of a Symposium on Statistical Hydrology*, pp. 371–382. U.S. Department of Agriculture.

Sklar M (1959). "Fonctions De Repartition an Dimensions et Leurs Marges." *Publications de l'Institut de Statistique de l'Université de Paris 8*. doi:10.2139/ssrn.4198458.

Solomatine DP, Ostfeld A (2008). "Data-Driven Modelling: Some Past Experiences and New Approaches." *Journal of Hydroinformatics*, **10**(1), 3–22. doi:10.2166/hydro.2008.015.

Stedinger JR, Vogel RM, Foufoula-Georgiou E (1993). "Frequency Analysis of Extreme Events." In DR Maidment (ed.), *Handbook of Hydrology*, chapter 18. McGraw-Hill.

Strobl C, Leisch F (2024). "Against the "One Method Fits All Data Sets" Philosophy for Comparison Studies in Methodological Research." *Biometrical Journal*, **66**(1), 2200104. doi:10.1002/bimj.202200104.

Tarasova L, Merz R, Kiss A, Basso S, Blöschl G, Merz B, Viglione A, Plötner S, Guse B, Schumann A, *et al.* (2019). "Causative Classification of River Flood Events." *Wiley Interdisciplinary Reviews: Water*, **6**(4), e1353. doi:10.1002/wat2.1353.

The Global Runoff Data Centre (2024). URL https://portal.grdc.bafg.de/applications/public.html?publicuser=PublicUser#dataDownload/Home.

Todorovic P, Rouselle J (1971). "Some Problems of Flood Analysis." *Water Resources Research*, **7**(5), 1144–1150. doi:10.1029/wr007i005p01144.

Toth E (2016). "Estimation of Flood Warning Runoff Thresholds in Ungauged Basins with Asymmetric Error Functions." *Hydrology and Earth System Sciences*, **20**(6), 2383–2394. doi:10.5194/hess-20-2383-2016.

Tyralis H, Papacharalampous G, Khatami S (2023). "Expectile-Based Hydrological Modelling for Uncertainty Estimation: Life After Mean." *Journal of Hydrology*, **617**(Part B), 128986. doi:10.1016/j.jhydrol.2022.128986.

Van Lanen HAJ, Laaha G, Kingston DG, Gauster T, Ionita M, Vidal JP, Vlnas R, Tallaksen LM, Stahl K, Hannaford J, Delus C, Fendekova M, Mediero L, Prudhomme C, Rets E, Romanowicz RJ, Gailliez S, Wong WK, Adler MJ, Blauhut V, Caillouet L, Chelcea S, Frolova N, Gudmundsson L, Hanel M, Haslinger K, Kireeva M, Osuch M, Sauquet E, Stagge JH, Van Loon AF (2016). "Hydrology Needed to Manage Droughts: The 2015 European Case." *Hydrological Processes*, **30**(17), 3097–3104. doi:10.1002/hyp.10838.

Vásquez C, Klik A, Stumpp C, Laaha G, Strauss P, Özcelik NB, Pistotnik G, Yin S, Dostal T, Gaona G, Strohmeier S (2024). "Rainfall Erosivity across Austria's Main Agricultural Areas: Identification of Rainfall Characteristics and Spatiotemporal Patterns." *Journal of Hydrology: Regional Studies*, **53**, 101770. doi:10.1016/j.ejrh.2024.101770.

Veatch W, Villarini G (2022). "Modeling Riverine Flood Seasonality with Mixtures of Circular Probability Density Functions." *Journal of Hydrology*, **613**(Part A), 128330. doi:10.1016/j.jhydrol.2022.128330.

Vogel RM, Papalexiou SM, Lamontagne JR, Dolan FC (2024). "When Heavy Tails Disrupt Statistical Inference." *The American Statistician*, pp. 1–15. doi:10.1080/00031305.2024.2402898.

Volpi E, Grimaldi S, Aghakouchak A, Castellarin A, Chebana F, Papalexiou SM, Aksoy H, Bárdossy A, Cancelliere A, Chen Y, Deidda R, Haberlandt U, Eris E, Fischer S, Francés F, Kavetski D, Rodding Kjeldsen T, Kochanek K, Langousis A, Mediero Orduña L, Montanari A, Nerantzaki SD, Ouarda TBMJ, Prosdocimi I, Ragno E, Rajulapati CR, Requena AI, Ridolfi E, Sadegh M, Schumann A, Sharma A (2024). "The Legacy of STAHY: Milestones, Achievements, Challenges, and Open Problems in Statistical Hydrology." *Hydrological Sciences Journal*, **69**(14), 1913–1949. doi:10.1080/02626667.2024.2385686.

Vrac M, Naveau P (2007). "Stochastic Downscaling of Precipitation: From Dry Events to Heavy Rainfalls." *Water Resources Research*, **43**(7), 2006WR005308. doi:10.1029/2006wr005308.

Waylen P, Woo MK (1982). "Prediction of Annual Floods Generated by Mixed Processes." *Water Resources Research*, **18**(4), 1283–1286. doi:10.1029/wr018i004p01283.

Wilks DS, Wilby RL (1999). "The Weather Generation Game: A Review of Stochastic Weather Models." *Progress in Physical Geography: Earth and Environment*, **23**(3), 329–357. doi:10.1177/030913339902300302.

WMO (2008). *Manual on Low-Flow Estimation and Prediction*, volume No. 50 of *Open Hydrology Report*. World Meteorological Organization, Geneva.

Yan L, Xiong L, Liu D, Hu T, Xu CY (2017). "Frequency Analysis of Nonstationary Annual Maximum Flood Series Using the Time-Varying Two-Component Mixture Distributions." *Hydrological Processes*, **31**(1), 69–89. doi:10.1002/hyp.10965.

Yan L, Xiong L, Ruan G, Xu CY, Yan P, Liu P (2019). "Reducing Uncertainty of Design Floods of Two-Component Mixture Distributions by Utilizing Flood Timescale to Classify Flood Types in Seasonally Snow Covered Region." *Journal of Hydrology*, **574**, 588–608. doi:10.1016/j.jhydrol.2019.04.056.

Yan L, Zhang L, Xiong L, Yan P, Jiang C, Xu W, Xiong B, Yu K, Ma Q, Xu CY (2023). "Flood Frequency Analysis Using Mixture Distributions in Light of Prior Flood Type Classification in Norway." *Remote Sensing*, **15**(2), 401. doi:10.3390/rs15020401.

Yevjevich VM (1967). "An Objective Approach to Definitions and Investigations of Continental Hydrologic Droughts." *Hydrology Papers 23*, Colorado State University.

Affiliation:

Gregor Laaha
Institute of Statistics
BOKU University
A-1190 Vienna, Austria
E-mail: gregor.laaha@boku.ac.at
URL: https://short.boku.ac.at/stat

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 54 *Submitted:* 2024-11-26
2025 *Accepted:* 2025-04-01