A Bayesian hierarchical approach to regional frequency analysis

Benjamin Renard

WATER RESOURCES RESEARCH, VOL. ???, XXXX, DOI:10.1029/,

A Bayesian Hierarchical Approach To Regional Frequency Analysis

B. Renard[1]

B. Renard, Cemagref, UR HHLY, Hydrology-Hydraulics, 3 bis quai Chauveau - CP 220, F-69336 Lyon, France. (benjamin.renard@cemagref.fr)

[1]UR HHLY, Hydrology-Hydraulics, 3 bis
quai Chauveau, CP 220, F-69336 Lyon,
France.

Abstract. Regional Frequency Analysis (RFA) has a long history in Hydrology, and numerous distinct approaches have been proposed over the year to perform the estimation of some hydrologic quantity at a regional level. However, most of these approaches still rely on strong hypotheses that may limit their application and complicate the quantification of predictive uncertainty. The objective of this paper is to propose a general Bayesian hierarchical framework to implement RFA schemes that avoid these difficulties. The proposed framework is based on a two-level hierarchical model. The first level of the hierarchy describes the joint distribution of observations. An arbitrary marginal distribution, whose parameters may vary in space, is assumed for at-site series. The joint distribution is then derived by means of an elliptical copula, therefore providing an explicit description of the spatial dependence between data. The second level of the hierarchy describes the spatial variability of parameters using a regression model that links the parameter values with covariates describing site characteristics. Regression errors are modeled with a Gaussian spatial field which may exhibit spatial dependence. This framework enables performing prediction at both gauged and ungauged sites and, importantly, rigorously quantifying the associated predictive uncertainty. A case study based on annual maxima of daily rainfall demonstrates the applicability of this hierarchical approach. Although numerous avenues for improvement can already be identified (amongst which the inclusion of temporal covariates to model time variability), the proposed

model constitutes a general framework for implementing flexible RFA schemes and quantifying the associated predictive uncertainty.

1. Introduction

1.1. Standard implementations of regional frequency analysis

The purpose of Regional Frequency Analysis (RFA) is to estimate the distribution of some hydrologic variable (e.g. annual maximum rainfall or runoff) using data from several sites. Compared with standard at-site frequency analysis, RFA attempts to improve the precision of estimates by sharing the information stemming from similar sites. Moreover, RFA enables estimation at ungauged or poorly gauged sites by transferring the information arising from neighboring gauging stations.

Numerous approaches have been proposed to implement RFA schemes. Amongst them, the index flood method proposed by *Dalrymple* [1960] is still widely used in engineering practice. It is based on a scale invariance hypothesis: within an homogeneous region, distributions from all sites are assumed identical, up to a scale factor, termed the index flood. The implementation of the index flood method can be summarized in three steps [e.g., *Hosking and Wallis*, 1997]: (i) delineation of an homogeneous region; (ii) estimation of the common regional distribution, based on scaled at-site data (i.e. divided by the index flood); (iii) transfer of information to ungauged or poorly gauged site using a regression model linking the index flood values with site characteristics.

The index flood method is widely used due to its ease of implementation and its robustness. However, its basic implementation is affected by several limitations:

- The delineation of homogeneous regions, where the scale invariance assumption holds, is far from obvious.
- The scale invariance assumption might simply be too restrictive is some cases.

• In most cases, the index flood is defined as the mean or the median of at-site data, but the physical reasons behind this choice are unclear.

• The regional distribution is estimated by pooling scaled at-site data, and treating them as is they were independent, which is rarely the case.

• Standard regression methods like ordinary least squares might be statistically inefficient because index flood values are statistics (as opposed to observations), and are therefore affected by estimation errors that may be dependent in space and whose properties may vary from site to site.

• The previous points make the quantification of the total predictive uncertainty challenging.

A wealth of research has been carried out over the years, either to improve the implementation of the index flood method, or to generalize it by abandoning some of its most restrictive assumptions (in particular, scale invariance). A non-exhaustive list of examples includes the work by *Ouarda et al.* [2001] on the concept of homogeneous region, the studies by *Stedinger* [1983] and *Hosking and Wallis* [1988] on estimating the regional distribution with spatially dependent data, or the extension to peak-over-threshold series of *Madsen and Rosbjerg* [1997] and *Ribatet et al.* [2007].

The transfer of information from gauged to ungauged site and the quantification of the associated predictive uncertainty has been a topic of particular attention [e.g., *Stedinger and Tasker*, 1985, 1986; *Reis et al.*, 2005; *Kjeldsen and Jones*, 2009a; *Micevski and Kuczera*, 2009]. Indeed, this transfer relies on a regression model linking the index flood values (estimated at gauged sites) and site characteristics. *Stedinger and Tasker* [1985, 1986] introduced the Generalized Least Squares (GLS) approach to account for both the het-

eroscedastic and spatially dependent nature of sampling errors (i.e., errors in estimating the index flood at gauged sites) and the existence of regression errors. *Robson and Reed* [1999] and *Kjeldsen and Jones* [2009a] further generalized the GLS approach by considering spatially dependent regression errors.

Despite these advances, GLS-based transfer of information from gauged to ungauged site still relies on the following two-step procedure:

1. Local estimation: an index flood is first chosen (e.g. the at-site mean or median), and is estimated at each gauged site. This estimation is affected by sampling errors, which are spatially dependent and whose variance varies from site to site. Consequently, the covariance matrix $\hat{\mathbf{\Sigma}}$ of sampling errors is also estimated;

2. Regional estimation: A regression model is estimated to link the index flood value with site/catchment characteristics. Importantly, the estimation of the regression model accounts for the existence of sampling errors in a GLS framework, and is hence performed conditionally on $\hat{\mathbf{\Sigma}}$. Also note that estimating a regression model is not restricted to estimating the regression coefficients, but also involves estimating the properties of regression errors, i.e. their covariance matrix $\hat{\mathbf{\Gamma}}$. This is of primary importance since this matrix plays an important role in the predictive uncertainty at ungauged sites.

Such a two-step procedure might be problematic in the context of quantifying the total predictive uncertainty. Indeed, estimates at step 2 are conditional on estimates at step 1. In particular, the covariance matrix $\hat{\mathbf{\Sigma}}$ is an estimate, and may itself be in error (see e.g. the discussion in *Stedinger and Tasker* [1985] and *Kroll and Stedinger* [1998] on desirable properties of $\hat{\mathbf{\Sigma}}$). Such error may then propagates to step 2. Consequently, it would be desirable to avoid separating the inference process in two separate steps: this can be

achieved using hierarchical models (see following section 1.2).

In addition to these issues related to the estimation procedure, the assumptions underlying index flood approaches remain questionable. Indeed, the scale invariance assumption is convenient because it merges all spatial variability into a single parameter (the index-flood parameter). However, this assumption forces the coefficient of variation of data to remain constant throughout an homogeneous region: this might be too restrictive in some regions, or it might force to drastically reduce the spatial extent of the region to ensure the scale invariance hypothesis is met. Several alternatives have therefore been proposed to move beyond this restrictive framework, in particular region of influence approaches [*Burn*, 1990] and recent developments [*Kjeldsen and Jones*, 2009b], normalized quantile regression [*Fill and Stedinger*, 1998] or empirical Bayes procedures [*Kuczera*, 1982a, b, 1983] (see also the discussion provided by *Griffis and Stedinger* [2007]).

1.2. Bayesian hierarchical models

An alternative approach, based on Bayesian hierarchical models, has been explored more recently. In particular, *Wikle et al.* [1998] proposed a general hierarchical framework to describe the spatial variability of the distribution of some environmental variable. The principle of a hierarchical model is to use several modeling layers. For instance, a first layer may assume that the data follow some distribution with unknown parameters, while a second layer may model the variability of those parameters in space, using some regression model. This closely corresponds to the successive steps involved in the standard implementation of RFA approaches. However, the main advantage of a hierarchical model is that all unknown quantities can be inferred simultaneously, therefore accounting for possible interactions between estimation errors made at different layers and yielding

a more rigorous quantification of the predictive uncertainty. In other words, hierarchical models allow describing the local variability of data *together* with their regional coherence, without separating the inference process in several steps. Moreover, such models are more general than the model underlying the index flood approach, since they do not require assuming scale invariance.

Several applications of Bayesian hierarchical models in a hydrological context have been proposed in the literature. For instance, *Cooley et al.* [2007] described the spatial variability of extreme rainfall, while *Aryal et al.* [2009] extended this description to both spatial and temporal variability. Similarly, *Lima and Lall* used Bayesian hierarchical models to describe daily rainfall occurrences [*Lima and Lall*, 2009] or runoff extremes [*Lima and Lall*, 2010] in a regional context. In addition to these hydrological applications, similar Bayesian hierarchical models have been used in other fields, e.g. for extreme wind speed modeling [*Coles and Casson*, 1998; *Casson and Coles*, 1998, 2000].

Despite improving the standard implementation of RFA approaches, all Bayesian hierarchical models described above rely on an assumption of conditional independence: data are assumed spatially independent given the values of their distribution's parameters. This would be a valid assumption if most spatial covariation in the data was explained by the spatial covariation in the parameters. However, it is questionable since spatial dependence between data in the one hand and parameters in the other hand arise from distinct processes, as noted by *Cooley et al.* [2007]: in a nutshell, data dependence can be interpreted as *weather* spatial dependence, while dependence between parameters (also termed process dependence) relates to *climate* spatial dependence. Weather should exhibit spatial dependence (at least within a short distance range), even if the climate were

perfectly known.

Examples of spatial hierarchical models explicitly accounting for intersite dependence in the observations are very few. *Perreault* [2000] proposed such a model to detect a regional step-change in annual runoff. Alternatively, *Micevski et al.* [2006] and *Micevski* [2007] proposed a Bayesian hierarchical regional flood model accounting for data dependence. However, in both cases, the explicit description of data dependence was rooted to a particular distributional assumption: Gaussian data were assumed by *Perreault* [2000], while *Micevski et al.* [2006] and *Micevski* [2007] used a mixture of log-normal distributions. Unfortunately, Gaussian-related assumptions may be too restrictive for other hydrologic variables or in other geographical contexts.

1.3. Objectives

Building on previous work described in the preceding sections, this paper aims to derive a general Bayesian hierarchical framework for RFA. In particular, this framework should enable an explicit description of spatial dependence between data, without relying on Gaussian-related distributional assumptions. This is achieved by means of the elliptical copula family [*Genest and Favre*, 2007], which constitutes a convenient tool to model dependence in a highly dimensional and non-Gaussian context.

The paper is organized as follows. Section 2 describes the two-level hierarchical framework, with level 1 modeling the joint distribution of observations and level 2 modeling the variation of the distribution' parameters in space. Section 3 describes the inference of the hierarchical model and its use for prediction at both gauged and ungauged sites. Section 4 illustrates the application of the proposed framework for the regional estimation of

extreme rainfall. Avenues for further improvement are identified and discussed in section 5, before summarizing the main results in section 6.

2. A Bayesian hierarchical modeling framework

2.1. Data level

Let $Y(s,t)$ denote the variable of interest at site s and time t. For instance, $Y(s,t)$ may represent the annual maximum daily rainfall at site s and year t. Let us further assume that observations are available at sites $\tilde{s}_1$,..., $\tilde{s}_M$ and times t=1:T. The ($T\times M$) observation matrix is denoted by $\tilde{\boldsymbol{y}} = (\tilde{\boldsymbol{y}}(\tilde{s}_1), ..., \tilde{\boldsymbol{y}}(\tilde{s}_M))$, with $\tilde{\boldsymbol{y}}(\tilde{s}_i) = (\tilde{y}(\tilde{s}_i,t))_{t=1:T}$ denoting the time series of observations at a given site $\tilde{s}_k$. The shorthand notation $\tilde{\boldsymbol{y}}(t) = (\tilde{y}(\tilde{s}_i,t))_{i=1:M}$ is also used to denote the (spatial) vector of observations at a given time t. Note that this notation assumes that all sites share the same observation period. Although considering non-concomitant observation periods would not affect the modeling hypotheses, it would certainly create tedious notation and complicate the model implementation (see discussion in section 5.1).

2.1.1. At-site distribution

At an arbitrary site s within the area of study, and at any time t, $Y(s,t)$ is assumed to be a realization from a given distribution whose parameters vary in space (see left-hand-side of Figure 1):

$$Y(s,t) \sim p(\boldsymbol{\theta}(s)) \tag{1}$$

RFA involves estimating the parameter vector $\boldsymbol{\theta}(s)$ at any (gauged or ungauged) site s. In order to avoid confusion with other parameters that will be introduced later on, the term "D-parameters" is systematically used to denote the parameters $\boldsymbol{\theta}(s) = (\theta_k(s))_{k=1:D}$

of the parent distribution. Note that D-parameters are allowed to vary in space but not in time. An extension of the framework for allowing time-varying D-parameters is possible but is left for future work, since this paper mainly focuses on spatial aspects.

2.1.2. Joint distribution

In a spatial context, the derivation of the likelihood of observations $\tilde{\boldsymbol{y}}$ requires knowing the multivariate distribution of the spatial observation vector $\tilde{\boldsymbol{y}}(t)$ at any time t. An assumption often made (explicitly or implicitly) by some regional estimation methods is that the data are spatially independent. In this case, the multivariate pdf is simply equal to the product of marginal pdfs in (1). Implications of this assumption are discussed by e.g. *Stedinger* [1983], *Hosking and Wallis* [1988], *Madsen and Rosbjerg* [1997] and *Renard and Lang* [2007]. In particular, these authors demonstrate that the variance (i.e. the uncertainty) of estimated quantities is underestimated when data do not support the independence assumption.

One of the main objectives of the framework presented in this paper is to overcome this limitation, by explicitly modeling spatial dependence. The approach taken to achieve this objective has to account for two important points: (i) it has to be applicable for an arbitrary choice of marginal distribution in (1); (ii) since regional analysis may involve hundreds of sites, it has to remain practical with high-dimensional data.

Point (i) above makes copulas a natural candidate to model dependence in the context of this framework. Indeed, the copula theory is based on the description of the dependence structure independently of marginal distributions [e.g., *Favre et al.*, 2004]. Point (ii) suggests focusing on the elliptical copula family, because of its ability to describe high-dimensional datasets. A thorough description of the elliptical copula family can be found

in the paper by *Genest and Favre* [2007], and a summary of the main characteristics of two particular members (the Gaussian and the Student copulas) is given in Appendix A. A M-dimensional elliptical copula is parameterized by a $(M \times M)$ symmetric dependence matrix $\mathbf{\Sigma}$ describing pairwise dependences. Additional parameters $\boldsymbol{\eta}$ may be used for some members of the family (e.g. the Student copula) to describe the strength of tail dependence. The multivariate distribution is finally derived by combining this dependence structure with the marginal distributions given by equation (1) (see Appendix A for details). An example of application of an elliptical copula (the Gaussian copula) to regional frequency analysis is given by *Renard and Lang* [2007].

Formally, it is assumed that the multivariate distribution of data from any set of M sites can be derived from an M-dimensional elliptical copula, with pairwise dependence matrix $\mathbf{\Sigma}$, additional dependence parameters $\boldsymbol{\eta}$ and marginal distributions $\{p(\boldsymbol{\theta}(s_1)), ..., p(\boldsymbol{\theta}(s_M))\}$:

$$(Y(s_1,t), ..., Y(s_M,t)) \sim EC_M(\mathbf{\Sigma}, \boldsymbol{\eta}, \{\boldsymbol{\theta}(s_1), ..., \boldsymbol{\theta}(s_M)\}) \quad (2)$$

The symbol EC_M stands for "M-dimensional Elliptical Copula", and analytical formulas for the corresponding joint pdf are given in Appendix A .

In order to simplify the model, it is further assumed that the dependence between data from two sites solely depends on the inter-site distance. An analogy can be drawn with the common treatment of stationary and isotropic spatial random fields in geostatistics [e.g., *Chiles and Delfiner*, 1999]. It follows that the elements of the pairwise dependence matrix $\mathbf{\Sigma}$ can be expressed as a function of the inter-site distance, parameterized by some vector $\boldsymbol{\psi}$ (see Figure 1):

$$\Sigma(i,j) = \Psi(\|s_i - s_j\|\,;\boldsymbol{\psi}) \tag{3}$$

Note that the inter-site distance is not necessarily Euclidean in the bi-dimensional xy-space (geographical coordinates). For instance, a distance in the tri-dimensional space xy+elevation may be useful if inter-site dependence is lower for sites having contrasted elevations. More generally, the distance may depend on covariates other than geographical coordinates [e.g., *Cooley et al.*, 2007; *Blanchet and Davison*, 2011]. However, for the sake of simplicity, this is not made explicit in the notation of equation (3). This topic is further discussed in section 5.8.

A valid dependence-distance function $\Psi(.;\boldsymbol{\psi})$ must ensure the positive-definiteness of the dependence matrix $\boldsymbol{\Sigma}$. Such a function can be chosen amongst the numerous covariogram models (e.g., exponential, spherical, Gaussian) existing in geostatistics. By analogy with the covariogram used in geostatistics, the dependence-distance model in (3) is termed "dependogram". Note that a distinct naming convention is used because the elements of the matrix $\boldsymbol{\Sigma}$ are not equal to the covariances between data pairs, but rather to the dependence coefficients of the elliptical copula.

It is stressed that the use of an elliptical copula to model spatial dependence is motivated by practical considerations. Although the elliptical family is quite flexible (in particular, it encompasses tail-dependent and tail-independent models), there can be no guarantee that an elliptical copula will be able to model the data at hand. In particular, a multivariate distribution derived from an elliptical copula is not compatible with the family of multivariate extreme value distributions [*Mikosch*, 2005]. It follows that the use of the elliptical copula model in extrapolation to estimate low probabilities (e.g.,

$Pr(Y(s_1,t) > u \cap ... \cap Y(s_M,t) > u)$ for some large u) is likely to yield inaccurate results. However, the aim of the framework presented in this paper is not to estimate the probability of extreme multivariate events, but rather to estimate the distribution of the variable of interest at any site, while accounting for the existence of dependences at observed levels. The implications of the elliptical copula assumption are more thoroughly discussed in section 5.2.

2.2. Process level

Section 2.1 described the construction of the multivariate distribution of data by specifying the following three components: (i) the at-site distribution (1); (ii) the elliptical copula (2) used to model spatial dependence; (iii) the dependogram (3). The second level of the hierarchical framework aims to describe the variation of the D-parameters in space, by means of a Gaussian spatial process whose mean depends on covariates describing the site (or catchment) characteristics (see right-hand-side of Figure 1).

2.2.1. D-parameter regression model

Let $\boldsymbol{\theta}(s) = (\theta_k(s))_{k=1:D}$ be the D-parameter vector. A regression model is used to describe the spatial variation of each D-parameter $\theta_k(s)$ as follows:

$$g_k(\theta_k(s)) = h_k(\boldsymbol{x}_k(s); \boldsymbol{\beta}_k) + \epsilon_k(s) \quad (4)$$

Equation (4) uses the following components:

1. One-to-one function g_k is termed the "link function" by analogy with generalized linear models [e.g., *Dobson*, 2001]. The identity function may be used in most cases. However, alternative functions might be useful for some D-parameters, e.g. a logarithm

function to ensure positivity or a logit function for a D-parameter comprised between zero and one.

2. Vector $\boldsymbol{x}_k(s)$ represents the set of covariates used to describe the characteristics of site (or catchment) s (e .g., elevation, distance to sea, catchment size, etc.). As in any regression framework, the choice of relevant covariates is of primary importance, and preliminary analyzes are often useful e.g. to eliminate highly correlated covariates.

3. Function $h_k(.;\boldsymbol{\beta}_k)$ is the regression function, parameterized by a vector of regression parameters $\boldsymbol{\beta}_k$. The most common choice of regression function is the linear function $h_k(\boldsymbol{x}_k(s);\boldsymbol{\beta}_k) = \boldsymbol{x}_k(s)\boldsymbol{\beta}_k$, but alternative functions may be used.

4. $\epsilon_k(s)$ is the residual of the regression. It stems from the imperfect nature of the regression model, and is therefore termed "regression error". In the hierarchical modeling context described in this paper, regression errors are treated as latent variables, i.e. unobserved variables that need to be inferred.

2.2.2. Regression errors

Regression errors $\boldsymbol{\epsilon}_k = (\epsilon_k(s))_{s=1:M}$ are commonly assumed to be spatially independent [e.g., *Stedinger and Tasker*, 1985, 1986; *Reis et al.*, 2005]. However, recent work by *Kjeldsen and Jones* [2009a] suggests that regression errors can be significantly dependent, especially for nearby sites and/or regression models with poor predictive ability. Ignoring spatial dependences between regression errors possibly affects regional frequency analysis in two distinct ways: (i) it may impact the accuracy and/or precision of estimates for the regression parameters; (ii) it may impact the predictive variance (i.e. the uncertainty in quantities estimated at ungauged sites) by not taking advantage of the regression errors estimated as nearby sites. Consequently, following *Kjeldsen and Jones* [2009a], regression

errors are assumed to be a realization from a spatial Gaussian field with mean zero and covariance matrix $\mathbf{\Gamma}_k$:

$$(\epsilon_k(s_1), ..., \epsilon_k(s_M)) \sim N_M(0; \mathbf{\Gamma}_k) \tag{5}$$

A covariogram model is used to relate pairwise covariances to inter-site distances:

$$\Gamma_k(i,j) = \Upsilon_k(\|s_i - s_j\|\,; \boldsymbol{\upsilon}_k) \tag{6}$$

In a Bayesian hierarchical context, the Gaussian distribution used to describe regression errors is termed the hyper-distribution. Parameters $\boldsymbol{\upsilon}_k$ defining the covariance matrix are termed the hyper-parameters (see Figure 1). Note that as in previous section 2.1.2, the inter-site distance is not necessarily Euclidean and can be interpreted in a wider sense. Moreover, note that regression errors $\epsilon_k(s)$ and $\epsilon_q(s)$ related to two distinct D-parameters $\theta_k(s)$ and $\theta_q(s)$ will be described by independently using equations (5-6) twice. This implies that an assumption of independence between regression errors $\epsilon_k(s)$ and $\epsilon_q(s)$ is effectively made. This assumption will be further discussed subsequently (see sections 4 and 5.3).

2.3. Remarks

In the hierarchical framework presented in previous sections, two distinct dependence structures are used: the dependogram (3) aims at describing dependence between observations, while the covariogram (6) aims at describing dependence between the parameters of their distribution (or more accurately, between the errors of the regression linking the D-parameters with catchment/site characteristics). As explained in the introduction, the former structure relates to weather spatial dependence, while the latter relates to climate

spatial dependence [see also the discussion in *Cooley et al.*, 2007]. These dependence structures play distinct roles in the context of regional frequency analysis: climate spatial dependence allows transferring information from gauged to ungauged sites, while the main effect of weather dependence is to diminish the information content of data collected at nearby sites. Consequently, both dependence structures are likely to have an impact on predictions and should therefore be accounted for. Note that the assumption of conditional independence frequently made in similar Bayesian hierarchical frameworks (see introduction) corresponds to forcing the dependence matrix $\mathbf{\Sigma}$ in equation (3) to unity. Moreover, the regression model of equation (4) yields several interesting particular cases when the regression errors are forced to zero:

1. A purely local D-parameter (i.e. whose value is site-specific) can be obtained with the regression model $\theta_k(\tilde{s}) = \beta_k^{(\tilde{s})}$. This corresponds to introducing as many regression parameters $\beta_k^{(\tilde{s})}$ as there are sites, and use them to model a site effect. An obvious drawback of this approach is that the estimation at ungauged site s is not directly possible since the site effect $\beta_k^{(s)}$ is unknown.

2. A regional D-parameter (i.e. having an identical value for all sites within the region) corresponds to the regression model $\theta_k(\tilde{s}) = \beta_k$. This is a rather strong assumption, which may yield an underestimation of the predictive uncertainty. Indeed, in the absence of regression errors, the uncertainty in the estimation of θ_k is identical for all sites (gauged or not).

3. The strong assumption that parameter $\theta_k(\tilde{s})$ is identical for all sites can be relaxed by using a regression model $g_k(\theta_k(\tilde{s})) = h_k(\boldsymbol{x}_k(\tilde{s}); \boldsymbol{\beta}_k)$. This corresponds to the "covariate modeling" approach proposed by several authors [e.g., *Katz et al.*, 2002; *Maraun et al.*,

2009]. However, the possible underestimation of predictive uncertainty in the absence of regression errors still holds.

Lastly, the Bayesian hierarchical framework shares several similarities with the GLS approach proposed by *Stedinger and Tasker* [1985, 1986] and its latest developments by *Kjeldsen and Jones* [2009a]. Indeed, GLS also uses two spatial dependence structures: (i) the covariance matrix of sampling estimation errors (e.g., errors in estimating an index flood at gauged sites), which results from the spatially dependent nature of data; (ii) the covariance matrix of regression errors (termed model errors in the GLS approach). The framework presented in this paper therefore borrows from GLS the objective of modeling the spatial variability in the distinct sources of errors affecting RFA. However, it differs in its implementation. GLS first estimates the spatial variability of sampling errors, and then uses this estimation to fit the regression in a second step. By contrast, the proposed framework performs the inference in a single step, which facilitates the quantification of the total predictive uncertainty (see discussion in the introduction section). Moreover, the regression is applied on each parameter, while in general GLS is rather applied to a single hydrologic quantity (e.g. an index flood or a quantile, but see *Tasker and Stedinger* [1989] and *Griffis and Stedinger* [2007] for exceptions). This is a more general approach, at least in principle; however what level of model complexity can be identified given the limited information content of data remains an open question (this will be further discussed in section 5.5).

3. Estimation and prediction

3.1. Posterior distribution

The application of the hierarchical framework described in section 2 requires estimating the following unknown quantities:

1. The dependogram parameters $\boldsymbol{\psi}$
2. Additional parameters of the elliptical copula $\boldsymbol{\eta}$
3. The regression parameters $\boldsymbol{\beta}_k$, for $k = 1{:}D$
4. The regression errors $\boldsymbol{\epsilon}_k$, for $k = 1{:}D$ (latent variables)
5. The covariogram parameters $\boldsymbol{\upsilon}_k$, for $k = 1{:}D$ (hyper-parameters)

The posterior distribution of these quantities, given observations $\tilde{\boldsymbol{y}}$ and covariates $\boldsymbol{x}$, can be derived as follows:

$$p(\boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\beta}_{k=1:D}, \boldsymbol{\epsilon}_{k=1:D}, \boldsymbol{\upsilon}_{k=1:D} | \tilde{\boldsymbol{y}}, \boldsymbol{x})$$

$$\propto p(\tilde{\boldsymbol{y}} | \boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\beta}_{k=1:D}, \boldsymbol{\epsilon}_{k=1:D}, \boldsymbol{\upsilon}_{k=1:D}, \boldsymbol{x}) p(\boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\beta}_{k=1:D}, \boldsymbol{\epsilon}_{k=1:D}, \boldsymbol{\upsilon}_{k=1:D} | \boldsymbol{x}) \tag{7a}$$

$$= p(\tilde{\boldsymbol{y}} | \boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\beta}_{k=1:D}, \boldsymbol{\epsilon}_{k=1:D}, \boldsymbol{x}) p(\boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\beta}_{k=1:D}) p(\boldsymbol{\epsilon}_{k=1:D}, \boldsymbol{\upsilon}_{k=1:D}) \tag{7b}$$

$$= p(\tilde{\boldsymbol{y}} | \boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\beta}_{k=1:D}, \boldsymbol{\epsilon}_{k=1:D}, \boldsymbol{x}) p(\boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\beta}_{k=1:D}) p(\boldsymbol{\epsilon}_{k=1:D} | \boldsymbol{\upsilon}_{k=1:D}) p(\boldsymbol{\upsilon}_{k=1:D}) \tag{7c}$$

The following assumptions have been made to derive this posterior distribution:

1. Equation (7a) is a direct application of Bayes theorem.
2. Equation (7b) assumes: (i) the prior distribution does not depend on covariates $\boldsymbol{x}$; (ii) prior independence between $\boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\beta}_{k=1:D}$ in the one hand, and $\boldsymbol{\epsilon}_{k=1:D}, \boldsymbol{\upsilon}_{k=1:D}$ in the other hand. This assumption aims at isolating the hierarchical components of the model, i.e. $\boldsymbol{\epsilon}_{k=1:D}$ and $\boldsymbol{\upsilon}_{k=1:D}$. Moreover, it is noted that the likelihood of observations can be derived without using the hyper-parameters $\boldsymbol{\upsilon}_{k=1:D}$. Indeed, regression errors $\boldsymbol{\epsilon}_{k=1:D}$

suffice to apply the regression model (4) and hence to derive D-parameters (see detailed derivation of the likelihood in following equation (9)).

3. Equation (7c) is an application of conditional probability rules.

The posterior distribution (7) is made up of the following components:

1. The terms $p(\boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\beta}_{k=1:D})$ and $p(\boldsymbol{\upsilon}_{k=1:D})$ are priors for the inferred parameters and hyper-parameters, respectively.

2. The term $p(\boldsymbol{\epsilon}_{k=1:D}|\boldsymbol{\upsilon}_{k=1:D})$ represents the hierarchical part of the model. Assuming regression errors $\boldsymbol{\epsilon}_k$ related to different D-parameters θ_k are mutually independent yields:

$$p(\boldsymbol{\epsilon}_{k=1:D}|\boldsymbol{\upsilon}_{k=1:D}) = \prod_{k=1}^{D} f_N(\epsilon_k(\tilde{s}_1), ..., \epsilon_k(\tilde{s}_M)|\mathbf{0}, \boldsymbol{\Gamma}_k(\boldsymbol{\upsilon}_k)) \quad (8)$$

where $\boldsymbol{\Gamma}_k(\boldsymbol{\upsilon}_k)$ is the covariance matrix derived from the covariogram model (6) and $f_N(\boldsymbol{z}|\boldsymbol{\mu}, \boldsymbol{\Gamma})$ represents the pdf of a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Gamma}$. Note that the assumption that there is no cross-correlation between regression errors related to different D-parameters may be restrictive. This assumption will be evaluated in the case study (section 4) and will be further discussed in section 5.3.

3. The term $p(\tilde{\boldsymbol{y}}|\boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\beta}_{k=1:D}, \boldsymbol{\epsilon}_{k=1:D}, \boldsymbol{x})$ is the likelihood of observations. Its derivation requires further explanation. At a given time step t, the likelihood of (spatial) observations $\tilde{\boldsymbol{y}}(t)$ can be computed as follows:

$$p(\tilde{\boldsymbol{y}}(t)|\boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\beta}_{k=1:D}, \boldsymbol{\epsilon}_{k=1:D}, \boldsymbol{x}) =$$

$$f_{EC}\left(\tilde{\boldsymbol{y}}(t)|\boldsymbol{\Sigma}(\boldsymbol{\psi}), \boldsymbol{\eta}, \{[\theta_k(\boldsymbol{x}_k(\tilde{s}_1), \boldsymbol{\beta}_k, \epsilon_k(\tilde{s}_1))]_{k=1:D}, ..., [\theta_k(\boldsymbol{x}_k(\tilde{s}_M), \boldsymbol{\beta}_k, \epsilon_k(\tilde{s}_M))]_{k=1:D}\}\right) \quad (9)$$

In equation (9), $\boldsymbol{\Sigma}(\boldsymbol{\psi})$ is the dependence matrix of the elliptical copula, derived from the dependogram model (3). For a given observation site $\tilde{s}_j$, the kth D-parameter $\theta_k(\boldsymbol{x}_k(\tilde{s}_j), \boldsymbol{\beta}_k, \epsilon_k(\tilde{s}_j))$ is derived by applying the regression equation (4), i.e.:

$$\theta_k(\boldsymbol{x}_k(\tilde{s}_j), \boldsymbol{\beta}_k, \epsilon_k(\tilde{s}_j)) = g_k^{-1}\left(h_k(\boldsymbol{x}_k(\tilde{s}_j), \boldsymbol{\beta}_k) + \epsilon_k(\tilde{s}_j)\right) \tag{10}$$

Lastly, $f_{EC}(\boldsymbol{z}|\boldsymbol{\Sigma}, \boldsymbol{\eta}, \{\boldsymbol{\theta}(s_1), ..., \boldsymbol{\theta}(s_M)\})$ represents the multivariate pdf of a M-dimensional vector $\boldsymbol{z}$ derived from the elliptical copula with pairwise dependence matrix $\boldsymbol{\Sigma}$, additional dependence parameters $\boldsymbol{\eta}$ and marginal distributions $\{p(\boldsymbol{\theta}(s_1)), ..., p(\boldsymbol{\theta}(s_M))\}$, as detailed in Appendix A.

Assuming temporal independence between observations $\tilde{\boldsymbol{y}}(t)$, the likelihood of the whole observation matrix $\tilde{\boldsymbol{y}}$ is simply obtained as follows:

$$p(\tilde{\boldsymbol{y}}|\boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\beta}_{k=1:D}, \boldsymbol{\epsilon}_{k=1:D}, \boldsymbol{x}) = \prod_{t=1}^{T} p(\tilde{\boldsymbol{y}}(t)|\boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\beta}_{k=1:D}, \boldsymbol{\epsilon}_{k=1:D}, \boldsymbol{x}) \tag{11}$$

3.2. Inference

The posterior distribution (7) poses a computational challenge because its dimension grows with the number of sites. This is due to the explicit modeling of regression errors through latent variables. Consequently, the number of quantities to be inferred from the posterior can amount to hundreds. This is typical of hierarchical models with latent variables used to describe unobserved processes [e.g., *Clark*, 2005]. A Markov Chain Monte Carlo (MCMC) sampler is used to address this difficulty. It is stressed that MCMC sampling from high-dimensional posteriors is challenging but by no means insurmountable. Successful examples are provided by e.g. *Crainiceanu et al.* [2003]; *Vrugt et al.* [2008];

Thyer et al. [2009]; *Reichert and Mieleitner* [2009] in Hydrology, not to mention numerous applications in other fields [e.g., *Clark*, 2003, in ecology][*Storz and Beaumont*, 2002, in genetics,etc.].

The MCMC sampler used in this paper is made up of two stages. Stage one makes use of an adaptive block Metropolis algorithm with univariate Gaussian jump distributions (i.e. components of the parameter vector are updated one at a time). The adaption strategy adjusts the jump variances to produce an adequate jump rate [see *Renard et al.*, 2006, for a detailed description]. This first sampler is used to perform a preliminary exploration of the posterior distribution properties (notably in terms of posterior covariance). In the second stage, a standard Metropolis sampler [*Metropolis and Ulam*, 1949; *Metropolis et al.*, 1953] is used, with a Gaussian jump distribution whose covariance matrix is specified using the preliminary exploration performed at stage one. Convergence is assessed by evolving four parallel chains and verifying that the Gelman-Rubin criteria [*Gelman et al.*, 1995] are close to one for all inferred quantities.

Note that additional computing efficiency might be achieved in some cases by using conjugate priors. However, this possibility is not investigated in this paper.

3.3. Prediction at gauged site

Once inference has been performed using the MCMC strategy outlined above, the next step is to use the model to predict some quantity of interest. In this section, focus is on prediction at a gauged site $\tilde{s}$. A typical quantity of interest is the p-quantile of the distribution of observations at site $\tilde{s}$, which can be directly derived from D-parameters $\boldsymbol{\theta}(\tilde{s})$.

In a Bayesian context, another interesting byproduct of the posterior distribution is the

predictive distribution of a (future) observation. In general, if $\boldsymbol{\Theta}$ denotes the vector of all parameters subject to inference, the pdf of this predictive distribution, evaluated at some value w, is mathematically defined as follows [*Gelman et al.*, 1995]:

$$p(w|\tilde{\boldsymbol{y}}) = \int p(w|\boldsymbol{\Theta})p(\boldsymbol{\Theta}|\tilde{\boldsymbol{y}})d\boldsymbol{\Theta} \tag{12}$$

In practice, this integration is not performed analytically but is approximated using the MCMC replicates from the posterior (7). The approximation algorithm simply consists in generating a value from the at-site distribution (1) for each MCMC replicate. The resulting sample is a realization from the predictive distribution and can therefore be used to estimate its characteristics (mean, variance, probability interval, etc.).

Let us assume that MCMC sampling generated a set of n_{sim} replicates from the posterior distribution, $(\boldsymbol{\psi}^{(j)}, \boldsymbol{\eta}^{(j)}, \boldsymbol{\beta}^{(j)}_{k=1:D}, \boldsymbol{\epsilon}^{(j)}_{k=1:D}, \boldsymbol{\upsilon}^{(j)}_{k=1:D})_{j=1:n_{sim}}$. These replicates can be used to make a prediction at gauged site $\tilde{s}$ using the following algorithm:

Do $j = 1 : n_{sim}$

1. compute D-parameters from the regression model (4), $\theta^{(j)}_k(\tilde{s}) = g_k^{-1}\left(h_k(\boldsymbol{x}_k(\tilde{s}), \boldsymbol{\beta}^{(j)}_k) + \epsilon^{(j)}_k(\tilde{s})\right)$ for $k = 1 : D$

2. compute derived quantity $z^{(j)} = z(\boldsymbol{\theta}^{(j)}(\tilde{s}))$ (e.g., the T-year quantile) or generate a value from the at-site distribution $w^{(j)} \sim p(\boldsymbol{\theta}^{(j)}(\tilde{s}))$

The samples $(z^{(j)})_{j=1:n_{sim}}$ and $(w^{(j)})_{j=1:n_{sim}}$ can be considered as realizations from the posterior distribution of the quantity of interest and from the predictive distribution, respectively.

3.4. Prediction at ungauged site

The fundamental difference with the prediction approach presented in previous section 3.3 is that the regression error $\epsilon_k(s)$ has not been inferred for an ungauged site s. However, the properties of the hyper-distribution (5) have been inferred. Since the hyper-distribution is the joint distribution of a vector of regression errors, it can be used to indirectly infer the properties of the regression error $\epsilon_k(s)$ for the target ungauged site s. More precisely, the predictive distribution of the regression error $\epsilon_k(s)$, given observed data $\tilde{\boldsymbol{y}}$ and covariates $\boldsymbol{x}$, can be computed as follows:

$$p(\epsilon_k(s)|\tilde{\boldsymbol{y}},\boldsymbol{x}) = \int p(\epsilon_k(s),\epsilon_k(\tilde{s}_1),...,\epsilon_k(\tilde{s}_M),\boldsymbol{\upsilon}_k|\tilde{\boldsymbol{y}},\boldsymbol{x})d\epsilon_k(\tilde{s}_1)...d\epsilon_k(\tilde{s}_M)d\boldsymbol{\upsilon}_k \tag{13a}$$

$$= \int p(\epsilon_k(s)|\epsilon_k(\tilde{s}_1),...,\epsilon_k(\tilde{s}_M),\boldsymbol{\upsilon}_k,\tilde{\boldsymbol{y}},\boldsymbol{x})p(\epsilon_k(\tilde{s}_1),...,\epsilon_k(\tilde{s}_M),\boldsymbol{\upsilon}_k|\tilde{\boldsymbol{y}},\boldsymbol{x})d\epsilon_k(\tilde{s}_1)...d\epsilon_k(\tilde{s}_M)d\boldsymbol{\upsilon}_k \tag{13b}$$

The second term in equation (13b) is the posterior distribution of $(\epsilon_k(\tilde{s}_1),...,\epsilon_k(\tilde{s}_M),\boldsymbol{\upsilon}_k)$. It can therefore be directly approximated using the MCMC replicates. The first term represents the distribution of the regression error $\epsilon_k(s)$ for the target ungauged site s, conditional on regression errors at gauged sites $\epsilon_k(\tilde{s}_1)...\epsilon_k(\tilde{s}_M)$ and hyper-parameters $\boldsymbol{\upsilon}_k$. Following equation (5), the joint distribution of $\epsilon_k(s)$ and $\epsilon_k(\tilde{s}_1)...\epsilon_k(\tilde{s}_M)$ (conditional on $\boldsymbol{\upsilon}_k$) is a multivariate Gaussian distribution with mean zero and covariance matrix $\boldsymbol{\Gamma}_k = \boldsymbol{\Gamma}_k(\boldsymbol{\upsilon}_k)$ derived from the covariogram (6). Let us partition this covariance matrix as follows:

$$\boldsymbol{\Gamma}_k = \begin{pmatrix} \sigma_k^2 & \boldsymbol{\Lambda}_k \\ \boldsymbol{\Lambda}_k^t & \boldsymbol{\Omega}_k \end{pmatrix} \tag{14}$$

where σ_k^2 is the marginal variance of regression errors, $\boldsymbol{\Omega}_k = \boldsymbol{\Omega}_k(\boldsymbol{\upsilon}_k)$ is the $M \times M$ covariance matrix of regression errors at gauged sites $\epsilon_k(\tilde{s}_1)...\epsilon_k(\tilde{s}_M)$, and $\boldsymbol{\Lambda}_k = \boldsymbol{\Lambda}_k(\boldsymbol{\upsilon}_k)$

is the $1 \times M$ vector of covariances between $\epsilon_k(s)$ and $\epsilon_k(\tilde{s}_1)...\epsilon_k(\tilde{s}_M)$. Using a well-known formula for conditional Gaussian distributions, it follows that the conditional distribution in equation (13b) is a univariate Gaussian distribution with mean $\mu_{k,cond}$ and variance $\sigma^2_{k,cond}$:

$$
\begin{aligned}
&p(\epsilon_k(s)|\epsilon_k(\tilde{s}_1),...,\epsilon_k(\tilde{s}_M),\boldsymbol{\upsilon}_k,\tilde{\boldsymbol{y}},\boldsymbol{x}) = N(\mu_{k,cond};\sigma^2_{k,cond}) \\
&\mu_{k,cond} = \boldsymbol{\Lambda}_k\boldsymbol{\Omega}_k^{-1}(\epsilon_k(\tilde{s}_1),...,\epsilon_k(\tilde{s}_M))^t \\
&\sigma^2_{k,cond} = \sigma^2_k - \boldsymbol{\Lambda}_k\boldsymbol{\Omega}_k^{-1}\boldsymbol{\Lambda}_k^t
\end{aligned} \tag{15}
$$

The conditional distribution in equation (15) complements the regression to transfer information from gauged to ungauged sites. Indeed, the vector $\boldsymbol{\Lambda}_k$ acts as a weight vector favoring gauged sites nearby the target ungauged site s. If $\epsilon_k(s)$ is independent from all $\epsilon_k(\tilde{s}_j)$'s, the conditional distribution is equal to the marginal distribution of regression errors, i.e. a Gaussian distribution with mean $\mu_{k,cond} = 0$ and variance $\sigma^2_{k,cond} = \sigma^2_k$. In this case, the transfer of information to ungauged site s does not favor any particular gauged site $\tilde{s}$.

The explanations given above lead to the following algorithm for prediction at an ungauged site s:

Do $j = 1 : n_{sim}$

1. compute $\boldsymbol{\Omega}_k^{(j)} = \boldsymbol{\Omega}_k(\boldsymbol{\upsilon}_k^{(j)})$; $\boldsymbol{\Lambda}_k^{(j)} = \boldsymbol{\Lambda}_k(\boldsymbol{\upsilon}_k^{(j)})$; $\sigma_k^{2(j)} = \sigma_k^2(\boldsymbol{\upsilon}_k^{(j)})$.

2. compute $\mu_{k,cond}^{(j)}$ and $\sigma_{k,cond}^{(j)}$ according to equation (15).

3. generate regression error $\epsilon_k^{(j)}(s)$ from the conditional distribution (15) for $k = 1 : D$ (Gaussian distribution with mean $\mu_{k,cond}^{(j)}$ and standard deviation $\sigma_{k,cond}^{(j)}$).

4. compute D-parameters from the regression model (4), $\theta_k^{(j)}(s) = g_k^{-1}\left(h_k(\boldsymbol{x}_k(s), \boldsymbol{\beta}_k^{(j)}) + \epsilon_k^{(j)}(s)\right)$ for $k = 1 : D$

5. compute derived quantity $z^{(j)} = z(\boldsymbol{\theta}^{(j)}(s))$ or generate a value from the at-site distribution $w^{(j)} \sim p(\boldsymbol{\theta}^{(j)}(s))$

4. Case study: Mediterranean extreme rainfall

The application of the Bayesian hierarchical approach is illustrated with a case study involving extreme rainfall data. The main objectives of this application are to demonstrate the feasibility of a Bayesian hierarchical approach and to assess the impact of some modeling hypotheses (in particular the choices of the data dependence model and the regression model). Note that a synthetic case study was also performed to verify the internal consistency of the modeling framework (not shown).

4.1. Data

Annual maxima from 87 series of daily rainfall are used in this study (Figure 2). Raingauges, whose elevations range from 1 to 1102 m., are located in the French Mediterranean area. This region is characterized by intense rainfall in autumn and is delimited by three mountainous areas: the Pyreneans (South-West), the Alps (East) and the Cevennes (Center). The thin lines in Figure 2 represent six homogeneous regions defined by *Pujol et al.* [2007].

Sixty raingauges are used for estimation, the remaining 27 series being used for model validation. Data are available over the period 1955-2004. Years with more than 15 days of missing values are treated as missing data. Since the treatment of missing data during estimation is not obvious, the 60 estimation series were chosen to ensure the completeness

of annual maxima series over the period 1955-2004 (this point is discussed in more details in section 5.1).

4.2. Model specification

4.2.1. Data level

Let $(y(\tilde{s}_i,t))_{i=1:60,t=1:50}$ denote the observed annual maxima. At any site s, the random variable $(Y(s,t))$ is assumed to follow a Generalized Extreme Value (GEV) distribution with parameters $(\mu(s),\lambda(s),\xi(s))$. The pdf of a GEV distribution with location parameter μ, scale parameter λ and shape parameter ξ is:

$$p(y|\mu,\lambda,\xi) = \frac{1}{\lambda}\left[1-\xi\left(\frac{y-\mu}{\lambda}\right)\right]^{\frac{1}{\xi}-1} exp\left\{-\left[1-\xi\left(\frac{y-\mu}{\lambda}\right)\right]^{1/\xi}\right\}$$
$$\lambda>0, \xi\neq 0, 1-\xi\left(\frac{y-\mu}{\lambda}\right)>0 \quad (16)$$

The joint distribution of observations $\tilde{\boldsymbol{y}}(t)$ is derived using three distinct assumptions on the dependence structure. The first assumption corresponds to using a Gaussian copula [e.g., *Renard and Lang*, 2007] with dependence matrix $\boldsymbol{\Sigma}$:

$$(Y(\tilde{s}_1,t),...,Y(\tilde{s}_M,t)) \sim GCop_M(\boldsymbol{\Sigma}, \{\mu(\tilde{s}_i),\lambda(\tilde{s}_i),\xi(\tilde{s}_i)\}_{i=1:M}) \quad (17)$$

The second assumption corresponds to using a Student copula with dependence matrix $\boldsymbol{\Sigma}$ and tail dependence coefficient ν:

$$(Y(\tilde{s}_1,t),...,Y(\tilde{s}_M,t)) \sim SCop_M(\boldsymbol{\Sigma}, \nu, \{\mu(\tilde{s}_i),\lambda(\tilde{s}_i),\xi(\tilde{s}_i)\}_{i=1:M}) \quad (18)$$

Lastly, the third assumption corresponds to assuming spatial independence between data. It can be viewed as a special case of Equation (17) with the dependence matrix $\boldsymbol{\Sigma}$

forced to unity.

In both the Student and Gaussian copula cases, the dependogram is formulated as a weighted sum of two exponential functions (with $\psi_1 < \psi_2$) [e.g., *Kjeldsen and Jones*, 2009a]:

$$\Sigma(i,j) = \psi_0 exp(-\psi_1 \|\tilde{s}_i - \tilde{s}_j\|) + (1-\psi_0) exp(-\psi_2 \|\tilde{s}_i - \tilde{s}_j\|) \quad (19)$$

4.2.2. Process level

The variation of D-parameters $(\mu(s), \lambda(s), \xi(s))$ in space is described using several modeling assumptions. First, an index-flood-like approach, assuming scale invariance, is used. This assumption induces strong constraints on the parameters of at-site distributions. More precisely, a GEV distribution complying with the scale invariance assumption can be reparameterized as follows [e.g., *Ribatet et al.*, 2007]:

$$Y(s,t) \sim GEV(\delta(s)\mu, \delta(s)\lambda, \xi) \quad (20)$$

Equation (20) states that (i) the shape parameter is constant throughout the region; (ii) the ratio between the location and the scale parameters is constant throughout the region, which implies that annual maxima from all sites have the same coefficient of variation. In other terms, the whole spatial variability of annual maxima is accounted for by the index flood parameter $\delta(s)$, with other parameters μ, λ, ξ being assumed regional.

Two distinct regression models are used in this index flood approach. The first model, $\mathcal{M}_1$, does not use any covariate and simply describes the spatial variation of the index flood D-parameter $\delta(s)$ using a Gaussian spatial field:

$$log(\delta(\tilde{s}_i)) = \epsilon(\tilde{s}_i); (\epsilon(\tilde{s}_1), ..., \epsilon(\tilde{s}_M)) \sim N(0, \mathbf{\Gamma}) \quad (21)$$

The second model, $\mathcal{M}_2$, uses the raingauge elevation as covariate. Preliminary investigations indicate that the effect of elevation depends on the region. Consequently, the model uses region-specific relationships between the index flood D-parameter $\delta(\tilde{s}_i)$ and elevation:

$$log(\delta(\tilde{s}_i)) = \beta_0^{(j)} + \beta_1^{(j)} * elevation(\tilde{s}_i) + \epsilon(\tilde{s}_i)$$
$$(\epsilon(\tilde{s}_1), ..., \epsilon(\tilde{s}_M)) \sim N(0, \mathbf{\Gamma}) \quad (22)$$

where the superscript (j) is used to denote the region of site $\tilde{s}_i$. In order to ensure the identifiability of the model, the following additional constraint is applied:

$$\sum_{j=1}^{6} \beta_0^{(j)} = 0 \quad (23)$$

In a second step, the index-flood approach described above is made less restrictive by abandoning the scale invariance assumption. More precisely, the following at-site distribution is assumed:

$$Y(s,t) \sim GEV(\mu(s), \lambda(s), \xi) \quad (24)$$

Compared to the index-flood equation (20), equation (24) does not assume a constant ratio between the location and the scale parameters throughout the region. However, it still assumes a constant shape parameter. The spatial variation of the location $\mu(s)$ and scale $\lambda(s)$ D-parameters is described using region-specific relationships with elevation:

$$log(\mu(\tilde{s}_i)) = \beta_0^{(j)} + \beta_1^{(j)} * elevation(\tilde{s}_i) + \epsilon_\mu(\tilde{s}_i)$$

$$(\epsilon_\mu(\tilde{s}_1), ..., \epsilon_\mu(\tilde{s}_M)) \sim N(0, \mathbf{\Gamma}_\mu)$$

$$log(\lambda(\tilde{s}_i)) = \alpha_0^{(j)} + \alpha_1^{(j)} * elevation(\tilde{s}_i) + \epsilon_\lambda(\tilde{s}_i)$$

$$(\epsilon_\lambda(\tilde{s}_1), ..., \epsilon_\lambda(\tilde{s}_M)) \sim N(0, \mathbf{\Gamma}_\lambda) \quad (25)$$

The model defined by equations (24)-(25) is noted $\mathcal{M}_3$.

In all models $\mathcal{M}_1$-$\mathcal{M}_3$, the covariance matrixes $\mathbf{\Gamma}$ in equations (21), (22) and (25) are parameterized as follows:

$$\Gamma(i,j) = \sigma^2 \left[\upsilon_0 exp(-\upsilon_1 \left\|\tilde{s}_i - \tilde{s}_j\right\|) + (1-\upsilon_0) exp(-\upsilon_2 \left\|\tilde{s}_i - \tilde{s}_j\right\|)\right] \quad (26)$$

Finally, vague priors are specified for all parameters by using uniform distributions with large support. The only exception is the shape parameter, for which a Gaussian prior with mean zero and standard deviation 0.3 is specified: this is similar to the "Geophysical prior" used by *Martins and Stedinger* [2000]. Prior specification is further discussed in section 5.6.

4.3. Estimation and prediction

In this section, inference is performed using the regression model $\mathcal{M}_3$ coupled with a Gaussian copula assumption for describing data dependence (equation (17)). This model is noted $\mathcal{M}_3$-GCop.

4.3.1. Parameter estimates

Figure 3a shows the posterior pdf of the regional shape parameter ξ. The median value of about -0.12 corresponds to an heavier tail than the Gumbel distribution: this is consistent with previous studies in the same area [*Neppel et al.*, 2007]. Moreover, this estimation is quite precise: this is a consequence of the hypothesis that this parameter is constant throughout the region.

Figure 3b-c shows the posterior pdfs of a few regression errors, for both location (b) and scale (c) regressions (*latent variables* in the vocabulary of hierarchical modeling). Recall that those errors are assumed mutually independent in the inference framework (see section 3). This assumption can be evaluated here by computing the correlation between series of estimated regression errors $\hat{\epsilon}_\mu(\tilde{s}_1), ..., \hat{\epsilon}_\mu(\tilde{s}_M)$ and $\hat{\epsilon}_\lambda(\tilde{s}_1), ..., \hat{\epsilon}_\lambda(\tilde{s}_M)$ (maximum-posterior estimates are used). A correlation of about 0.52 is found, which suggests that the assumption that both error processes are not cross-correlated might not be realistic. This is further discussed in section 5.3.

The standard deviations of regression errors are represented in Figure 3d. Those terms (*hyper-parameters* in the vocabulary of hierarchical modeling) are of primary importance since they control the predictive uncertainty at ungauged sites. In this case study, the estimated hyper-standard deviations are similar for both location and scale regressions. They roughly correspond to a standard error of 15% in the regressions.

4.3.2. Data and parameter dependences

Figure 4 shows the estimated dependence-distance relationships, for both the data dependence model (19) and the regression errors dependence model (26). The dependence between data (Figure 4a) is precisely estimated, and confirm that data are not spatially independent. Conversely, the dependence between regression errors (Figure 4b-c) appears

more difficult to estimate, with large posterior intervals denoting a lower precision. However, the strength of dependence is rather limited here, with the correlograms dropping to near-zero values at relatively short distances of about 10-20 km.

4.3.3. Prediction

The 100-year daily rainfall is predicted on a 40*50 grid (yielding 10*10 km cells) using the procedure detailed in section 3.4. Since the highest raingauge has an elevation of only 1102 m, high-elevation areas from the Alps and the Pyreneans (whose maximal elevations reach 4810 m and 3404 m, respectively) were excluded from the prediction (white areas in Figure 5). The left panel of Figure 5 shows the 100-year daily rainfall estimated on the grid using the posterior median. An area with higher $R_{0.99}$ values (reaching 400 mm at some grid points) is located on the Cevennes mountain range: this area is well-known to be affected by the highest rainfall intensities in France. Moreover, the rainfall quantile abruptly drops to smaller values ($\approx$ 100-150 mm) downwind of the Cevennes.

The Bayesian framework used in this paper enables a direct assessment of the uncertainties affecting predictions. The right panel of Figure 5 shows the uncertainty in predicted 100-year daily rainfall, measured by the posterior coefficient of variation. This uncertainty appears relatively uniform over the region, and mostly corresponds to 15-20% coefficients of variations. Note however the high-uncertainty area appearing on the foothill of the Alps (East): this is due to the low number of calibration sites in this region (only three sites), which does not enable a precise estimation of the elevation effect.

4.4. Impact of the data dependence model

In this section, the impact of the model used to describe spatial dependence between data is evaluated. To this aim, inference is performed with the three dependence models

591 described in section 4.2.1, coupled with the regression model $\mathcal{M}_3$. The three resulting
592 models are noted $\mathcal{M}_3$-Inde (independence assumption), $\mathcal{M}_3$-GCop (Gaussian copula as-
593 sumption) and $\mathcal{M}_3$-SCop (Student copula assumption).

594 4.4.1. Parameter estimates

595 The regional shape parameter is only moderately impacted by the data dependence
596 model (Figure 6a). The posterior pdf of model $\mathcal{M}_3$-SCop is slightly shifted and is more
597 variable compared with models $\mathcal{M}_3$-Inde and $\mathcal{M}_3$-GCop. A similar observation can be
598 made for the hyper-standard deviation of location regression errors (Figure 6b). Posterior
599 pdfs for the hyper-standard deviation of scale regression errors (Figure 6c) have a similar
600 variance but a slightly different mode.

601 4.4.2. Data and parameter dependences

602 Figure 7a shows the estimated dependence between data. The dependogram drops
603 immediately to zero for $\mathcal{M}_3$-Inde due to the spatial independence assumption. Dependo-
604 grams for $\mathcal{M}_3$-GCop and $\mathcal{M}_3$-SCop are virtually indistinguishable. However, the Student
605 copula also depends of an additional parameter controlling the strength of asymptotic
606 dependence. The posterior pdf of this parameter is shown in Figure 7a, and corresponds
607 to a rather limited tail dependence.

608 Stronger differences are observed for the dependence between regression errors (Figure
609 7b-c). In particular, the dependogram for $\mathcal{M}_3$-GCop drops faster to zero for location
610 regression errors (Figure 7b). Moreover, the choice of the data dependence model appears
611 to impact the uncertainty in the estimation of the regression errors correlograms (for both
612 location and scale regressions, Figure 7b-c): $\mathcal{M}_3$-Inde yields wider posterior intervals than
613 $\mathcal{M}_3$-SCop and $\mathcal{M}_3$-GCop.

4.4.3. Prediction

Results described in previous sections 4.4.1 and 4.4.2 show that the impact of the model for data dependence is quite complex. A more integrated assessment can be made by comparing the prediction of 100-year quantiles obtained with the three models. Figure 8 maps the relative difference (in percent) between (i) $\mathcal{M}_3$-Inde (top panels) and $\mathcal{M}_3$-SCop (bottom panels) estimates in the one hand; and (ii) $\mathcal{M}_3$-GCop estimates in the other hand. The Gaussian Copula is therefore considered as a benchmark model, and the two other dependence assumptions are compared to this benchmark.

Overall, assuming spatial independence between data yields only minor change in $R_{0.99}$ estimates (top left panel). Exceptions are located in the Cevennes ($R_{0.99}$ values are $\approx$ 10-15% higher), and on the foothill of the Alps ($R_{0.99}$ values are $\approx$ 10-35% higher in the North, and $\approx$ 10-35% lower in the South). Simarly, changes in the uncertainty of $R_{0.99}$ estimates are minor (top right panel). However, uncertainty reductions dominate, with decreases in the range 0-25%. This is consistent with the expected behavior when spatial dependence is ignored (see discussion in section 2.1.2): the variance of estimates may be underestimated, i.e. this decrease in uncertainty may be unduly optimistic. Note that once again, the foothill of the Alps is an exception to this overall decrease.

Overall, replacing the Gaussian copula model by a Student copula also yields minor change in $R_{0.99}$ estimates (bottom left panel), except in the Alps region where strong increases are observed ($\approx$ 40%, culminating at $\approx$ 100% for a couple of pixels). In terms of uncertainty (bottom right panel), increases now dominate, but are still moderate (in the range 0-20%), with the Alps showing an opposite behavior (uncertainty reductions in the range 20-40%).

4.5. Impact of the regression model

In this section, the impact of the regression model is evaluated. Inference is performed with the three regression models described in section 4.2.2, coupled with the Gaussian copula model for data dependence. The three resulting models are noted $\mathcal{M}_1$-GCop (scale invariance assumption, no elevation effect), $\mathcal{M}_2$-GCop (scale invariance assumption with elevation effect) and $\mathcal{M}_3$-GCop (elevation effect, scale invariance is not assumed).

4.5.1. Parameter estimates

The assumption of scale invariance notably impacts the shape parameter estimates (Figure 9a), with a marked shift between scale-invariant models $\mathcal{M}_1$-GCop and $\mathcal{M}_2$-GCop in the one hand, and model $\mathcal{M}_3$-GCop in the other hand. Conversely, the completeness of the regression model seems to be the primary factor of influence for the hyper-standard deviations (Figure 9b-c): model $\mathcal{M}_1$-GCop, which ignores the elevation effect, yields a markedly higher hyper-standard deviation than models $\mathcal{M}_2$-GCop and $\mathcal{M}_3$-GCop, which use elevation as covariate. This is an expected result, since improving the regression model results in decreasing the standard deviations of regression errors. Note that in Figure 9b-c, the posterior pdfs for $\mathcal{M}_1$-GCop and $\mathcal{M}_2$-GCop are related to the regression for the index flood parameter $\delta(s)$, as defined in equation (21). Since this regression acts on both the location and scale D-parameters (see Equation (20)), those pdfs are repeated in both panels b and c.

4.5.2. Data and parameter dependences

Figure 10a shows that data dependence is virtually identical with all regression models. This is consequence of using a copula formalism, with dependence being modeled independently of marginal distributions. Since the regression model only applies on marginal distributions, it does not impact the estimation of dependence between observed data.

Conversely, the dependence between regression errors is strongly impacted by the choice of a regression model (Figure 10b-c). The amount of dependence is markedly higher with simple model $\mathcal{M}_1$-GCop because the whole spatial variability in the distribution of annual maxima is accounted for by the regression errors. Models $\mathcal{M}_2$-GCop and $\mathcal{M}_3$-GCop yield lower dependences, suggesting that the elevation effect explains a major part of the spatial variability. This result is consistent with the findings of *Kjeldsen and Jones* [2009a] who observed a decrease in the dependence between regression errors when the regression model is improved. Lastly, model $\mathcal{M}_3$-GCop yield higher uncertainties than model $\mathcal{M}_2$-GCop. This might be due to the estimation of two distinct regression models for location and scale D-parameters in model $\mathcal{M}_3$-GCop, while model $\mathcal{M}_2$-GCop uses a single regression acting on the index flood parameter $\delta(s)$.

4.5.3. Prediction

Maps of 100-year quantiles are compared in a similar way to section 4.4.3. Figure 11 maps the relative difference (in percent) between (i) $\mathcal{M}_1$-Gcop (top panels) and $\mathcal{M}_2$-GCop (bottom panels) estimates in the one hand; and (ii) $\mathcal{M}_3$-GCop estimates in the other hand. The regression model $\mathcal{M}_3$ is therefore considered as a benchmark model, and the two other regression models are compared to this benchmark.

The simple index-flood model $\mathcal{M}_1$ (which ignores the elevation effect) yields marked differences with the benchmark model $\mathcal{M}_3$ in terms of $R_{0.99}$ values (upper left panel): higher values are observed in the Cevennes (up to $\approx +100\%$) and in the Alps (up to $\approx +30\%$), while lower values are observed in the south-eastern corner of the domain (up to $\approx$ -50%). Marked differences also appear for the uncertainty in $R_{0.99}$ estimates (upper right panel), with an overall increase exceeding $\approx 50\%$, except in the Alps where the uncer-

tainty decreases (up to $\approx 50\%$). Note however that the difference between $\mathcal{M}_1$ and $\mathcal{M}_3$ uncertainties is not uniform and follows and interesting spatial pattern: uncertainties are similar for both models nearby calibration sites, while $\mathcal{M}_1$ estimates become far more uncertain than $\mathcal{M}_3$ estimates in poorly-gauged areas. This is a consequence of regression errors showing a significant amount of spatial dependence with model $\mathcal{M}_1$ (see section 4.5.2 and Figure 10b-c). This spatial dependence improves the efficiency of the tranfer of information from gauged sites to nearby ungauged sites. However, when moving further away from the calibration sites, this dependence vanishes and the uncertainty becomes primarily controlled by the hyper-standard deviation of regression errors. This hyper-standard deviation is markedly higher for model $\mathcal{M}_1$ (see section 4.5.1 and Figure 9b-c), which explains its higher uncertainty in poorly-gauged areas.

Qualitatively similar results are found for the index-flood model $\mathcal{M}_2$ (which includes an elevation effect). However, differences with the model $\mathcal{M}_3$ in terms of $R_{0.99}$ values are smaller (in the range $\approx \pm 20\%$, lower left panel). The uncertainty in $R_{0.99}$ estimates is also larger with model $\mathcal{M}_2$ than with model $\mathcal{M}_3$ (lower right panel), with increases in the range $\approx$ 30-40% being quite evenly distributed in space.

4.5.4. Assessment of the scale invariance assumption

The scale invariance assumption underlying models $\mathcal{M}_1$ and $\mathcal{M}_2$ can be appraised by computing the coefficients of variation (CV) of data from each site. As explained in section 4.2.2, scale invariance implies that the CV remains constant throughout the studied area. Figure 12 shows the empirical CV computed on each site (sites are reordered by increasing CV). The dashed horizontal lines show a 90% posterior interval of the CV resulting from model $\mathcal{M}_2$-GCop. Numerous empirical CVs are well outside this interval, which casts

doubts on the validity of the scale invariance assumption. Conversely, model $\mathcal{M}_3$-GCop (vertical bars) is able to track more closely the varying CVs at different sites.

4.6. Validation

The validation of predictions arising from the Bayesian hierarchical framework is performed by using tools and concepts borrowed from the field of probabilistic forecasting. Such tools are of particular interest when the quantification of uncertainty is a primary concern, which is the case in this paper. In particular, the reliability and the precision of predictions are evaluated: reliability refers to the ability to derive predictive distributions that are consistent with observations from validation sites, while precision refers to the amount of uncertainty in predictions. Both concepts are complementary: predictions can be reliable but not precise (uncertainties are large, but predictions remain consistent with observations) or alternatively precise but not reliable (e.g. due to an underestimation of uncertainties). An ideal predictive framework would yield predictions that are reliable and as precise as possible.

4.6.1. Reliability of predictions

The reliability of predictions can be assessed by comparing annual maxima from the 27 validation sites with their predictive distribution (derived using the "ungauged site" procedure described in section 3.4). More precisely, this comparison is based on the predictive QQ-plot used by e.g. *Dawid* [1984], *Gneiting et al.* [2007], *Laio and Tamea* [2007] or *Thyer et al.* [2009]. Let F_s be the cdf of the predictive distribution at site s, and $\{Y(s,t)\}_{t=1:T_s}$ validation data at site s. Under the assumption that validation data are realizations from the predictive distribution, the p-values $\{F_s(Y(s,t))\}_{t=1:T_s}$ are realizations from a uniform distribution on [0;1]. This can be evaluated with a QQ-plot

comparing the empirical cdf of p-values with the cdf of a uniform distribution. This comparison can be performed individually for each validation site, or by pooling p-values from all validation sites.

Figure 13 shows the site-specific predictive QQ-plots for each model (first five panels), then compares the pooled-sites-predictive QQ-plots of the five models in a single plot (lower right panel). Overall, the reliability of the predictive distributions appears acceptable with all regression models, although some specific sites show signs of departure from the 1:1 line. Moreover, when all sites are pooled together, all models yield nearly-diagonal pp-plots (lower right panel), suggesting that all methods provide equally reliable predictions. This result might appear surprising at first sight, given the differences between models highlighted in previous sections 4.4 and 4.5. However, there is no contradiction in this statement: in the framework of probabilistic prediction, several predictive distributions can be equally reliable but distinct from each other. In particular, the fact that the overall reliability is similar does not mean that, for a given site, the predictive distributions will themselves be similar. This is illustrated in Figure 14, where the model $\mathcal{M}_2$-GCop yields a predictive distribution markedly different from the other four models. Moreover, the probabilistic predictions may also differ in their precision (also termed "sharpness" in the field of probabilistic prediction [see *Gneiting et al.*, 2007, for further discussion]).

4.6.2. Precision of predictions

Given that all models were found equally reliable, one would favor the one yielding the most precise predictions. In order to evaluate the predictive precision, Figure 15 shows the posterior coefficient of variation of the estimated 100-year rainfall $R_{0.99}$, as a function of the distance between the validation site and the nearest calibration site. In general,

index-flood models $\mathcal{M}_1$-GCop and $\mathcal{M}_2$-GCop appear less precise than models based on the regression $\mathcal{M}_3$. Also note that model $\mathcal{M}_2$-GCop yields very imprecise predictions (CV>0.3) for three validation sites, which are all located in the same region (South-Eastern Mediterranean coast).

Moreover, an apparent trend suggests that the $\mathcal{M}_1$-predictions are more precise for validation sites located close to a calibration site than for isolated sites. For this model, Figure 15 suggests that if a calibration site is available at about 10 km, the predictive uncertainty is divided by almost two compared with isolated sites. This confirms an observation previously made in section 4.5.3, and illustrates a positive outcome of modeling spatial dependence between regression errors. On the other hand, Figure 15 also shows that when the regression model is improved ($\mathcal{M}_3$), the predictive precision is as good as model $\mathcal{M}_1$ nearby calibration sites, despite the fact that spatial dependence between regression errors is negligible (see section 4.5.2). Moreover, $\mathcal{M}_3$-predictions are more precise for isolated sites. This important observation suggests that improving the regression to make errors as spatially independent as possible is the most sensible strategy.

5. Discussion

The case study described in section 4 demonstrates the feasibility of a Bayesian hierarchical approach to regional frequency analysis. Despite this encouraging preliminary investigation, numerous avenues for improvement can already be identified. This section discusses the main issues to be addressed.

5.1. Treatment of missing data

In the case study of section 4, calibration sites were selected in order to avoid missing values in the calibration dataset. This was made for the sake of simplicity, but it cannot be considered as an acceptable practical solution. Indeed, most regional analyses make use of at-site data with different lengths. Restricting the calibration dataset to years shared by all sites results in a loss of information. Moreover, data from poorly-gauged sites still yield a substantial information on the at-site distribution. Combining this information with the regional information transfered from nearby sites is likely to improve the inference.

The treatment of missing data is challenging due to the explicit modeling of spatial dependence between data. Indeed, the likelihood equation (9) requires writing the joint distribution of a (spatial) vector $\tilde{\boldsymbol{y}}(t)$ at any time t. A first possibility would be to consider that the size of vector $\tilde{\boldsymbol{y}}(t)$ varies with t, depending on the available data. Unfortunately, this complicates the implementation of the model and increases its computational cost, due to the manipulation and inversion of the dependence matrix $\boldsymbol{\Sigma}$, whose dimension would also vary with t in this case. An alternative and possibly more efficient approach would be to consider models for missing data and data augmentation algorithms [see e.g. *Gelman et al.*, 1995].

5.2. Sensitivity to the model for data dependence

As stressed in section 2.1.2, the use of an elliptical copula to model data dependence corresponds to a parametric assumption on the structure of dependence. As such, it might be an inappropriate model for some data, especially in the case of extreme data [*Mikosch*, 2005]. This is a crucial issue if the dependence model is to be used for estimating rare multivariate events (e.g., Pr(annual maxima from all raingauges larger than 100 mm)).

The literature suggests that such estimations are highly sensitive to the choice of the dependence model [e.g. *Coles et al.*, 1999; *Renard and Lang*, 2007]. However, the impact of the dependence model on the estimation at ungauged site is not as clear. Indeed, prediction at ungauged site does not directly use the data dependence model (see section 3.4). Yet, the latter may indirectly influence predictions when inappropriate, by inducing bias in parameter estimates and/or inadequate quantification of uncertainties (posterior variance).

The preliminary investigation carried out in section 4.4 suggests that the choice of a data dependence model is of second order importance compared with the choice of an appropriate regression model to describe spatial variability. However, this result is based on a single case study and can therefore not be considered as a generality. Consequently, the sensitivity of predictions to the data dependence model needs to be further evaluated. In particular, the impact of the spatial dependence model might be more pronounced for datasets showing a higher level of asymptotic dependence. Including multivariate extreme models recently proposed in the literature [e.g. *Keef et al.*, 2009; *Padoan et al.*, 2010] would also constitute an improvement.

5.3. The role of spatial- and cross-dependence between regression errors

The framework developed in this paper follows the recent work by *Kjeldsen and Jones* [2009a] in assuming that regression errors are spatially dependent. The case study of section 4 illustrates the benefit of this assumption: when the regression model is poor and fails to capture key relationships with spatial covariates (model $\mathcal{M}_1$), spatial dependence between regression errors is significant and improves the predictive precision nearby calibration sites. However, this spatial dependence quickly vanishes when the regression

model is improved (models $\mathcal{M}_2$ and $\mathcal{M}_3$). This illustrates the relationship between the quality of the regression and the existence of spatial dependence in regression errors: in a nutshell, spatial dependence acts as a surrogate for the spatial variability "missed" by the regression model. This suggests that improving the regression to make regression errors as spatially independent as possible is a better strategy than attempting to refine the description of spatial dependence. However, in cases where no satisfying regression can be uncovered, the explicit modeling of spatial dependence appears beneficial in terms of predictive precision.

Moreover, cross-correlation between regression errors related to distinct D-parameters is neglected in this framework. Unfortunately, the case study suggests that such cross-correlation exists (see section 4.3.1). Its impact on predictions is unclear at this stage and requires further evaluation.

An explicit modeling of cross-correlation was adopted by *Tasker and Stedinger* [1989], in the case of spatially independent regression errors. This paper somehow considers the opposite option: spatial dependence between regression errors is modeled, but their cross-correlation is ignored. Which of these two options should be favored is likely case-specific, and depends on the relative strength of spatial- and cross-correlation. Moreover, it is worth noting that both types of dependence can be constrained (to some extent) by improving the model. In the one hand, improving the regression model decreases spatial dependence as discussed above. In the other hand, cross-correlation can be decreased by reparameterizing the marginal distributions: for instance, the GEV distribution could be parameterized in terms of location, CV and shape parameters (instead of location, scale and shape, see the discussion in *Stedinger and Griffis* [2011]), which may limit cross-

correlation between location and scale regression errors. Which of these two strategies offers the most flexibility remains an open question, that will be addressed in future work. Lastly, the third strategy would be to explicitly model both spatial- and cross-correlation in regression errors. However, this would require a description of the dependence between spatial random fields, which is a challenging task. Solutions might exist and be borrowed from the field of geostatistics [e.g. using cokriging tools, see *Goovaerts*, 1998]. However, improving the model to make either spatial- or cross-correlation negligible seems easier to implement and might be an acceptable solution in many cases.

5.4. Validation procedures

As any predictive statistical model, a model constructed within the hierarchical framework presented in this paper needs to be thoroughly validated based on data that were not used for estimation. In the case study of section 4, this was achieved by comparing the predictive distribution (derived in an ungauged site context) with validation data. Such a comparison assesses the overall reliability of the predictive distribution, and may detect systematic predictive biases (e.g. the predictive mean is significantly smaller that the mean of validation data). However, it might not be sufficient to assess the reliability of predicted extreme quantiles, which are of primary interest in frequency analysis. Consequently, more specialized validation procedures are needed to open predicted extreme quantiles to direct validation and scrutiny. The validation tools recently proposed by *Garavaglia et al.* [2011] might be of particular interest in this respect.

5.5. Model complexity and identifiability

The case study of section 4 compares several models differing in their flexibility (and, as a consequence, in their complexity). In particular, models $\mathcal{M}_1$ and $\mathcal{M}_2$ are based on the scale invariance assumption. This limits the complexity of the model by merging all spatial variability into a single D-parameter (the index-flood parameter $\delta(s)$), the other D-parameters being assumed constant over the region. However, this assumption seems unrealistic in the studied region. The alternative model $\mathcal{M}_3$ describes the spatial variations in two D-parameters, namely the location and the scale parameters, while the third D-parameter (shape) remains constant. This is similar to the approach evaluated by *Stedinger and Lu* [1995], and it was found to yield reliable and more precise predictions in this case study. Consequently, it would be tempting to go even further in terms of model complexity, by using three distinct regression models for the location, scale and shape parameters.

However, as in any statistical model, a trade-off between descriptive and predictive power has to be found. Model complexity may come at the cost of reduced predictive ability. Moreover, non-identifiability issues may arise if the information content of the data does not support the inference of several spatial processes. Establishing an acceptable trade-off between the flexibility of the model and the level of complexity that can be identified from the data is a very challenging task. In particular, it is linked with the preceding discussion on validation procedures: introducing additional complexity is only beneficial if it demonstrably improves predictive reliability and/or precision, yet the power of validation procedures to detect model failures may be too limited, especially for extreme quantiles. Consequently, in the case study presented in this paper, developing more efficient validation tools seems to be a prerequisite before attempting to model spatial variations in the

shape parameter.

Lastly, it is stressed that the high dimensionality of models resulting from the hierarchical framework presented herein does not necessarily lead to over-parameterized or non-identifiable models. In the context of Bayesian hierarchical modeling, the issue of model complexity is far from obvious [see *Spiegelhalter et al.*, 2002, for a thorough discussion on this topic] and requires specialized criteria for model selection [e.g. *Cooley et al.*, 2007].

5.6. Prior specification

The case study of section 4 uses mostly non-informative priors. This is because the aim of this case study is to understand some properties of the proposed inference approach, rather than to seek an optimal estimation of rainfall extremes in this particular region. In this context, using precise priors might exert a strong, case-specific leverage on the conclusions. However, the specification of precise and accurate priors might indeed be beneficial to the inference. In particular, the following investigations would be of interest:

- Prior specification depends on the particular choices of marginal distributions and regression functions. It is therefore difficult to derive general specification guidelines. In particular, the fact that D-parameters are not inferred directly, but are only defined through the regression function (4) complicates prior specification. Methods for expressing prior knowledge (either based on expertise or on external data not used for inference) in the form of a proper prior distribution need to be developed.

- Using conjugate priors has the potential to ease computations by replacing (at least partly) MCMC sampling by explicit formulas. However, the choice of a conjugate family depends on the choice of marginal distributions and regression functions, making it difficult to provide general guidelines. The benefit of using conjugate priors could be

first evaluated for the hyper-parameters of the Gaussian hyper-distributions, for which conjugate families are known [e.g. *Gelman et al.*, 1995].

- The sensitivity of the inference to priors should be further evaluated, especially for complex models that are more prone to identifiability issues (see discussion in section 5.5).

5.7. Modeling variability in time

Since this paper mainly focuses on spatial variability, at-site distributions were assumed to be constant in time (see equation (1)). A natural extension of the framework would be to allow temporal variations in at-site distributions, by including covariates varying with time. The simplest application would be to include a linear trend in one (or possibly several) D-parameter(s). The covariate could be in this case the time itself, or a large-scale climatic index (e.g. NAO or PDO indexes). This would be an important development since the strong natural variability of at-site data (in particular, extreme data) limits the detectability of climatic effects. Studying the impact of climate variability/change at a regional scale is likely to improve this assessment [*Renard et al.*, 2008; *Aryal et al.*, 2009].

5.8. Application to runoff or other hydrological variables

Although the hierarchical approach presented in this paper can in principle be applied to runoff data, it is likely to yield sub-optimal predictions if standard dependence models are used. This is because runoff data are structured by the hydrologic network. Consequently, euclidean distances (e.g. between gauging stations or between catchment centroids) are in general a sub-optimal predictor of intersite dependences. Moreover, predictions should be constrained to ensure the consistency between estimates at upstream and downstream nested catchments. Several authors proposed specialized approaches to model network

dependences [e.g. *Gottschalk*, 1993a, b; *Sauquet et al.*, 2000; *Sauquet*, 2006; *Skoien et al.*, 2006]. The applicability of such approaches within the Bayesian hierarchical framework presented in this paper will be evaluated in future work.

More generally, the choice of a relevant distance to explain spatial dependences is a challenging task, but is also a promising avenue to improve predictions. This choice heavily depends on the spatial properties of the hydrological variable. For rainfall, a xy-Euclidean distance (as used in the case study of section 4) may seem reasonable at first sight, although accounting for elevation may improve predictions. Alternatively, consider applying the hierarchical approach presented in this paper to snow depth values: a meaningful dependence-distance relationship is very unlikely to be derived if elevation is neglected in the definition of the inter-site distance. Recent work by Blanchet et al. [*Blanchet et al.*, 2009; *Blanchet and Lehning*, 2010; *Blanchet and Davison*, 2011] addresses this issue, and proposes practical solutions to derive meaningful distances. Those solutions might be extended to hydrological variables other than snow depth.

6. Conclusion

Regional frequency analysis has a long history in Hydrology, and is widely applied in practice to estimate the distribution of a hydrologic variable at ungauged or poorly gauged sites. However, despite numerous methodological developments over the years [e.g. *Stedinger*, 1983; *Stedinger and Tasker*, 1985, 1986; *Hosking and Wallis*, 1988; *Madsen and Rosbjerg*, 1997; *Reis et al.*, 2005; *Ribatet et al.*, 2007; *Kjeldsen and Jones*, 2009a], most implementations still rely on several hypotheses that complicate the quantification of predictive uncertainty; moreover, unrealistic hypotheses may yield unreliable predictions. The objective of this paper was therefore to propose a general Bayesian hierarchical

framework to overcome some of these limitations, and to evaluate its applicability based on a case study. This framework builds on previous work by several authors [in particular, *Wikle et al.*, 1998; *Perreault*, 2000; *Micevski et al.*, 2006; *Micevski*, 2007; *Cooley et al.*, 2007; *Lima and Lall*, 2009; *Aryal et al.*, 2009; *Lima and Lall*, 2010], who explored the usefulness of spatial and temporal Bayesian hierarchical models. The main features of the proposed framework are the following:

1. At-site data can be modeled with any distribution, with the sole requirement that the same distribution (but with different parameters) is used for all sites.

2. Intersite dependence is explicitly modeled by means of an elliptical copula.

3. The variation of parameters in space is described with a regression model linking parameter values and covariates.

4. Regression errors are modeled by means of a Gaussian spatial field, which allows transferring estimates from gauged to ungauged sites, while quantifying the associated predictive uncertainty.

A case study based on extreme rainfall data in Mediterranean France demonstrated the applicability of the framework and its reliable estimation of predictive uncertainty. Although numerous improvements remain to be implemented (see discussion section), these encouraging results warrant further research to develop this framework. In particular, the inclusion of temporal covariates would allow describing the evolution of hydrologic variables in both space and time. Importantly, this improved understanding of hydrologic variability would be achieved with a rigorous quantification of the associated uncertainties.

Appendix A: The Gaussian and Student Elliptical copulas

964 The following notation is used:

965 1. $F_1(y), ..., F_H(y)$ are the marginal cdfs of a H-dimensional random vector $(Y_1, ..., Y_H)$.

966 2. $f_1(y), ..., f_H(y)$ are the corresponding marginal pdfs.

967 3. $\phi(y)$ is the cdf of the standard normal distribution.

968 4. $\psi(y)$ is the corresponding pdf.

969 5. $\tau_\nu(y)$ is the cdf of the Student's t distribution with ν degrees of freedom

970 6. $t_\nu(y)$ is the corresponding pdf

971 7. $\Phi_\Sigma(y_1, ..., y_H)$ is the joint cdf of a H-dimensional Gaussian distribution, with mean
972 0 and covariance matrix Σ

973 8. $\Psi_\Sigma(y_1, ..., y_H)$ is the corresponding joint pdf

974 9. $\Theta_{\Sigma,\nu}(y_1, ..., y_H)$ is the joint cdf of a H-dimensional Student distribution, with mean
975 0, covariance matrix Σ and ν degrees of freedom

976 10. $T_{\Sigma,\nu}(y_1, ..., y_H)$ is the corresponding joint pdf

977 The Gaussian and the Student copulas build the joint cdf of the random vector
978 $(Y_1, ..., Y_H)$ as follows:

$$F_{Gaussian}(y_1, ..., y_H) = \Phi_\Sigma(u_1, ..., u_H) \tag{A1}$$

$$F_{Student}(y_1, ..., y_H) = \Theta_{\Sigma,\nu}(v_1, ..., v_H) \tag{A2}$$

979 where $u_i = \phi^{-1}(F_i(y_i))$ and $v_i = \tau_\nu^{-1}(F_i(y_i))$.

980 The corresponding joint pdfs can be obtained by differentiating the above cdfs:

$$f_{Gaussian}(y_1, ..., y_H) = \frac{(\prod_{i=1}^{H} f_i(y_i))}{(\prod_{i=1}^{H} \psi(u_i))} \times \Psi_\Sigma(u_1, ..., u_H) \tag{A3}$$

$$f_{Student}(y_1, ..., y_H) = \frac{(\prod_{i=1}^{H} f_i(y_i))}{(\prod_{i=1}^{H} t_\nu(v_i))} \times T_{\Sigma,\nu}(v_1, ..., v_H) \quad \text{(A4)}$$

Acknowledgments. The suggestions by Jery Stedinger, two anonymous reviewers and the associate editor significantly improved the content of this paper and are gratefully acknowledged. This work is funded by the ANR research project ExtraFlo. Meteo France is gratefully acknowledged for providing the data. Many thanks to Luc Neppel for his help.

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, *Journal of Hydrometeorology*, *10*(1), 241–253, doi:10.1175/2008JHM1007.1.

Blanchet, J., and A. C. Davison (2011), Spatial modelling of extreme snow depth, *Annals of Applied Statistics. In Press.*

Blanchet, J., and M. Lehning (2010), Mapping snow depth return levels: smooth spatial modeling versus station interpolation, *Hydrology and Earth System Sciences*, *14*(12), 2527–2544, doi:10.5194/hess-14-2527-2010.

Blanchet, J., C. Marty, and M. Lehning (2009), Extreme value statistics of snowfall in the swiss alpine region, *Water Resources Research*, *45*.

Burn, D. (1990), Evaluation of regional flood frequency analysis with a region of influence approach, *Water Resources Research*, *26*(10), 2257–2265.

Casson, E., and S. Coles (1998), Extreme hurricane wind speeds: Estimation, extrapolation and spatial smoothing, *Journal of Wind Engineering and Industrial Aerodynamics*, *74-76*, 131–140.

Casson, E., and S. Coles (2000), Simulation and extremal analysis of hurricane events, *Applied Statistics*, *49*, 227–245.

Chiles, J.-P., and P. Delfiner (1999), *Geostatistics: Modeling Spatial Uncertainty*, Wiley Series in Probability and Statistics.

Clark, J. (2005), Why environmental scientists are becoming bayesians, *Ecology letters*, *8*, 2–14.

Clark, J. S. (2003), Uncertainty and variability in demography and population growth: A hierarchical approach, *Ecology*, *84*(6), 1370–1381.

Coles, S., and E. Casson (1998), Extreme value modelling of hurricane wind speeds, *Structural Safety*, *20*, 283–296.

Coles, S., J. Heffernan, and J. A. Tawn (1999), Dependence measures for extreme value analyses, *Extremes*, *2*, 339–365.

Cooley, D., D. Nychka, and P. Naveau (2007), Bayesian spatial modeling of extreme precipitation return levels, *Journal of the American Statistical Association*, *102*(479), 824–840.

Crainiceanu, C. M., J. R. Stedinger, D. Ruppert, and C. T. Behr (2003), Modeling the us national distribution of waterborne pathogen concentrations with application to cryptosporidium parvum, *Water Resources Research*, *39*(9).

Dalrymple, T. (1960), Flood frenquency analysis, *U.S. Geol. Surv. Water Supply*, *1543A*.

Dawid, A. P. (1984), Statistical-theory - the prequential approach, *Journal of the Royal Statistical Society Series a-Statistics in Society*, *147*, 278–292.

Dobson, A. (2001), *An Introduction to Generalised Linear Models*, Texts in statistical science series.

Favre, A. C., S. El Adlouni, L. Perreault, N. Thiemonge, and B. Bobee (2004), Multivariate hydrological frequency analysis using copulas, *Water Resources Research*, *40*(1).

Fill, H. D., and J. R. Stedinger (1998), Using regional regression within index flood procedures and an empirical bayesian estimator, *Journal of Hydrology*, *210*(1-4), 128 – 145, doi:10.1016/S0022-1694(98)00177-2.

Garavaglia, F., M. Lang, E. Paquet, J. Gailhard, R. Garçon, and B. Renard (2011), Reliability and robustness of rainfall compound distribution model based on weather pattern sub-sampling, *Hydrology and Earth System Sciences*, *15*(2), 519–532, doi:10.5194/hess-15-519-2011.

Gelman, A., J. Carlin, H. Stern, and D. Rubin (1995), *Bayesian data analysis*, Texts in Statistical Science, Chapman and Hall.

Genest, C., and A. C. Favre (2007), Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data, *Water Resources Research*, *43*.

Gneiting, T., F. Balabdaoui, and A. E. Raftery (2007), Probabilistic forecasts, calibration and sharpness, *Journal of the Royal Statistical Society Series B-Statistical Methodology*, *69*, 243–268.

Goovaerts, P. (1998), Ordinary cokriging revisited, *Mathematical geology*, *30*(1), 21–42.

Gottschalk, L. (1993a), Correlation and covariance of runoff, *Stochastic Hydrology and Hydraulics*, *7*, 85–101, doi:10.1007/BF01581418.

Gottschalk, L. (1993b), Interpolation of runoff applying objective methods, *Stochastic Hydrology and Hydraulics*, *7*, 269–281, doi:10.1007/BF01581615.

Griffis, V., and J. Stedinger (2007), Evolution of flood frequency analysis with bulletin 17, *JOURNAL OF HYDROLOGIC ENGINEERING*, *12*, 283–297.

Hosking, J., and J. R. Wallis (1988), The effect of intersite dependence on regional flood frequency analysis, *Water Resources Research*, *24*, 588–600.

Hosking, J., and J. R. Wallis (1997), *Regional Frequency Analysis: an approach based on L-Moments*, Cambridge University Press, Cambridge, UK.

Katz, R. W., M. B. Parlange, and P. Naveau (2002), Statistics of extremes in hydrology, *Advances in Water Resources*, *25*(8-12), 1287 – 1304, doi:10.1016/S0309-1708(02)00056-8.

Keef, C., C. Svensson, and J. A. Tawn (2009), Spatial dependence in extreme river flows and precipitation for great britain, *Journal of Hydrology. To appear.*, doi: 10.1016/j.jhydrol.2009.09.026.

Kjeldsen, T. R., and D. A. Jones (2009a), An exploratory analysis of error components in hydrological regression modeling, *Water Resources Research*, *45*.

Kjeldsen, T. R., and D. A. Jones (2009b), A formal statistical model for pooled analysis of extreme floods, *Hydrology research*, *40*(5), 465–480.

Kroll, C. N., and J. R. Stedinger (1998), Regional hydrologic analysis: Ordinary and generalized least squares revisited, *Water Resources Research*, *34*(1), 121–128.

Kuczera, G. (1982a), Combining site-specific and regional information: An empirical bayes approach, *Water Resources Research*, *18*(2), 306–314.

Kuczera, G. (1982b), Robust flood frequency models, *Water Resources Research*, *18*(2), 315–324.

Kuczera, G. (1983), Effect of sampling uncertainty and spatial correlation on an empirical bayes procedure for combining site and regional information, *Journal of Hydrology*, *65*, 373–398.

Laio, F., and S. Tamea (2007), Verification tools for probabilistic forecasts of continuous hydrological variables, *Hydrology and Earth System Sciences*, *11*(4), 1267–1277.

Lima, C. H. R., and U. Lall (2009), Hierarchical bayesian modeling of multisite daily rainfall occurrence: Rainy season onset, peak, and end, *Water Resources Research.*, *45*.

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, *Journal of Hydrology*, *383*(3-4), 307–318.

Madsen, H., and D. Rosbjerg (1997), The partial duration series method in regional index-flood modeling, *Water Resources Research*, *33*(4), 737–746.

Maraun, D., T. Osborn, and H. Rust (2009), The influence of synoptic airflow on uk daily precipitation extremes. part i: Observed spatio-temporal relationships, *Climate Dynamics*, pp. 1–15, doi:10.1007/s00382-009-0710-9.

Martins, E., and J. Stedinger (2000), Generalized maximum likelihood gev quantile estimators for hydrologic data, *Water Resources Research*, *36*(3), 737–744.

Metropolis, N., and S. Ulam (1949), The monte carlo method, *Journal of the American Statistical Association*, *44*, 335–341.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *Journal of chemical physics*, *21*, 1087–1092.

Micevski, T. (2007), Nonhomogeneity in eastern australian flood frequency data: Identification and regionalisation, Ph.D. thesis, University of Newcastle, Australia.

Micevski, T., and G. Kuczera (2009), Combining site and regional flood information using a bayesian monte carlo approach, *Water Resources Research*, *45*.

Micevski, T., G. Kuczera, and S. Franks (2006), A bayesian hierarchical regional flood model, in *Proc. 30th Hydrology and Water Resources Symposium*, edited by E. Australia, Launceston, Tasmania, Australia.

Mikosch, T. (2005), How to model multivariate extremes if one must?, *Statistica Neerlandica*, *59*(3), 324–338.

Neppel, L., P. Arnaud, and J. Lavabre (2007), Extreme rainfall mapping: Comparison between two approaches in the mediterranean area, *Comptes Rendus Geoscience*, *339*(13), 820–830, doi:10.1016/j.crte.2007.09.013.

Ouarda, T., C. Girard, G. Cavadias, and B. Bobee (2001), Regional flood frequency estimation with canonical correlation analysis, *Journal of Hydrology*, *254*(1-4), 157–173.

Padoan, S., M. Ribatet, and S. Sisson (2010), Likelihood-based inference for max-stable processes, *Journal of the American Statistical Association*, *105*, 263–277.

Perreault, L. (2000), Analyse bayesienne retrospective d'une rupture dans les sequences de variables aleatoires hydrologiques, Ph.D. thesis, ENGREF, INRS-Eau.

Pujol, N., L. Neppel, and R. Sabatier (2007), Regional tests for trend detection in maximum precipitation series in the french mediterranean region, *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *52*(5), 956–973.

Reichert, P., and J. Mieleitner (2009), Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters, *Water Resources Research*, *45*.

Reis, D. S., J. R. Stedinger, and E. S. Martins (2005), Bayesian generalized least squares regression with application to log pearson type 3 regional skew estimation, *Water Re-*

sources Research, *41*(10).

Renard, B., and M. Lang (2007), Use of a gaussian copula for multivariate extreme value analysis: some case studies in hydrology, *Advances in Water Resources*, *30*(4), 897–912.

Renard, B., V. Garreta, and M. Lang (2006), An application of bayesian analysis and mcmc methods to the estimation of a regional trend in annual maxima, *Water Resources Research*, *42*(12).

Renard, B., et al. (2008), Regional methods for trend detection: assessing field significance and trend consistency, *Water Resources Research.*, *44*.

Ribatet, M., E. Sauquet, J.-M. Gresillon, and T. Ouarda (2007), A regional bayesian pot model for flood frequency analysis, *Stochastic environmental research and risk assessment*, *21*(4), 327–339.

Robson, A., and D. Reed (1999), *Flood Estimation Handbook. Volume 3: Statistical procedures for flood frequency estimation*, vol. 3, Wallingford.

Sauquet, E. (2006), Mapping mean annual river discharges: Geostatistical developments for incorporating river network dependencies, *Journal of Hydrology*, *331*(1-2), 300 – 314, doi:10.1016/j.jhydrol.2006.05.018.

Sauquet, E., L. Gottschalk, and E. Leblois (2000), Mapping average annual runoff: a hierarchical approach applying a stochastic interpolation scheme, *Hydrological sciences journal*, *45*(6).

Skoien, J. O., R. Merz, and G. Blschl (2006), Top-kriging - geostatistics on stream networks, *Hydrology and Earth System Sciences*, *10*(2), 277–287, doi:10.5194/hess-10-277-2006.

Spiegelhalter, D. J., N. G. Best, B. R. Carlin, and A. van der Linde (2002), Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society Series B-Statistical Methodology*, *64*, 583–616.

Stedinger, J., and L. Lu (1995), Appraisal of regional and index flood quantile estimators, *Stochastic Hydrology and Hydraulics*, *9*(1), 49–75.

Stedinger, J. R. (1983), Estimating a regional flood frequency distribution, *Water Resources Research*, *19*, 503–510.

Stedinger, J. R., and V. W. Griffis (2011), Getting from here to where? flood frequency analysis and climate, *Journal of the American Water Resources Association (JAWRA)*, *47*(3), 506–513, doi:10.1111/j.1752-1688.2011.00545.

Stedinger, J. R., and G. D. Tasker (1985), Regional hydrologic analysis: 1. ordinary, weighted and generalized least squares compared, *Water Resources Research*, *21*(9), 14211432 [Correction, Water Resour. Res.,22(5), 844, 1986.].

Stedinger, J. R., and G. D. Tasker (1986), Regional hydrologic analysis: 2. model-error estimators, estimation of sigma and log-pearson type 3 distributions, *Water Resources Research*, *22*(10), 1487 1499.

Storz, J. F., and M. A. Beaumont (2002), Testing for genetic evidence of population expansion and contraction: An empirical analysis of microsatellite dna variation using a hierarchical bayesian model, *Evolution*, *56*(1), 154–166.

Tasker, G. D., and J. R. Stedinger (1989), An operational gls model for hydrologic regression, *Journal of Hydrology*, *111*, 361:375.

Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological

modelling: a case study using bayesian total error analysis, *Water Resources Research*, *45*, doi:10.1029/2008WR006825.

Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with markov chain monte carlo simulation, *Water Resour. Res.*, *44*.

Wikle, C. K., L. M. Berliner, and N. Cressie (1998), Hierarchical bayesian space-time models, *Environmental and Ecological Statistics*, *5*, 117–154, doi:10.1023/A:1009662704779.

Figure 1. Schematic of the hierarchical modeling framework

Figure 2. Raingauges location. The thin lines represent six homogeneous regions defined by *Pujol et al.* [2007].

Figure 3. Posterior pdfs of some inferred quantities: (a) regional shape parameter ξ; (b) errors ϵ_μ in the regression for the location D-parameter (only 7 distributions are shown for readability); (c) errors ϵ_λ in the regression for the scale D-parameter (d) hyper-standard deviations σ_μ (solid black line) and σ_λ (dashed red line), corresponding to the standard deviations of location/scale regression errors.

Figure 4. Estimated dependence structures. (a) Data dependence. The dots represent estimated pairwise dependences for all available pairs of sites; (b) dependence in regression errors ϵ_μ for the location D-parameter; (c) dependence in regression errors ϵ_λ for the scale D-parameter.

Figure 5. Estimation of the 100-year daily rainfall $R_{0.99}$. Left: $R_{0.99}$ point-estimates (posterior median); Right: uncertainty in estimating $R_{0.99}$ (measured by the posterior coefficient of variation).

Figure 6. Posterior pdfs of some inferred quantities with three distinct models for data dependence. (a) Regional shape parameter ξ; (b) hyper-standard deviations σ_μ (standard deviations of location regression errors); (c) hyper-standard deviations σ_λ (standard deviations of scale regression errors).

Figure 7. Estimated dependence structures with three distinct models for data dependence. Solid black line = $\mathcal{M}_3$-Inde; dashed red line = $\mathcal{M}_3$-GCop; dotted blue line = $\mathcal{M}_3$-SCop. Shaded areas represent 90% posterior intervals. (a) Data dependence; (b) dependence in regression errors ϵ_μ for the location D-parameter; (c) dependence in regression errors ϵ_λ for the scale D-parameter.

Figure 8. Impact of the data dependence model on predictions of $R_{0.99}$. Maps show relative differences with the benchmark model $\mathcal{M}_3$-GCop. Top panels = $\mathcal{M}_3$-Inde; bottom panels = $\mathcal{M}_3$-SCop. Left = $R_{0.99}$ point-estimates (posterior median); Right = uncertainty in estimating $R_{0.99}$ (measured by the posterior coefficient of variation).

Figure 9. Posterior pdfs of some inferred quantities with three distinct regression models. (a) Regional shape parameter ξ; (b) hyper-standard deviations σ_μ (standard deviations of location regression errors); (c) hyper-standard deviations σ_λ (standard deviations of scale regression errors).

Figure 10. Estimated dependence structures with three distinct regression models. Dotted blue line = $\mathcal{M}_1$-GCop; solid black line = $\mathcal{M}_2$-GCop; dashed red line = $\mathcal{M}_3$-GCop. Shaded areas represent 90% posterior intervals. (a) Data dependence; (b) dependence in regression errors ϵ_μ for the location D-parameter; (c) dependence in regression errors ϵ_λ for the scale D-parameter.

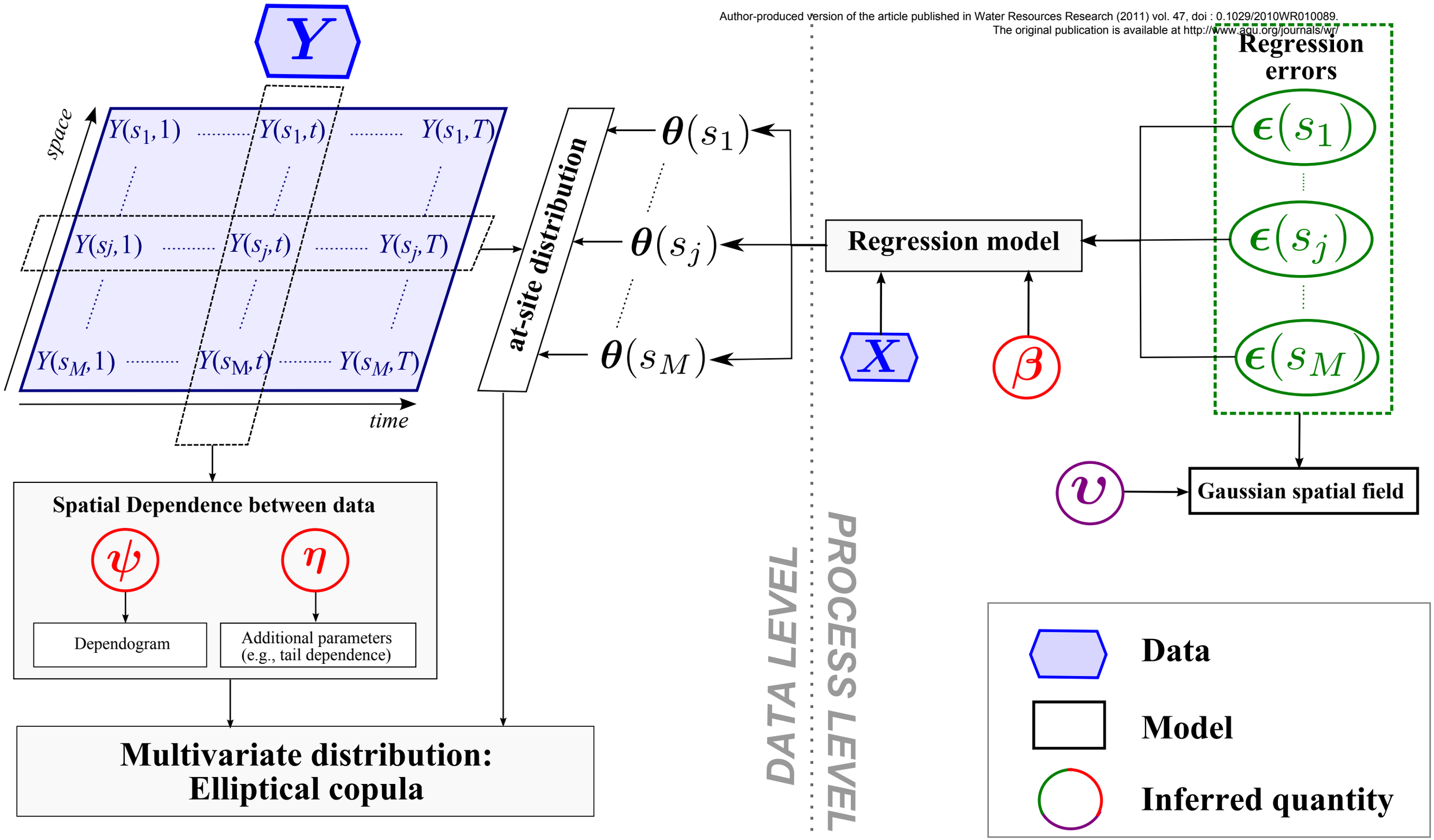
Figure 11. Impact of the regression model on predictions of $R_{0.99}$. Maps show relative differences with the benchmark model $\mathcal{M}_3$-GCop. Top panels = $\mathcal{M}_1$-GCop; bottom panels = $\mathcal{M}_2$-GCop. Left = $R_{0.99}$ point-estimates (posterior median); Right = uncertainty in estimating $R_{0.99}$ (measured by the posterior coefficient of variation).

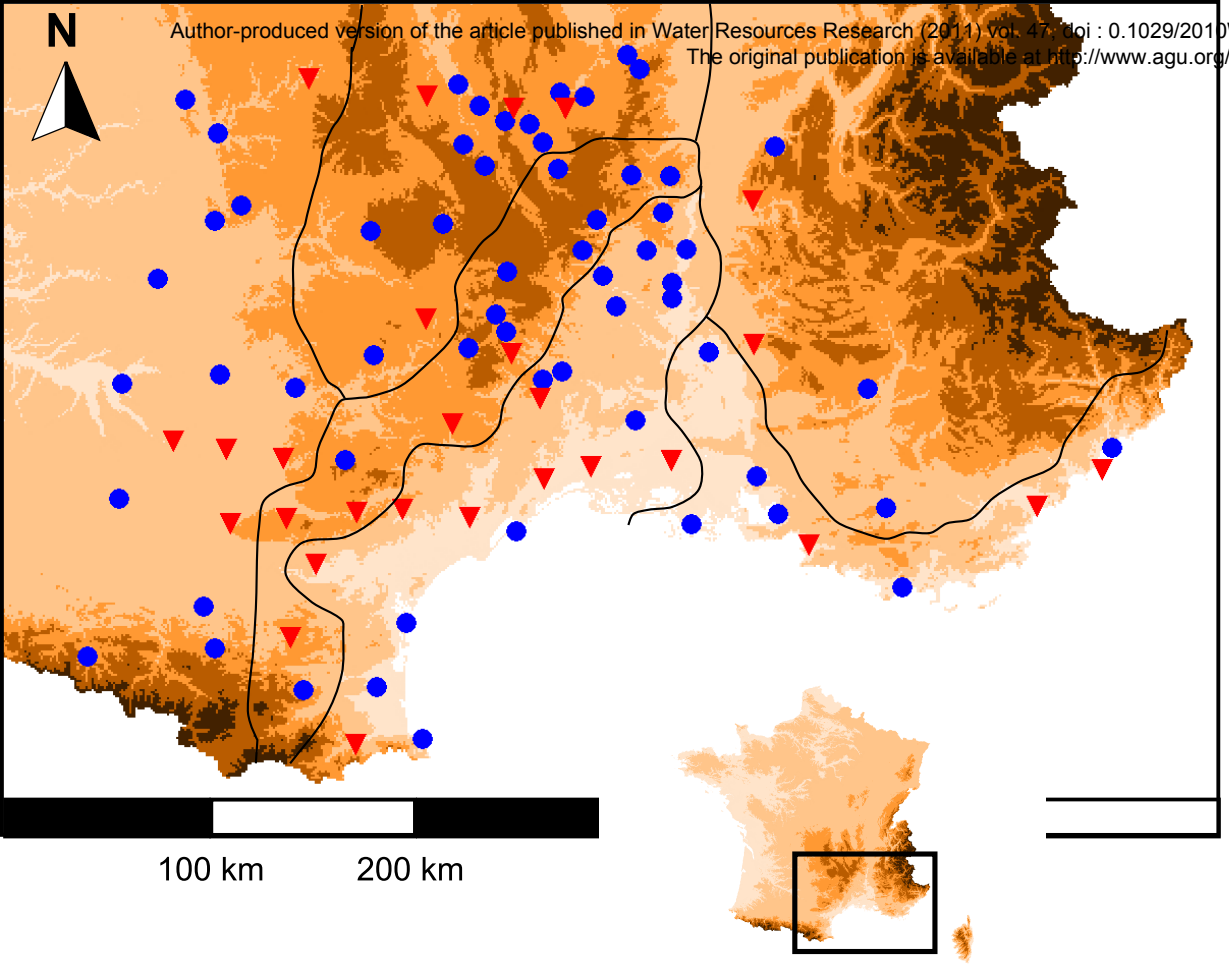
Figure 12. Evaluation of the scale invariance hypothesis. Points = empirical coefficients of variation of data; dashed horizontal lines = coefficient of variation estimated from $\mathcal{M}_2$-GCop (90% posterior interval); vertical bars = coefficient of variation estimated from $\mathcal{M}_3$-GCop (90% posterior intervals).

Figure 13. Predictive QQ-plots at validation sites for each of the five studied models. The lower right panel corresponds to merging data from all validation sites together.

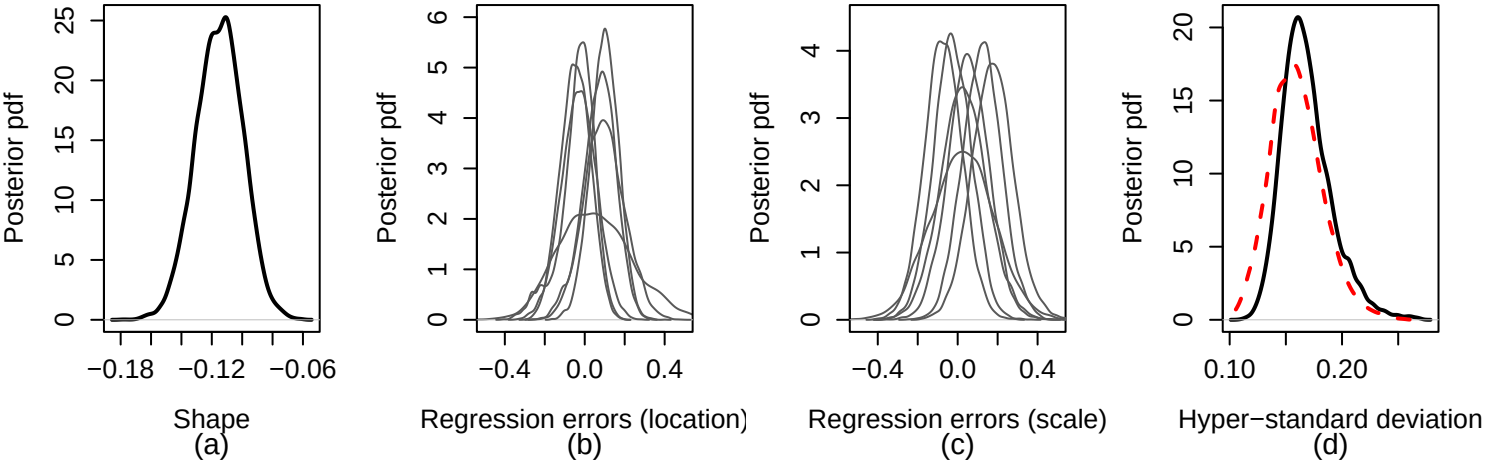
Figure 14. Comparison of predictive distributions obtained at one particular validation site.

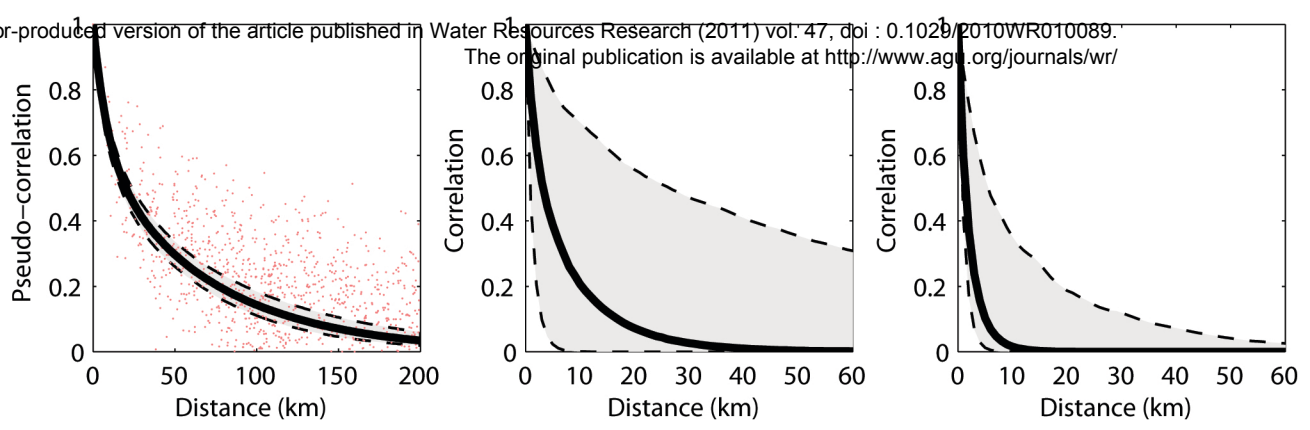
Figure 15. Posterior coefficients of variation of $R_{0.99}$ estimated at validation sites, as a function of the distance between the validation site and the nearest calibration site.



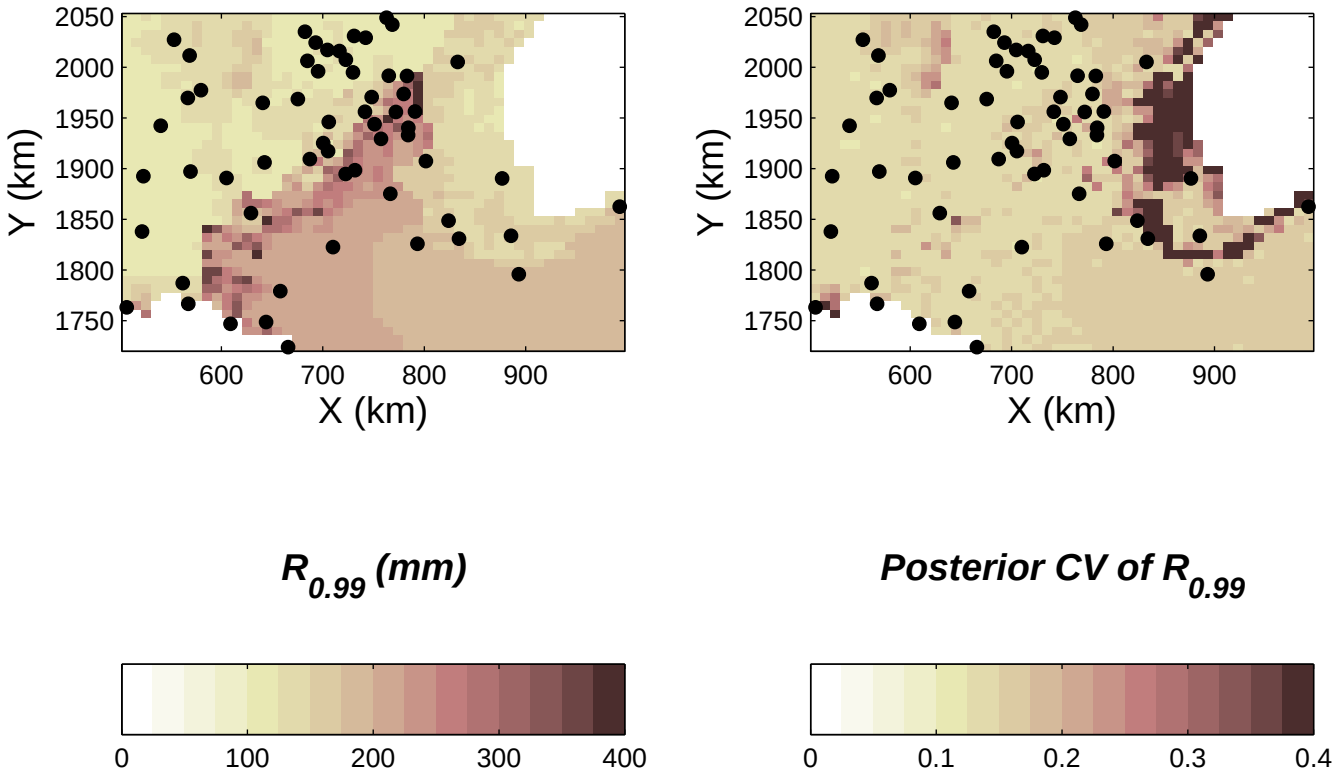


Author-produced version of the article published in Water Resources Research (2011) vol. 47, doi : 0.1029/2010WR010089.
The original publication is available at http://www.agu.org/journals/wr/

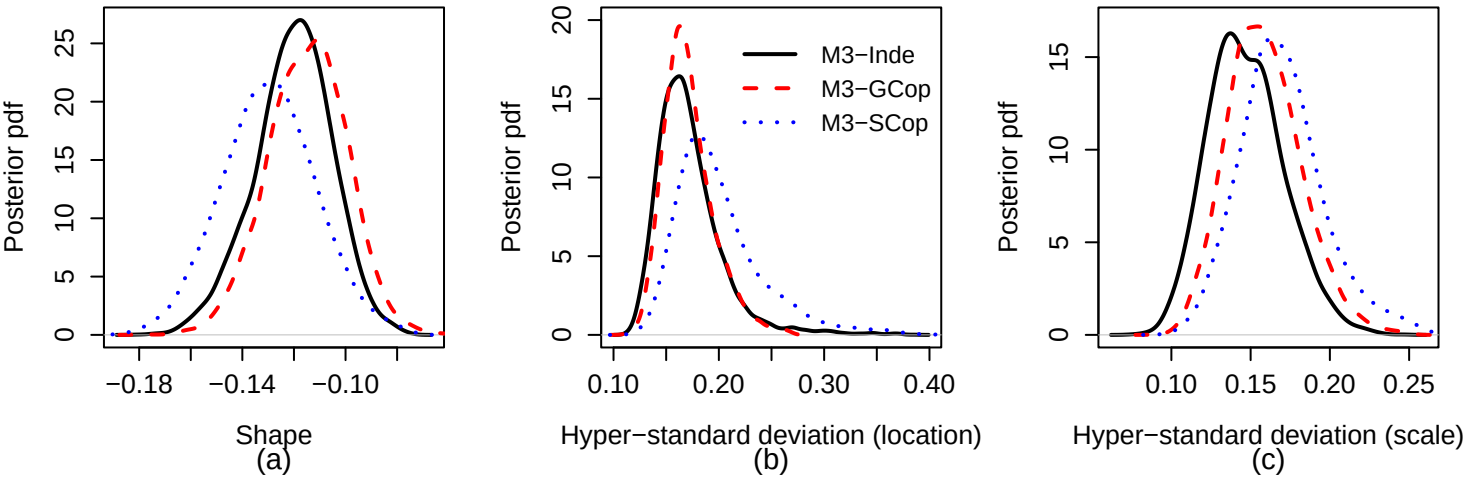




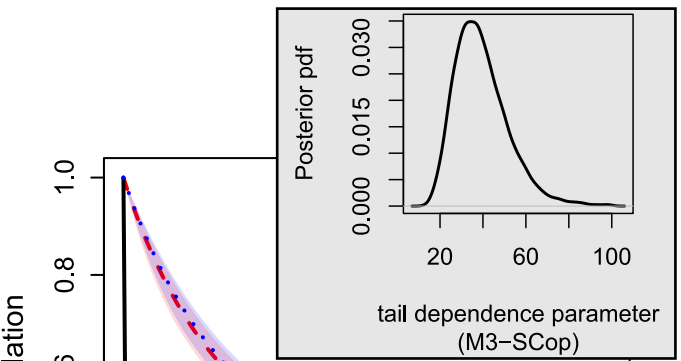
(a) (b) (c)

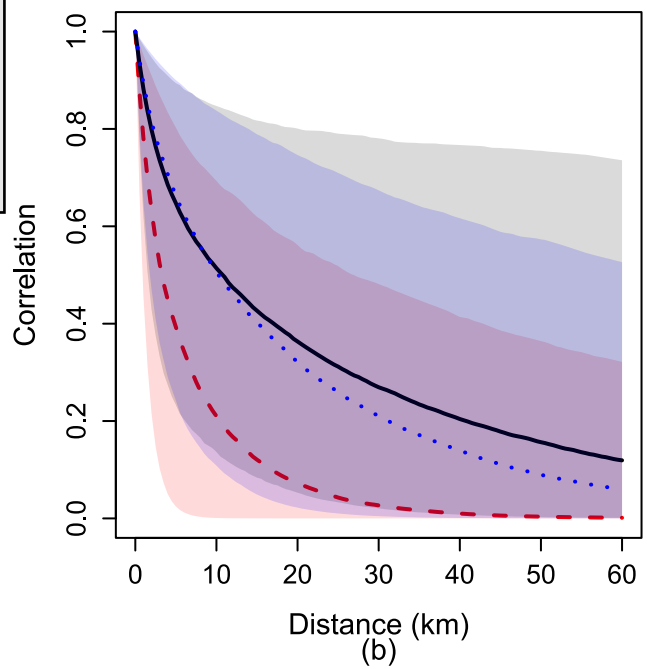


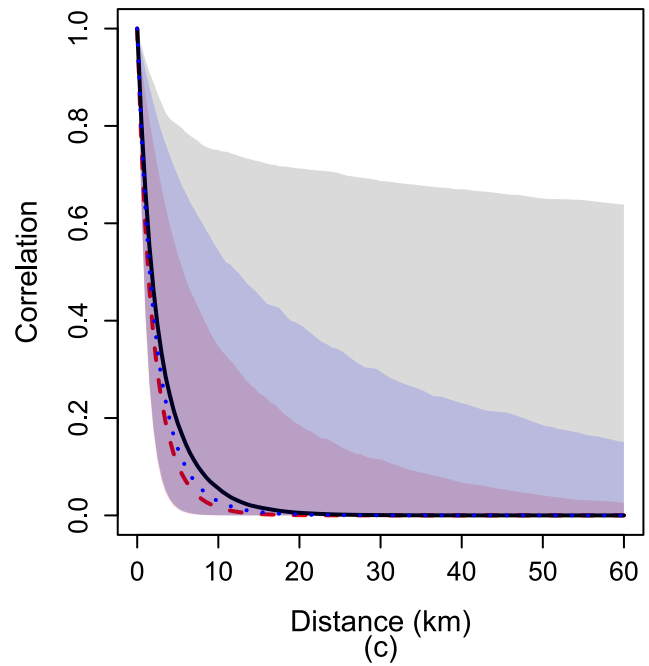
Author-produced version of the article published in Water Resources Research (2011) vol. 47, doi : 0.1029/2010WR010089.
The original publication is available at http://www.agu.org/journals/wr/

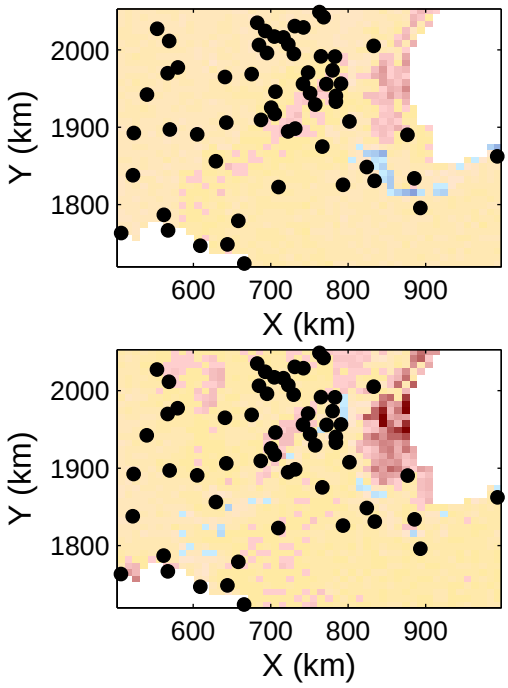


Author-produced version of the article published in Water Resources Research (2011) vol. 47, doi : 0.1029/2010WR010089.
The original publication is available at http://www.agu.org/journals/wr/

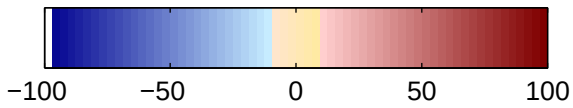


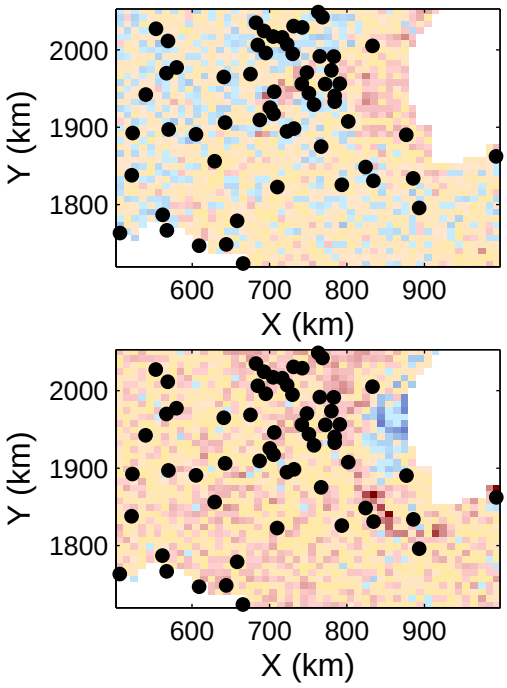




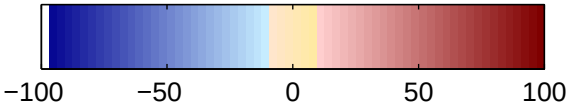


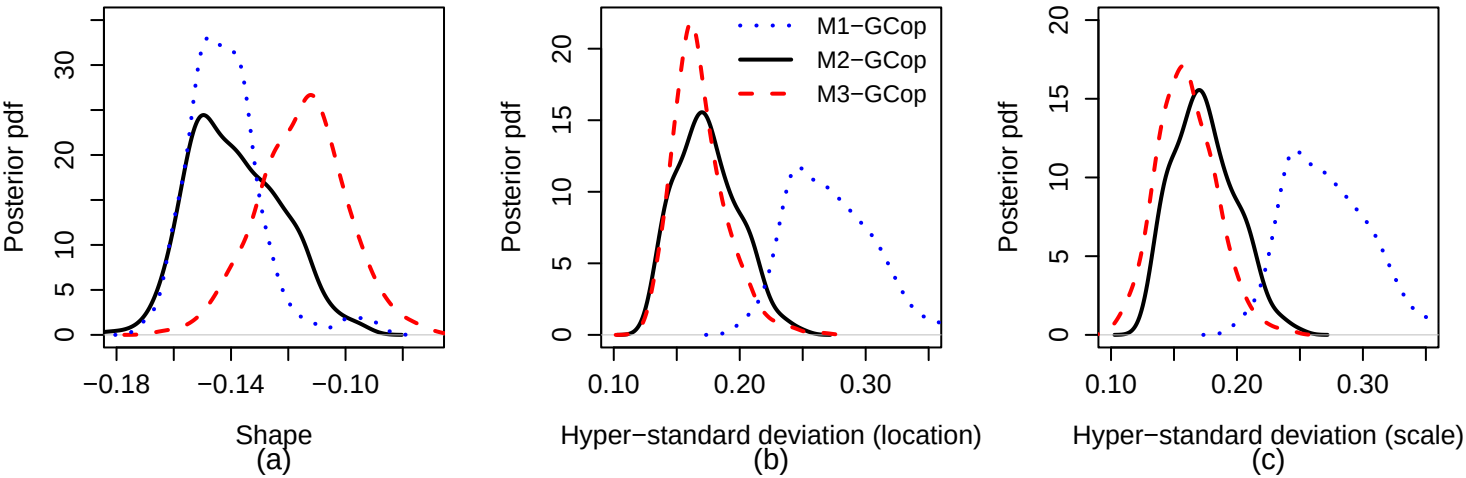
Relative change in $R_{0.99}$ (%)

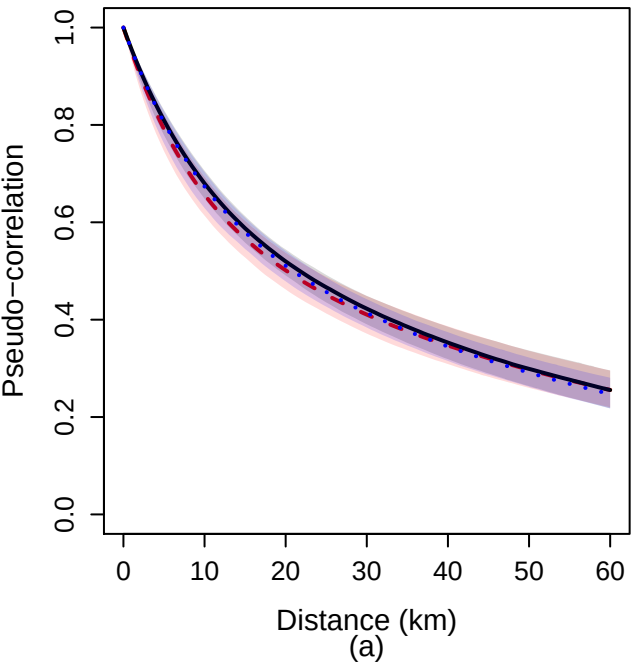


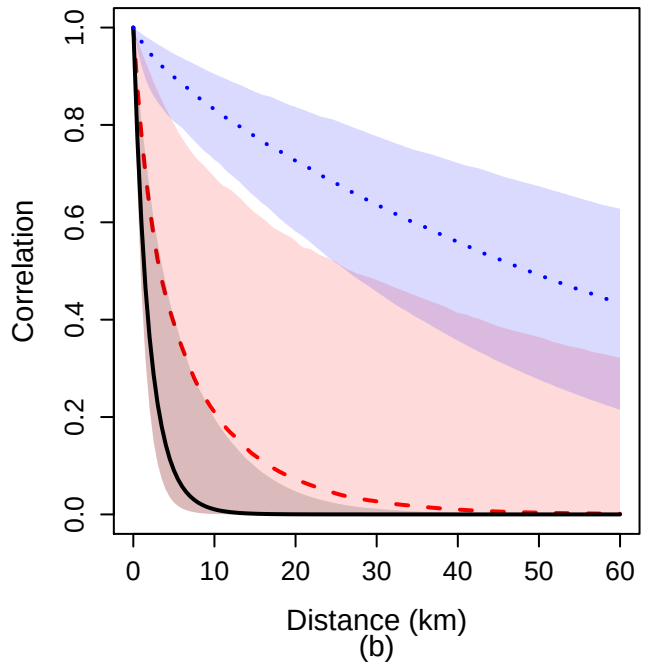


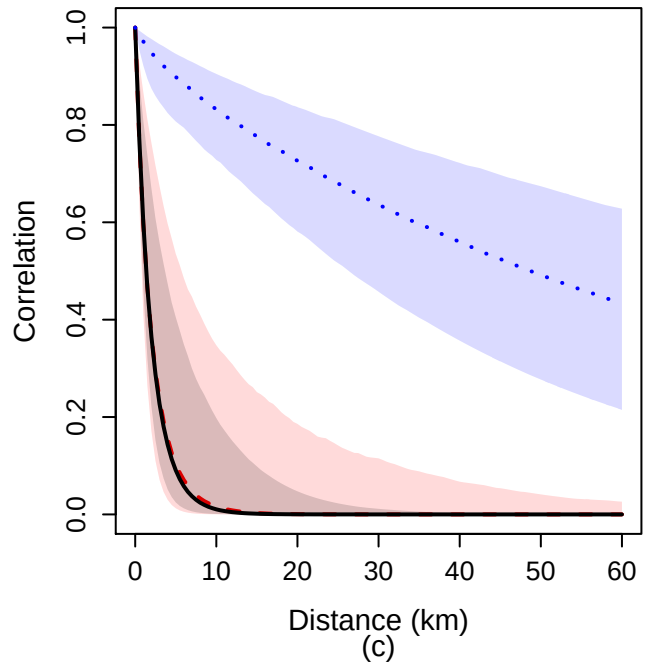
Relative change in $R_{0.99}$ uncertainty (%)

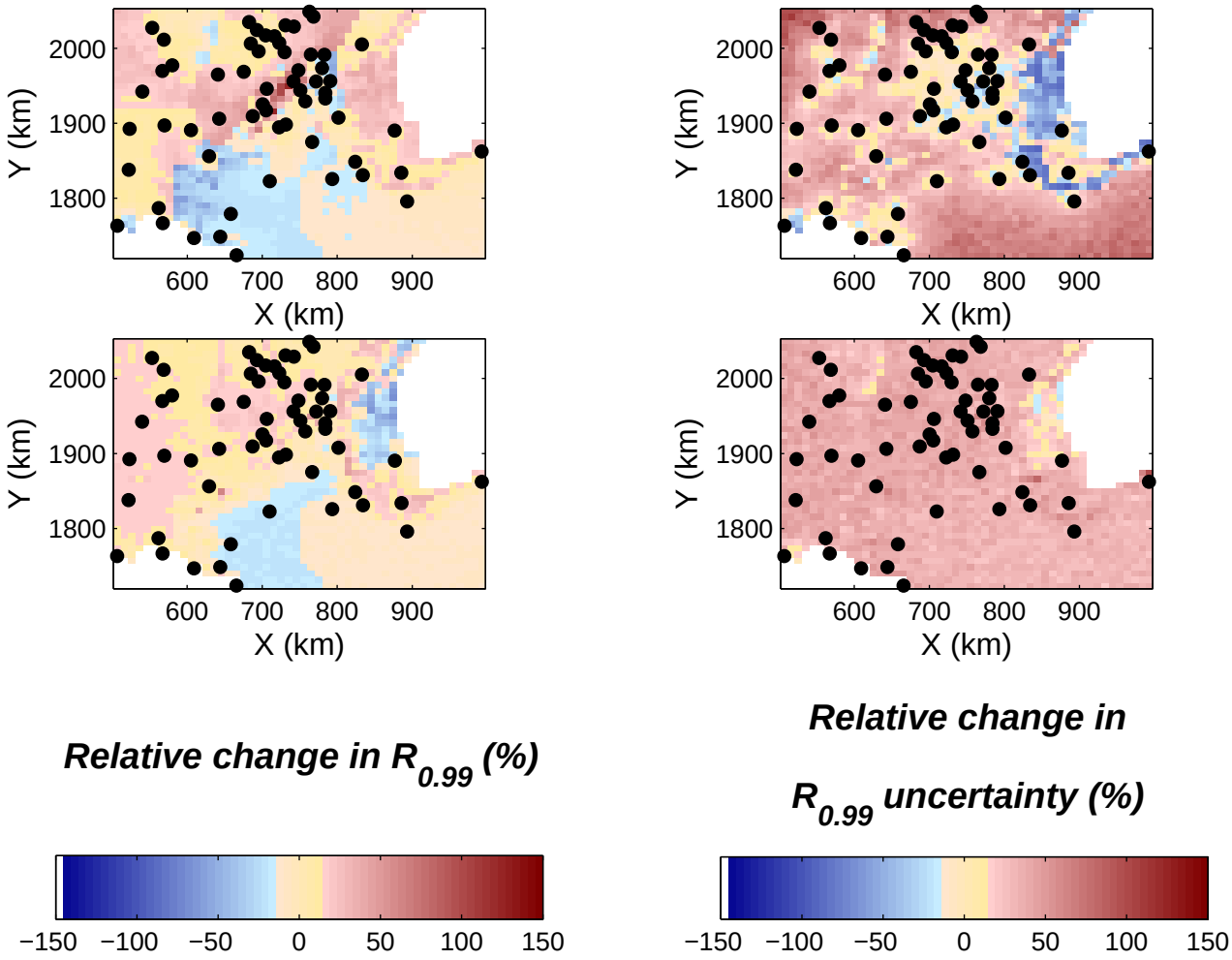




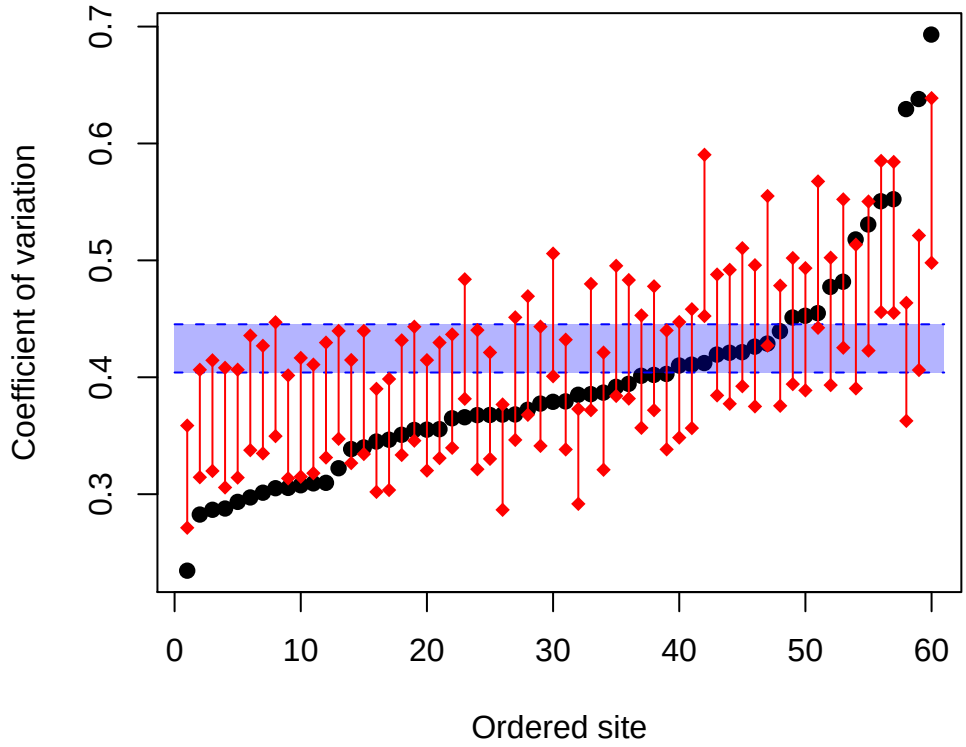


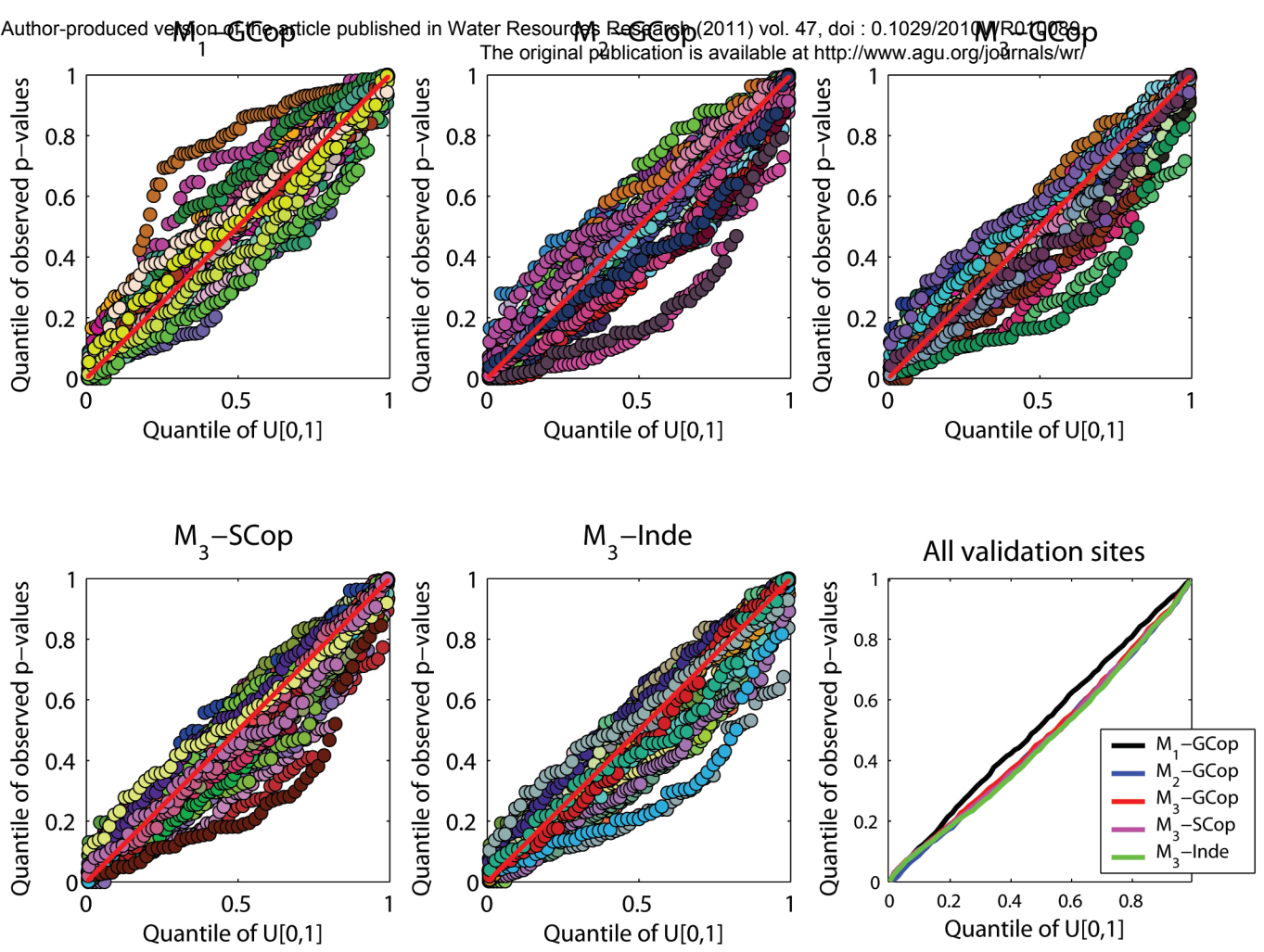




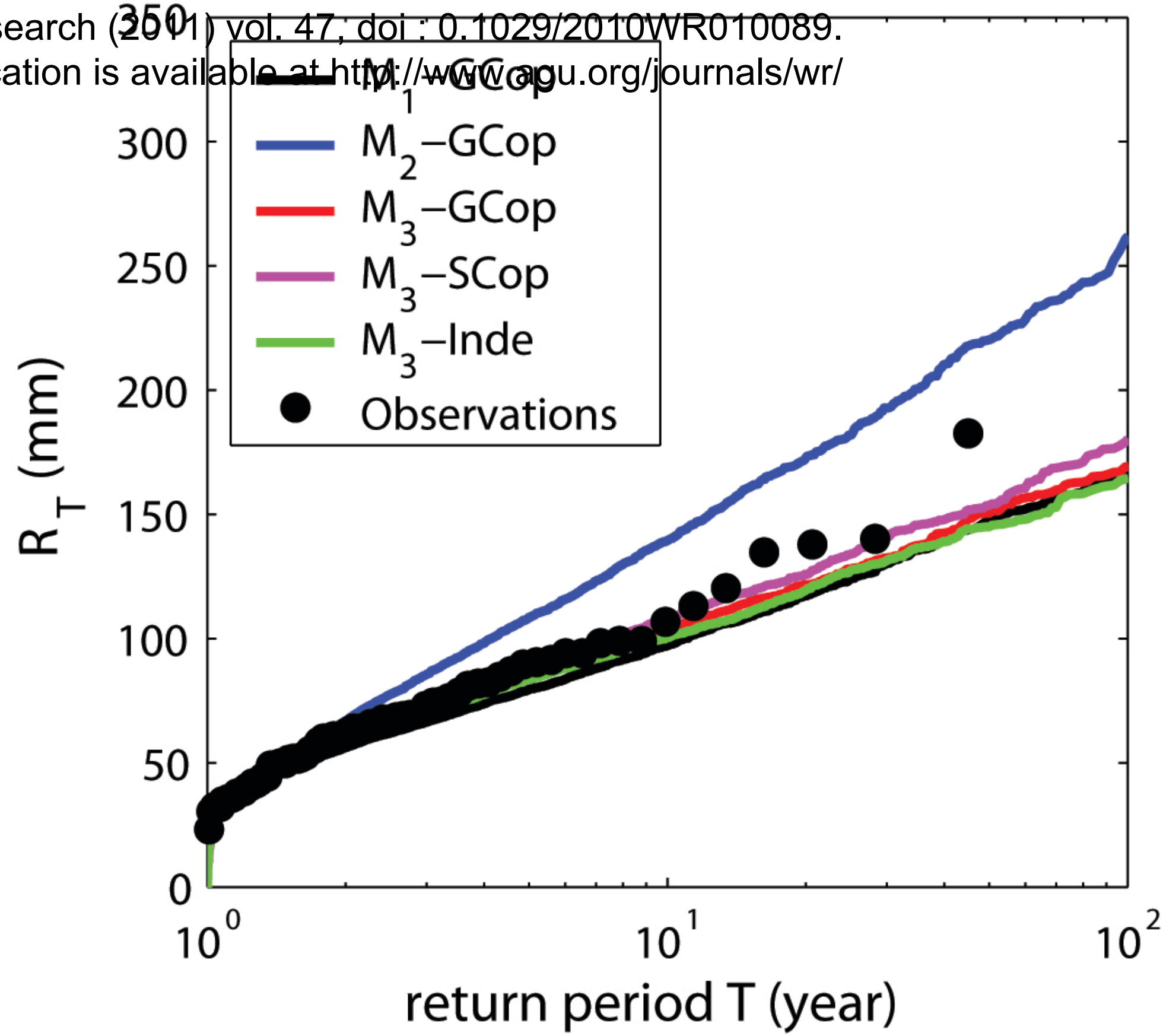


ticle published in Water Resources Research (2011) vol. 47, doi : 0.1029/2010WR010089.
The original publication is available at http://www.agu.org/journals/wr/





search (2011) vol. 47, doi : 0.1029/2010WR010089.
cation is available at http://www.agu.org/journals/wr/



d in Water Resources Research (2011) vol. 47, doi : 0.1029/2010WR010089.
The original publication is available at http://www.agu.org/journals/wr/