

## RESEARCH ARTICLE

10.1002/2013WR014849

# Incorporating spatial dependence in regional frequency analysis

Zhuo Wang<sup>1</sup>, Jun Yan<sup>1,2</sup>, and Xuebin Zhang<sup>3</sup>

<sup>1</sup>Department of Statistics, University of Connecticut, Storrs, Connecticut, USA, <sup>2</sup>Center for Environmental Sciences and Engineering, University of Connecticut, Storrs, Connecticut, USA, <sup>3</sup>Climate Research Division, Environment Canada, Downsview, Ontario, Canada

### Key Points:

- Exploiting spatial dependence improves the efficiency of RFA
- Larger gain in efficiency is possible when the spatial dependence is stronger
- Estimation could be biased if the dependence model is severely misspecified

### Correspondence to:

J. Yan,  
jun.yan@uconn.edu

### Citation:

Wang, Z., J. Yan, and X. Zhang (2014), Incorporating spatial dependence in regional frequency analysis, *Water Resour. Res.*, 50, 9570–9585, doi:10.1002/2013WR014849.

Received 8 OCT 2013

Accepted 25 OCT 2014

Accepted article online 30 OCT 2014

Published online 23 DEC 2014

**Abstract** The efficiency of regional frequency analysis (RFA) is undermined by intersite dependence, which is usually ignored in parameter estimation. We propose a spatial index flood model where marginal generalized extreme value distributions are joined by an extreme-value copula characterized by a max-stable process for the spatial dependence. The parameters are estimated with a pairwise likelihood constructed from bivariate marginal generalized extreme value distributions. The estimators of model parameters and return levels can be more efficient than those from the traditional index flood model when the max-stable process fits the intersite dependence well. Through simulation, we compared the pairwise likelihood method with an L-moment method and an independence likelihood method under various spatial dependence models and dependence levels. The pairwise likelihood method was found to be the most efficient in mean squared error if the dependence model was correctly specified. When the dependence model was misspecified within the max-stable models, the pairwise likelihood method was still competitive relative to the other two methods. When the dependence model was not a max-stable model, the pairwise likelihood method led to serious bias in estimating the shape parameter and return levels, especially when the dependence was strong. In an illustration with annual maximum precipitation data from Switzerland, the pairwise likelihood method yielded remarkable reduction in the standard errors of return level estimates in comparison to the L-moment method.

## 1. Introduction

Natural extremes, such as extreme rainfall or extreme temperature, have profound impact on both the environment and the society. Regional frequency analysis (RFA) is widely used in characterizing the frequency of the extreme events. It is a technique that based on the regionalization concept that trades space for time to obtain adequate estimation of model parameters based on data from a low-density network with short record length [see e.g., Ouarda, 2013, for a recent review]. It uses data from a number of sites that are identified to be in a homogeneous region in certain sense (see section 2 for specific definition) to estimate the quantiles of the variables of interest at each site in the region. That is, short records from different sites within a homogeneous region are pooled to improve the estimation efficiency. Widely used in water resources research, the RFA approach has various models and has been extended to accommodate temporal nonstationarity [e.g., Ouarda et al., 2006; Cunderlik and Ouarda, 2006; Leclerc and Ouarda, 2007] and multivariate analysis [e.g., Ouarda et al., 2000; Javelle et al., 2002; Chebana and Ouarda, 2009]. Nevertheless, our focus is the stationary index flood model with marginal generalized extreme value (GEV) distributions. We improve the efficiency of this type of RFA by incorporating spatial dependence and compare its performance with competing methods to better understand its advantages and limitations.

In an index flood model, the marginal distributions are identical apart from a site-specific scaling factor. Two popular methods are available for parameter estimation, neither of which needs to specify the intersite dependence. The L-moment method [e.g., Hosking and Wallis, 1997] estimates the parameters by solving the equations that match the sample L-moments with the population moments. It first estimates the site-specific scaling factor with at-site data, and then uses it to scale the data at each site. The scaleless data are then pooled to estimate the parameters of the shared scaleless distribution by matching the L-moments. Properties of the L-moments were studied by Hosking [1990]. L-moments are

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

more robust to sampling variability than conventional moments, and their existence only requires existence of the mean. The second method is the independence likelihood method that adds up the marginal loglikelihood from all sites, ignoring the intersite dependence, and then maximizes it (R. L. Smith, Regional estimation from spatially dependent data, University of Chapel Hill, unpublished data, 1990b). The independence likelihood method in RFA gives the most efficient estimator for larger samples, and it can incorporate covariates into model parameters [e.g., *Buishand*, 1991; *Northrop*, 2004], which is necessary in many cases where temporal or spatial nonstationarity is present. Both the L-moment method and the independence likelihood method are robust to intersite dependence at the cost of low efficiency when the dependence is strong.

The impact of intersite dependence on many aspects of RFA is a fundamental issue that has been actively investigated. Intersite correlation does not introduce bias but increases the variance in predicting regional mean or moments [e.g., *Matalas and Langbein*, 1962; *Stedinger*, 1983]. For index flood models, *Hosking and Wallis* [1988] reported similar findings in predicting flood quantiles. Intersite dependence was found in general to increase the variance of the estimator in other contexts such as estimation of regional exceeding probability of a flood level [*Troutman and Karlinger*, 2003] or a regional envelope curve [*Castellarin et al.*, 2005]. Intersite correlation is part of the model in probabilistic regional envelope curves [*Castellarin*, 2007; *Viglione et al.*, 2012]. It has been used in regression analysis with generalized least squares to estimate the parameters of a model of the target quantity as a function of basin characteristics [*Griffis and Stedinger*, 2007]. For testing regional homogeneity with the heterogeneity measures of *Hosking and Wallis* [1993], intersite dependence reduces the power of the tests [*Castellarin et al.*, 2008]. Most of the existing simulation studies generated data from meta-Gaussian models, which essentially use the normal copula for the dependence structure [e.g., *Hosking and Wallis*, 1988]. For extreme observations, however, the Pearson correlation coefficient may not be a good dependence measure [*Embrechts et al.*, 2002] and the normal copula may not be a good dependence model [*Genest and Favre*, 2007; *Gudendorf and Segers*, 2010]. Smith (unpublished data, 1990b) reported a study where the intersite dependence was modeled with an extreme-value copula, but it was an exchangeable Gumbel copula which does not allow the dependence to weaken as the distance between two sites increases.

Spatial extreme modeling has made progress recently in the statistics literature; see *Davison et al.* [2012] for a recent review. Max-stable processes extend the multivariate extreme value distribution to the infinite dimensional setting [*de Haan*, 1984], with marginal distribution of any dimension being multivariate extreme-value. These models provide a natural modeling framework for spatial extremes. A pairwise likelihood approach has been used in parameter estimation due to the unavailability of the joint multivariate density function [e.g., *Padoan et al.*, 2010; *Davison and Gholamrezaee*, 2012]. The pairwise likelihood approach has a robust feature that its validity in inference only needs the correct specification of the bivariate joint density of all the pairs, instead of the full joint density. A spatial index model retains the marginal GEV distributions and uses a max-stable process model for the dependence structure. The pairwise likelihood approach with pairwise bivariate generalized extreme value distributions can potentially increase the efficiency for marginal GEV parameter and return level estimation. Similar efficiency improvement with max-stable process model has recently been reported in detection of nonstationarity in precipitation extremes [*Westra and Sisson*, 2011].

The rest of the article is organized as follows. The index flood model with GEV distributions is reviewed in Section 2, along with two existing estimation methods: L-moment and independence likelihood. Spatial extreme models, their application in RFA with index flood model, and parameter estimation with a pairwise likelihood method are introduced in Section 3. A large scale simulation study that compares the performance of proposed method with the L-moment method and the independence likelihood method is reported in Section 4. All three methods are illustrated in an example of Swiss annual maximum daily precipitation in Section 5. A discussion concludes in Section 6.

## 2. Index Flood Model With GEV Distribution

### 2.1. Model

The index flood model is a widely used RFA model with the homogeneity assumption being that all the sites have an identical distribution up to a site-specific scaling factor known as the index variable. It

originated from applications to flood data in hydrology, but the method can be used with any kind of data [Hosking and Wallis, 1997, p. 6]. Examples of application to precipitation are Kysely and Picek [2007] and Ngongondo et al. [2011]. Let  $Q_s(\cdot)$  be the quantile function of the distribution at site  $s$ ; i.e.,  $Q_s(u) = \inf \{y \in R : u \leq F_s(y)\}$ , where  $F_s$  is the distribution function at site  $s$ . The index flood procedure assumes that for all site  $s$  in a homogeneous region,  $Q_s(u) = c_s q(u)$ , where  $c_s$  is a site specific index variable, and  $q(\cdot)$  is called the regional growth curve, the scale-free quantile function shared by all sites. Within the region, the  $T$ -year return level at any site  $s$  which is the upper  $1/T$ -quantile, is proportional to the return level of the scale-free distribution:  $Q_s(1 - 1/T) = c_s q(1 - 1/T)$ .

The GEV distribution is often used to model the regional growth curve. It can be obtained as the limit distribution of properly normalized maximum of a sequence of independent and identically distributed random variables. The cumulative distribution function of a GEV distribution is

$$F(y; \mu, \sigma, \xi) = \begin{cases} \exp \{-[1 + \xi(y - \mu)/\sigma]^{-1/\xi}\}, & 1 + \xi(y - \mu)/\sigma > 0, \quad \xi \neq 0; \\ \exp \{-\exp[-(y - \mu)/\sigma]\}, & \xi = 0, \end{cases} \quad (1)$$

where  $\mu$ ,  $\sigma$  and  $\xi$  are the location, scale and shape parameters, respectively. Let  $GEV(\mu, \sigma, \xi)$  denote this distribution. The shape parameter  $\xi$  controls the tail behavior of the distribution. The distribution is known as the Gumbel distribution when  $\xi = 0$ . The case with  $\xi > 0$  has heavy tail is of most interest since real data of extreme events often exhibits heavy tail. The quantile function of  $GEV(\mu, \sigma, \xi)$  is the inverse function of  $F$  in (1):

$$Q(u; \mu, \sigma, \xi) = \begin{cases} \mu + (\sigma/\xi)\{(-\ln u)^{-\xi} - 1\}, & \xi \neq 0; \\ \mu + \sigma\{-\ln(-\ln u) - 1\}, & \xi = 0. \end{cases} \quad (2)$$

The  $T$ -year return level is then  $Q(1 - 1/T)$ .

When the GEV distribution is used in a index flood model, the location parameter can be used as the index variable. In particular, let  $c_s = \mu_s$  and let  $Z$  be a  $GEV(1, \gamma, \xi)$  variable. It is straightforward to show that the distribution of  $Y_s = c_s Z$  is  $GEV(\mu_s, \mu_s \gamma, \xi)$ . Therefore, the homogeneity for this index flood model means that the ratio of the scale parameter to the location parameter of the GEV parameters is a constant ( $\gamma$ ), and that the shape parameters at all sites are the same [e.g., Buishand, 1991; Hanel et al., 2009]. Note that the index flood model only specifies the marginal GEV distributions; no spatial dependence is specified.

## 2.2. Existing Estimation Methods

Suppose that we observe annual maxima of a variable of interest at  $m$  sites over  $n$  years. Let  $Y_{s,t}$ ,  $s = 1, \dots, m$  and  $t = 1, \dots, n$ , be the record in year  $t$  from site  $s$ . The data from year to year are assumed to be independent, but within the same year, spatial dependence exists across the sites. For ease of presentation, the notations are for balanced data where all sites have the same length of records, but the methods can be easily adapted to use varying length of records. Let  $\mu_s$ ,  $\sigma_s = \mu_s \gamma$ , and  $\xi_s$  be the location, scale, and shape parameters, respectively, of the GEV distribution at site  $s$ . The parameters to be estimated are  $\beta = (\mu_1, \dots, \mu_m, \gamma, \xi)$ . Two existing estimation methods are the L-moment method [Hosking and Wallis, 1988] and the independence likelihood method [Hanel et al., 2009; Smith, unpublished data, 1990b], neither of which requires the specification of the spatial dependence. As both methods target small samples, asymptotic variance estimator of the parameter estimator is not expected to work well, which is observed in our simulation studies in Section 4. A parametric bootstrap procedure with preserved spatial dependence is used to assess the uncertainty of the estimator.

### 2.2.1. L-Moment

The L-moment method proceeds as follows. First, for each site  $s$ , estimate the GEV parameters  $(\mu_s, \sigma_s, \xi_s)$  using data from this site with the L-moment method, and let  $\hat{\mu}_s$  be the estimate of  $\mu_s$ . Then use  $\hat{\mu}_s$  to scale the data at each site  $s$  by letting  $X_{s,t} = Y_{s,t} / \hat{\mu}_s$ . Apply the L-moment method to the pooled, scaled data  $X_{s,t}$ ,  $s = 1, \dots, m$ ,  $t = 1, \dots, n$ , to fit a GEV distribution with location 1, scale  $\gamma$  and shape  $\xi$ . The only extra difficulty in the last step is that the location parameter of the GEV distribution is restricted to be 1. With two unknown parameters  $(\gamma, \xi)$ , the estimating equations match the first two sample L-moments  $(l_1, l_2)$  with their population counterparts:

$$l_1 = 1 - \gamma \{1 - \Gamma(1 - \xi)\} / \xi,$$

$$l_2 = -\gamma(1 - 2^\xi) \Gamma(1 - \xi) / \xi,$$

where  $\xi < 1$ . The solutions to the equations are the L-moment estimates  $(\hat{\gamma}, \hat{\xi})$ . The implicit restriction  $\xi < 1$  is required for the existence of the L-moments (finite mean), which makes the L-moment method more efficient than the likelihood method for small samples when  $\xi < 1$  is true [e.g., Coles and Dixon, 1999]. When the homogeneity assumption is not valid, bias may be introduced, but RFA may still be more accurate than single site analysis [Lettenmaier et al., 1987; Hosking and Wallis, 1997].

**2.2.2. Independence Likelihood**

Denote  $f_{s,t}(\cdot; \beta)$  as the probability density function of the GEV distribution at site  $s$  in year  $t$ . The independence likelihood method estimates  $\beta$  by maximizing the log-likelihood function pretending that the sites are independent. That is,  $\hat{\beta}$  is the maximizer of

$$L_{i,n}(\beta) = \sum_{t=1}^n \sum_{s=1}^m \log f_{s,t}(Y_{s,t}; \beta). \tag{3}$$

Similar to the L-moment method, this method only assumes correct specification of the marginal GEV distribution at each site. No spatial dependence is taken into account in the point estimation. For large samples, the variance of the estimator has a sandwich form under certain regularity conditions and can be consistently estimated by a sandwich estimator (Smith, unpublished data, 1990b). The sandwich variance adjusts for the unspecified spatial dependence. For small samples, however, a bootstrap procedure that preserves the spatial dependence can be used [Heffernan and Tawn, 2004]. Modification of the likelihood method to improve its small sample performance has been obtained by adding a penalty on the shape parameter [Coles and Dixon, 1999] or, equivalently, imposing a prior distribution on it [Martins and Stedinger, 2000]. We do not consider them here because they introduce the complexity of penalty form selection or prior specification.

**3. Spatial Index Flood Model**

Spatial extreme models have gained much focus in the statistics literature. Can one exploit spatial extreme modeling in RFA to improve efficiency? If so, the spatial dependence can be used in a positive way for better efficiency instead of as nuisance that reduces the efficiency.

**3.1. Max-Stable Process**

By Sklar’s Theorem, the distribution function  $H$  of a  $p$ -dimensional continuous random vector  $(X_1, \dots, X_p)$  with marginal distribution  $F_1, \dots, F_p$ , respectively, can be uniquely represented as

$$H(x_1, \dots, x_p) = C\{F_1(x_1), \dots, F_p(x_p)\}, \quad (x_1, \dots, x_p) \in \mathbb{R}^p.$$

where  $C : [0, 1]^p \rightarrow [0, 1]$ , called a *copula*, is a  $p$ -dimensional distribution function with standard uniform marginals [Sklar, 1959]. When  $H$  is a multivariate extreme value distribution, the corresponding copula  $C$  must be an extreme-value copula, which satisfies a max-stable property [Gudendorf and Segers, 2010]. If all the margins are transformed to unit Fréchet distribution with distribution function  $G(x) = e^{-1/x}$ , the max-stable property means

$$\Pr(Z_1 \leq kz_1, \dots, Z_p \leq kz_p)^k = \Pr(Z_1 \leq z_1, \dots, Z_p \leq z_p), \quad z_i > 0, \quad i=1, \dots, p, \quad k > 0.$$

Max-stability is a defining property for max-stable processes whose marginal copula in any dimension is an extreme-value copula.

In a recent review, Davison et al. [2012] gave a spectral characterization of max-stable processes that unifies the characterizations in de Haan [1984] and Schlather [2002]. Consider a spatial domain  $\mathcal{X} \subset \mathbb{R}^2$ . Let  $\{W(x) : x \in \mathcal{X}\}$  be a nonnegative stationary stochastic process on  $\mathcal{X}$  with  $E\{W(x)\} = 1$  and  $W_1, W_2, \dots$  be independent copies of  $W$ . Let  $\zeta_1, \zeta_2, \dots$  be the points of a Poisson process on  $\mathbb{R}^+$  with intensity  $s^{-2}ds$ . Then,

$$Z(x) = \max_j \zeta_j W_j(x), \tag{4}$$

is a stationary max-stable process on  $\mathcal{X}$  with unit Fréchet marginal distributions. Different forms of  $W(x)$  lead to different parametric max-stable models.

It is often desirable to measure the extremal dependence of  $m$  sites. The extremal coefficient is such a measure. Consider a max-stable process  $Z$  defined in (4) at sites  $\{x_1, \dots, x_m \in \mathcal{X}\}$ . The extremal coefficient of the  $m$  sites is

$$\Pr [Z(x_1) \leq z, \dots, Z(x_m) \leq z] = \exp \left( -\frac{\theta_m}{z} \right). \tag{5}$$

It can be interpreted as the effective sample size of the  $m$  variables. The upper bound of  $\theta_m$  is  $m$ , meaning complete independence, while the lower bound is 1, meaning complete dependence. Specifically, for two sites  $x_1$  and  $x_2$ , a bivariate extremal coefficient function  $\theta(h)$  can be defined as

$$\Pr [Z(x_1) \leq z, Z(x_2) \leq z] = \exp \left( -\frac{\theta(h)}{z} \right), \tag{6}$$

where  $h = x_1 - x_2$  [Schlather and Tawn, 2003]. In a spatial context, the bivariate extremal coefficient of two sites is often modeled to increase from 1 to 2 as the distance  $\|h\|$  between the two sites increases from zero to infinity. When  $\theta(h) = \theta(\|h\|)$  — the dependence measure depends only on distance instead of direction — the model is isotropic.

### 3.2. Parametric Max-Stable Models

We consider three isotropic models that are used in the simulation study. The first model is obtained by taking  $\mathcal{X} = \mathbb{R}^2$  and  $W_j(x) = g(x - X_j)$ , where  $g$  is a bivariate density function and  $X_1, X_2, \dots$  are the points of a homogeneous Poisson process with unit rate on  $\mathcal{X}$ . The special case where  $g$  is the normal density with mean zero and covariance matrix  $\Sigma$  is known as the Smith model (R. L. Smith, Max-stable processes and spatial extremes, University of Surrey, unpublished data, 1990a). The bivariate marginal distribution function at two sites  $x_i$  and  $x_j$  is

$$\Pr [Z(x_i) \leq z_i, Z(x_j) \leq z_j] = \exp \left\{ -\frac{1}{z_i} \Phi \left( \frac{a}{2} + \frac{1}{a} \log \frac{z_i}{z_j} \right) - \frac{1}{z_j} \Phi \left( \frac{a}{2} + \frac{1}{a} \log \frac{z_j}{z_i} \right) \right\}, \tag{7}$$

where  $\Phi$  is the cumulative distribution function of the standard normal variable,  $a^2 = \Delta^\top \Sigma^{-1} \Delta$ , and  $\Delta = x_i - x_j$ . The bivariate density function can be obtained by differentiating the distribution function [e.g., Padoan et al., 2010]. The bivariate extremal coefficient function is  $\theta(h) = 2\Phi(\sqrt{h^\top \Sigma^{-1} h}/2)$ , with range from 1 to 2, providing full range of dependence level. A limitation of the Smith model is that the storms generated from it have shapes that are too regular compared to the reality.

The second model we consider is the Schlather model [Schlather, 2002]. It is obtained by taking  $W(x) = \max \{0, 2\sqrt{\pi}\epsilon(x)\}$ , where  $\epsilon(x)$  is a stationary Gaussian process with unit variance and correlation function  $\rho$ , a function of the euclidean distance between two sites  $x_i$  and  $x_j$ . The bivariate marginal distribution function is

$$\Pr [Z(x_i) \leq z_i, Z(x_j) \leq z_j] = \exp \left\{ -\frac{1}{2} \left( \frac{1}{z_i} + \frac{1}{z_j} \right) \left( 1 + \sqrt{1 - 2[\rho(h) + 1] \frac{z_i z_j}{(z_i + z_j)^2}} \right) \right\}, \tag{8}$$

where  $h = \|x_i - x_j\|$ . The bivariate extremal coefficient is  $\theta(h) = 1 + \sqrt{(1 - \rho(h))/2}$ , with a range from 1 to  $1 + \sqrt{1/2}$ , or about 1.707. Therefore, the Schlather model does not provide full range. It cannot be used to model sites that are completely independent. Models of correlation function are standard in spatial statistics [e.g., Banerjee et al., 2004, Table 2.1], and contain parameters that characterizing the strength of the spatial dependence.

The third model we consider is a geometric Gaussian process obtained by taking  $W(x) = \exp \{ \delta \epsilon(x) - \delta^2/2 \}$ , where  $\epsilon(x)$  is again a stationary Gaussian process with unit variance and correlation function  $\rho$ , and  $\delta^2 > 0$  is the variance of  $W(x)$  on the log scale [Davison et al., 2012, p. 172]. The bivariate marginal distribution is the same as (7) for the Smith model, except that  $a^2 = 2\delta^2(1 - \rho(h))$ . The bivariate extremal coefficient

function is  $\theta(h) = 2\Phi(\delta\sqrt{(1-\rho(h))/2})$ . The range of  $\theta(h)$  is from 1 to  $2\Phi(\delta/\sqrt{2})$ . The upper bound is 1.96, quite close to 2, if  $\delta^2 = 8$ . As  $\delta \rightarrow \infty$ ,  $\theta(h)$  approaches 2 for any  $\rho$ .

### 3.3. Application to RFA

#### 3.3.1. Spatial Index Flood Model

To incorporate the spatial dependence in the index flood model, we assume that the dependence structure among the sites is an extreme-value copula described by a max-stable process model. This assumption is in addition to the homogeneity assumption for the index flood model with GEV margins in Section 2. We keep using  $Y_{s,t}$ ,  $s = 1, \dots, m$ ,  $t = 1, \dots, n$ , as the observed data at site  $s$  in year  $t$ . The spatial index flood model completely specifies the joint distribution of  $\{Y_{s,t} : s = 1, \dots, m\}$  for each  $t$ . As seen from Sklar's theorem, it is sufficient to specify the marginal models and the spatial dependence structure. Under the setup of index flood model, the marginal distribution at each site  $s$  is still GEV  $(\mu_s, \mu_s\gamma, \xi)$ . The extreme-value copula is specified by a parametric max-stable process with dependence parameter  $\alpha$ . Although the model is fully specified and easy to understand, the joint density is unavailable except for lower dimensions ( $m = 2, 3$ ) for certain parametric models. The bivariate marginal density of two sites depends on  $\alpha$  in addition to the marginal GEV parameters. Let  $F_s(\cdot; \beta)$  be the cumulative distribution function of the GEV distribution at site  $s$ , let  $G$  be the cumulative distribution function of the unit Fréchet distribution, and let  $G^{-1}$  be the inverse function of  $G$ . The bivariate density of site  $i$  and  $j$ ,  $f_{i,j}$ , is

$$f_{i,j}(y_i, y_j; \beta, \alpha) = g_{i,j}(z_i, z_j; \alpha) |J(y_i, y_j; \beta)|,$$

where  $g_{i,j}(z_i, z_j; \alpha)$  is the bivariate marginal density of the max-stable process model with unit Fréchet margins,  $z_i = G^{-1}\{F_i(y_i; \beta)\}$ , and

$$|J(y_i, y_j; \beta)| = \left| \frac{d}{dy_i} G^{-1}\{F_i(y_i; \beta)\} \frac{d}{dy_j} G^{-1}\{F_j(y_j; \beta)\} \right|.$$

#### 3.3.2. Pairwise Likelihood Estimation

Inferences about max-stable process models have been mostly based on the composite likelihood approach [Padoan et al., 2010; Davison and Gholamrezaee, 2012]. The composite likelihood approach constructs an objective function, known as the composite likelihood, by putting together pieces of tractable likelihood, such as lower dimensional marginal densities [Lindsay, 1988]. The composite likelihood is maximized to give the maximum composite likelihood estimator (MCLE) as if it were a likelihood. Under mild conditions, correct specification of the pieces in the composite likelihood leads to consistency and asymptotic normality of the MCLE. It has wide applications where the full joint distribution is unavailable or intractable but lower-order marginal or conditional distributions are known [e.g., Varin, 2008; Varin et al., 2011]. When the pieces in the composite likelihood are pairwise bivariate densities, the composite likelihood is also called pairwise likelihood. The independence likelihood in Smith (unpublished data, 1990b) is also a composite likelihood constructed from the univariate marginal GEV distributions.

The dependence parameter  $\alpha$  and marginal parameter  $\beta$  are estimated jointly in the pairwise likelihood method. The pairwise likelihood is constructed with the bivariate density of all the site-pairs within the same years:

$$L_{P,n}(\beta, \alpha) = \sum_{t=1}^n \sum_{i=1}^{m-1} \sum_{j=i+1}^m \log f_{i,j}(Y_{i,t}, Y_{j,t}; \beta, \alpha). \tag{9}$$

When the record lengths are different across sites, it can be constructed from all the available pairs within each year. Let  $(\hat{\beta}_n, \hat{\alpha}_n)$  be the maximizer of the pairwise log-likelihood (9). Under certain regularity conditions,  $(\hat{\beta}_n, \hat{\alpha}_n)$  is consistent to the true parameter vector and is asymptotically normally distributed [e.g., Padoan et al., 2010]. The variance of  $(\hat{\beta}_n, \hat{\alpha}_n)$  can be estimated by a sandwich estimator, which can only give valid inference when the sample size  $n$  is large. For small to moderate sample sizes, as is often the case with RFA, a bootstrap variance estimator is preferred. Heffernan and Tawn [2004] proposed a bootstrap procedure that preserves the dependence structure for multivariate extremes. This procedure has been applied in a nonstationary index flood model [Hanel et al., 2009], and is used here.

## 4. Simulation Study

A simulation study was conducted to compare the performance of the three estimation methods for index flood model: L-moment, independence likelihood, and pairwise likelihood. Unlike the other two methods,

which do not need to specify spatial dependence, the pairwise likelihood method incorporates spatial dependence through the extra specification of an extreme-value dependence model. It has the potential of being more efficient when the dependence model is correctly specified, but risks severe bias otherwise. The L-moment method has been found to be unbiased regardless of the spatial dependence [Hosking and Wallis, 1988]. Nevertheless, existing studies all used normal copulas, which provide no extremal dependence, in generating data. The performance of the L-moment method for data with extremal dependence as generated from max-stable processes has not previously been assessed. Our simulation design reflects these needs.

#### 4.1. Design

We considered data from  $m$  sites over  $n$  years in a study region  $\mathcal{X}=[0, 10]^2$ . Data from different years were independent, but within the same year, data from different sites were generated with spatial dependence. The center point (5, 5) is included so that parameter estimates and return level estimates are compared at this point across scenarios. The additional  $m-1$  sites were randomly generated in the region. The marginal distribution at site  $s$  is GEV  $(\mu_s, \mu_s\gamma, \xi)$  with  $\gamma=0.3$ ,  $\xi=0.2$ ,  $\mu_s=37$  for  $s=(5, 5)$ , and  $\mu_s$  for other sites randomly generated from a normal distribution  $N(43.0, 4.4^2)$  and rounded to an integer. The parameters of this normal distribution were the sample mean and sample variance of the L-moment estimates of the  $\mu_s$ 's from an extreme rainfall data in Southern Ontario analyzed in Wang *et al.* [2014]. The values of  $\mu_s$  ranged from 33 to 51.

Four factors were considered in the experimental design: the spatial dependence model, the spatial dependence level, the number of sites  $m$ , and the length of the record  $n$ . Four spatial dependence models were used to generate data, including three extreme-value models and one nonextreme-value model. The three parametric isotropic extreme-value models were the Smith model, the Schlather model, and the geometric Gaussian model, abbreviated as SM, SC, and GG, respectively. The nonextreme-value model is a Gaussian copula, which is also known as meta-Gaussian model, abbreviated as GA. For each model, three levels of dependence were used: weak, moderate, and strong, abbreviated as W, M, and S, respectively. The SM model had  $\Sigma=\tau I_2$ , where  $I_2$  is the identity matrix of dimension 2, with  $\tau$  chosen to be 4, 16, and 64 for the W, M, and S dependence, respectively. Observations at two sites with distance over  $2\tau^{1/2}$  would be close to independent. In our study region, these choices correspond to the cases where two sites are almost independent if their distance exceeds 4, 8, and 16, respectively. The SC model had a Gaussian correlation function  $\rho(h)=\exp[-(\|h\|/\phi)^2]$  with range parameter  $\phi > 0$  chosen such that the resulting bivariate extremal coefficient function matches that from the SM model as close as possible. It is a special case of the power exponential correlation family with smooth parameter fixed at 2 as in the R package SpatialExtremes [Ribatet and Singleton, 2013]. Through nonlinear least squares, the values of  $\phi$  were tuned to be 2.942, 5.910 and 13.153 for W, M, and S dependence, respectively. For the GG model, a bigger  $\delta^2$  offers fuller range of dependence level for two sites, but the data generating function for this model in the R package SpatialExtremes works well only for  $\delta^2 < 10$ . As a compromise,  $\delta^2$  was fixed at 8. The GG model also had a Gaussian correlation structure, and similarly through nonlinear least squares, the range parameter was set to be 7.134, 14.780, and 31.149 for W, M, and S dependence, respectively. For the GA model, an exponential correlation function  $\rho(h)=\exp[-(\|h\|/\phi)]$  was used with range parameter  $\phi$  to be 6, 12, and 20, which were chosen so that the fitted exponential correlation curves of the empirical correlation of the score functions of  $\mu$  and  $\xi$  are close to those of SM model. Two levels of  $m$  were considered,  $m \in \{10, 20\}$ . When  $m = 20$ , 10 additional sites were generated and added to those sites used in the case of  $m = 10$ . Finally, two levels of  $n$ ,  $n \in \{10, 25\}$ , were considered. This design led to 48 scenarios.

For each scenario, we generated 1000 data sets. For each data set, we estimated the GEV parameters and  $T$ -year return level  $Q_T$  for  $T \in \{50, 100, 500\}$ . The three methods, L-moment, independence likelihood, and pairwise likelihood, are abbreviated as LM, IL, and PL, respectively. In optimization, the IL estimator used the LM estimators as starting values, and the PL estimator used the IL estimators as starting values. Given the large number of parameters in the model, we maximized the likelihood with an iterative procedure that maximizes the objective function with respect to one parameter at a time while the other parameters are held constant. The procedure is iterated over all parameters until convergence. It was reported to give better estimation when the number of parameters is large [Blanchet and Davison, 2011]. In contrast to the LM and IL method, the PL method needed to specify a dependence model, which may be correct or incorrect.

We studied its performance under correct specification and misspecification of the dependence structure within the spatial extreme model and with nonextreme-value copula (GA model).

## 4.2. Results

The center site [5,5] which was presented in all replicates was used to do the comparison across various scenarios. As the results from the three spatial extreme models were similar, we use the GG model to represent the three extreme-value models. Since sample size  $n = 10$  was too small for the likelihood methods to be numerically reliable, the results for  $n = 10$  were based on trimmed data where 2% from each tail were excluded in the summary. The results for  $n = 25$  were stable and included all 1000 replicates. The PL method had correct specification of the dependence model for data generated from the extreme-value models (SM, SC, and GG). For data generated from the GA model, the PL method was obtained under the specification of a GG model with Gaussian correlation structure. For each method, we report the relative bias and the relative root mean squared error (RMSE) for the GEV parameters and return levels at the center point.

### 4.2.1. When PL Is Correctly Specified

We first look at the results for data generated from extreme-value dependence models (Figure 1 for the GG model). The bias decreases for all methods as  $n$  goes from 10 to 25. At  $n = 25$ , the bias of the PL method is quite small, with the largest relative magnitude of 7.8%. The bias of the LM method, however, remains high, and it is bigger especially with stronger dependence level. More sites did not help, especially for strong dependence. The relative bias is 17.5% for  $\xi$  and 20.6% for  $Q_{500}$  under strong dependence level and  $m = 20$ . This behavior of the LM method is in contrast to the existing result that intersite dependence does not introduce bias in RFA [Hosking and Wallis, 1988]. It may be explained by that the simulation here was done with data generated from max-stable processes, which ensures that all marginal copulas are extreme-value copulas. In existing studies, however, data were generated mostly with normal copulas, which is not an extreme-value copula.

The RMSE for all methods decreases as  $n$  increases as expected. Between  $m = 10$  and  $m = 20$ , little difference was observed in RMSE. We have also tried  $m = 5$  (not reported here) and found that the relative RMSEs did decrease when  $m$  increased from 5 to 10. This indicates that increasing the number of sites helps increase the efficiency for smaller  $m$ , but only up to a certain point, an observation consistent with the findings in Hosking and Wallis [1988]. As the dependence gets stronger, the RMSE increases for all methods, but the magnitude of the change is the smallest for the PL method. This is because under correct specification, the PL method incorporates spatial dependence in the estimation while the other two methods do not. For both sample sizes, the PL method is a clear winner among the three. The comparison between the LM method and the IL method is mixed. For  $n = 25$ , the LM method is less efficient in parameter estimates but more efficient in some return level estimates than the IL method. This is possible because the return levels are nonlinear transformations of the parameters; see equation (2). For  $n = 10$ , the LM performs better than the IL method in most of the return level estimates, even though it is less efficient in estimating  $\gamma$  and comparable in estimating  $\mu$  and  $\xi$ . Further investigation revealed that the variance of the LM estimator is much smaller than that of the IL estimator, which compensates the larger bias (especially in  $\xi$ ) of the LM estimator. This makes sense since it is known that L-moment estimator has a restriction  $\xi < 1$ , which makes it more efficient than the likelihood method for small samples when the restriction is true. In an earlier version under a slightly different simulation design, we had  $n \in \{50, 100\}$  (not reported here) and found that for large sample sizes, the efficiency order was PL, IL, and LM from the highest to the lowest. Among the three GEV parameters, the efficiency gain of the PL method relative to the IL method was always the greatest for the shape parameter  $\xi$  (in one case it was as large as 2.76); the RE for the  $\mu$  and  $\gamma$  were close to 1. That is, the efficiency gain of the PL method is mostly realized in  $\xi$ , which controls the tail behavior of the GEV distribution. This leads to the efficiency gain of the PL method in estimating the return levels, especially for longer return periods such as 500 years.

### 4.2.2. Misspecification of PL Within Spatial Extreme Models

The efficiency gain in the PL method relative to the IL method comes with a cost: one needs to specify the spatial dependence model. We first look at the results for cases where misspecification is within the class of spatial extreme models; that is, one max-stable model is misspecified as another max-stable model. Figure 2 summarizes the relative efficiency (RE) in mean squared error of the PL method in estimating the GEV



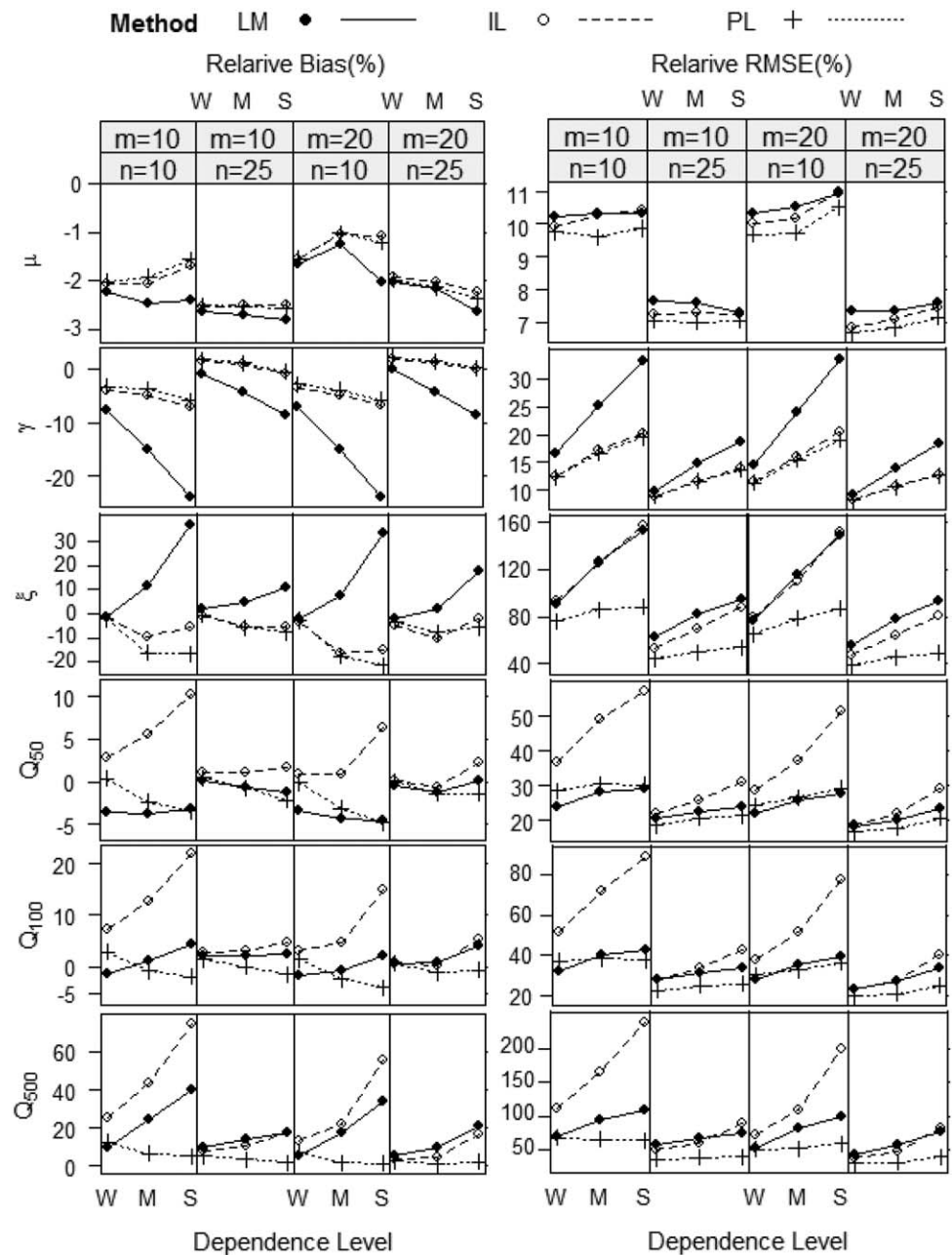
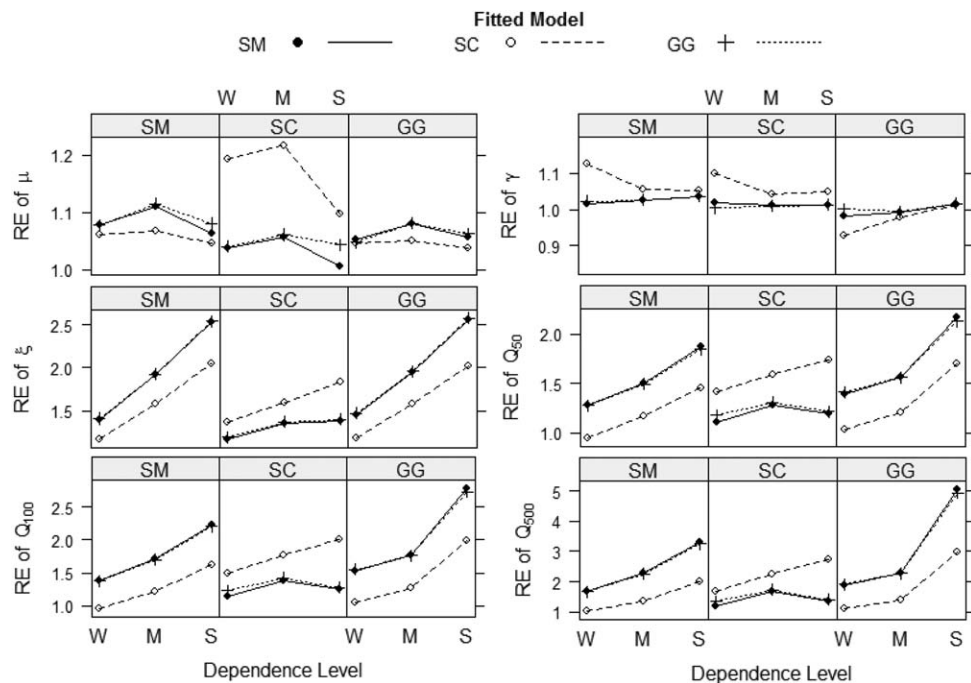


Figure 1. Relative bias (%) and relative RMSE (%) for three methods with data from the GG model.

parameters and return levels, using the IL method as the reference (RE is the ratio of MSE of IL over MSE of PL), under both correct specification and misspecification within spatial extreme models for  $n = 25$  and  $m = 10$ . When a SM model was misspecified as a GG model, or vice versa, the resulting estimator was almost as efficient as that under correct specification, and was much more efficient than the IL estimator. This is because the SM model and the GG model are very similar models, as evident from their similar bivariate distributions given in Section 3.2. When a SM or GG model was misspecified as a SC model, the resulting PL estimator is comparable with the IL estimator for weak dependence, but more efficient than the IL estimator for moderate or strong dependence. This is as expected, because in contrast to the other two models, the SC model does not provide full range of dependence and, therefore, cannot accommodate weak dependence or close to independence. When a SC model was misspecified as a SM or GG model, the resulting PL estimator remained competitive compared to the IL estimator, especially for the cases with stronger



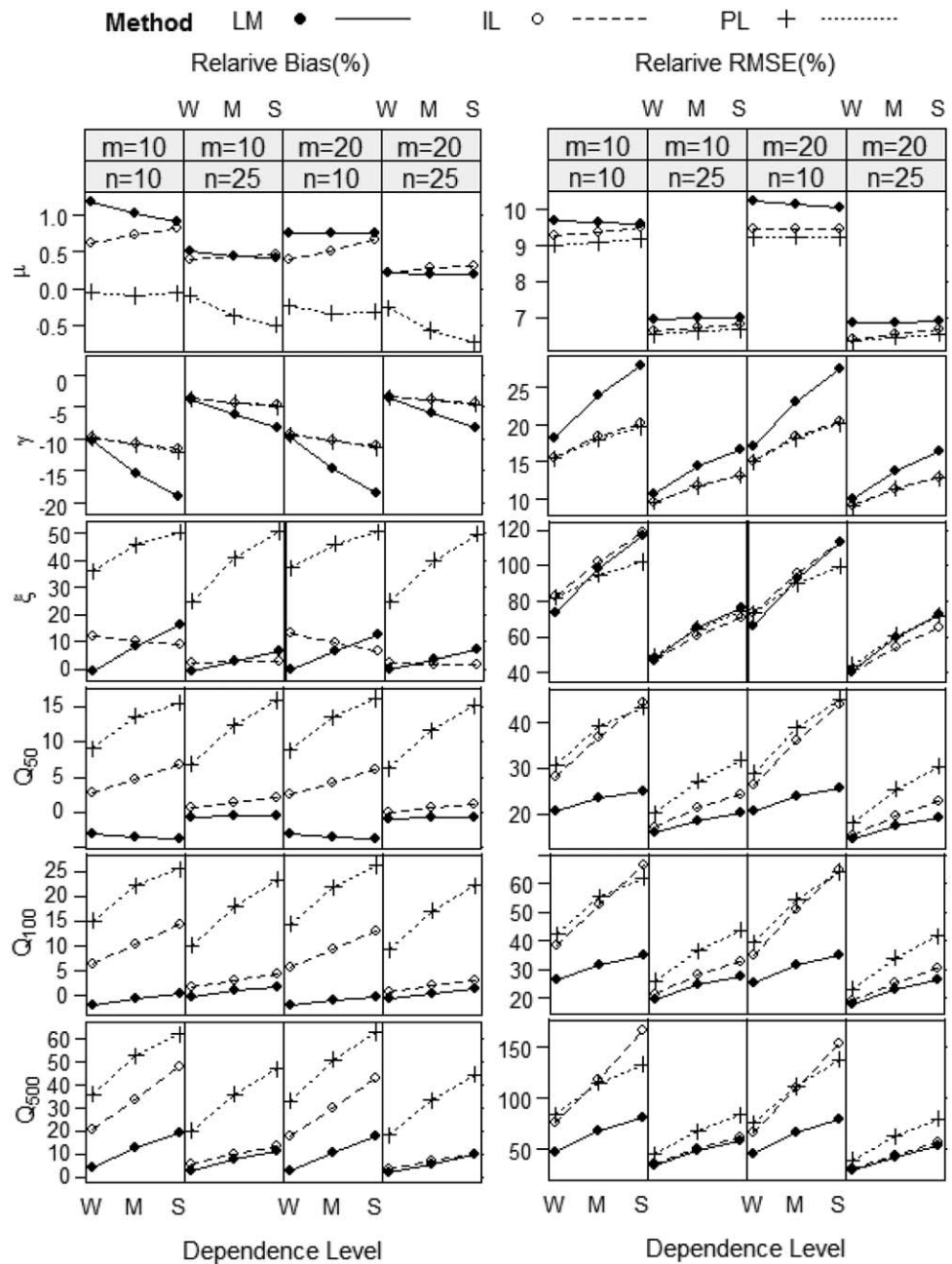
**Figure 2.** Relative efficiency (RE) of PL method (with the IL method as reference) under correct specification and misspecification within the class of extreme-value dependence models with  $n = 25$  and  $m = 10$ . The grouped variable is the model that generated the data, and the line in each panel represents the corresponding fitted model.

spatial dependence. Among the three parametric models, both the SM model and GG model offer full range of dependence level and high efficiency under misspecification, but since the SM model gives too regular shapes of extreme observations to be observed in practice [e.g., Schlather, 2002], we recommend using the GG model. This is also why we presented the results for data generated from the GG model only in the main text.

**4.2.3. Misspecification of PL Under Nonextreme-Value Model**

What if the true dependence model is a nonextreme-value copula but we fit an extreme-value dependence model? Figure 3 summarizes the results for data generated from the GA model, with the PL method specified under a GG model. The misspecified PL estimator has small bias in  $\mu$  and  $\gamma$ , but large bias in  $\xi$  (as high as 50% for  $m = 20$  and  $n = 25$ ), which is much larger than that of the LM estimator or IL estimator. The bias increases as the dependence level gets stronger, and having more sites do not help. This large bias played a major part in the RMSE of the PL estimator; the RE for  $\xi$  and all return levels are as small as 0.51. The LM estimator is better than the IL estimator for all return levels. The performance of the LM method is similar to what was reported in Hosking and Wallis [1988]—its relative bias is only alarmingly noticeable (9.8% for  $Q_{500}$ ) under the strong dependence level. This is reasonable because this scenario is the closest to the data generation scheme of Hosking and Wallis [1988], where a relative bias up to 5% for  $Q_{1000}$  was reported. Therefore, the spatial dependence does not affect the LM estimator under normal copula as much as it does under extreme-value copulas. For instance, the bias of the LM estimator in estimating  $Q_{500}$  with data from the GG model is about twice as much as that with data from the GA model: 33.6% versus 17.5% when  $n = 10$  and 20.6% versus 9.8% when  $n = 25$ .

In summary, incorporating spatial dependence in the index flood model through max-stable processes may improve the efficiency of RFA, but at the cost of having to specify the dependence model. The LM method and the IL method do need to do so, which makes them attractive when no evidence supports extreme-value dependence. Misspecification of the PL method can lead undesired large bias. If, however, extreme-value copulas are known to correctly specify the dependence structure or provide adequate fit to the data through a goodness-of-fit test, then the PL method may be preferred by exploiting the dependence structure to give more efficient RFA. Essentially, it is still a story of bias-variance trade-off. To reap the potential efficiency gain in practice, one must check the goodness-of-fit of max-stable processes, which has been

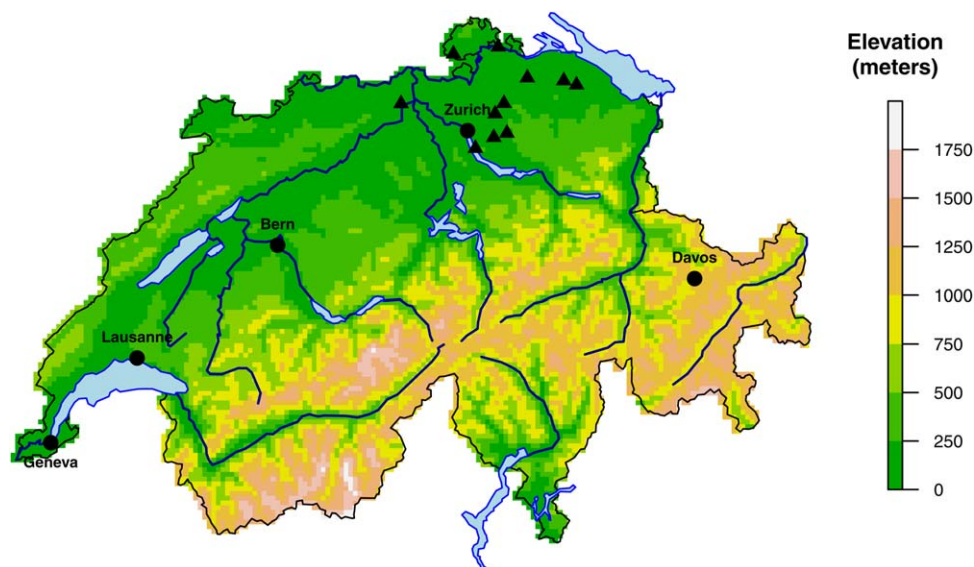


**Figure 3.** Relative Bias (%) and relative RMSE (%) for three methods with data from the GA model. The PL method using a GG model specification.

studied recently [Kojadinovic et al., 2014], in addition to the goodness-of-fit tests for the marginal GEV models and the homogeneity assumption on the GEV parameters.

### 5. Illustration

For illustration, we applied the index flood model with all three estimation methods to the Swiss rainfall data that has been analyzed by many authors in modeling spatial extremes [e.g., Davison et al., 2012]. The data consist of summer maximum daily precipitation (mm) for 51 stations over the years of 1962–2008 in the Plateau region of Switzerland; it is available in the R package SpatialExtremes [Ribatet and Singleton, 2013]. To make a more realistic RFA, we used only the last 25 years of data from 1984 to 2008 ( $n = 25$ ). We



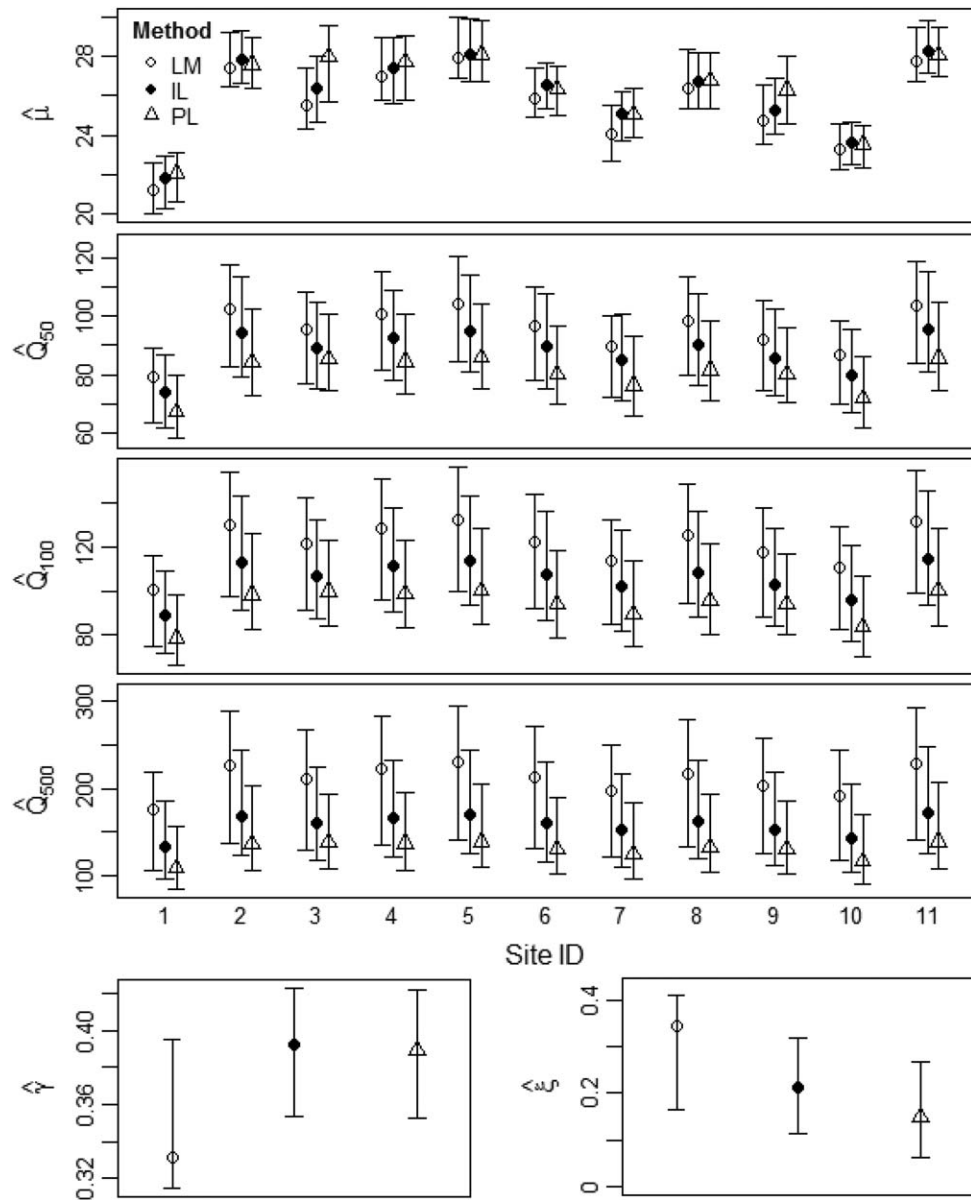
**Figure 4.** Elevation map of Switzerland with the 11 stations that were used in the Swiss rainfall analysis. The 11 stations are marked by triangles, and the dots represent cities in Switzerland.

further filtered the sites in attempt to enhance the chance that the PL method gives reliable inferences. As we pointed in the discussion of the simulation study, the higher efficiency of the PL method is only achievable when the bivariate marginal distributions are correctly specified; otherwise, the PL method could lead to serious bias. The flood-index model we considered assumes that all the marginal distributions are GEV distributions with the same shape parameter and the same ratio of location parameter and scale parameter. These assumptions are shared by all three methods, but the PL method assumes additionally that the dependence structure can be captured by a max-stable (e.g., a geometric Gaussian) process. The goodness-of-fit of the geometric Gaussian process on this data has been checked graphically [Davison *et al.*, 2012]. A formal goodness-of-fit test for max-stable process models was not rejected for this data [Kojadinovic *et al.*, 2014]. The test, however, was applied on the whole region globally, and may have low power in detecting local lack-of-fit. Therefore, we further applied the test for bivariate extreme-value dependence [Kojadinovic *et al.*, 2011] on all the pairs.

We filtered the sites by three tests on the model assumptions: (1) the goodness-of-fit of univariate GEV distribution was not rejected at any single site by a Kolmogorov–Smirnov test; (2) the homogeneity hypothesis ( $\sigma_s/\mu_s$  and  $\xi_s$  are both constant) was not rejected by a nonparametric bootstrap test procedure which preserves the spatial dependence [Heffernan and Tawn, 2004]; and (3) the hypothesis of bivariate extreme-value dependence was not rejected for any pair of the sites by a nonparametric test proposed by Kojadinovic *et al.* [2011]. Note that the first two are needed by all three methods, but the other one is only needed by the PL method. The 3rd test was an additional measure on model specification check given that the geometric Gaussian process has been known to fit this data well [Davison *et al.*, 2012; Kojadinovic *et al.*, 2014]. For a different data set, it will be necessary to run model diagnosis and global goodness-of-fit test too. This process ended up with 11 sites; see map in Figure 4.

We fitted the index flood model with marginal GEV distribution to the 25 year data of the 11 sites with all three methods. The PL method was carried out under the same GG model that was used in the simulation study; that is, it had a Gaussian correlation function with a single range parameter and  $\delta^2=8$  was fixed. The standard errors of all the parameter estimates were obtained with a spatial-dependence-preserving bootstrap procedure [Heffernan and Tawn, 2004] with 1000 bootstrap samples.

Figure 5 summarizes the point estimate and 95% confidence interval (CI) for each parameter and three site-specific return levels (50, 100, and 500 years) from the three methods. The bounds of the 95% CI were the 2.5% and 97.5% percentiles of the 1000 bootstrap estimates, respectively. The point estimates from the three methods are similar for  $\mu_s$ 's, but quite different for  $\gamma$  and  $\xi$ . For  $\gamma$ , the estimates are 0.331 (s.e. 0.021), 0.392 (s.e. 0.018), and 0.389 (s.e. 0.018) for LM, IL, and PL, respectively. For  $\xi$ , the estimates are 0.345 (s.e.



**Figure 5.** Estimated parameters and return levels (in mm) along with their 95% confidence intervals from the bootstrap procedure for the Swiss rainfall data. The PL method used a geometric Gaussian model with a Gaussian correlation function for the spatial dependence.

0.063), 0.212 (s.e. 0.053), and 0.148 (s.e. 0.054) for LM, IL, and PL, respectively. The shape parameter is estimated to be significantly different from zero regardless of the method, suggesting the tails of the GEV distributions are heavier than the tail of the Gumbel distribution. The differences have a drastic effect on return level estimates. Consider for example the 11th site. The estimates of  $Q_{500}$  for this site are 228.39 (s.e. 40.17), 171.28 (s.e. 30.44), and 139.27 (s.e. 24.52), respectively, from the LM method, the IL method, and the PL method. The reduction in the standard error of the PL method is remarkable. The standard error of  $Q_{500}$  from the PL method is about 40% smaller than that from the LM method, and 20% smaller than that from the IL method. The bounds of the 95% bootstrap CIs are asymmetric around the point estimates, which are most notable for  $\gamma$  and  $\xi$ . The asymmetry appears to be in opposite direction for the LM method and the PL method, making the overlaps of the CIs to be bigger than those symmetric CIs constructed from the bootstrap standard errors. The CIs for the return levels have the shortest length from the PL method, followed by the IL method and then the LM method. We emphasize that the short CIs from the PL method does come at a cost — we had to specify the dependence model with a geometric Gaussian process with a

**Table 1.** Point Estimate and Bootstrap Standard Error for the Full 47 Years Data Analysis (Abbreviated as Full), the Average Point Estimate and Average Bootstrap Standard Error Based on 100 Subsets of 25 Years (Abbreviated as Ave), and the Standard Deviation of the 100 Point Estimates

	Point Estimate						Standard Error						Standard Error of 100 Point Estimates		
	LM		IL		PL		LM		IL		PL		LM	IL	PL
	Full	Ave	Full	Ave	Full	Ave	Full	Ave	Full	Ave	Full	Ave			
$\mu$	20.5	20.7	20.4	20.5	20.5	20.6	0.43	0.55	0.42	0.55	0.41	0.55	1.31	1.28	1.26
	26.5	26.6	26.5	26.6	26.4	26.5	0.49	0.65	0.46	0.61	0.45	0.60	1.70	1.61	1.53
$Q_{50}$	68.9	67.8	66.0	65.4	61.6	61.3	4.4	6.1	4.0	5.7	3.6	5.3	12.4	12.2	9.2
	88.9	87.3	85.7	84.8	79.2	79.0	5.8	8.1	5.4	7.7	4.9	7.1	16.8	15.4	11.6
$Q_{100}$	84.4	83.7	79.3	78.7	72.5	72.2	6.9	10.1	6.0	8.8	5.4	8.0	19.7	17.6	12.6
	108.9	107.8	102.8	102.1	93.2	93.0	9.2	13.2	8.1	11.8	7.2	10.6	26.3	22.5	16.1
$Q_{500}$	133.9	137.6	119.0	119.8	103.3	103.8	17.8	29.2	14.1	21.6	12.1	18.7	51.5	38.4	24.8
	172.8	177.5	154.4	155.3	132.9	133.7	23.5	38.3	18.8	28.6	16.0	24.7	68.0	49.3	31.9
$\gamma$	0.338	0.327	0.358	0.353	0.355	0.351	0.014	0.021	0.012	0.017	0.012	0.017	0.030	0.026	0.026
$\xi$	0.274	0.256	0.223	0.205	0.178	0.168	0.045	0.062	0.037	0.053	0.038	0.053	0.133	0.097	0.071

Gaussian correlation structure. The estimate of the range parameter in the dependence model is  $\hat{\phi} = 2377.44$  (s.e. 92.31). We have gone through the extra steps in filter the sites to be included in the analysis beyond the known model checking in the literature on this data, which turned out to be worth it.

The systematically lower return level estimates from the PL method than those from the LM method may be explained from two perspectives. First, the return level is a function of all three parameters of the GEV distribution but is most sensitive to the shape parameter. The location parameter estimates are similar across different methods. The scale parameter estimates and the shape parameter estimates, however, tend to compensate each other: higher scale parameter estimate is accompanied with lower shape parameter estimate. The LM method has lower  $\hat{\gamma}$ , hence lower scale estimate, and higher  $\hat{\xi}$ , which led to the higher return level estimates. Second, the shape parameter seems to have the most room for efficiency improvement as seen in the simulation (Figure 2 in the manuscript). The efficiency gain of the PL method, assuming that the PL is correctly specified, is only on average if replicates were available. For a single data set, the truth is unknown, and the lower point estimate of the shape parameter from the PL method than that from the IL method is quite likely, noting that the 95% bootstrap confidence intervals from the two methods overlap by large.

The analysis so far is based on the last 25 years data and the availability of the whole 47 years data enabled us to compare the performance of the three methods more thoroughly in other ways. We randomly selected 100 subsets of 25 year data, and for each subset we ran the same analysis as we did for the last 25 years. The same analysis was also repeated on the full 47 years of data. Table 1 summarizes results from these analyses. Site specific estimates are only presented for two sites with the smallest or the largest  $\hat{\mu}_s$  from the full data analysis. Point estimates and bootstrap standard errors are reported for the full data analysis. For the 100 subset analyses, we reported the average of the point estimates, the average of the bootstrap standard errors, and the standard deviation of the point estimates. It is reassuring that the point estimates from the full data analysis are very close to the average of those from the 100 subset analyses for all parameters and all three methods. For the estimates of the shape parameter and the return levels, the PL method always has the smallest standard error while the LM method always has the largest standard error, regardless of the full data analysis or the average of the subset analyses. This is consistent to the results from the last 25 years of data. The standard deviation of the 100 subset point estimates of the shape parameter and the return levels has the same pattern, but the difference in the magnitudes is even more obvious. For instance, the standard deviation of the 100 subset estimates  $Q_{500}$  is 68.0, 49.3, and 31.9 for LM, IL, and PL, respectively.

Another interesting finding is about the ratio of the standard errors from the full data analysis and those from the subset analyses. More data are associated with smaller standard errors in theory. For most parameter estimates and all methods, the ratios of the standard error from the full 47 years data to that from the average of the 25 years subsets are close to  $\sqrt{25/47} \approx 0.73$ . This is expected to happen for large sample, which suggests that the large sample results approximate the finite sample results quite well even for sample size of 25 in this data analysis. We also studied the ratios of the 95% bootstrap CI lengths, most of which were higher than 0.73 due to the asymmetry in the CIs.

## 6. Discussion

This paper explores the idea of incorporating intersite dependence in RFA with index flood models to improve the efficiency in estimation. The efficiency gain comes at the cost of having to specify the dependence model in addition to the usual specifications such as marginal distributions and the regional homogeneity assumption. When the dependence model is correctly specified, smaller standard errors and narrower confidence intervals can be obtained for model parameter and return levels. Misspecification of the dependence model, however, may result in serious bias, especially when the true dependence model is not of extreme-value type and the dependence is strong. This makes it important to check the goodness-of-fit for the dependence structure, in addition to the usual check for marginal goodness-of-fit and regional homogeneity, to reap the efficiency gain. The L-moment method and the independence method may sometimes be preferable because they have no need for dependence model specification. The L-moment method implicitly constrains the shape parameter to be less than 1 for the existence of the L-moments, which gives efficient estimator for small samples when the constraint does hold. As extreme-value copula can be very different from nonextreme-value copula (e.g., normal copula), it has more effect on the bias of the L-moment estimator than does the normal copula for typical record lengths in RFA, a result that has not previously been reported. The independence likelihood method may have unreasonable estimates in small samples unless it is modified to impose a similar constraint, but for large samples, it is more efficient than the L-moment method.

Spatial dependence in RFA is often a nuisance because the goal of an RFA is usually to estimate marginal return levels. With marginal GEV distributions, it is desirable to improve the efficiency without specifying a spatial dependence model. Specification and selection of a working dependence model can be avoided by a combined estimating equation approach based on data contrasts for clustered data [Stoner and Leroux, 2002]. For spatial data, estimating equation approaches have also been applied for marginal models with no need to correctly specify the dependence structure [Yasui and Lele, 1997; Clayton and Lin, 2005; Lin, 2008, 2010]. The marginal score equations can be combined in certain way to improve the efficiency [Nikolopoulos et al., 2011]. Further research in this direction may benefit not just RFA but general spatial extreme modeling.

### Acknowledgments

Z. Wang's graduate assistantship was partially supported by a contract with the Environment Canada. The Swiss rainfall data are publicly available from the open source R package SpatialExtremes (<http://CRAN.R-project.org/package=SpatialExtremes>).

### References

- Banerjee, S., B. P. Carlin, and A. E. Gelfand (2004), *Hierarchical Modeling and Analysis for Spatial Data*, 452 pp., Chapman and Hall, N. Y.
- Blanchet, J., and A. C. Davison (2011), Spatial modeling of extreme snow depth, *Ann. Appl. Stat.*, 5(3), 1699–1725.
- Buishand, T. A. (1991), Extreme rainfall estimation by combining data from several sites, *Hydrol. Sci. J.*, 36, 345–365.
- Castellarin, A. (2007), Probabilistic envelope curves for design flood estimation at ungauged sites, *Water Resour. Res.*, 43, W04406, doi:10.1029/2005WR004384.
- Castellarin, A., R. M. Vogel, and N. C. Matalas (2005), Probabilistic behavior of a regional envelope curve, *Water Resour. Res.*, 41, W06018, doi:10.1029/2004WR003042.
- Castellarin, A., D. Burn, and A. Brath (2008), Homogeneity testing: How homogeneous do heterogeneous cross-correlated regions seem?, *J. Hydrol.*, 360(1–4), 67–76.
- Chebana, F., and T. Ouarda (2009), Index flood-based multivariate regional frequency analysis, *Water Resour. Res.*, 45, W10435, doi:10.1029/2008WR007490.
- Clayton, M. K., and P.-S. Lin (2005), Analysis of binary spatial data by quasi-likelihood estimating equations, *Ann. Stat.*, 33(2), 542–555.
- Coles, S. G., and M. J. Dixon (1999), Likelihood-based inference for extreme value models, *Extremes*, 2, 5–23.
- Cunderlik, J. M., and T. B. Ouarda (2006), Regional flood-duration–frequency modeling in the changing environment, *J. Hydrol.*, 318(1), 276–291.
- Davison, A. C., and M. M. Gholamrezaee (2012), Geostatistics of extremes, *Proc. R. Soc. A*, 468, 581–608.
- Davison, A. C., S. A. Padoan, and M. Ribatet (2012), Statistical modeling of spatial extremes, *Stat. Sci.*, 27(2), 161–186.
- de Haan, L. (1984), A spectral representation for max-stable processes, *Ann. Probab.*, 12, 1194–1204.
- Embrechts, P., A. McNeil, and D. Straumann (2002), Correlation and dependence in risk management: Properties and pitfalls, in *Risk Management: Value at Risk and Beyond*, edited by M. A. H. Dempster, pp. 176–223, Cambridge Univ. Press, Cambridge, U. K.
- Genest, C., and A.-C. Favre (2007), Everything you always wanted to know about copula modeling but were afraid to ask, *J. Hydrol. Eng.*, 12, 347–368.
- Griffis, V., and J. Stedinger (2007), The use of GLS regression in regional hydrologic analyses, *J. Hydrol.*, 344(1–2), 82–95, doi:10.1016/j.jhydrol.2007.06.023.
- Gudendorf, G., and J. Segers (2010), Extreme-value copulas, in *Copula Theory and Its Applications*, edited by P. Jaworski et al., pp. 127–145, Springer Berlin Heidelberg.
- Hanel, M., T. A. Buishand, and C. Ferro (2009), A nonstationary index flood model for precipitation extremes in transient regional climate model simulations, *J. Geophys. Res.*, 114, D15107, doi:10.1029/2009JD011712.
- Heffernan, J. E., and J. A. Tawn (2004), A conditional approach for multivariate extreme values (with discussion), *J. R. Stat. Soc., Ser. B*, 66(3), 497–546.
- Hosking, J. R. M. (1990), L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *J. R. Stat. Soc., Ser. B*, 52, 105–124.

- Hosking, J. R. M., and J. R. Wallis (1988), The effect of intersite dependence on regional flood frequency analysis, *Water Resour. Res.*, *24*, 588–600.
- Hosking, J. R. M., and J. R. Wallis (1993), Some statistics useful in regional frequency analysis, *Water Resour. Res.*, *29*, 271–281.
- Hosking, J. R. M., and J. R. Wallis (1997), *Regional Frequency Analysis: An Approach Based on L-moments*, 224 pp., Cambridge Univ. Press, Cambridge, U. K.
- Javelle, P., T. B. Ouarda, M. Lang, B. Bobée, G. Galéa, and J.-M. Grésillon (2002), Development of regional flood-duration–frequency curves based on the index-flood method, *J. Hydrol.*, *258*(1), 249–259.
- Kojadinovic, I., J. Segers, and J. Yan (2011), Large-sample tests of extreme-value dependence for multivariate copulas, *Can. J. Stat.*, *39*(4), 703–720.
- Kojadinovic, I., H. Shang, and J. Yan (2014), A class of goodness-of-fit tests for spatial extremes models based on max-stable processes, *Stat. Interface*, in press.
- Kysely, J., and J. Picek (2007), Regional growth curves and improved design value estimates of extreme precipitation events in the Czech Republic, *Clim. Res.*, *33*(3), 243–255.
- Leclerc, M., and T. B. Ouarda (2007), Non-stationary regional flood frequency analysis at ungauged sites, *J. Hydrol.*, *343*(3), 254–265.
- Lettenmaier, D. P., J. Wallis, and E. Wood (1987), Effect of regional heterogeneity on flood frequency estimation, *Water Resour. Res.*, *23*, 313–323.
- Lin, P.-S. (2008), Estimating equations for spatially correlated data in multi-dimensional space, *Biometrika*, *95*(4), 847–858.
- Lin, P.-S. (2010), A working estimating equation for spatial count data, *J. Stat. Plann. Inference*, *140*(9), 2470–2477.
- Lindsay, B. G. (1988), Composite likelihood methods, in *Statistical Inference from Stochastic Processes*, edited by N. U. Prabhu, pp. 221–239, Am. Math. Soc., Providence, R. I.
- Martins, E. S., and J. R. Stedinger (2000), Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, *Water Resour. Res.*, *36*, 737–744, doi:10.1029/1999WR900330.
- Matalas, N. C., and W. B. Langbein (1962), Information content of the mean, *J. Geophys. Res.*, *67*(9), 3441–3448.
- Ngongondo, C. S., C.-Y. Xu, L. M. Tallaksen, B. Alemaw, and T. Chirwa (2011), Regional frequency analysis of rainfall extremes in Southern Malawi using the index rainfall and l-moments approaches, *Stochastic Environ. Res. Risk Assess.*, *25*(7), 939–955.
- Nikoloulopoulos, A. K., H. Joe, and N. R. Chaganty (2011), Weighted scores method for regression models with dependent data, *Biostatistics*, *12*, 653–665.
- Northrop, P. (2004), Likelihood-based approaches to flood frequency estimation, *J. Hydrol.*, *292*, 96–113.
- Ouarda, T., J. Cunderlik, A. St-Hilaire, M. Barbet, P. Bruneau, and B. Bobée (2006), Data-based comparison of seasonality-based regional flood frequency methods, *J. Hydrol.*, *330*(1), 329–339.
- Ouarda, T. B., M. Haché, P. Bruneau, and B. Bobée (2000), Regional flood peak and volume estimation in Northern Canadian Basin, *J. Cold Reg. Eng.*, *14*(4), 176–191.
- Ouarda, T. B. M. J. (2013), Hydrological frequency analysis, regional, in *Encyclopedia of Environmetrics*, 2nd ed., edited by A. El-Shaarawi and W. Piegorisch, John Wiley, Chichester, U. K., doi:10.1002/9780470057339.vnn043.
- Padoan, S. A., M. Ribatet, and S. A. Sisson (2010), Likelihood-based inference for max-stable processes, *J. Am. Stat. Assoc.*, *105*(489), 263–277.
- Ribatet, M., and R. Singleton (2013), *SpatialExtremes: Modelling Spatial Extremes*, R package version 1.9-1. [Available at <http://CRAN.R-project.org/package=SpatialExtremes>.]
- Schlather, M. (2002), Models for stationary max-stable random fields, *Extremes*, *5*(1), 33–44.
- Schlather, M., and J. A. Tawn (2003), A dependence measure for multivariate and spatial extreme values: Properties and inference, *Biometrika*, *90*(1), 139–156.
- Sklar, A. W. (1959), Fonctions de répartition à  $n$  dimension et leurs marges, *Publ. Inst. Stat. Univ. Paris*, *8*, 229–231.
- Stedinger, J. R. (1983), Estimating a regional flood frequency distribution, *Water Resour. Res.*, *19*, 503–510, doi:10.1029/WR019i002p00503.
- Stoner, J. A., and B. G. Leroux (2002), Analysis of clustered data: A combined estimating equations approach, *Biometrika*, *89*(3), 567–578.
- Troutman, B. M., and M. R. Karlinger (2003), Regional flood probabilities, *Water Resour. Res.*, *39*(4), 1095, doi:10.1029/2001WR001140.
- Varin, C. (2008), On composite marginal likelihoods, *Adv. Stat. Anal.*, *92*(1), 1–28.
- Varin, C., N. Reid, and D. Firth (2011), An overview of composite likelihood methods, *Stat. Sinica*, *21*(1), 5–42.
- Viglione, A., A. Castellarin, M. Rogger, R. Merz, and G. Blöschl (2012), Extreme rainstorms: Comparing regional envelope curves to stochastically generated events, *Water Resour. Res.*, *48*, W01509, doi:10.1029/2011WR010515.
- Wang, Z., J. Yan, and X. Zhang (2014), Incorporating spatial dependence in regional frequency analysis, *Tech. Rep. 9*, Dep. of Stat., Univ. of Conn., Storrs, Conn.
- Westra, S., and S. A. Sisson (2011), Detection of non-stationarity in precipitation extremes using a max-stable process model, *J. Hydrol.*, *406*(1), 119–128.
- Yasui, Y., and S. Lele (1997), A regression method for spatial disease rates: An estimating function approach, *J. Am. Stat. Assoc.*, *92*, 21–32.