Sun, X., Lall, U., Merz, B., Nguyen, D. (2015): Hierarchical Bayesian clustering for nonstationary flood frequency analysis: Application to trends of annual maximum flow in Germany. - *Water Resources Research*, *51*, 8, pp. 6586–6601.

DOI: http://doi.org/10.1002/2015WR017117

AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2015WR017117

Key Points:
- Develop a nonstationary frequency analysis approach for large area data sets
- Investigate the trends in annual maximum daily stream flow over Germany

Correspondence to:
X. Sun,
xs2226@columbia.edu

Citation:
Sun, X., U. Lall, B. Merz, and N. V. Dung (2015), Hierarchical Bayesian clustering for nonstationary flood frequency analysis: Application to trends of annual maximum flow in Germany, *Water Resour. Res.*, *51*, 6586–6601, doi:10.1002/2015WR017117.

Received 20 FEB 2015
Accepted 22 JUL 2015
Accepted article online 27 JUL 2015
Published online 24 AUG 2015

Hierarchical Bayesian clustering for nonstationary flood frequency analysis: Application to trends of annual maximum flow in Germany

Xun Sun[1], Upmanu Lall[1,2], Bruno Merz[3,4], and Nguyen Viet Dung[3]

[1]Columbia Water Center, Columbia University, New York, New York, USA, [2]Department of Earth and Environmental Engineering, Columbia University, New York, New York, USA, [3]Section Hydrology, GFZ German Research Center for Geosciences, Potsdam, Germany, [4]Institute of Earth and Environmental Science, University of Potsdam, Potsdam, Germany

Abstract Especially for extreme precipitation or floods, there is considerable spatial and temporal variability in long term trends or in the response of station time series to large-scale climate indices. Consequently, identifying trends or sensitivity of these extremes to climate parameters can be marked by high uncertainty. When one develops a nonstationary frequency analysis model, a key step is the identification of potential trends or effects of climate indices on the station series. An automatic clustering procedure that effectively pools stations where there are similar responses is desirable to reduce the estimation variance, thus improving the identification of trends or responses, and accounting for spatial dependence. This paper presents a new hierarchical Bayesian approach for exploring homogeneity of response in large area data sets, through a multicomponent mixture model. The approach allows the reduction of uncertainties through both full pooling and partial pooling of stations across automatically chosen subsets of the data. We apply the model to study the trends in annual maximum daily stream flow at 68 gauges over Germany. The effects of changing the number of clusters and the parameters used for clustering are demonstrated. The results show that there are large, mainly upward trends in the gauges of the River Rhine Basin in Western Germany and along the main stream of the Danube River in the south, while there are also some small upward trends at gauges in Central and Northern Germany.

1. Introduction

It is widely recognized that regional frequency analysis of hydroclimatic extremes can reduce uncertainty in precipitation/flood frequency estimates at locations with short records [*Kysely et al.*, 2011; *Sun et al.*, 2014]. A variety of different methods have been developed for regional flood frequency analysis under the paradigm of stationarity [e.g., *Lang et al.*, 1999; *Madsen and Rosbjerg*, 1997; *Meshgi and Khalili*, 2009; *Ribatet et al.*, 2007; *Sankarasubramanian and Srinivasan*, 1999]. The importance of addressing potential sources of nonstationarity in frequency analysis is now well recognized and two main approaches have emerged. One seeks to identify trends (secular or periodic) in individual series of extremes in the context of a formal model (e.g., the Generalized Extreme Value distribution) through the use of time as a covariate with an appropriate basis function to represent the temporal variation of one or more of the parameters of the distribution. The second considers a causal framework where the model parameters depend on climate or land use covariates, which may themselves change in space and time. The first case is primarily a diagnostic model, since extrapolating a trend model far into the future without a causal framework is not likely to yield good results. The second case is diagnostic and allows one to assess sensitivity to the proposed causal variables, and to then potentially predict into the future or at other locations, using possible scenarios of the causal variables. In both cases, uncertainty in estimation and the homogeneity of the parameters estimated across a relatively large region raises questions as to identifiability and reliability of the model. For instance, over a large enough region, trends in extreme precipitation may switch from positive to negative, or the response to an El Niño condition may switch from enhancement to suppression of extremes. The automatic identification of such heterogeneities while reducing the uncertainty of estimates in a formal modeling framework is the goal of this paper. A general hierarchical Bayesian model which explicitly considers finite mixtures (or clusters) and partial pooling of local parameters in

each cluster is developed and applied to historical data at a country scale from a network of stations in Germany.

Many frequency analysis models that incorporate spatial and/or temporal covariates have been developed [*Cannon*, 2010; *Chavez-Demoulin and Davison*, 2005; *Cooley et al*., 2007; *Delgado et al*., 2012; *El Adlouni et al*., 2007; *Katz et al*., 2002; *Khaliq et al*., 2006; *Leclerc and Ouarda*, 2007; *Nasri et al*., 2013; *Ouarda and El-Adlouni*, 2011; *Renard et al*., 2006; *Sun et al*., 2014]. With this framework, the effect of temporal trend, seasonality, climate variability, spatial variability, etc. on the frequency and severity of hydro-meteorological events can be quantitatively characterized [e.g., *Aryal et al*., 2009; *Begueria et al*., 2011; *Chen et al*., 2014; *Cunderlik and Burn*, 2003; *Kysely et al*., 2010; *Leclerc and Ouarda*, 2007; *Lima et al*., 2015; *Lima and Lall*, 2009, 2010; *Renard et al*., 2008; *Rust et al*., 2009]. As indicated earlier, the need to reduce estimation uncertainty at a given station using regional analysis has been recognized, and a common challenge for regional models is how to shrink the estimation variance by proper pooling information across sites. A possible solution is to consider a region where the climate effects or temporal trends may be homogeneous. However, it is not obvious how to a priori identify a region where the trends or climate effects are homogeneous. Hence, most authors [e.g., *Srinivas et al*., 2008] only apply regional models in small areas where the process may be considered to be homogeneous, and large domain are divided into subregions using clustering techniques, such as K-means [*MacQueen*, 1967], Fuzzy C-means [*Bezdek et al*., 1984] and Gaussian mixture clustering [*McLachlan and Peel*, 2004]. However, these applications typically cluster on the mean and/or standard deviation of each station, and not on the climate sensitivity or trend, and are not suitable for the highly skewed data, typical of hydro-climatic extremes [*Bernard et al*., 2013]. Some Bayesian clustering models have been developed to consider spatiotemporal variation based on Expectation-Maximization (EM) algorithm or Dirichlet process [e.g., *Xiong and Yeung*, 2004; *Johnson et al*., 2013; *Nieto-Barajas and Contreras-Cristán*, 2014], but these models have been limited to a Gaussian framework.

The model presented in this paper improves on existing models by

1. Considering probability distributions suitable for extremes, such as the Generalized Extreme value distribution;
2. Allowing the local and/or regional parameters of these distributions to depend on specified spatial or temporal covariates using appropriate basis functions;
3. Seeking subregions or clusters of the larger domain, while allowing a flexible multilevel model in each subregion where the degree of homogeneity in each subregion can be explicitly modeled through partial pooling across the sites selected for that subregion. The clustering criteria can consider a user specified subset of parameters (e.g., mean, standard deviation, shape, climate effect, temporal trend) of interest;
4. Developing a unified inference framework where classification into subregions and the estimation of all local and regional parameters of the extreme value distributions is part of the same inference process while formally modeling parameter uncertainty. Here the parameter estimation includes the parameters of the model that expresses the dependence on covariates, as well as the hyper-parameters of the prior distributions introduced; and
5. Using this framework to identify trends as well as climate dependence across the region.

The application of the model to the data from Germany was motivated by the availability of a complete data set across the country, and by past investigations that led to inconclusive arguments as to the presence of trends across the nation, thus providing an opportunity to explore whether the techniques developed could provide a useful interpretation of the data.

The paper is organized as follows. The next section introduces the hierarchical Bayesian clustering model. Section 3 provides a case study of trends on annual maximum German flow. The conclusion and further extensions of the model are discussed in section 4.

2. Hierarchical Bayesian Clustering Model for Frequency Analysis

2.1. Full Pooling and Partial Pooling Model in a Homogeneous Region

For a homogeneous region, consider that the observations follow a distribution with parameters varying with time and space:

$$Y(s,t) \sim D\left(\vec{\theta}(s,t)\right) \tag{1}$$

where D is a prespecified distribution and $\vec{\theta} = \left(\theta^{(l)}\right)_{l=1,\ldots,L}$ are the associated distribution parameters, and L is the number of distribution parameters.

Each distribution parameter can potentially be modeled as a function of the covariates. For example, a component of $\vec{\theta}(s,t)$ can be modeled as $\theta(s,t) = \mu_0(s) + \mu_1 t$, where $\mu_0(s)$ is a site-specific (local) regression parameter, μ_1 is a regional regression parameter which is common for all sites, and t is the temporal covariate. More generally, the regression function can be written as follows:

$$\theta^{(l)}(s,t) = R\left(\vec{\boldsymbol{\beta}}_{loc}^{(l)}, \vec{\boldsymbol{\beta}}_{reg}^{(l)}; \vec{\boldsymbol{x}}(s,t)\right) \tag{2}$$

where R is a general regression function (for instance it could be written in terms of a suitably chosen set of basis functions, including splines, if a nonlinear dependence on covariates is to be explored); $\vec{\boldsymbol{\beta}}_{loc}^{(l)}$ is the collection of local regression parameters, and $\boldsymbol{\beta}_{reg}^{(l)}$ is a collection of regional regression parameters, and $\vec{\boldsymbol{x}}(s,t)$ refers to a set of covariates defined with respect to site and time indices s and t, respectively.

In a hierarchical Bayesian model, pooling of information across gauges for regressions can be considered at three stages: no pooling, full pooling and partial pooling [*Devineni et al.*, 2013]. In a no pooling model, each gauge is modeled independently, thus the regression parameters only involve local parameters $\vec{\boldsymbol{\beta}}_{loc}^{(l)}$. Consequently, case studies using a no pooling model consider a series of local analyses. In a full pooling model, the regression parameters are typically assumed to have a regional effect, which is common over all stations. However, some regression parameters may still remain site-specific. Thus both local and regional regression parameters ($\vec{\boldsymbol{\beta}}_{loc}^{(l)}$ and $\vec{\boldsymbol{\beta}}_{reg}^{(l)}$) are involved. In a partial pooling model, the regression parameters are allowed to vary by site, but are assumed to be drawn from a common hyper-distribution D' across the region. The aim is to shrink the local parameter toward the common regional mean, but an estimation of the across site variance or covariance is also included. This hyper-distribution describes the second level of the hierarchical Bayesian model. A schematic of the partial pooling model is presented in Figure 1. The level 2 model may be used for each local parameter, if desired. As a consequence, the no pooling model and the full pooling model are two end points of the partial pooling model. We stratify the local parameters $\vec{\boldsymbol{\beta}}_{loc}^{(l)}$ into two groups $\vec{\boldsymbol{\beta}}_{loc,P}^{(l)}$ and $\vec{\boldsymbol{\beta}}_{loc,N}^{(l)}$, with the former having a second level model:

$$\vec{\boldsymbol{\beta}}_{loc,P}^{(l)} \sim D'^{(l)}\left(\vec{\boldsymbol{\alpha}}^{(l)}\right) \tag{3}$$

where $\vec{\boldsymbol{\alpha}}^{(l)}$ is the hyper-parameters, and $D'^{(l)}$ refers to the corresponding density function.

For any site s, the likelihood function of the hierarchical Bayesian model can be written as follows:

$$f_H(s) = \prod_{t=1}^{T} f_D\left(y(s,t)|\vec{\theta}(s,t)\right) \prod_{l=1}^{L} f_{D'^{(l)}}\left(\vec{\boldsymbol{\beta}}_{loc,P}^{(l)}|\vec{\boldsymbol{\alpha}}^{(l)}\right) \tag{4}$$

In particular, the likelihood function of a full pooling model for any site s is $f_H(s) = \prod_{t=1}^{T} f_D\left(y(s,t)|\vec{\theta}(s,t)\right)$. The complete likelihood function can be calculated by taking the product of $f_H(s)$ among all stations.

2.2. An Example

Consider a model for the linear temporal trend of annual maximum data $Y(s,t)$, which is assumed to be described by a GEV distribution at each site, with the location parameter defined by a linear function of the time index t. One example of the no pooling (equation (5)), full pooling (equation (6)) and partial pooling (equation (7)) models can be established as follows:

$$\text{No pooling}: \quad Y(s,t) \sim GEV(\mu_0(s) + \mu_1(s)t, \sigma_0(s), \xi(s)) \tag{5}$$

$$\text{Full pooling}: \quad Y(s,t) \sim GEV(\mu_0(s) + \mu_1 t, \sigma_0(s), \xi) \tag{6}$$

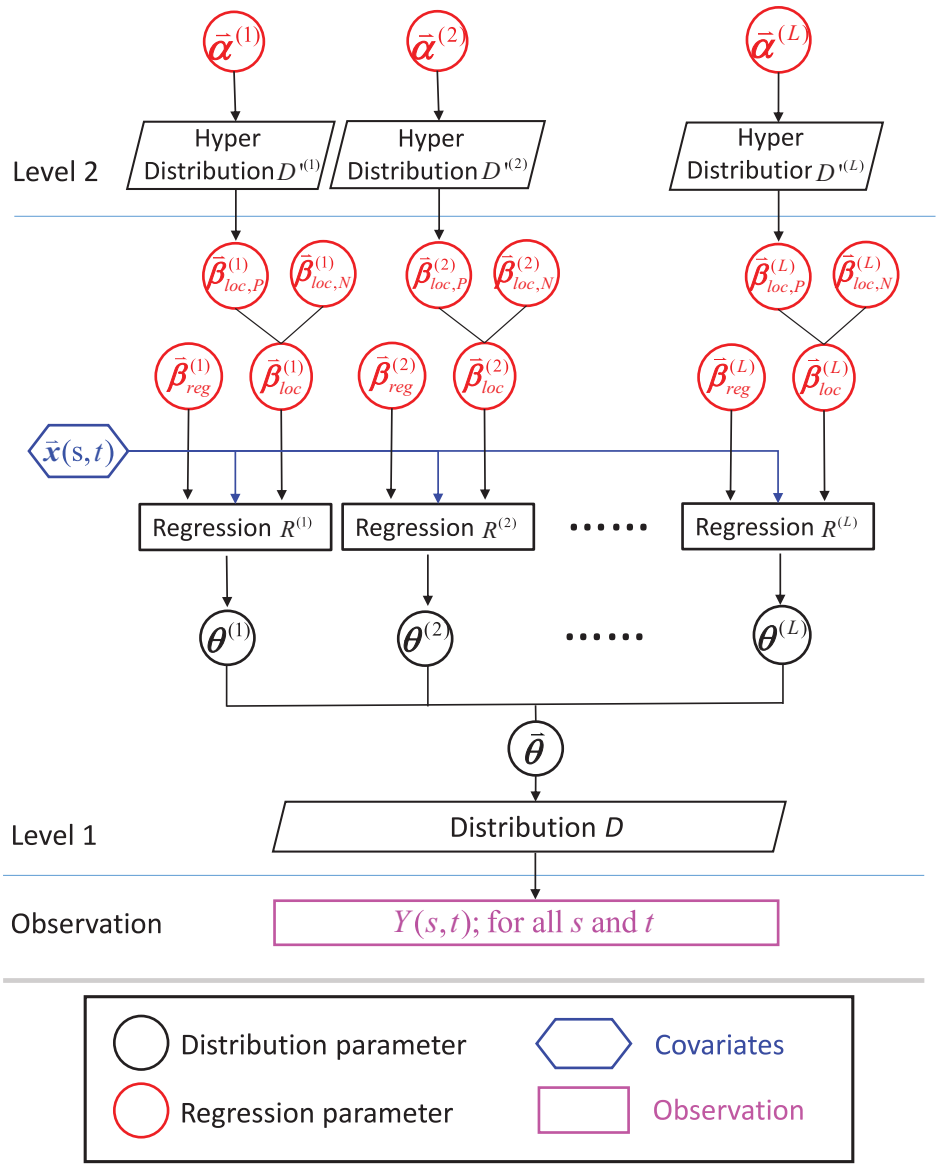

Figure 1. Schematic of the hierarchical Bayesian model. Level 2 applies in the partial pooling case where the local regression parameters follow a hyper-distribution. The no pooling model and full pooling model are particular cases of the partial pooling model.

$$\text{Patial pooling}: \begin{cases} \text{Level 1}: Y(s,t) \sim GEV(\mu_0(s)+\mu_1(s)t, \sigma_0(s), \xi) \\ \text{Level 2}: \mu_1(s) \sim N\left(\mu_\mu, \sigma_\mu\right) \end{cases} \tag{7}$$

where $\mu_0(s)$ is the intersection on the location parameter, $\mu_1/\mu_1(s)$ is the slope that characterizes the temporal trend, $\sigma_0(s)$ is the scale parameter and $\xi/\xi(s)$ is the shape parameter. The difference between μ_1 and $\mu_1(s)$ (or ξ and $\xi(s)$) is that the former is a regional parameter while the latter is a site-specific parameter. μ_μ and σ_μ are two hyper-parameters in the second level model of equation (7).

These three models use the same regression function on the location parameter of the GEV distribution to describe the temporal trend, while they differ in the settings of μ_1 and ξ which are regional, site-specific or having a second level model. Equations (5)–(7) constitute one example of parameter settings. Depending on each specific case study, the regression parameters are flexible to be regional, site-specific or to have a second level model.

In the no pooling model (equation (5)), both μ_1 and ξ are site-specific, and thus estimated locally. In the full pooling model (equation (6)), both are regional and estimated by using all data. In the partial pooling model (equation (7)), the temporal slope μ_1 of each site is assumed to come from a Gaussian distribution with hyper-parameters μ_μ and σ_μ. The idea of using a level 2 model for μ_1 is that the level 2 model can improve the precision when estimating the slope μ_1, while still allowing variation on the slope across stations rather than considering the same slope for all stations. ξ is still regional in equation (7), because it requires more data to obtain a precise estimation due to the large uncertainty [*Coles*, 2001, p. 106].

2.3. Hierarchical Bayesian Clustering Model

The hierarchical Bayesian clustering model is now introduced to deal with observations which are not necessarily in a homogenous region. We assume that the stations can be divided into K clusters, and the observations in each cluster follow a partial pooling model as described in section 2.1. The membership of each station to each cluster is not prescribed a-priori. Each station has a probability π_k to belong to cluster C_k. This leads to a mixture structure across clusters:

$$\text{For any station } s,\ \boldsymbol{Y}(s) \sim \sum_{k=1}^{K} \pi_k f_{H_k}(s) \tag{8}$$

where k is the cluster number. $\boldsymbol{Y}(s)=(Y(s,1),\ldots,Y(s,T))$ is the observation of site s, and $f_{H_k}(s)$ is the likelihood function of the hierarchical Bayesian model for cluster k. In equation (8), it is implicit that the membership of each station should be consistent for all time steps.

We reformulate some notations of the parameters to better represent the mixture structure. We denote the distribution parameter of each cluster by $\vec{\theta}_k=\left(\theta_k^{(l)}\right)_{l=1,\ldots,L}$. Due to the mixture structure, the covariates in the regression function are limited to temporal covariates. Thus the regression function applied to each $\theta_k^{(l)}$ is a function of temporal covariate $\vec{\boldsymbol{x}}(t)$ and the regression parameters:

$$\theta_k^{(l)}=R^{(l)}\left(\vec{\boldsymbol{\beta}}_{loc,P}^{(l)},\vec{\boldsymbol{\beta}}_{loc,N}^{(l)},\left(\vec{\boldsymbol{\beta}}_{reg}^{(l)}\right)_k;\vec{\boldsymbol{x}}(t)\right) \tag{9}$$

where $\vec{\boldsymbol{\beta}}_{loc,P}^{(l)}$ and $\vec{\boldsymbol{\beta}}_{loc,N}^{(l)}$ are local regression parameters with the former having a second level model. $\left(\vec{\boldsymbol{\beta}}_{reg}^{(l)}\right)_k$ are regional parameters which are identical over all stations inside the cluster k. The level 2 model for $\vec{\boldsymbol{\beta}}_{loc,P}^{(l)}$ is the following:

$$\vec{\boldsymbol{\beta}}_{loc,P}^{(l)} \sim D'^{(l)}\left(\left(\vec{\boldsymbol{\alpha}}^{(l)}\right)_k\right) \tag{10}$$

where $D'^{(l)}$ is the hyper-distribution, and $\left(\vec{\boldsymbol{\alpha}}^{(l)}\right)_k$ is the associated hyper-parameters.

The full likelihood of this hierarchical Bayesian clustering model can be calculated as follows:

$$\begin{aligned}\Lambda&=\prod_{s=1}^{S}\sum_{k=1}^{K}\pi_k f_{H_k}(s)\\&=\prod_{s=1}^{S}\sum_{k=1}^{K}\pi_k\prod_{t=1}^{T}f_D\left(y(s,t)|\vec{\boldsymbol{\beta}}_{loc,N}^{(l)},\vec{\boldsymbol{\beta}}_{loc,P}^{(l)},\left(\vec{\boldsymbol{\beta}}_{reg}^{(l)}\right)_k,\vec{\boldsymbol{x}}(t)\right)\prod_{l=1}^{L}f_{D'^{(l)}}\left(\vec{\boldsymbol{\beta}}_{loc,P}^{(l)}|\left(\vec{\boldsymbol{\alpha}}^{(l)}\right)_k\right)\end{aligned} \tag{11}$$

Taking the example of section 2.2 and considering K clusters as a candidate for the classification problem, the explicit formula for each cluster k can be extended from equation (7):

$$\begin{aligned}&\text{Level 1}: Y(s,t) \sim GEV(\mu_0(s)+\mu_1(s)t,\sigma_0(s),\xi_k)\\&\text{Level 2}: \mu_1(s) \sim N\left(\mu_{\mu_k},\sigma_{\mu_k}\right)\end{aligned} \tag{12}$$

where μ_0 and σ_0 are site-specific parameters without second level modeling, μ_1 is site-specific parameter which has a second level model, and ξ_k is the regional parameter. In this model $\xi_k,\mu_{\mu_k},\sigma_{\mu_k}$ are three parameters that are associated with cluster k.

2.4. Parameter Inference

In this model, the regression parameters and hyper-parameters are estimated in Bayesian framework with Markov chain Monte Carlo (MCMC) methods. No-U-Turn sampler [*Homan and Gelman*, 2014] is used as the sampling algorithm in MCMC. The posterior pdf of the regression parameters is given as follows:

$$f\left(\left(\vec{\boldsymbol{\beta}}^{(l)}_{loc,N}\right)_{l=1,\ldots,L},\left(\vec{\boldsymbol{\beta}}^{(l)}_{loc,P}\right)_{l=1,\ldots,L},\left(\vec{\boldsymbol{\beta}}^{(l)}_{reg}\right)_{k=1,..,K;l=1,\ldots L},\left(\vec{\boldsymbol{\alpha}}^{(l)}\right)_{k=1,..,K;l=1,\ldots,L},\vec{\pi}_{\,k=1,..,K}|Y(s,t)\right)$$
$$\propto \Lambda * f\left(\left(\vec{\boldsymbol{\beta}}^{(l)}_{loc,N}\right)_{l=1,\ldots,L},\left(\vec{\boldsymbol{\beta}}^{(l)}_{reg}\right)_{k=1,..,K;l=1,\ldots L},\left(\vec{\boldsymbol{\alpha}}^{(l)}\right)_{k=1,..,K;l=1,\ldots,L},\vec{\pi}_{\,k=1,..,K}\right) \tag{13}$$

where $f\left(\left(\vec{\boldsymbol{\beta}}^{(l)}_{loc,N}\right)_{l=1,\ldots,L},\left(\vec{\boldsymbol{\beta}}^{(l)}_{reg}\right)_{k=1,..,K;l=1,\ldots L},\left(\vec{\boldsymbol{\alpha}}^{(l)}\right)_{k=1,..,K;l=1,\ldots,L},\vec{\pi}_{\,k=1,..,K}\right)$ is the prior joint pdf, and Λ is the likelihood function calculated in equation (11). To the specific model used in section 3, we associate a Dirichlet distribution with identical parameters (e.g., a vector of 2 with length k in this study) as the prior for π_k, and flat priors (normal or uniform distribution with large variance) for the rest parameters. For the initial values, we fit a stationary GEV distribution first using maximum likelihood to obtain an approximate value of the location, scale and shape parameters for each station. These were then used as the starting point for the intercept of location parameter μ_0, scale σ_0 and shape ξ of each station in our clustering model. The slope parameters μ_1 and dependence parameter λ is set to 0 initially. For the hyper-parameters, the mean of μ_1 is set to 0 initially. Four chains using different initial value of π_k, standard deviation of μ_1 and shape parameter ξ, and 30,000 simulations (burn-in the first half simulations) are run for all parameters. Convergence is investigated by GR index [*Gelman and Rubin*, 1992], which should be smaller than 1.2 for each parameter.

2.5. Spatial Dependence

The clustering model can increase the power of detection for weak signals by appropriately grouping stations together. However, if the spatial dependence is ignored, uncertainties will be under-estimated. This is because the same information shared by the correlated stations will be used repeatedly, which falsely reduces the estimation uncertainties.

In a hierarchical Bayesian model, a formal way to consider spatial dependence is to model the spatial data with a multivariate distribution, where the dependence is explained by the covariance matrix [e.g., *Chen et al*., 2014; *Devineni et al*., 2013]. This requires well-known characterization of the multivariate distribution, such as multi-Normal distribution. However, if the multivariate form of a distribution is not well known (e.g., GEV distribution), Gaussian or Student copulas can be used to obtain the joint distribution, where the dependence is considered through the covariance matrix of a multivariate Gaussian or Student distribution [e.g., *Renard*, 2011; *Sun et al*., 2014]. Unfortunately, how to apply these two approaches to hierarchical Bayesian clustering model is not immediately obvious, since the membership of each station is unknown during the process. Thus, it is not clear how a multivariate distribution can be applied. Therefore, we seek to model the spatial dependence through the streamflow network [*Jensen*, 1996; *Pearl*, 2014]. A specific approach to this study is presented in section 3.2.2.

2.6. Number of Clusters

In a case study, the number of clusters K is generally unknown a priori. In the Bayesian framework, the identification of the number of clusters K has two possibilities. One is to consider K as a regression parameter whose posterior distribution is estimated along with the other regression parameters. However, varying K in the mixture model is usually computationally intensive and analytic expressions for computation in non-Gaussian frameworks are lacking. The other is to fit models for varying choice of K and use some model selection techniques to choose the best number K based on the goodness-of-fit and model complexity criterion, such as Bayes factors [*Kass and Raftery*, 1995], Deviance Information Criterion (DIC) [*Spiegelhalter et al*., 2002], the Watanable-Akaike Information Criterion (WAIC) [*Watanabe*, 2010]. Unfortunately, these criterion have theoretical or practical limitations for mixture models [*Gelman et al*., 2014, chap. 22.4]. There are also many other approaches for choosing the number of clusters based on different measures of homogeneity or performance under cross validation. The "silhouette coefficient" [*Rousseeuw*, 1987] method can be

used to evaluate the tightness of clusters, and this was used in the applications here to guide the selection of the number of clusters.

3. Case Study

3.1. Case Study Area, Flood Characteristics and Data

We study the trends in flood magnitude at 68 streamflow gauges distributed across Germany. These gauges have different flood regimes [*Beurton and Thieken*, 2009]. Catchments in the western and central part of Germany have floods in winter, and annual maximum flow occurs only rarely in other seasons. Gauges in North and East Germany are also dominated by winter floods, however with a considerable share of spring and summer floods. The southern region is dominated by summer floods generated by snowmelt events in the catchments of the alpine Danube tributaries.

There are a number of flood trend studies for different catchments or subregions of Germany, however, only two studies covered Germany completely. *Petrow and Merz* [2009] applied the Mann-Kendall test, considering field significance and auto-correlation, for different flood indicators for 145 gauges in the period 1951–2002. They found, depending on the flood indicator used, significant changes at the 10% significance level for 10% to 45% of the gauges. Gauges in the Rhine and Danube catchments in west and south Germany had a considerably higher fraction of changes compared to the Weser and Elbe catchments in North and East Germany. Most of the significant changes were upward. The spatial and seasonal coherence of significant changes suggested climate as driver of the detected trends. In a related study, *Petrow et al.* [2009] clustered the gauges according to their flood seasonality and compared flood changes with changes in flood-prone atmospheric circulation patterns. They found, in particular for flood-prone circulation patterns, a tendency for higher persistence supporting the conclusion that flood changes in the second half of the 20th century in Germany may be dominated by climate effects.

Our study investigates a longer time period, i.e., the hydrological years 1934–2005 (1 November 1934 to 31 October 2005). Daily discharge time series were obtained from the water authorities of different federal states of Germany. The data are part of the hydrometric observation network of the water authorities in Germany, hence, the observations are regularly checked and can be assumed to be of good quality. For this joint period, annual maximum daily stream flow from 68 catchments was derived. The catchments cover

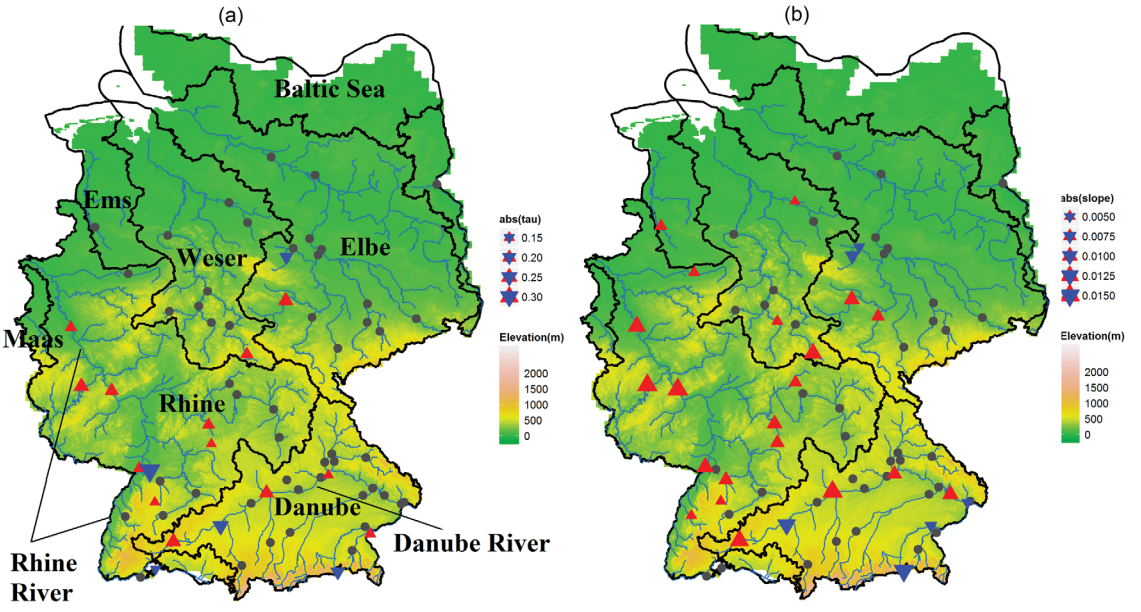

Figure 2. (a) Mann Kendall trend test with a significance level of 0.1 for annual maximum stream flow, where red (blue) triangles represent the stations with significant positive (negative) trend. The size of triangles shows the absolute value of the Kendall rank correlation coefficient. (b) No pooling model (equation (5)) for annual maximum stream flow, where red (blue) triangles represent the stations where μ_1 is significantly positive (negative), i.e., 90% credibility interval of μ_1 is larger (smaller) than 0. The size of triangles shows the magnitude of slope μ_1.

almost the complete area of Germany (Figure 2). Seven gauges had missing values for one complete year, and five gauges had data gaps of more than 1 year. These gaps were filled by correlating the daily time series to the station with the highest correlation coefficient. The catchment area varies between 135 and 144,232 km^2 with a median of 4,151 km^2.

The annual maximum streamflow data were preprocessed by standardizing the data by subtracting the mean and dividing by the standard deviation of each series. An alternate model, that considers scaling of the at-site mean and standard deviation in terms of the drainage area [e.g., *Lima and Lall*, 2010] could have been pursued. However, since the focus of this paper was on demonstrating a clustering approach for trend identification, a simpler model was pursued.

3.2. Models

Considering the large area and the variability of flood regimes in Germany, the homogeneity assumption across the 68 stations may not hold, and thus exploring whether a single model or a clustered model applies is of interest. We apply the hierarchical Bayesian clustering model (section 2.3), to see whether this model can improve the precision of the estimates and obtain more robust information on the trends in the extreme flows compared with the local analysis (using no pooling models). A step by step development (from no pooling, to full pooling, and then to partial pooling) of the clustering model will be demonstrated in this study.

Similar to the examples described in section 2.2, the same regression functions are consistently used throughout the case study for the no pooling and clustering models (for both full pooling and partial pooling models). A time-varying GEV distribution is applied to the extreme data by assuming that the regression function on the location parameter has a temporal trend [e.g., *Coles*, 2001; *Katz et al.*, 2002]:

$$Y(t) \sim GEV(\mu_0 + \mu_1 t, \sigma_0, \xi) \tag{14}$$

where $Y(t)$ is the observation of time t, μ_0, μ_1, σ_0 and ξ are the regression parameters that need to be estimated, in which μ_1 is the parameter that characterizes the temporal trend. Thus the no pooling model is the same as presented in equation (5).

3.2.1. Clustering Model Ignoring Spatial Dependence

We first consider clustering models without considering spatial dependence across the gauges. Two full pooling and two partial pooling models (Table 1) are considered for each cluster k. The third column of Table 1 lists the regional parameters, which are used to cluster the stations. Each station belongs to cluster k with probability π_k.

Table 1. Regional Models for Clustering

Model	Distribution and Regressions	Parameters for Clustering
Full Pooling Model		
M1	$Y(s,t) \sim GEV\left(\mu_0(s) + \mu_{1,k} t, \sigma_0(s), \xi(s)\right)$	$\mu_{1,k}$
M2	$Y(s,t) \sim GEV\left(\mu_0(s) + \mu_{1,k} t, \sigma_0(s), \xi_k\right)$	$\mu_{1,k}, \xi_k$
Partial Pooling Model Ignoring Spatial Dependence		
M3	Level 1 : $Y(s,t) \sim GEV(\mu_0(s) + \mu_1(s)t, \sigma_0(s), \xi(s))$ Level 2 : $\mu_1(s) \sim N\left(\mu_{\mu_k}, \sigma_{\mu_k}\right)$	$\mu_{\mu_k}, \sigma_{\mu_k}$
M4	Level 1 : $Y(s,t) \sim GEV(\mu_0(s) + \mu_1(s)t, \sigma_0(s), \xi_k)$ Level 2 : $\mu_1(s) \sim N\left(\mu_{\mu_k}, \sigma_{\mu_k}\right)$	$\mu_{\mu_k}, \sigma_{\mu_k}, \xi_k$
Partial Pooling Model Considering Spatial Dependence		
M5	Level 1 : $Y(s,t) \sim GEV\left(\mu_0(s) + \mu_1(s)t + \sum_{v \in V_s} \lambda_v y(s_v, t), \sigma_0(s), \xi(s)\right)$ Level 2 : $\mu_1(s) \sim N\left(\mu_{\mu_k}, \sigma_{\mu_k}\right)$	$\mu_{\mu_k}, \sigma_{\mu_k}$
M6	Level 1 : $Y(s,t) \sim GEV\left(\mu_0(s) + \mu_1(s)t + \sum_{v \in V_s} \lambda_v y(s_v, t), \sigma_0(s), \xi_k\right)$ Level 2 : $\mu_1(s) \sim N\left(\mu_{\mu_k}, \sigma_{\mu_k}\right)$	$\mu_{\mu_k}, \sigma_{\mu_k}, \xi_k$

In the full pooling model (Table 1: M1, M2), stations within the same cluster have the same temporal trend. In the partial pooling model (Table 1: M3, M4), the trend of stations within each cluster is assumed to come from a Gaussian distribution at the second level of the model. In M1 and M3, the clustering is based only on the trend slope, while in M2 and M4 both the trend slope and the shape parameter of the distribution are considered.

In both the full pooling and partial pooling model, we pool the information through the trend and/or shape parameter. The motivation of considering these two parameters for clustering is that, for the standardized extreme data used in the case study, the shape parameter determines the tail behavior of the extreme value distribution, and is difficult to estimate. Pooling this parameter across sites, or spatially modeling it may offer improvements in the uncertainty of extremes across all sites. Likewise modeling the trend as indexed to space is useful. The models exemplified in this study are used to demonstrate the usefulness of the clustering model. We therefore do not attempt to compare the other combinations of settings on different regression parameters.

3.2.2. Clustering Model With Spatial Dependence

The spatial dependence is considered across the stream flow network. We assume that the dependence between the stations along one river can be summarized through the most immediate upstream stations, on any contributing tributaries. Thus we add an additional term to the location parameter of the GEV distribution to represent the contribution of the immediate upstream stations. As an illustration, considering model M3 as the original model, and station B and C being the immediate upstream stations of A, the level 1 model of M3 becomes:

$$\begin{gathered}\text{At Station A:}\\ Y(s_A,t)\sim GEV\left(\mu_0(s_A)+\mu_1(s_A)t+\lambda_{B,A}y(s_B,t)+\lambda_{C,A}y(s_C,t),\sigma_0(s_A),\xi(s_A)\right)\end{gathered} \tag{15}$$

where $y(s_B,t)$, $y(s_C,t)$ are the annual maximum flows for station B and C; and $\lambda_{B,A}$ and $\lambda_{C,A}$ are the dependence parameter characterizing the influence of station B and C on station A, which need to be estimated. The complete partial pooling model is:

$$\begin{gathered}\text{Level 1}: Y(s,t)\sim GEV\left(\mu_0(s)+\mu_1(s)t+\sum_{v\in V_s}\lambda_{v,s}y(s_v,t),\sigma_0(s),\xi(s)\right)\\ \text{Level 2}: \mu_1(s)\sim N\left(\mu_{\mu_k},\sigma_{\mu_k}\right)\end{gathered} \tag{16}$$

where an additional term $\sum_{v\in V_s}\lambda_{v,s}y(s_v,t)$ is added to the location parameter compared with M3. V_s is the collection of the immediate upstream stations of station s, $y(s_v,t)$ is the observation of an immediate upstream station s_v, and $\lambda_{v,s}$ is the corresponding dependence parameter. As a part of regression, $\lambda_{v,s}$ is estimated simultaneously with the other parameters. We denote equation (16) by M5, which is extended from M3 to account for the spatial dependence. Similarly, the extension of M4 is denoted by M6. The detailed equations are presented in Table 1.

In this structure, the slope parameter $\mu_1(s)$ no longer represents the trend of the station s. Instead, it represents the net trend after subtracting the effect of immediate upstream stations. Consequently, the membership of the stations identified in model M5 and M6 could be different from the membership identified in model M3 and M4, and the trend at the examined station needs to be reconstructed by accounting both its net trend and the trend of the immediate upstream stations along with their dependence parameters.

3.3. Results

3.3.1. Local Model

Figure 2 shows the results from the local analysis. For comparison the widely used Mann-Kendall test has been applied as well (Figure 2a). It shows that for 26.5% (18 out of 68 stations) of all stations trends are significant at the 0.1 significance level. Most of the significant trends are positive (13 stations). The Rhine catchment shows the largest share of significant trends (50% stations in the catchment). The overall result for the no-pooling model is similar (Figure 2b). However, more stations (39.7%; 27 out of

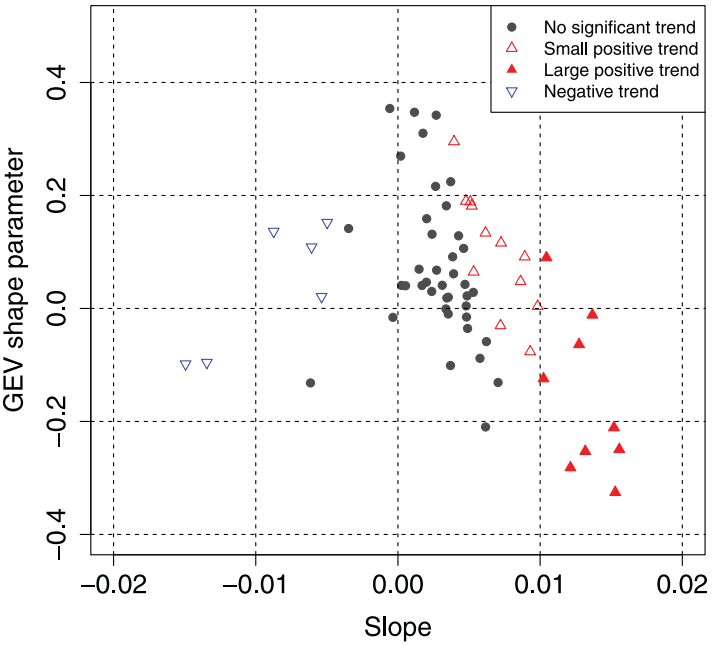

Figure 3. Scatterplot of the slope (trend) parameter μ_1 and shape parameter ξ in the local analysis. Red triangles and blue triangles are stations with significant positive and negative μ_1 values in Figure 2b, respectively.

68), particularly in the South and West, are found to have a significant trend compared to the Mann-Kendall test.

The scatterplot between the estimated slope μ_1 and shape ξ (Figure 3) shows that most of the stations with near zero trend have a positive shape parameter (grey dots), while most of the stations with large positive trend (solid red triangles) have negative shape values. This indicates that considering the shape parameter may help to better classify the trend.

3.3.2. Full Pooling Clustering Models

Table 2 summarizes the results for the full pooling models with *K* equal to 2, 3 and 4. In general, the full pooling models identify more clusters than the partial pooling models. This is expected because the partial pooling models allow more variation within each cluster in the trend parameter μ_1, and accounting for the uncertainty of the trend at each station, a fewer number of classes are sufficient in the partial pooling model.

For the full pooling model, two clusters are identified from M1. One cluster contains the stations with negative trend, and the other contains the remaining stations, which is a mixture of no trend and positive trend stations. When the shape parameter is also considered for clustering (M2), three clusters are identified. The new cluster contains the stations with large positive trend. This corresponds to the finding in Figure 3 that stations with large trend slope tend to have a negative shape parameter. This highlights the utility of considering the shape parameter for clustering in this case study.

Table 2. Summary of the Clustering Results for Models in Table 1[a]

	K=2	K=3	K=4
Full Pooling Model			
M1	Two clusters are detected: 1. Negative trend (μ_1) 2. The rests (no trend and positive trend are mixed)	Same as K=2	Same as K=3
M2	Two clusters are detected: 1. Large positive trend (μ_1=0.01) with negative shape (ξ) 2. Near zero trend with positive shape (ξ)	Three clusters are detected: 1. Negative trend with near zero shape 2. Large positive trend (μ_1=0.01) with negative shape 3. Small trend (μ_1=0.003) with positive shape	Same as K=3
Partial Pooling Model Ignoring Spatial Dependence			
M3	All stations are in the same cluster.	Same as K=2.	Same as K=3
M4	Two clusters are detected: 1. Small trend (μ_μ=0.003) with positive shape 2. Small trend (μ_μ =0.005) with negative shape	Same as K=2.	Same as K=3
Partial Pooling Model Considering Spatial Dependence			
M5	All stations are in the same cluster.	Same as K=2.	/
M6	Two clusters are detected: 1. Small net trend (μ_μ =0.002) with negative shape 2. Small net trend (μ_μ =0.003) with positive shape	Same as K=2	/

[a]The values of the trend in the parentheses is the median of posterior pdf of $\mu_{1,k}$ for the full pooling models and μ_{μ_k} for the partial pooling models.

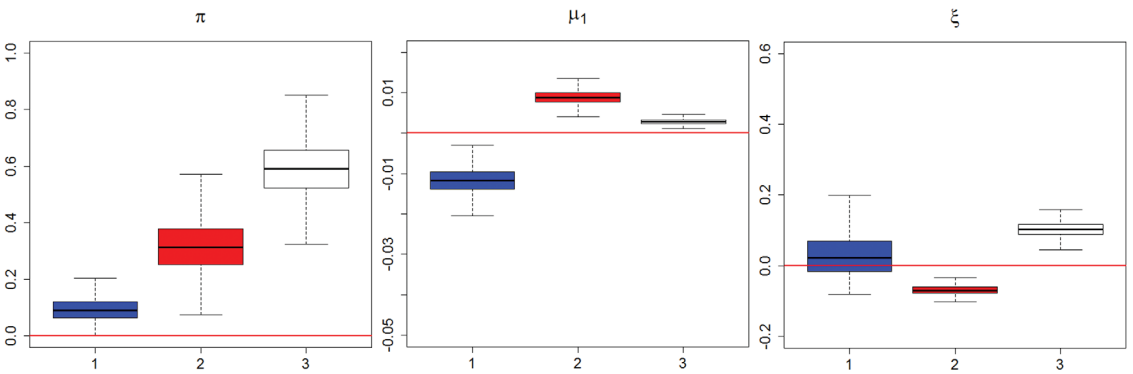

Figure 4. Posterior distribution of π_k, μ_{1k} and ξ_k for the full pooling model *M2* (K=3). The two "hinges" of the box are the first and third quartile. The lower and upper extremes are $\pm 1.58\ IQR/\sqrt{\text{sample size}}$ as the classical definition of the box.

Figure 4 shows the posterior distribution of π_k, μ_{1k} and ξ_k for M2 (K=3). Based on the median posterior probabilities, we note that about 10%, 30% and 60% of the stations belong to clusters 1, 2 and 3 respectively. The posterior distributions of μ_{1k} and ξ_k are clearly separated across the clusters, with largest uncertainty for cluster 1, as expected given the smallest set of stations that belong to it. According to the posterior distribution of all regression parameters, the posterior probability of station s belonging to cluster k can be calculated:

$$\text{Prob}(s \in \textit{cluster}\ k) = \frac{\pi_k\ f_{H_k}(s)}{\sum_{k=1}^{K} \pi_k f_{H_k}(s)} \tag{17}$$

Figure 5a illustrates the membership of each station according to the largest posterior probability for each station belonging to cluster k for model M2 (K=3) using the modal parameters (defined as the parameter set providing the largest posterior distribution in the MCMC iterations). According to the estimation results (Table 2 (M2, K=3)), blue dots, red dots and grey dots in Figure 5a represent the cluster with negative trends, large positive trends and weak positive trends, respectively. The cluster with large positive trends contains most of the stations along the Danube River and in the Rhine catchment. To compare this result

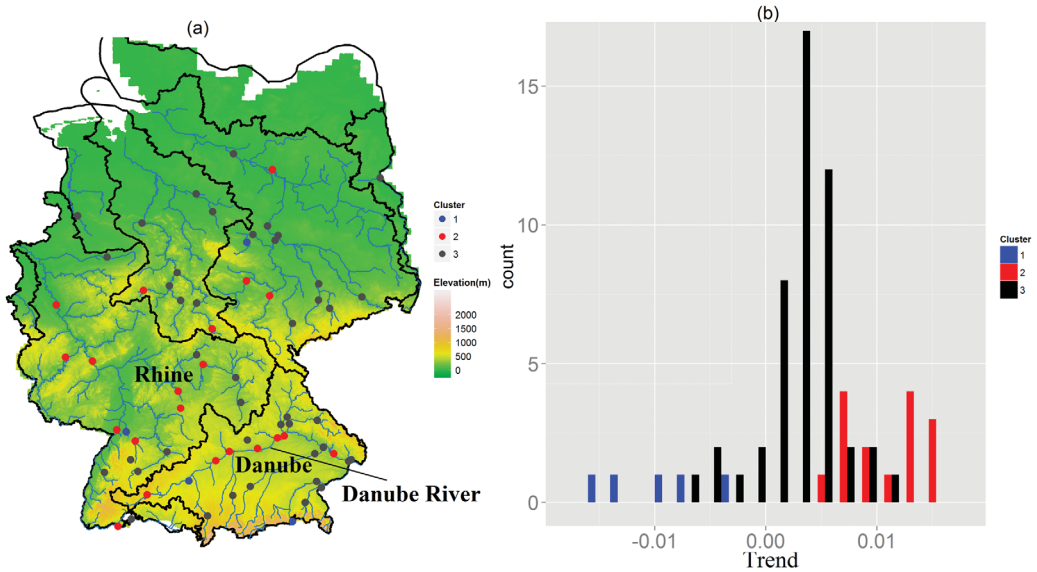

Figure 5. (a) Clusters for the full pooling model *M2* (K=3) determined by the largest posterior belonging probability for each station. (b) Histogram of μ_1 estimated in the local analysis for each cluster of *M2* (K=3).

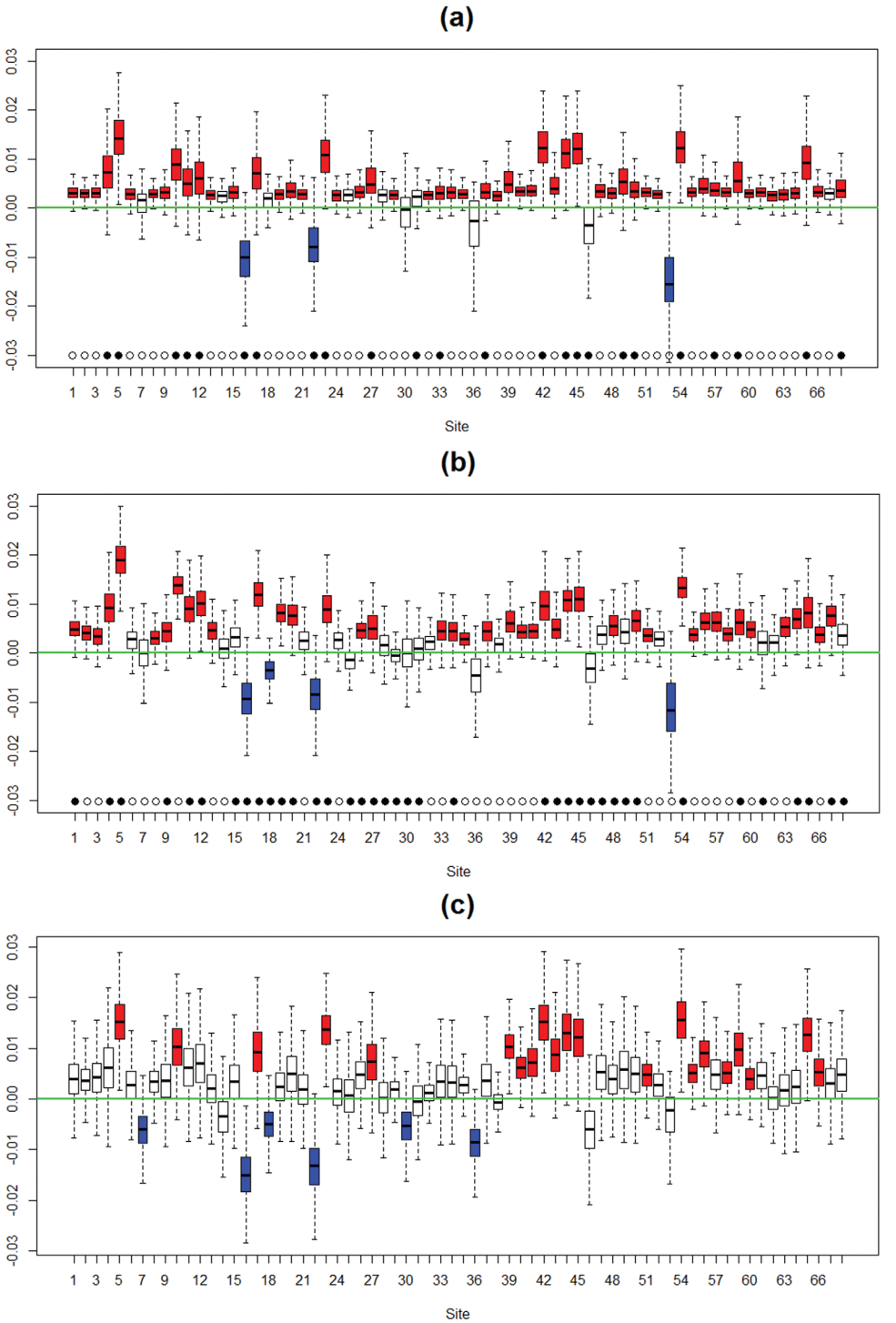

Figure 6. (a) Boxplot plot of trend parameter μ_1 of each station estimated in *M4* ($K=2$). (b) Boxplot of the trend estimated in *M6* ($K=2$), which is the temporal trend reconstructed through the flow network, rather than the μ_1 estimated in *M6* ($K=2$). (c) Boxplot plot of trend parameter μ_1 of each station for no pooling model. Red boxes represent the stations with significant positive trend, while the blue boxes are for the stations with significant negative trend. The solid and hollow dots in Figures 6a and 6b represent the cluster of each station.

with the local analysis, we plot the histogram (Figure 5b) of the temporal trend μ_1 estimated in the local model (equation (5)) against the membership of each station illustrated in Figure 5a. The histograms in blue, black and red show the trends with negative values, near zero values and large positive values,

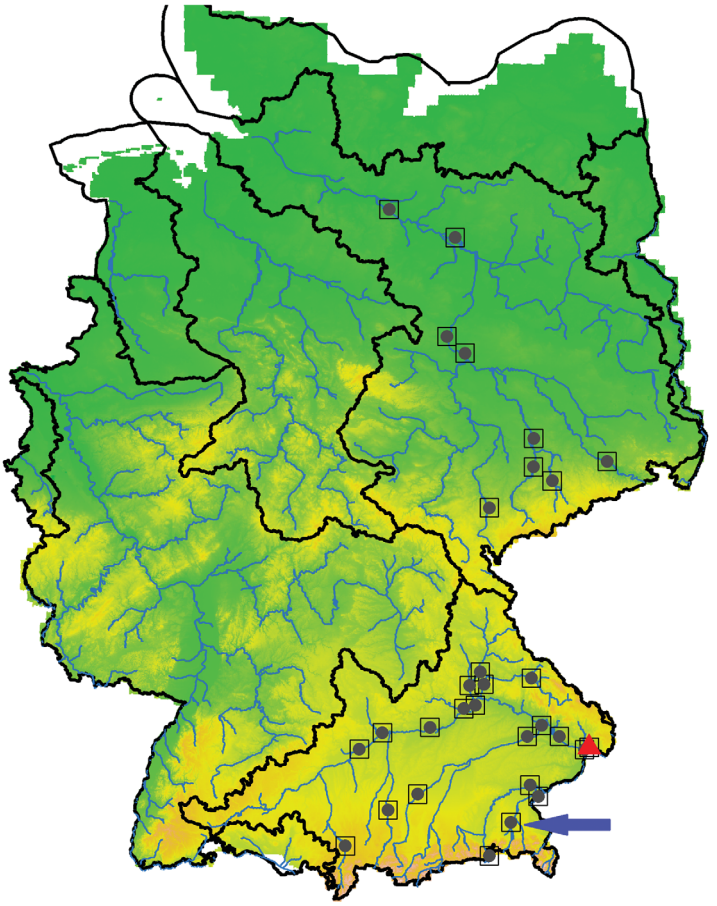

Figure 7. Correlation between station Achleiten (red triangle) and the other stations. The squares show the stations whose empirical correlation with station Achleiten on the original data is larger than 0.4. Grey dots show the same, but on the residual of M4. The pointed station is the only station whose empirical correlation with station Achleiten on the residual of M6 is larger than 0.4.

respectively, which is consistent with the result of model M2 (K=3). Thus model M2 (K=3) provides a good classification when clustering on the temporal trend and the shape parameter.

The main drawback of the full pooling model is that it does not provide a site specific value for the trend, because stations within each cluster share the same trend. To determine whether trend exists at a site, one needs to verify whether the 90% coverage of the posterior distribution of the estimated trend does not across zero. The partial pooling model presented next explores this issue.

3.3.3. Partial Pooling Models

In the partial pooling models, all stations are merged into a single cluster when only the slope is used for clustering (M3 and M5), while two clusters are identified when using both slope and shape for clustering (M4 and M6). The reduction in the number of clusters is expected since we now allow heterogeneity of response within each cluster. Figure 6a shows the posterior distribution of the trend μ_1 for each station for M4, which ignores the spatial dependence. It indicates that most stations with large positive and negative trends are in one cluster and the remaining stations are in the other cluster (with some exceptions like station 36 and 53). This classification is consistent with that found in the full pooling model, while allowing variation of trends across stations in a cluster.

As discussed in section 3.2.2, when spatial dependence exists and is ignored, estimation uncertainties will be under-estimated due to the strong correlation between stations. As a result, for the stations with small trends, the under-estimation of uncertainty will cause these trends to be falsely considered significant. To justify the approach for spatial dependence described in section 3.2.2, we investigate the residuals of different models. If the correlation between residuals is low, spatial dependence is appropriately described by the covariates in the regression. We firstly verify the correlation between the stations on our original data

Table 3. Summary of Model Performance

	Full Pooling Model		Partial Pooling Model Ignoring Spatial Dependence		Partial Pooling Model Considering Spatial Dependence	
	M1	M2	M3	M4	M5	M6
Identify clusters	√	√		√		√
Detect at-site trends			√	√	√	√
Reduce uncertainty	√	√	√	√	√	√
Model Spatial dependence					√	√

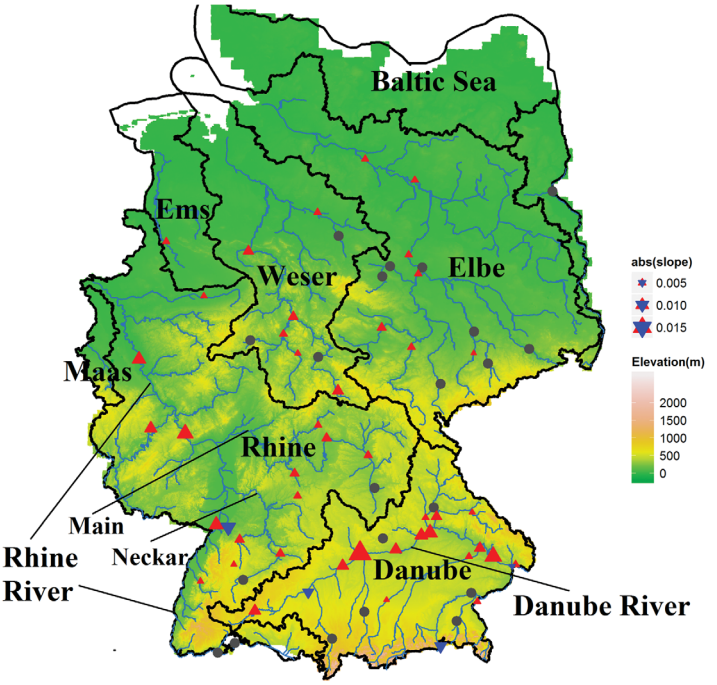

Figure 8. The reconstructed temporal trend from the partial pooling model including inter-site dependence *M6* ($K=2$). Red triangles are the stations with significant positive trend, while the blue triangles are for stations with significant negative trend. Grey dots are the stations without significant trend.

set. If strong correlation exists, it is necessary to consider spatial dependence. Then we verify the correlation between the residuals of M4, which is data removing the temporal trend estimated in M4. In the end, we check the correlation of residuals in M6. Figure 7 shows the correlation between the station Achleiten and other stations. The squares show the stations whose empirical correlation with station Achleiten on original data is larger than 0.4, while grey dots show the same but on the residuals of M4. It is found that more than 20 stations have a large correlation with station Achleiten on both original data set and the residuals of M4. It indicates that it is necessary to consider spatial dependence, since the statistically significant correlation cannot be explained by temporal trends. However, there is only one station that has a larger correlation on the residual of M6. This indicates that the spatial dependence of the stations can be mostly explained through the immediate upstream stations, which suggests that M6 is an appropriate model for considering the spatial dependence.

Figure 6b shows the posterior distribution of the reconstructed trend from M6. The uncertainty of μ_1 is larger in Figure 6b than in Figure 6a. This confirms our argument that ignoring spatial dependence leads the under-estimation of estimation uncertainties. The membership shown in Figure 6a is based on the flow observed at each station, while the membership in Figure 6b is based on the flow subtracting the contribution of immediate upstream. This is the why the membership identified is different in Figures 6a and 6b. Compared with the local analysis (Figure 6c), the estimation uncertainty in the clustering model (Figure 6b) is notably reduced, leading to a better identification of the trend.

Table 3 summarizes the performance of the six models introduced in Table 1. Among the six models, M6 covering all the four points on the list, is therefore more suitable than the other models for this case study. In fact, the order of introducing M1 to M6 shows an exploratory demonstration of developing specific clustering models for a specific case study.

As a result, from the partial pooling model with spatial dependence (M6, Figure 8), 45 out of 68 (66.2%) stations detect a significant trend, which is many more than that detected using a no-pooling model. Large, consistent and significant temporal trends are found for the stations along the Danube River, but the trend on upstream branches is generally small or not significant. The Rhine and its tributaries Neckar and Main also show a high fraction of stations with significant positive trends. There are also some small, however significant positive trends at the stations in the North, i.e., at the rivers Elbe, Weser and Ems.

4. Discussion and Conclusions

We presented a regional hierarchical Bayesian clustering model to study the temporal trend annual maximum daily stream flow across Germany. Under this model, a variety of possible modeling structures, frequency distributions, and regression functions can be explored, and compared in terms of the maximum posterior criteria as well as the resulting solutions and what they highlight. Unlike some methods [e.g., *Pujol*

et al., 2007] where stations are clustered first using some criterion, and then the analysis is done for each cluster, the hierarchical Bayesian model presented allows the clustering and the parameter estimation to proceed simultaneously. Compared with the local analysis, the clustering model reduced the estimation uncertainty by transferring information across gauges with similar characteristics.

A GEV distribution with a temporal linear regression function on the location parameter was applied to the example data set for Germany, on standardized annual maximum flows. The trend and/or the shape parameter were used for clustering. It was found that including the shape parameter as a criteria for clustering led to a better identification of clusters. Both the full pooling and partial pooling models provided generally consistent clusters and identified the significant trends in the southern and western stations. The partial pooling model with the consideration of spatial dependence is the most appropriate model among the discussed models. According to this model, significant trends were detected on the gauges alone the Danube River and Rhine River, and small to moderate trends for gauges in Central and Northern Germany.

The hierarchical clustering model developed in this paper is a flexible way to incorporate different components into clustering. These components are not limited to the climate effect, temporal trend or shape. Other components like the mean, standard deviation or seasonality can also be employed for clustering. Aswe demonstrated in this paper, adding an appropriate component will strength the identification of clusters. However, if too many components are used for clustering, it is challenging to obtain convergence of the MCMC in a high dimensional space. Thus an exploratory analysis is necessary to point to the components which should be used for clustering for regional frequency analysis models in nonhomogeneous regions. An extension of the model can include using inter-site dependence for clustering [*Bernard et al*., 2013].

In this example application, we demonstrated that the partial pooling model with Markovian spatial dependence on the river network has the best results. In general, when comparing no pooling, full pooling and partial pooling model, the no pooling (local) model has the largest uncertainty due to the low signal-to-noise ratio in a single site data. Both full pooling and partial pooling models are regional approaches which strengthen the signal-to-noise ratio by grouping the stations in the space. When choosing between full pooling and partial pooling, in a specific case the issue is whether trend or climate effects are believed to be identical or to be random with a common mean and variance over the domain. The former provides only regional estimation, while the latter retain also site-specific estimation. In both regional models, considering spatial dependence can avoid under-estimate uncertainties.

A few comments about possible extensions and applications of the model are presented in closing. In this paper, we used the immediate upstream stations to consider inter-site dependence in the regional models. However, annual maximum stream flow of some gauges is highly correlated even though these gauges are not nested, i.e., they belong to different watersheds. This indicates that large-scale climate variations may play an important role in the variation of maximum stream flow across Germany. Further improvement on modeling spatial dependence can focus on jointly considering the distribution of spatial data [e.g., *Devineni et al*., 2013; *Sun et al*., 2014], which will be able to account the dependence is nonnested stations. This is not evident in the current model because the membership of stations are unknown prior to the estimation. In subsequent work, the clustering model developed in this paper will be used to study the consistency of climate effects on extreme stream flow over different regions. The clustering model described in this paper can also be used to predict the new values on the gauged stations following by the mathematical formula [*Gelman et al*., 2014, p. 7] and practical guidance [*Renard*, 2011] in Bayesian framework.

Acknowledgments

We thank the Federal Agency for Cartography and Geodesy in Germany (BKG) for provision of the digital elevation model of Germany. We are grateful for providing the discharge data: Bavarian State Office of Environment (LfU), Baden-Württemberg Office of Environment, Measurements and Environmental Protection (LUBW), Brandenburg Office of Environment, Health and Consumer Protection (LUGV), Saxony State Office of Environment, Agriculture and Geology (SMUL), Saxony-Anhalt Office of Flood Protection and Water Management (LHW), Thüringen State Office of Environment and Geology (TLUG), Hessian Agency for the Environment and Geology (HLUG), Rhineland Palatinate Office of Environment, Water Management and the Factory Inspectorate (LUWG), Saarland Ministry for Environment and Consumer Protection (MUV), Office for Nature, Environment and Consumer Protection North Rhine-Westphalia (LANUV NRW), Lower Saxony Office for Water Management, Coast Protection and Nature Protection (NLWKN), Water and Shipping Management of the Fed. Rep. (WSV), prepared by the Federal Institute for Hydrology (BfG). The data sets are curated by the third author and data requests should be sent to bmerz@gfz-potsdam.de. This work was supported in part by a grant from the American International Group (CU12-2326). U. Lall was supported by an IPA from the US Army Corps of Engineers.

References

Aryal, S. K., B. C. Bates, E. P. Campbell, Y. Li, M. J. Palmer, and N. R. Viney (2009), Characterizing and modeling temporal and spatial trends in rainfall extremes, *J. Hydrometeorol.*, *10*(1), 241–253.

Begueria, S., M. Angulo-Martinez, S. M. Vicente-Serrano, J. Ignacio Lopez-Moreno, and A. El-Kenawy (2011), Assessing trends in extreme precipitation events intensity and magnitude using non-stationary peaks-over-threshold analysis: A case study in northeast Spain from 1930 to 2006, *Int. J. Climatol.*, *31*(14), 2102–2114.

Bernard, E., P. Naveau, M. Vrac, and O. Mestre (2013), Clustering of maxima: Spatial dependencies among heavy rainfall in France, *J. Clim.*, *26*(20), 7929–7937.

Beurton, S., and A. H. Thieken (2009), Seasonality of floods in Germany, *Hydrol. Sci. J.*, *54*(1), 62–76.

Bezdek, J. C., R. Ehrlich, and W. Full (1984), FCM: The fuzzy c-means clustering algorithm, *Comput. Geosci.*, *10*(2), 191–203.

Cannon, A. J. (2010), A flexible nonlinear modelling framework for nonstationary generalized extreme value analysis in hydroclimatology, *Hydrol. Processes*, *24*(6), 673–685.

Chavez-Demoulin, V., and A. C. Davison (2005), Generalized additive modelling of sample extremes, *J. R. Stat. Soc., Ser. C*, *54*(1), 207–222.

Chen, X., Z. Hao, N. Devineni, and U. Lall (2014), Climate information based streamflow and rainfall forecasts for Huai River basin using hierarchical Bayesian modeling, *Hydrol. Earth Syst. Sci.*, *18*(4), 1539–1548.
Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, vol. 208, Springer, London, U. K.
Cooley, D., D. Nychka, and P. Naveau (2007), Bayesian spatial modeling of extreme precipitation return levels, *J. Am. Stat. Assoc.*, *102*(479), 824–840.
Cunderlik, J. M., and D. H. Burn (2003), Non-stationary pooled flood frequency analysis, *J. Hydrol.*, *276*(1-4), 210–223.
Delgado, J. M., B. Merz, and H. Apel (2012), A climate-flood link for the lower Mekong River, *Hydrol. Earth Syst. Sci.*, *16*(5), 1533–1541.
Devineni, N., U. Lall, N. Pederson, and E. Cook (2013), A tree-ring-based reconstruction of Delaware River basin streamflow using hierarchical Bayesian regression, *Journal of Climate*, *26*(12), 4357–4374.
El Adlouni, S., T. B. M. J. Ouarda, X. Zhang, R. Roy, and B. Bobee (2007), Generalized maximum likelihood estimators for the nonstationary generalized extreme value model, *Water Resour. Res.*, *43*.
Gelman, A., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Stat. Sci.*, *7*(4), 457–472.
Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2014), *Bayesian Data Analysis*, 3rd ed., Chapman and Hall, London, U. K.
Homan, M. D., and A. Gelman (2014), The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo, *J. Mach. Learning Res.*, *15*(1), 1593–1623.
Jensen, F. V. (1996), *An Introduction to Bayesian Networks*, UCL Press, London, U. K.
Johnson, D. S., R. R. Ream, R. G. Towell, M. T. Williams, and J. D. L. Guerrero (2013), Bayesian clustering of animal abundance trends for inference and dimension reduction, *J. Agric. Biol. Environ. Stat.*, *18*(3), 299–313.
Kass, R. E., and A. E. Raftery (1995), Bayes factors, *Jo. Am. Stat.l Assoc.*, *90*(430), 773–795.
Katz, R. W., M. B. Parlange, and P. Naveau (2002), Statistics of extremes in hydrology, *Adv. Water Resour.*, *25*(8-12), 1287–1304.
Khaliq, M. N., T. B. M. J. Ouarda, J. C. Ondo, P. Gachon, and B. Bobee (2006), Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review, *J. Hydrol.*, *329*(3-4), 534–552.
Kysely, J., J. Picek, and R. Beranova (2010), Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold, *Global Planet. Change*, *72*(1-2), 55–68.
Kysely, J., L. Gaal, and J. Picek (2011), Comparison of regional and at-site approaches to modelling probabilities of heavy precipitation, *Int. J. Climatol.*, *31*(10), 1457–1472.
Lang, M., T. B. M. J. Ouarda, and B. Bobee (1999), Towards operational guidelines for over-threshold modeling, *J. Hydrol.*, *225*(3-4), 103–117.
Leclerc, M., and T. B. M. J. Ouarda (2007), Non-stationary regional flood frequency analysis at ungauged sites, *J. Hydrol.*, *343*(3-4), 254–265.
Lima, C. H., U. Lall, T. J. Troy, and N. Devineni (2015), A climate informed model for nonstationary flood risk prediction: Application to Negro River at Manaus, Amazonia, *J. Hydrol.*, *522*, 594–602.
Lima, C. H. R., and U. Lall (2009), Hierarchical Bayesian modeling of multisite daily rainfall occurrence: Rainy season onset, peak, and end, *Water Resour. Res.*, *45*, W07422, doi:10.1029/2008WR007485.
Lima, C. H. R., and U. Lall (2010), Spatial scaling in a changing climate: A hierarchical Bayesian model for non-stationary multi-site annual maximum and monthly streamflow, *J. Hydrol.*, *383*(3-4), 307–318.
MacQueen, J. (1967), Some methods for classification and analysis of multivariate observations, paper presented at the fifth Berkeley symposium on mathematical statistics and probability, *1*, 281–297, Oakland, Calif.
Madsen, H., and D. Rosbjerg (1997), The partial duration series method in regional index-flood modeling, *Water Resour. Res.*, *33*(4), 737–746.
McLachlan, G., and D. Peel (2004), *Finite Mixture Models*, John Wiley, N. Y.
Meshgi, A., and D. Khalili (2009), Comprehensive evaluation of regional flood frequency analysis by L- and LH-moments. II. Development of LH-moments parameters for the generalized Pareto and generalized logistic distributions, *Stochastic Environ. Res. Risk Assess.*, *23*(1), 137–152.
Nasri, B., S. El Adlouni, and T. B. Ouarda (2013), Bayesian Estimation for GEV-B-Spline Model, *Open J. Stat.*, *3*(02), 118.
Nieto-Barajas, L. E., and A. Contreras-Cristán (2014), A Bayesian nonparametric approach for time series clustering, *Bayesian Anal.*, *9*(1), 147–170.
Ouarda, T. B. M. J., and S. El-Adlouni (2011), Bayesian nonstationary frequency analysis of hydrological variables, *J. Am. Water Resour. Assoc.*, *47*(3), 496–505.
Pearl, J. (2014), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Francisco.
Petrow, T., and B. Merz (2009), Trends in flood magnitude, frequency and seasonality in Germany in the period 1951–2002, *J. Hydrol.*, *371*(1), 129–141.
Petrow, T., J. Zimmer, and B. Merz (2009), Changes in the flood hazard in Germany through changing frequency and persistence of circulation patterns, *Nat. Hazards Earth Syst. Sci.*, *9*(4), 1409–1423.
Pujol, N., L. Neppel, and R. Sabatier (2007), Regional tests for trend detection in maximum precipitation series in the French Mediterranean region, *Hydrol. Sci. J.*, *52*(5), 956–973.
Renard, B. (2011), A Bayesian hierarchical approach to regional frequency analysis, *Water Resour. Res.*, *47*, W11513, doi:10.1029/2010WR010089.
Renard, B., M. Lang, and P. Bois (2006), Statistical analysis of extreme events in a non-stationary context via a Bayesian framework: Case study with peak-over-threshold data, *Stochastic Environ. Res. Risk Assess.*, *21*(2), 97–112.
Renard, B., et al. (2008), Regional methods for trend detection: Assessing field significance and regional consistency, *Water Resour. Res.*, *44*, W08419, doi:10.1029/2007WR006268.
Ribatet, M., E. Sauquet, J. M. Gresillon, and T. B. M. J. Ouarda (2007), A regional Bayesian POT model for flood frequency analysis, *Stochastic Environ. Res. Risk Assess.*, *21*(4), 327–339.
Rousseeuw, P. J. (1987), Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *J. Comput. Appl. Math.*, *20*, 53–65.
Rust, H. W., D. Maraun, and T. J. Osborn (2009), Modelling seasonality in extreme precipitation, *Eur. Phys. J.*, *174*, 99–111.
Sankarasubramanian, A., and K. Srinivasan (1999), Investigation and comparison of sampling properties of L-moments and conventional moments, *J. Hydrol.*, *218*(1-2), 13–34.
Spiegelhalter, D. J., N. G. Best, B. R. Carlin, and A. van der Linde (2002), Bayesian measures of model complexity and fit, *J. R. Stat. Soc., Ser. B*, *64*, 583–616.
Srinivas, V. V., S. Tripathi, A. R. Rao, and R. S. Govindaraju (2008), Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering, *J. Hydrol.*, *348*(1–2), 148–166.
Sun, X., M. Thyer, B. Renard, and M. Lang (2014), A general regional frequency analysis framework for quantifying local-scale climate effects: A case study of ENSO effects on Southeast Queensland rainfall, *J. Hydrol.*, *512*, 53–68.
Watanabe, S. (2010), Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory, *J. Mach. Learning Res.*, *11*, 3571–3594.
Xiong, Y., and D.-Y. Yeung (2004), Time series clustering with ARMA mixtures, *Pattern Recognition*, *37*(8), 1675–1689.