Depth and homogeneity in regional flood frequency analysis

F. Chebana* and T.B.M.J. Ouarda

Industrial Chair in Statistical Hydrology/Canada Research Chair on the Estimation of Hydrometeorological Variables,
INRS-ETE, 490 rue de la Couronne, Quebec (QC),
Canada G1K 9A9

*Corresponding author: Tel: (418) 654-2542
Fax: (418) 654-2600
Email: fateh_chebana@ete.inrs.ca

July 11th 2008

Abstract:

Regional frequency analysis (RFA) consists generally in two steps: 1) delineation of hydrological homogeneous regions, and 2) regional estimation. Existing regionalization methods which adopt this two-step approach suffer from two principal drawbacks. First, the restriction of the regional estimation to a particular region by excluding some sites can correspond to a loss of some information. Second, the definition of a region generates a border effect problem. In order to overcome these problems, a new method is proposed in the present paper. The proposed method is based on three elements: (i) a weight function to treat the border effect problem, (ii) a function to evaluate how "similar" each site is to the target one, and (iii) an iterative procedure to improve estimation results. Element (ii) is treated using the statistical notion of depth functions which is introduced to provide a ranking of stations in a multivariate context. Furthermore, the properties of depth functions meet the characteristics sought in RFA. It is shown that the proposed method is flexible and general and that traditional RFA methods represent special cases of the depth-based approach corresponding to particular weight functions. A comparison is carried out with the canonical correlation analysis (CCA) approach. Results indicate that the depth-based approach performs better than CCA both in terms of relative bias and relative root mean squares error.

1. Introduction

One of the problems encountered in hydrology is the lack of data, since the extreme events we want to estimate are rare and record lengths are short. Consequently, statistical inference is difficult in such sites. To overcome this problem, hydrologists have recourse to data from other sites that are hydrologically similar to the target one. The estimation of extreme hydrological events, such as floods, at sites where little or no data is available is the main aim of regional frequency analysis. Delineation of homogeneous hydrological regions and regional estimation are the two main steps in a regional flood frequency procedure. Several studies have focused on the delineation of homogeneous regions (e.g., Burn, 1990; Hosking and Wallis, 1993; Ouarda et al., 2006; and Chebana and Ouarda, 2007) and on regional estimation (e.g., Dalrymple, 1960; Stedinger and Tasker, 1986; Ouarda and Ashkar, 1994; Durrans and Tomic, 1996; Nguyen and Pandey, 1996; Madsen and Rosbjerg, 1997; Alila, 1999, 2000 and Chokmani and Ouarda, 2004). An intercomparison of various regional flood estimation procedures was presented by GREHYS (1996a,b) by coupling four methods for delineating homogenous regions and seven regional estimation methods.

In a regional estimation procedure, one is interested in maximizing the amount of transferred information. The delineation step corresponds usually to the exclusion of a number of sites which may lead to a loss of some relevant information. Furthermore, the definition of a region leads to the problem of the so-called “border effect”. This means that for two sites that are very close but which are located on each side of the region limits, one is excluded while the other one is included even though both sites offer similar information. This problem is not present when the limits correspond to natural borders.

These elements motivate the development of a new method that overcomes some of these drawbacks. The method developed in the present paper is based on the notion of depth function. The depth function is a statistical notion developed in the seventies and which receives increasing interest (e.g. Tukey, 1975; Liu, 1990; Liu and Singh, 1993; Rousseeuw and Hubert, 1999; Zuo and Serfling, 2000; Mizera, 2002; Mizera and Müller. 2004; Zuo and Cui, 2005 and Lin and Chen, 2006). The purpose for introducing the depth function is to provide an outward ordering of points in a multivariate context. The most important properties of a depth function are (i) affine invariance, (ii) maximality at center, (iii) monotonicity related to the deepest point, and (iv) vanishing at infinity. These properties fit the constraints of regional flood frequency analysis. Indeed, the affine invariance is useful to remove the scale effect when treating several variables. The center, where the depth function is maximal, represents the target site. The monotonicity related to the center point means that sites far from the target site are less important. Finally, the very far sites have no importance or no contribution; this is the vanishing at infinity property. Hence, the contribution of each site is related to its similarity to the target site.

The proposed procedure focuses on the estimation as a goal and avoids the delineation step. Note that the delineation step is only an intermediate technical tool to estimate quantiles. Different quantile estimation methods are proposed in the literature, such as the index flood method and regressive models (see GREHYS, 1996a,b). In the present paper, the estimation is based on the regressive model. The estimation of the regression parameters is obtained using a weighted least squares method. A key element is related to the choice of the weights. These weights are selected as functions of the site depth. Hence, the proposed method overcomes the problems related to the border effect by using smooth weight functions. It also reduces the problems related to the lack of data by using all available data in a more efficient manner.

Furthermore, traditional approaches represent special cases of depth-based approaches corresponding to particular weight functions. Finally, the non requirement of data normality, the availability of several kinds of depth functions and the smoothness of the weight functions provide this method with a high level of flexibility.

Based on hydrological variables, on one hand, and on physio-meteorological characteristics, on the other hand, regional regression is frequently integrated with the CCA approach (Ouarda et al., 2000, 2001). The CCA-regression provides flood quantile estimates at ungauged sites by using site physiographic characteristics. In order to study the performance of the proposed depth-based approach, it is compared to the CCA-regression approach with optimal neighborhoods (Ouarda et al., 2001). This comparison is based on a data set from 151 gauging sites in the southern part of the province of Quebec, Canada. The specific quantiles corresponding to 10- and 100-year return periods are estimated and a jackknife resampling procedure is used to evaluate the estimation errors.

The paper is organized as follows. Brief presentations of the depth functions, the CCA approach and the weighted least squares method are given in Section 2. Section 3 deals with the proposed methodology in its general form. A case study is presented in Section 4 and the developed approach is applied and compared to the CCA method in Section 5. Results and discussions are reported in Section 6. Conclusions and future promising work are presented in the last section.

2. Background

In this section, a brief description of the background material related to depth functions, the CCA method and weighted regression analysis is presented.

2.1 Depth function:

Tukey (1975) presented the pioneering work in which the depth notion was introduced. The author proposed a halfspace depth in order to define a multivariate analogous to the univariate rank and order statistics. Later, several depth functions were formulated in an ad-hoc manner. Zuo and Serfling (2000) standardized these definitions and classified existing examples in the literature.

For a given cumulative distribution function F on R^d $(d \geq 1)$ a corresponding depth function is any bounded, nonnegative function $D(x;F)$ which provides a F-based center-outward ordering of points x in $\mathbb{R}^d$ that satisfies the following properties:

i. *Affine invariance*: the depth of a point $x \in R^d$ should not depend on the underlying coordinate system or, in particular, on the scales of the underlying measurements.

ii. *Maximality at center*: for a distribution having a uniquely defined center (e.g., the point of symmetry with respect to some notion of symmetry), the depth function should attain its maximum value at this center.

iii. *Monotonicity relative to deepest point*: as a point $x \in R^d$ moves away from the deepest point (the point at which the depth function attains its maximum value; in particular, for a symmetric distribution, the center) along any fixed ray through the center, the depth at x should decrease monotonically.

iv. *Vanishing at infinity*: the depth of a point x should approach zero as its norm $\|x\|$ approaches infinity.

A formal definition of depth functions, based on these properties, is given in Zuo and Serfling (2000). In the following, we denote by Δ the class of cumulative distribution functions on R^d and by F_Z the cumulative distribution function of a given random vector Z. Let the mapping $D(.;.): R^d \times \Delta \to R$ be bounded, non-negative, and satisfy the following conditions:

i. $D(Ax+b; F_{AX+b}) = D(x; F_X)$ holds for any random vector X in R^d, any $d \times d$ nonsingular matrix A, and any d-vector b;

ii. $D(\theta; F) = \sup_{x \in R^d} D(x; F)$ holds for any $F \in \Delta$ having center θ;

iii. For any $F \in \Delta$ having deepest point θ, $D(x; F) \le D\left(\theta + \alpha(x-\theta)x; F\right)$ holds for $\alpha \in [0,\ 1]$; and

iv. $D(x; F)$ converges to 0 as the norm $\|x\|$ goes to infinity, for each $F \in \Delta$.

Then $D(.; F)$ is called a statistical depth function.

Several kinds of depth functions are introduced in the literature. Here we present some of the key ones:

1. *Mahalanobis depth:* It is defined on the basis of the Mahalanobis distance $d_A^2(x, y) = (x-y)' A^{-1} (x-y)$ between two points $x, y \in R^d$ with respect to a positive definite matrix A (Mahalanobis, 1936). The Mahalanobis depth is then given by:

$$MHD(x; F) = \frac{1}{1 + d_A^2(x, \mu)} \tag{1}$$

where F is a given distribution and μ and A are any corresponding location and covariance measures, respectively. Note that the Mahalanobis distance is used in the development of the CCA approach for regional flood frequency analysis (Ouarda et al., 2001). It is also important to note that the Mahalanobis depth function has values in the interval $[0,1]$. Hence, its values are more interpretable than those of the corresponding Mahalanobis distance.

2. *L^2 depth*: It is defined for a distribution F and $x \in R^d$ as:

$$L^2D(x;F) = \frac{1}{1 + E\left\|x - X\right\|_{\Sigma^{-1}}} \tag{2}$$

where Σ is the covariance matrix of F and $\left\|x\right\|_M = \sqrt{x'Mx}$.

3. *Simplicial volume depth* (Oja, 1983): It is given through the expression :

$$SVD^{\alpha}(x,F) = \left(1 + E\left[\left(\frac{\Delta\left(S\left[x, X_1, ..., X_d\right]\right)}{\sqrt{\det\left(\Sigma\right)}}\right)^{\alpha}\right]\right)^{-1} \quad \text{for} \quad x \in R^d \tag{3}$$

where $\Delta\left(S\left[x, X_1, ..., X_d\right]\right)$ denotes the volume of the d-dimensional simplex $S\left[x, X_1, ..., X_d\right]$, Σ is the covariance matrix of F and $\alpha > 0$. The quantity $\Delta\left(S\left[x, X_1, ..., X_d\right]\right)$ is a measure of the dispersion of the point cloud.

4. *Projection depth* (Liu, 1992): It is defined for $x \in R^d$ as :

$$PD\left(x;F\right) = \left(1 + \sup_{\|u\|=1} \frac{\left|u'x - \mathrm{Med}\left(u'X\right)\right|}{\mathrm{MAD}\left(u'X\right)}\right)^{-1} \tag{4}$$

where X has distribution F, Med denotes the univariate median, $\mathrm{MAD}(Y) = \mathrm{Med}\left(\left|Y - \mathrm{Med}(Y)\right|\right)$ represents the median absolute deviation for a random variable Y, and $\|\cdot\|$ is the Euclidian norm.

5. *Halfspace depth* (Tukey, 1975): It is defined for $x \in R^d$ with respect to a probability P on R^d as:

$$HD(x;P) = \inf\{P(H) : H \text{ a closed halfspace that contains } x\} \tag{5}$$

Note that, occasionally, some examples of depth functions do not meet some of the previous four properties in some special cases. For instance, the $L^2D(x;F)$ function only meets the above conditions under symmetric assumptions on the distribution F (see Zuo and Serfling, 2000).

A sample version of $D(x;F)$, denoted by $D_n(x) = D(x;\hat{F}_n)$, may be defined by replacing F with a suitable empirical function $\hat{F}_n$. The asymptotic properties of $D_n(x)$ are studied in several papers including Liu (1990), Arcones et al. (1994), Massé (2004) and Lin and Chen (2006). Liu and Singh (1993) established for the sample Mahalanobis depth function that $\sup_x |D_n(x) - D(x;F)|$ converges to zero almost surely as n goes to infinity, under suitable conditions on F. For convenience, the following notation is used for the Mahalanobis depth function in the next sections:

$$MHD_A(x;\mu) = \frac{1}{1 + d_A^2(x,\mu)} \tag{6}$$

Depth functions are applied in several fields. For instance, Caplin and Nalebuff (1988, 1991a,b) employed depth notions in econometric and social studies. They were also applied in industrial quality control by Liu and Singh (1993) and Liu (1995). Ghosh and Chaudhuri (2005) investigated the use of depth functions in nonparametric discrimination analysis. Mizera and Müller (2004) defined and studied the location-scale depth and gave some statistical applications.

The computation of some depth functions is complex and requires specific algorithms. For instance, Miller et al. (2003) developed an algorithm for the computation of the halfspace depth. However, to our knowledge, similar algorithms are not available for the projection depth. The Mahalanobis depth is among the simplest ones to evaluate if the parameter μ and the parameter matrix A are identified.

Liu et al. (1999) presented descriptive statistics, graphics and inference related to several depth functions. A detailed description of the theoretical background of depth functions is available in Zuo and Serfling (2000).

2.2 Canonical Correlation Analysis (CCA)

Canonical correlation analysis is concerned with the amount of linear relationship between two sets of variables. Consider two normal random vectors X (physiographical and meteorological variables) and Y (hydrological variables), CCA provides two sets of basis vectors (called canonical variables), one for X and the other for Y. The main property of such vectors is that the correlations between the projections of the variables onto these basis vectors are mutually

maximized (Muirhead, 1982). More precisely, let W and V be linear combinations of X and Y respectively, i.e.,

$$\begin{aligned} V &= a'X \\ W &= b'Y \end{aligned} \tag{7}$$

Let Σ be the covariance matrix of the variables X and Y, defined as:

$$\Sigma = \mathrm{cov}\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{bmatrix} \tag{8}$$

The correlation between W and V can then be calculated as:

$$\rho = \frac{a'\,\Sigma_{XY} b}{\sqrt{a'\,\Sigma_X a b'\,\Sigma_Y b}} \tag{9}$$

The goal of the CCA is to find the vectors a and b maximizing ρ subject to the constraint that W and V must have unit variances. Once the first pair of canonical variables is obtained, other pairs of canonical variables can be obtained in the uncorrelated directions to the previous ones by maximizing equation (9) subject to the constraint of unit variance. For more details concerning CCA application in regional flood frequency analysis, the reader is referred to Ouarda et al. (2001).

Based on the canonical hydrological variables W and physio-meteorological variables V, the Mahalanobis distance for an ungauged target-site with given physiographical characteristics $V=v_0$ is given by:

$$D^2 = d^2_{I_p - \Lambda'\Lambda}\left(W, \Lambda v_0\right) = \left(W - \Lambda v_0\right)' \left(I_p - \Lambda'\Lambda\right)^{-1} \left(W - \Lambda v_0\right) \tag{10}$$

where I_p is the $p \times p$ identity matrix, $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_p)$ with $\lambda_i = \mathrm{corr}(V_i, W_i), i = 1, ..., p$ and p is the rank of the covariance matrix Σ_{XY}.

To belong to the neighborhood of the target-site, at a $100(1-\alpha)\%$ confidence level, a site should have a distance D^2 less than $\chi^2_{\alpha,p}$ where $P\left(\chi_p{}^2 \le \chi^2_{\alpha,p}\right) = 1-\alpha$ and $\chi_p{}^2$ has a chi-squared distribution with p degrees of freedom. The Mahalanobis distance can also serve to indicate where, in the canonical space W that encompasses all possible realizations of the random variable W, would be found the realizations w of W for which V has realized v_0.

2.3 Weighted least squares estimation

Most commonly, the power product model given by equation (10) is integrated with the CCA approach and is used to describe the relationship between flood quantiles QT of a return period T, and the physio-meteorological and basin characteristics $A_1,...,A_r$ for a given region:

$$QT = \alpha_0 A_1^{\alpha_1} A_2^{\alpha_2} ... A_r^{\alpha_r} e \tag{11}$$

Taking s quantiles QT corresponding to s return periods, we construct a vector Y of the hydrological variables, that is $Y = \left(QT_1, QT_2, ..., QT_s\right)$. Then, using the log-transformation and the matrix form, we obtain the multivariate log-linear model:

$$\log Y = \left(\log X\right)\beta + \varepsilon \tag{12}$$

where $\log X = (1, \log A_1, \log A_2, ..., \log A_r)$ is the $(r+1)$ vector of the physio-meteorological variables, β is the $(r+1)\times s$ matrix of parameters and ε represents the error vector with:

$$E(\varepsilon) = 0 \text{ and } \mathrm{Var}(\varepsilon) = \Gamma \tag{13}$$

If the number of sites in the region is denoted by N, the parameter β can be estimated, using the weighted least squares estimation method, by:

$$\hat{\beta}_w = \arg\min_{\beta} \sum_{i=1}^{N} w_i \left(\log Y_i - \beta \log X_i\right)' \left(\log Y_i - \beta \log X_i\right)$$
$$= \left((\log X)' \Omega \log X\right)^{-1} (\log X)' \Omega \log Y \quad (14)$$

and the matrix Γ is estimated by:

$$\hat{\Gamma} = \frac{\left(\log Y - \hat{\beta}_w \log X\right)\left(\log Y - \hat{\beta}_w \log X\right)'}{N - r - 1} \quad (15)$$

where $\Omega = \mathrm{diag}(w_1, ..., w_N)$ is the diagonal matrix composed by the weight elements $w_1, ..., w_N$. A detailed description of multivariate regression analysis can be found in Rencher (2002).

3. Approach development

3.1 Description

Limited use of weighted least squares methods was made in the filed of regional flood frequency analysis. Madsen and Rosbjerg (1997) used weighted least squares (WLS) and generalized least squares (GLS) methods in a regional flood estimation procedure that combines the index-flood concept with an empirical Bayes method. In the WLS and GLS methods, the weights are related to the variance and covariance of the errors in the regression model.

The approach proposed in the present work is focused directly on quantile estimation using the weighted least squares method to estimate regression parameters and does not use any delineation technique. The choice of the weights in equation (14) is very important for parameter estimation and hence for the predicted value $\hat{Y}$. In the un-weighted estimation, all weights are equal to one. However, if a region or a neighborhood has been defined, weights correspond to zero if the site is excluded from the region and one if it is included in the region. In the proposed

approach the weights are chosen differently. They are related to a weight function and a depth function which will be developed in Section 3.2. This makes the methodology very flexible and more general.

A special attention is given to the choice and the evaluation of the depth function. A convenient depth function, which is related to the neighborhood approach, is the Mahalanobis depth function (6). The value of μ is generally unknown for the ungauged site, and must be estimated. The values of the depth function for the gauged sites are highly related to the quality of μ estimates. In other words, the problem here is how to get the hydrological «reference value» with respect to which the depths are computed. This reference value represents the deepest point. Therefore, in order to get an accurate estimate of this «reference value», the proposed approach utilizes an iterative estimation procedure based on the log-linear model. The iterative procedure requires a start point, a criterion and a stopping condition. The approach is described below in its general aspect and also with options for the iteration elements.

The iterative estimation procedure serves to improve the depth values and to make them more accurate. This iterative technique has some similarities with the so-called One-step estimator (see e.g., van der Vaart, 1998, pp. 71). Note that the iterative estimation procedure and the way the weights are selected represent two elements that differentiate the proposed approach from the WLS and GLS methods as applied in Madsen and Rosbjerg (1997). In the described methodology, all available sites in the data set are used, without any restriction to a region or a neighborhood. However, each site is associated to a weight related to its hydrological depth with

respect to the target-site. In that case the problem of the delineation of a region becomes rather a problem of a choice of weight and depth functions.

3.2 Computation algorithm

The computation algorithm is based on the following estimate of the parameter β. It is the estimator given in equation (14) with particular weights w_i. To this end, let $\varphi(.)$ be a positive increasing weight function, and D_N be a sample depth function. Then, the estimators (14) and (15) are given respectively by:

$$\begin{aligned}\hat{\beta}_{D_N,\varphi} &= \arg\min_{\beta} \sum_{i=1}^{N} \varphi\left(D_N(Y_i)\right)\left(\log Y_i - \beta \log X_i\right)' \left(\log Y_i - \beta \log X_i\right) \\ &= \left((\log X)' \Omega_{D_N,\varphi} \log X\right)^{-1} (\log X)' \Omega_{D_N,\varphi} \log Y\end{aligned} \tag{16}$$

and

$$\hat{\Gamma}_{D_N,\varphi} = \frac{\left(\log Y - \hat{\beta}_{D_N,\varphi} \log X\right)\left(\log Y - \hat{\beta}_{D_N,\varphi} \log X\right)'}{N - r - 1} \tag{17}$$

where $\Omega_{D,\varphi} = \mathrm{diag}\left(\varphi\left(D(Y_1)\right),...,\varphi\left(D(Y_N)\right)\right)$.

The weight function is assumed to be increasing, to ensure that the deeper is the site, the more important it is and hence it receives a higher weight. It is important to indicate that the matrix in (17) contains the inter-quantile correlation rather than the inter-site correlation. The latter is taken into account in the GLS approach (see e.g. Madsen and Rosbjerg, 1997). It would be useful, in future efforts, to focus on the integration of spatial correlation in the depth-based approach.

Suppose i_0 is the index of the target-site. The algorithm is composed of the following steps:

1. *Model:* Consider the model (12) relating $\log X$ and $\log Y$.
2. *Initial step:* In order to get a starting estimator $\hat{Y}_{1,i_0}$.

   - Use a preliminary approach to estimate the model parameters. One possible option is the uniform approach which allocates equal weights to all sites.

   - Give the predicted initial value $\hat{Y}_{1,i_0}$.

3. k^{th} step ($k = 2,3,\ldots$): For each site i different from the target one i_0,

   a. Consider the predicted value from the previous step (k-1), i.e. $\hat{Y}_{k-1,i_0}$ using the corresponding parameter estimators $\hat{\beta}_{k-1,i_0}$ and $\hat{\Gamma}_{k-1,i_0}$ given in (16) and (17) respectively, where the notation is adapted to the iteration context.

   b. Compute the depth of $\log Y_i$ with respect to $\log \hat{Y}_{k-1,i_0}$ between sites i and i_0.

   It is convenient to use the Mahalanobis depth $MHD_{\hat{\Gamma}_{k-1,i0}}\left(\log Y_i; \log \hat{Y}_{k-1,i_0}\right)$ given by (6).

   c. Compute the weight w_i corresponding to site i. It is given for a weight function φ and the Mahalanobis depth by $w_i = \varphi\left(MHD_{\hat{\Gamma}_{k-1,i0}}\left(\log Y_i; \log \hat{Y}_{k-1,i_0}\right)\right)$. Some examples of weight functions are presented in the next section.

   d. Estimate the model parameters using the weighted method.

   In the log-linear model, the estimators $\hat{\beta}_{k,i_0}$ and $\hat{\Gamma}_{k,i_0}$ are expressed through (16) and (17) respectively.

   e. Give the predicted value $\hat{Y}_{k,i_0}$ using X_{i_0} and $\hat{\beta}_{k,i_0}$ in model (12).

f. Assess the criterion to evaluate the model estimation quality. Several known criteria in regression analysis can be used, for instance, the relative bias (RB) and the relative root mean square error (RRMSE).

4. *Iteration:* Redo step 3 while the criterion is improving.
5. *End condition:* Stop the iteration whenever the criterion deteriorates or converges.
6. *Final result:* Get the estimators of the regression model parameters and hence the final predicted value.

3.3 Particular cases of the proposed approach

The proposed method is general and includes some known methods as special cases representing particular weight functions. Indeed:

1. The uniform approach which uses all sites corresponds to the weight function $\varphi_U \equiv 1$. In this case the estimator $\hat{\beta}$ is given by:

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{N} w_i \left(\log Y_i - \beta \log X_i\right)^2 = \arg\min_{\beta} \sum_{i=1}^{N} \left(\log Y_i - \beta \log X_i\right)^2 \qquad (18)$$

2. The traditional CCA approach with a given value of α corresponds to:

$$\varphi_{CCA}(x) = 1_{[C_{\alpha,p},1]}(x) = \begin{cases} 1 & \text{if } x \in [C_{\alpha,p},1] \\ 0 & \text{else} \end{cases} \qquad (19)$$

where $C_{\alpha,p} = \dfrac{1}{1+\chi^2_{\alpha,p}}$ and $\chi^2_{\alpha,p}$ is the χ^2_p quantile of order α. The function (19) can be written in an informal way as:

$$\varphi_{CCA}(\text{site depth}) = 1_{[C_{\alpha,p},1]}(\text{site depth}) = 1_{[0,\chi^2_{\alpha,p}]}(\text{site distance})$$
$$= \begin{cases} 1 & \text{if the site is inside the neighborhood} \\ 0 & \text{if the site is outside the neighborhood} \end{cases}$$

Therefore, the estimator $\hat{\beta}$ is the following:

$$\begin{aligned}\hat{\beta} &= \arg\min_{\beta} \sum_{i=1}^{N} w_i \left(\log Y_i - \beta \log X_i\right)^2 \\ &= \arg\min_{\beta} \sum_{\text{neighborhood sites } i} 1\left(\log Y_i - \beta \log X_i\right)^2 \\ &+ \arg\min_{\beta} \sum_{\text{non neighborhood sites } i} 0\left(\log Y_i - \beta \log X_i\right)^2\end{aligned} \tag{20}$$

3. The following weight function is used by Zuo et al. (2004):

$$\varphi_Z(x) = \begin{cases} \dfrac{\exp\left(-K\left(1-\left(x/C\right)\right)^2\right) - \exp(-K)}{1 - \exp(-K)} & \text{if } x < C \\ 1 & \text{elsewhere} \end{cases} \tag{21}$$

where C and K are positive constant coefficients.

4. With an extra coefficient $s \geq 1$ in equation (21), Lin and Chen (2006) used the following weight function:

$$\varphi_{LC}(x) = \begin{cases} \dfrac{\exp\left(-K\left(1-\left(x/C\right)^s\right)^2\right) - \exp(-K)}{1 - \exp(-K)} & \text{if } x < C \\ 1 & \text{elsewhere} \end{cases} \tag{22}$$

In functions (21) and (22) the constant C defines the support of the weight function and K represents the slope of the decay to zero. The CCA neighborhood approach may serve to guide the choice of the constant C. Indeed, the choice of a constant C may depend on the χ_p^2 quantile associated to the optimal α as $C_{\alpha,p} = \dfrac{1}{1+\chi^2_{\alpha,p}}$.

5. A simple linear function can also be used as a weight function:

$$\varphi_{Linear}(x) = \begin{cases} 0 & \text{if } x \le d_1 \\ \dfrac{x - d_1}{d_2 - d_1} & \text{if } d_1 \le x \le d_2 \\ 1 & \text{if } x \ge d_2 \end{cases} \quad (23)$$

with $d_2 > d_1 > 0$.

Special cases 1 and 2 along with the general depth-based approach are illustrated in Figure 1.

4. Case study

In this section, the approach proposed in Section 3 is applied on a real world data set and its performance is compared to that of the CCA approach. The case study on which the comparison is carried out concerns the hydrometric station network of the southern part of the province of Quebec, Canada. To be selected, each station in the data set must have a flood record of at least 15 years of data and its historical data must be homogenous, stationary and independent. The area of these catchments is larger than 200 km^2 and less than 100 000 km^2. Finally, a total of 151 stations located between the 45^o N and the 55^o N are selected. The geographical location of these stations is shown in Figure 2.

The variable selection is based on a previous study by Chokmani and Ouarda (2004). The selected variables are of three types: physiographical, meteorological, and hydrological. The physiographical variables are: basin area (AREA), mean basin slope (MBS) and the fraction of the basin area covered with lakes (FAL). The meteorological variables are annual mean total

precipitation (AMP) and annual mean degree days over 0^{o}C (AMD). The hydrological variables are represented by at-site flood quantiles QT corresponding to a return period T.

The data bases used in the present study and the at-site frequency analysis are presented in Kouider et al. (2002). The at-site flood frequency procedure includes the use of statistical tests of hypothesis to test for homogeneity, independence and stationarity, the fitting of several statistical distributions to the station data, and the use of goodness-of-fit tests to identify the most appropriate distribution for each station. It is important to mention that the approach proposed in the present paper does not require the use of a unique regional distribution. It is hence possible to use a different (most appropriate) distribution in each station. For more details concerning the at-site frequency analysis, the reader is referred to Kouider et al. (2002).

Eaton et al. (2002) reported that, when modeling the physical mechanism of drainage systems, scale effect should be eliminated from experiment data since it may have a negative impact on the results. To reduce the scale effect, flood quantiles QT are standardized by the basin area to obtain specific quantiles $QST = QT / AREA$. In this study, the 10-year (QS10) and the 100-year (QS100) specific flood quantiles are selected. The basic statistics of these variables are summarized in Table 1.

CCA application requires all variables to be transformed in order to be normalized and standardized. The appropriate normalizing transformations for the present data set were obtained by Chokmani and Ouarda (2004). A logarithmic transformation was used for the variables QS10, QS100, AREA, MBS, AMP and AMD, and a square-root transformation for FAL. The traditional

CCA procedure is applied with values of the coefficient α ranging in the interval [0, 1]. An optimal value of the coefficient α is selected according to minimum values of the relative bias (RB) and the relative root mean square error (RRMSE) of the jackknife resampling procedure as explained in Ouarda et al. (2001). The optimal value is found to be $\alpha = 0.25$ for the present case study. The scatter plots of sites in the hydrological canonical space (W1,W2) and the physio-meteorological canonical space (V1,V2) are illustrated in Figure 3.

5. Study methodology

The proposed depth-based approach described in Section 3 is applied to the above case study, and is compared to the CCA approach. Other methods are also considered in the comparison.

From equation (16), it can be seen that the depth computation method and the weight function are the two main elements of the estimation in the proposed approach. The depth can be computed in two ways: by the CCA Mahalanobis distance using directly equation (10) in equation (6); or by the iterative algorithm described in Section 3. Equation (6) indicates that the Mahalanobis depth and Mahalanobis distance are equivalent, that is, their values can be deduced from each other. Various combinations of depth computation and weight selection methods are considered in this study. The following methods are compared:

I. The uniform approach which uses all sites with the same importance.

II. The traditional CCA approach considered with the optimal value of α.

III. The depth-based approach considered with the following weight functions:

a. φ_Z with $K = 200$ and $C = 0.51$.

b. φ_{LC} with $s = 5$, $K = 200$ and $C = 0.51$.

c. φ_{LC} with $s = 2$, $K = 100$ and $C = 0.52$.

d. φ_{Linear} with $d_1 = 0.30$ and $d_2 = 0.80$.

IV. The CCA approach with iteration: It consists in the combination of the weight function φ_{CCA} with the optimal value of $\alpha = 0.25$ along with the Mahalanobis depth evaluated by the iterative algorithm. This combination can be seen as a special case of (III) with the specific weight function φ_{CCA}.

V. The depth-based approach without iteration: The depth is evaluated from the CCA Mahalanobis distance using equation (10). The weight function is φ_{LC} with the following coefficients:

a. $s = 2$, $K = 100$ and $C = 0.30$, which is similar in shape to φ_{CCA} with $\alpha = 0.25$.

b. $s = 2$, $K = 100$ and $C = 0.52$, which is one of the weight functions considered in (III).

The combinations (IV) and (V) are introduced to study the effect of the depth evaluation method and the weight function selection. The considered combinations are summarized in the left-hand part of Table 2 and the corresponding weight functions are illustrated in Figure 4.

In order to evaluate the performance of the various methods, a jackknife resampling procedure is used. It consists in considering each site as an ungauged one by removing it

temporarily from the region. The criteria employed to evaluate the performances of the approaches are the relative bias (RB) and the relative root mean square error (RRMSE) given respectively by:

$$RB = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i - \hat{y}_i}{y_i}\right) \quad (24)$$

$$RRMSE = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{y_i - \hat{y}_i}{y_i}\right)^2} \quad (25)$$

where y_i are the local realizations of the hydrological variable, $\hat{y}_i$ are the regional estimates and N is the number of sites in the data set.

6. Results and discussion

Results related to the various methods are summarized in Table 2. For all methods, results indicate that the RB and RRMSE are smaller for QS10 than QS100. Generally in frequency analysis, QST is more accurately estimated than QST' if T < T' since for small return periods, the corresponding quantile is close to the central body of the distribution. Hence, an important part of the data contributes to its estimation. Table 2 shows also that the results of the uniform method (I) are the worst. This confirms the need to use regional delineation techniques. The remaining methods are classified according to the depth evaluation procedure (iteration or direct CCA). The iterative depth evaluation leads to better results than the direct evaluation using CCA. Indeed, the results of the depth-based approach (III) are the best, and are followed by those of method (IV). In these two methods the depths are iteratively evaluated. Moreover, the differences in terms of RB and RRMSE are not significant between the various combinations in (III).

In methods (II) and (V), depths are evaluated directly using the CCA Mahalanobis distance. These methods lead to RB and RRMSE values that are larger than those obtained by methods (III) and (IV). In particular, the RB and RRMSE of methods (II) and (V) are significantly larger than those of methods (III). The RB and RRMSE of (IV) are slightly smaller than those of (II). In these last two methods, the same weight function φ_{CCA} is used. Hence, the CCA approach results can be slightly improved when depths are iteratively evaluated. However, the results from (V.b) are not satisfactory compared to the other methods except the uniform one (I). Note that combination (V.b) uses the same weight function than (III.c). Note also that the results of combinations (V.a) and (II) are very similar. In both these methods the depth is evaluated using the CCA Mahalanobis distance and the corresponding weight functions have similar shapes (see Figure 4). The results of combination (III.d) and the shape of the corresponding weight function φ_{Linear} suggest classifying the gauged sites into three classes: a class of sites to be excluded from the regional estimation procedure; another class of "intermediate" sites for which the contribution is gradually employed; and a last class of sites to be fully included in the estimation.

In the following, selected detailed results are presented. Figure 5 presents the evolution of the performance criteria as a function of the iteration number for methods (I), (II) and (III.c). For the uniform (I) and the traditional CCA (II) approaches, the criteria are represented by straight lines, since they are independent of the iteration number. For both QS10 and QS100, the criteria values of method (III.c) improve with the iteration number. Moreover, they seem to converge, after approximately 15 to 20 iterations, to the corresponding results given in Table 2. This illustrates the superiority of the depth-based approach over the traditional CCA approach. Note

that the computer running time of all the 40 iterations of (III.c) is comparable to that of the traditional CCA approach.

Figure 6 presents, for each site *i*, the relative error $\left(y_i - \hat{y}_i\right)/y_i$ related to the estimation of the specific quantiles QS10 and QS100 with respect to the basin area. It concerns the results of the CCA approach (II) and those of the 20th iteration of the depth-based approach (III.c). The 20th iteration is selected, as an example, since the criteria converge after 15 to 20 iterations. Using both estimation methods, large negative errors are observed for some sites such as number 46, 64, 66 and 148. In particular, sites 64 and 66 have small basin areas. It is generally observed that the relative errors obtained from the depth-based approach (III.c) are smaller than those obtained from the traditional CCA approach (II).

The following two elements can be used to explain some aspects related to the CCA approach:

1. Under the normality condition imposed on *X* and *Y*, it is implicitly assumed that the conditional canonical variable $\left(W \mid V = v_0\right)$ is $N\left(\Lambda v_0, I_p - \Lambda^2\right)$, see Ouarda et al. (2001). Consequently:

$$\mathrm{E}\left(W \mid V = v_0\right) = \Lambda v_0 \text{ and } \mathrm{Var}\left(W \mid V = v_0\right) = I_p - \Lambda^2 \qquad (26)$$

which suggests modeling the relationship between *W* and *V* by a linear regression model as follows:

$$W = \Lambda V + \varepsilon, \text{ with } \varepsilon \sim N\left(0, I_p - \Lambda^2\right) \qquad (27)$$

Usually, in the CCA approach, sites are presented in the hydrological canonical space (W1,W2) (see Figure 3). Following relation (27), it is of interest to present sites in the

canonical spaces (V1,W1) and (V2,W2). This is illustrated in Figure 7 for the considered data set. It shows that the relationship between V1 and W1 can be acceptably considered to be linear, and hence meets the model (27), whereas it is not the case for V2 and W2. The illustration in the space (V1,W1) is useful to get prior information about the estimation error for a given site.

2. The canonical hydrological value W_0 of the target-site is unknown. Hence, in the CCA approach, the Mahalanobis distance (10) is computed with respect to its expectation Λv_0 (it represents also its estimator using (27)). This influences the computation of the Mahalanobis distance, especially when the Mahalanobis distance $(W_0 - \Lambda v_0)' (I_p - \Lambda^2)^{-1} (W_0 - \Lambda v_0)$ between W_0 and Λv_0 is large. This is the case, for instance, for site number 66.

Figure 8 illustrates the Mahalanobis depths evaluated from (II) and from the 20th iteration of (III.c) for a selection of sites. The sites considered in Figure 8 are selected from Figures 6 and 7 to represent a wide variety of conditions. Indeed, from Figure 6, sites number 66, 122 and 148 are estimated with very high relative errors. These sites are located far from the straight line relating V1 and W1 as shown in Figure 7.a. However, sites number 4, 92 and 141 are accurately estimated as shown in Figure 6 and they are located near the regressive line in Figure 7.a. Figure 8 shows that the global shapes of the depth values by both methods are similar. However, differences in depth magnitude are observed, especially for site 122. This site has a very high value of basin area covered with lakes (FAL = 43%).

From the above results and analysis it appears that the first important element, for quantile regression estimation, is the accurate computation of depth values. The second important element is related to the selection of the weight function. Hence, the results of the classical delineation can be improved with accurate depth values (combination IV). However, generally less accurate depth values with an arbitrary weight function can not improve the classical results (combinations V). Consequently, the best combination of these two elements is a smooth weight function along with an iteratively evaluated Mahalanobis depth (combinations III).

7. Conclusions and future research directions

The present paper provides an adaptation of the statistical notion of "depth function" in regional flood frequency analysis. The depth-based approach is introduced in order to overcome some drawbacks of the classical methods and to improve their performances. Along with the flexibility and generality of the proposed method, it is shown that its use leads to improvements in traditional methods. Furthermore, the proposed methodology can be useful in other areas and disciplines of water resources where the regression model is applicable.

Introducing the statistical notion of depth function in regional flood frequency analysis leads to several directions for future work including the following:

– The determination of an optimal weight function remains an element to be developed. One option is to restrict the optimization problem to a specific parametric class of weight functions, such as φ_{LC} or φ_{Linear}. Then, according to the optimal value of the considered criterion, the coefficients of the weight function can be obtained. More flexible S-shaped

weight functions (such as the Gompertz function) represent a promising class to be considered for optimization.

- The estimation in the index flood model (Dalrymple, 1960) should be developed following a similar approach to the regression model. A comparison to the region of influence approach (Burn, 1990) is to be considered. Note that the region of influence approach uses special weight functions.
- In the present paper, the Mahalanobis depth function is considered for its simplicity and its link to the CCA approach. It is of interest to consider other depth functions and to compare the corresponding results. Note that Lin and Chen (2006) presented an estimator that is similar to the one given by equation (16) using the projection depth function (equation 4). They studied its asymptotic properties, including consistency, normality and robustness.
- The estimation of the uncertainty associated to regional quantiles represents an important topic that is not often treated in the literature. It would be of value, in future efforts, to study the uncertainty associated to the approach proposed in the present paper. It is important to mention that some of the methods in the literature (such as Generalized and Weighted least squares approaches) allow to carry out an analysis of the uncertainty (see e.g. Madsen and Rosbjerg, 1997). However, with these approaches, the estimation of the uncertainty of the T-year event represents only the uncertainty on the regressive part without integrating the uncertainty associated to the definition of the homogeneous regions. The depth-based method does not suffer from this problem and hence it requires only one global uncertainty evaluation.
- Information concerning the spatial correlation between the various sites should be integrated in the depth-based flood frequency procedure.

Acknowledgments

Financial support for this study was graciously provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada, and the Canada Research Chair Program. The authors wish to thank the Editor in Chief, the Associate Editor and the two anonymous reviewers for their useful comments which led to the improvement of the paper.

Bibliography

Arcones, M. A.; Chen, Z. and Giné, E. (1994) Estimators Related to *U*-Processes with Applications to Multivariate Medians: Asymptotic Normality. *Ann. Statist.*, **22**, 1460-1477.

Alila, Y. (1999) A hierarchical approach for the regionalization of precipitation annual maxima in Canada. *J. Geophys. Res.*, **104**, 31,645-31,655.

Alila, Y. (2000) Regional rainfall depth-duration-frequency equations for Canada. *Water Resour. Res.*, **36**, 1767-1778.

Burn, D. H. (1990) Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resour. Res.*, **26**, 2257-2265.

Caplin, A. and Nalebuff, B. (1988) On 64%-majority rule. *Econometrica*, **56**, 787-814.

Caplin, A. and Nalebuff, B. (1991a) Aggregation and social choice: A mean voter theorem. *Econometrica*, **59**, 1-23.

Caplin, A. and Nalebuff, B. (1991b) Aggregation and imperfect competition: On the existence of equilibrium. *Econometrica*, **59**, 25-59.

Chebana, F. and Ouarda, T.B.M.J. (2007) Multivariate *L*-moment homogeneity test. *Water Resour. Res.*, **43**, W08406, doi:10.1029/2006WR005639.

Chokmani, K. and Ouarda T.B.M.J. (2004) Physiographical space-based kriging for regional flood frequency estimation at ungauged sites. *Water Resour. Res.*, **40**, W12514, doi: 10.1029/2003WR002983.

Dalrymple, T. (1960) Flood frequency methods. *United States Geological Survey Water-Supply Paper*, **1543** A, 11-51.

Durrans, S. R. and Tomic, S. (1996) Regionalization of low-flow frequency estimates: an Alabama case study. *Water Resour. Bull.*, **32**, 23-37.

Eaton, B.; Church, M. and Ham, D. (2002) Scaling and regionalization of flood flows in British Columbia, Canada. *Hydrol. Processes*, **16**, 3245– 3263.

Ghosh, A. K. and Chaudhuri, P. (2005) On Maximum Depth and Related Classifiers. *Scand. J. Statist.*, **32**, 327–350.

GREHYS (1996a) Presentation and review of some methods for regional flood frequency analysis. *J. Hydrology,* **186**, 63-84.

GREHYS (1996b) Inter-comparison of regional flood frequency procedures for Canadian rivers. *J. Hydrology,* **186**, 85-103.

Hosking, J. R. M. and Wallis, J. R. (1993) Some statistics useful in regional frequency analysis. *Water Resour. Res.,* **29**, 271-282.

Kouider, A., Gingras, H.; Ouarda, T. B. M. J.; Ristic-Rudolf, Z. and Bobée, B. (2002) Analyse fréquentielle locale et régionale et cartographie des crues au Québec [In French], Rep. R-627-el, Eau, Terre, et Environ., INRS, Ste-Foy, Que., Canada.

Lin, L. and Chen, M. H. (2006) Robust estimating equation based on statistical depth. *Statist. Papers*, **47**, 263-278.

Liu, R. Y. (1990) On a notion of data depth based on random simplices. *Ann. Statist.,* **18**, 405-414.

Liu, R. Y. (1992) Data depth and multivariate rank tests. In L-l *Statistics and Related Methods* (Y. Dodge, ed.) 279 294. North-Holland, Amsterdam.

Liu, R. Y. and Singh, K. (1993) A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.*, **88**, 257-260.

Liu, R. Y. (1995) Control Charts for Multivariate Processes. *J. Amer. Statist. Assoc.*, **90**, 1380-1388.

Liu, R. Y.; Parelius, J. M. and Singh, K. (1999) Multivariate analysis by data depth: descriptive statistics, graphics and inference, (with discussion and a rejoinder by Liu and Singh). *Ann. Statist.,* **27**, 783-858.

Madsen, H. and Rosbjerg, D. (1997) Generalized least squares and empirical Bayes estimation in regional partial duration series index-flood modeling. *Water Resources Research*, **33**, 771-781.

Mahalanobis, P. C. (1936) On the generalized distance in statistics. *Proc*. *Nat*. *Acad*. *Sci*. *India,* **12,** 49-55.

Massé, J.-C. (2004) Asymptotics for the Tukey depth process, with an application to a multivariate trimmed mean. *Bernoulli*, **10**, 1-23.

Miller, K. ; Ramaswami, S. ; Rousseeuw, P. ; Sellares, T. ; Souvaine, D. ; Streinu, I. and Struyf, A. (2003) Efficient Computation of Location Depth Contours by Methods of Combinatorial Geometry. *Statistics and Computing,* **13**, 153-162.

Mizera, I. (2002) On depth and deep points: a calculus. *Ann. Statist.,* **30**, 1681–1736.

Mizera, I. and Müller, C. H. (2004) Location-Scale Depth (with discussion). *J. Amer*. *Statist*. *Assoc*., **99**, 949-966.

Muirhead, R. J. (1982) *Aspect of Multivariate Statistical Theory*. John Wiley, Hoboken, N. J.

Nguyen, V.-T.-V. and Pandey, G. (1996) A new approach to regional estimation of floods in Quebec. In: Delisle, C.E., Bouchard, M.A. (Eds.), Proceedings of the 49th Annual Conference of the CWRA, June 26–28, Quebec City. Collection Environnement de l'U. de M., 587–596.

Oja, H. (1983) Descriptive statistics for multivariate distributions. *Statist*. *Probab*. *Lett*., **1,** 327-332.

Ouarda, T. B. M. J. and Ashkar, F. (1994) Regional multiple regression flood frequency estimation by the Peaks-Over-Threshold method. Internal Report, Department of Mathematics, University of Moncton, Moncton, New Brunswick, Canada, Research report for the Strategic Grant No. STR0118482 of NSERC, 20 pp.

Ouarda, T. B. M. J.; Haché, M.; Bruneau, P. and Bobée, B. (2000) Regional Flood Peak and Volume Estimation in a Northern Canadian Basin. *ASCE J. Cold. Reg. Engrg.,* **14**, 176-191.

Ouarda, T. B. M. J.; Girard, C.; Cavadias, G. S. and Bobée, B. (2001) Regional flood frequency estimation with canonical correlation analysis. *J. Hydrology,* **254,** 157-173.

Ouarda, T. B. M. J.; Cunderlik, J. M.; St-Hilaire, A.; Barbet, M.; Bruneau, P.; Bobée, B. (2006) Data-based comparison of seasonality-based regional flood frequency methods. *J. Hydrology*, **330**, 329-339.

Rencher, A. C. (2002) *Methods of multivariate analysis*. Second edition. Wiley Series in Probability and Statistics. John Wiley & Sons*,* New York.

Rousseeuw, P. J. and Hubert, M. (1999) Regression depth. *J. Amer. Statist. Assoc.*, **94,** 389-433.

Stedinger, J.R. and Tasker, G. (1986) Regional hydrologic analysis, 2, Model-error estimators, estimation of sigma and log Pearson type 3 distributions. *Water Resour. Res.*, **22**, 1487-1499.

Tukey, J. (1975) Mathematics and picturing data. In *Proceedings of the 1975 International Congress of Mathematics,* **2,** 523-531.

van der Vaart, A.W. (1998) *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

Zuo, Y. and Serfling, R. (2000) General notions of statistical depth function. *Ann. Statist.,* **28** 461–482.

Zuo, Y.; Cui, H. J. and Young, D. (2004) Influence function and maximum bias of projection depth bused estimators. *Ann. Statist.,* **32**, 189-218.

Zuo, Y. and Cui, H. (2005) Depth weighted scatter estimators. *Ann. Statist.,* **33**, 381–413.

Table 1. Descriptive statistics of hydrological, physiographical and meteorological variables

Variable	Unit	Min	Mean	Max	Standard deviation
MBS	%	0.96	2.43	6.81	0.99
FAL	%	0.00	7.72	47.00	7.99
AMP	mm	646	988	1534	154
AMD	degree-day	8589	16346	29631	5382
AREA	km^2	208	6255	96600	11716
QS10	$m^3/s.km^2$	0.03	0.22	0.53	0.13
QS100	$m^3/s.km^2$	0.03	0.31	0.94	0.20

1 **Table 2. Quantile estimation results with the various methods**

2

				QS10		*QS100*	
	Combination approach	**Depth computation**	**Weight function**	**RB (%)**	**RRMSE (%)**	**RB (%)**	**RRMSE (%)**
(I)	Uniform	-	φ_U	-8.60	54.94	-10.93	64.05
(II)	Traditional CCA	CCA	φ_{CCA} optimal α	-7.54	44.62	-8.14	51.84
(III)	Depth-based	Iteration					
a)			φ_Z K = 200, C =0.51	-4.37	39.17	-3.51	44.92
b)			φ_{LC} s=5, K = 200, C =0.51	-4.47	38.78	-3.75	44.53
c)			φ_{LC} s=2, K = 100, C =0.52	-5.22	38.94	-4.68	44.72
d)			φ_{Linear} d_1= 0.30, d_2 =0.80	-3.83	38.78	-2.52	44.57
(IV)	CCA with iteration	Iteration	φ_{CCA} optimal α	-5.52	42.27	-4.98	49.02
(V)	Depth-based without iteration	CCA					
a)			φ_{LC} s =2, K= 100, C =0.30	-7.75	44.81	-8.38	51.82
b)			φ_{LC} s =2, K= 100, C =0.52	-8.26	47.63	-9.88	57.02

3

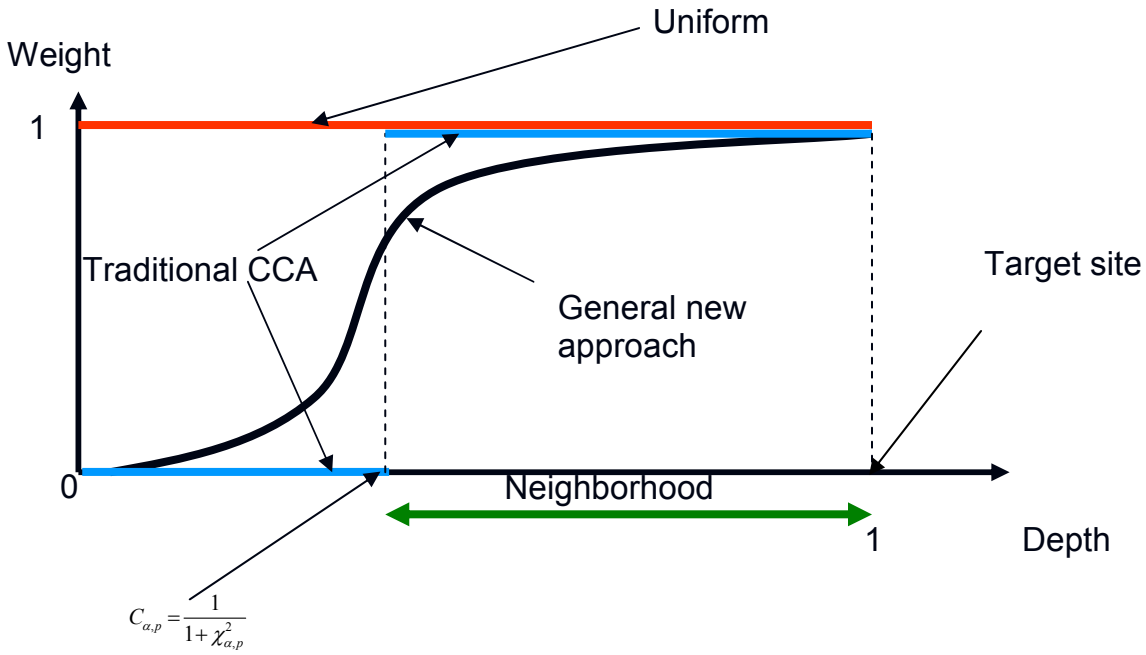

Figure 1. Illustration of different regionalization approaches with the corresponding weight functions

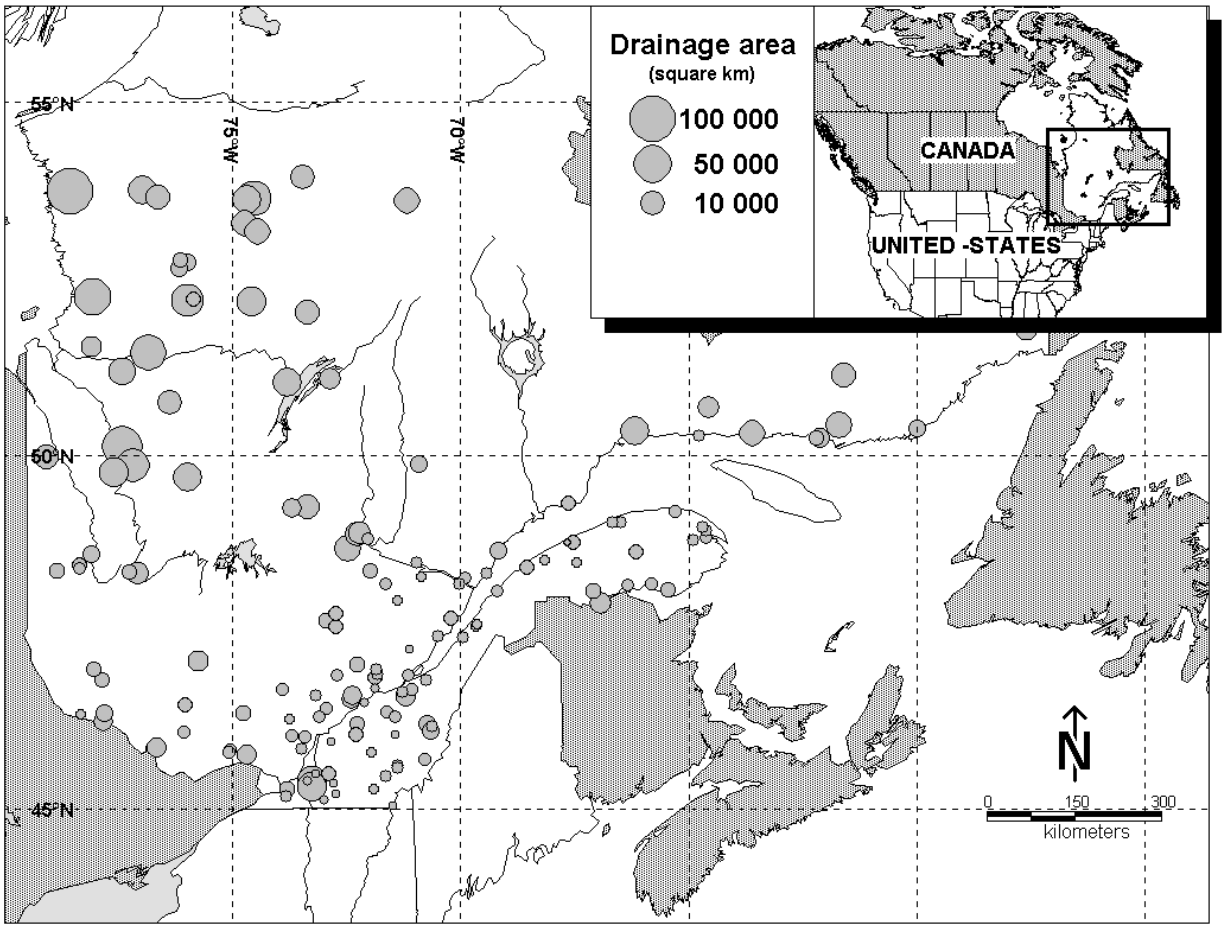

Figure 2. Geographical location of the studied sites in the province of Quebec, Canada

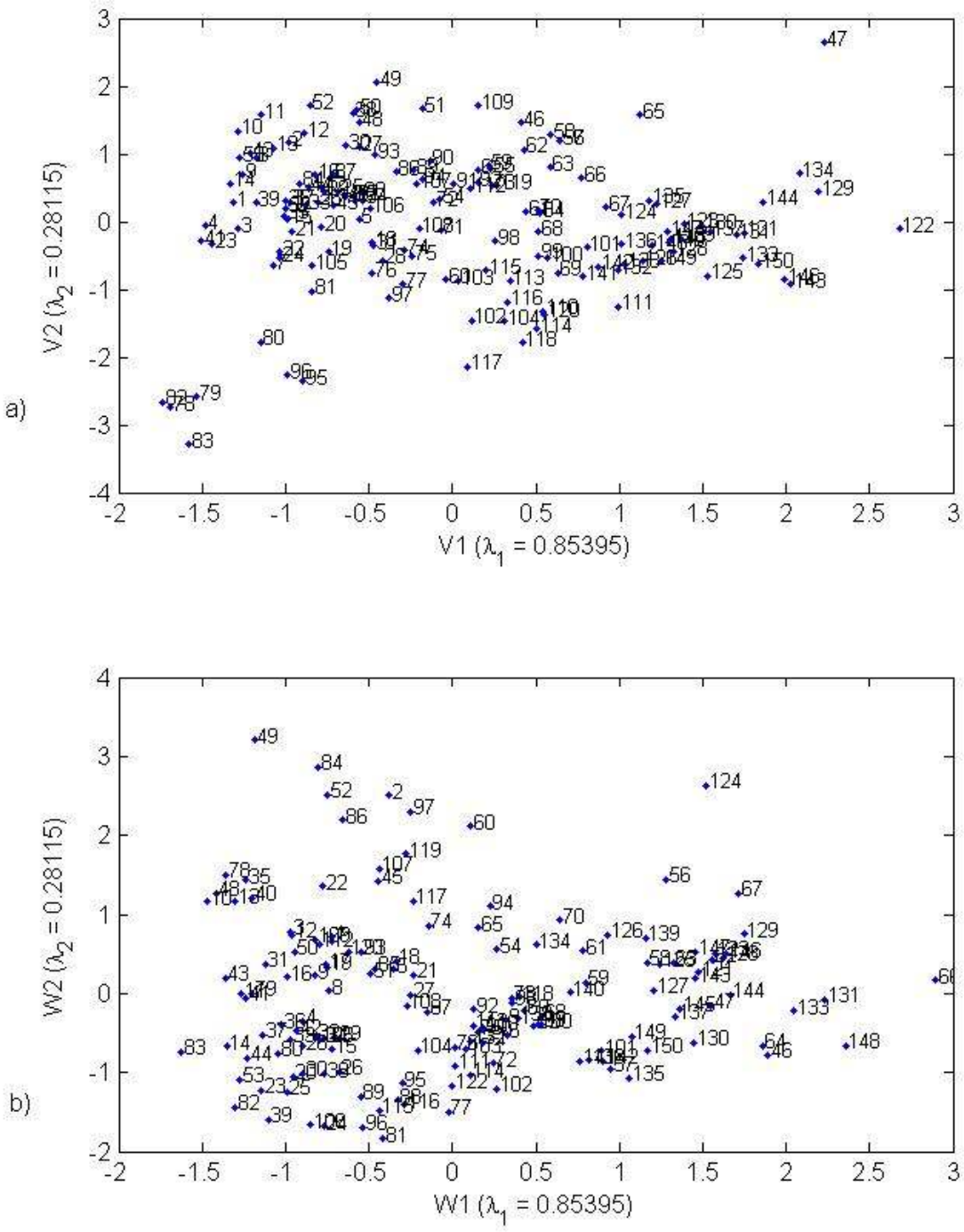

Figure 3. Data set in the canonical spaces (a) physio-meteorological (b) hydrological

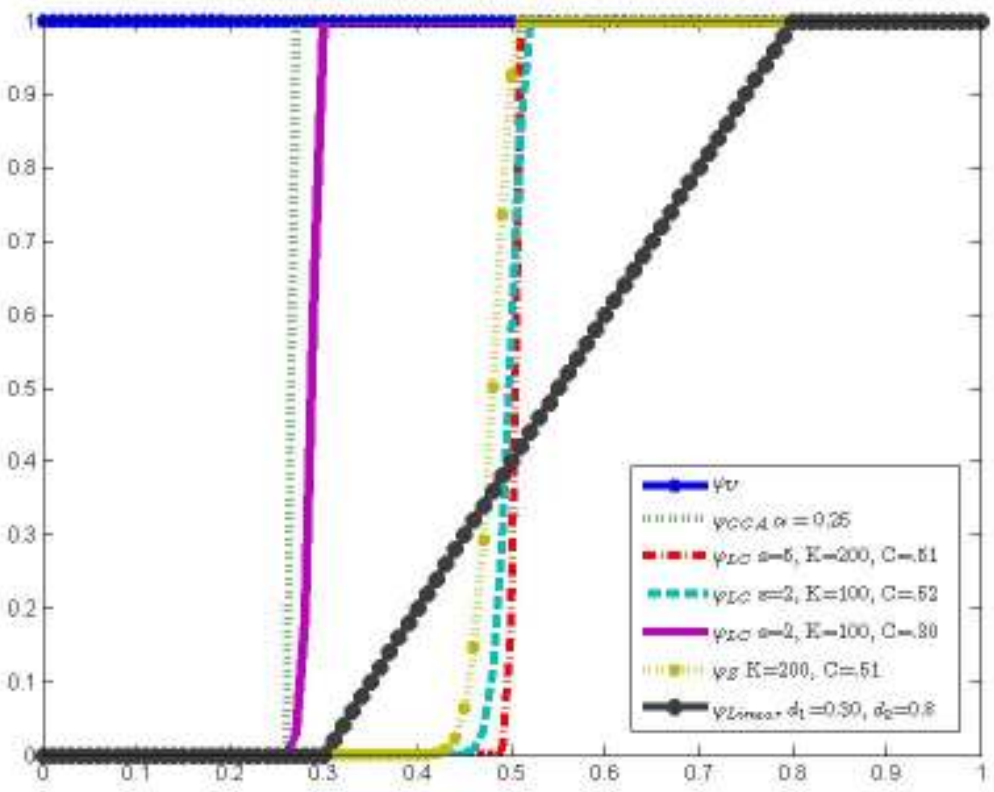

Figure 4. Weight functions used in the comparative study

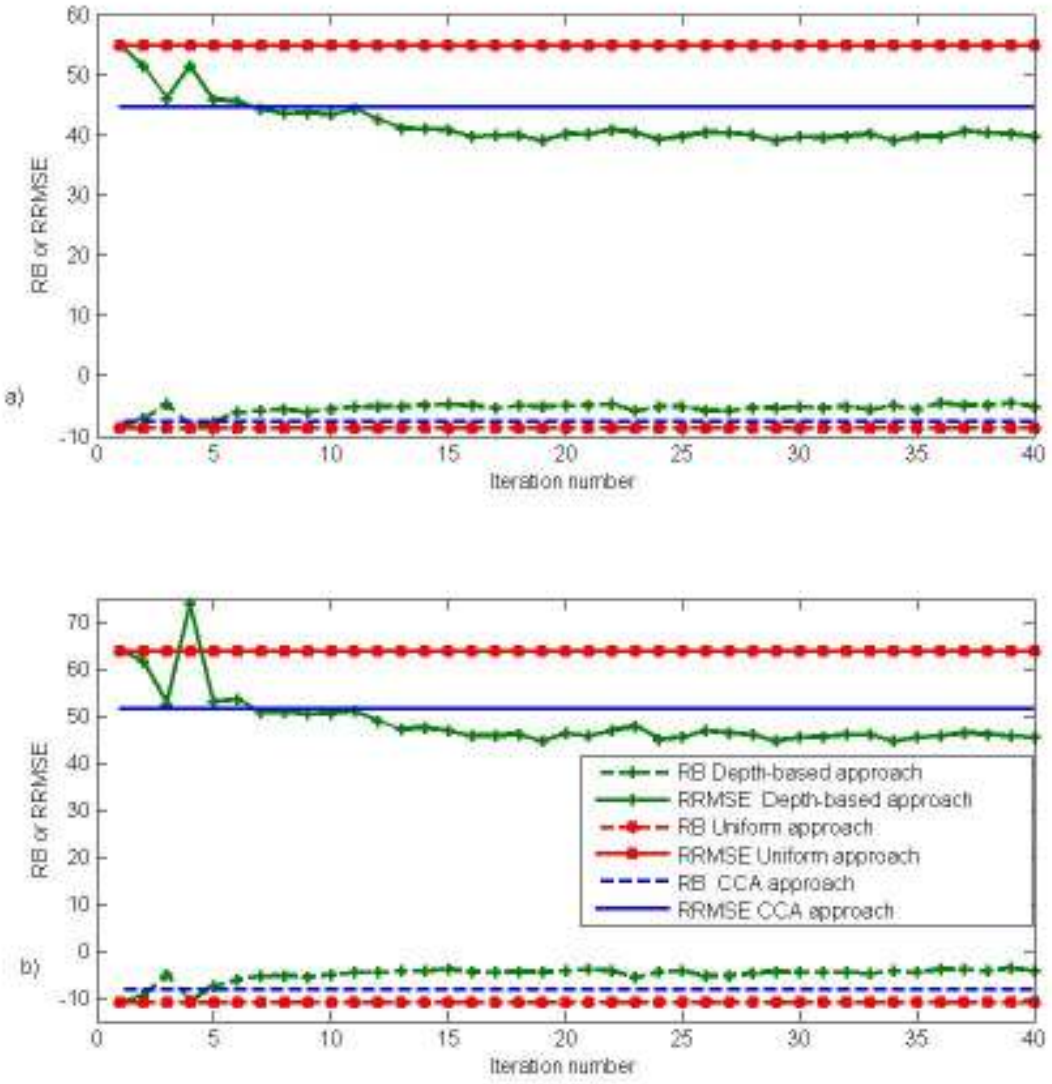

Figure 5. RB and RRMSE using the uniform (I), the CCA (II) and the depth-based (III.c) methods for the estimation of (a) QS10 and (b) QS100

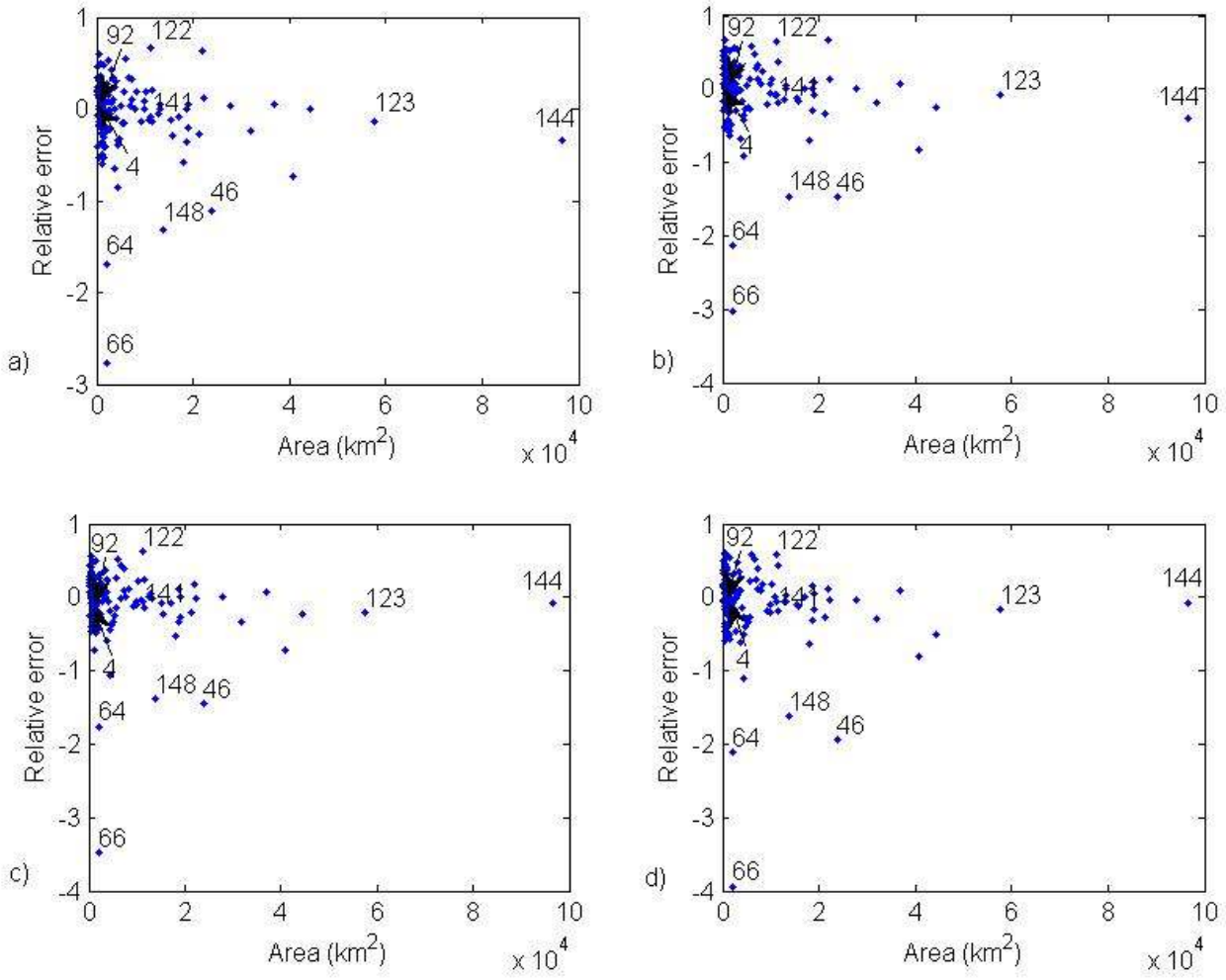

Figure 6. Quantile estimation relative errors of: a) QS10 from the 20th iteration of the method (III.c), b) QS100 from the 20th iteration of the method (III.c), c) QS10 from the CCA method (II), d) QS100 from the CCA method (II). Selected sites are indicated by their respective numbers

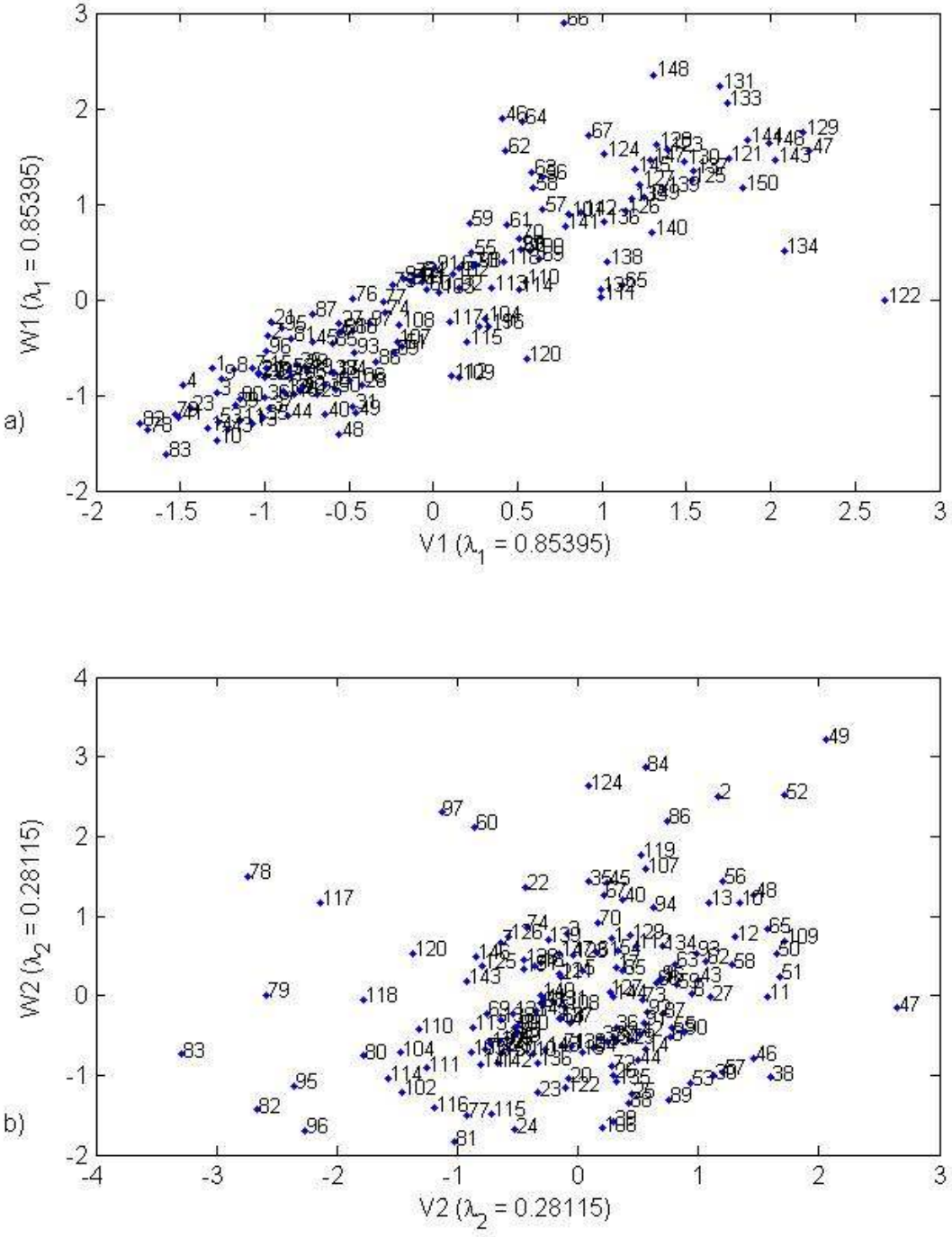

Figure 7. Scatter plot of the data set in the canonical spaces (V1,W1) and (V2,W2)

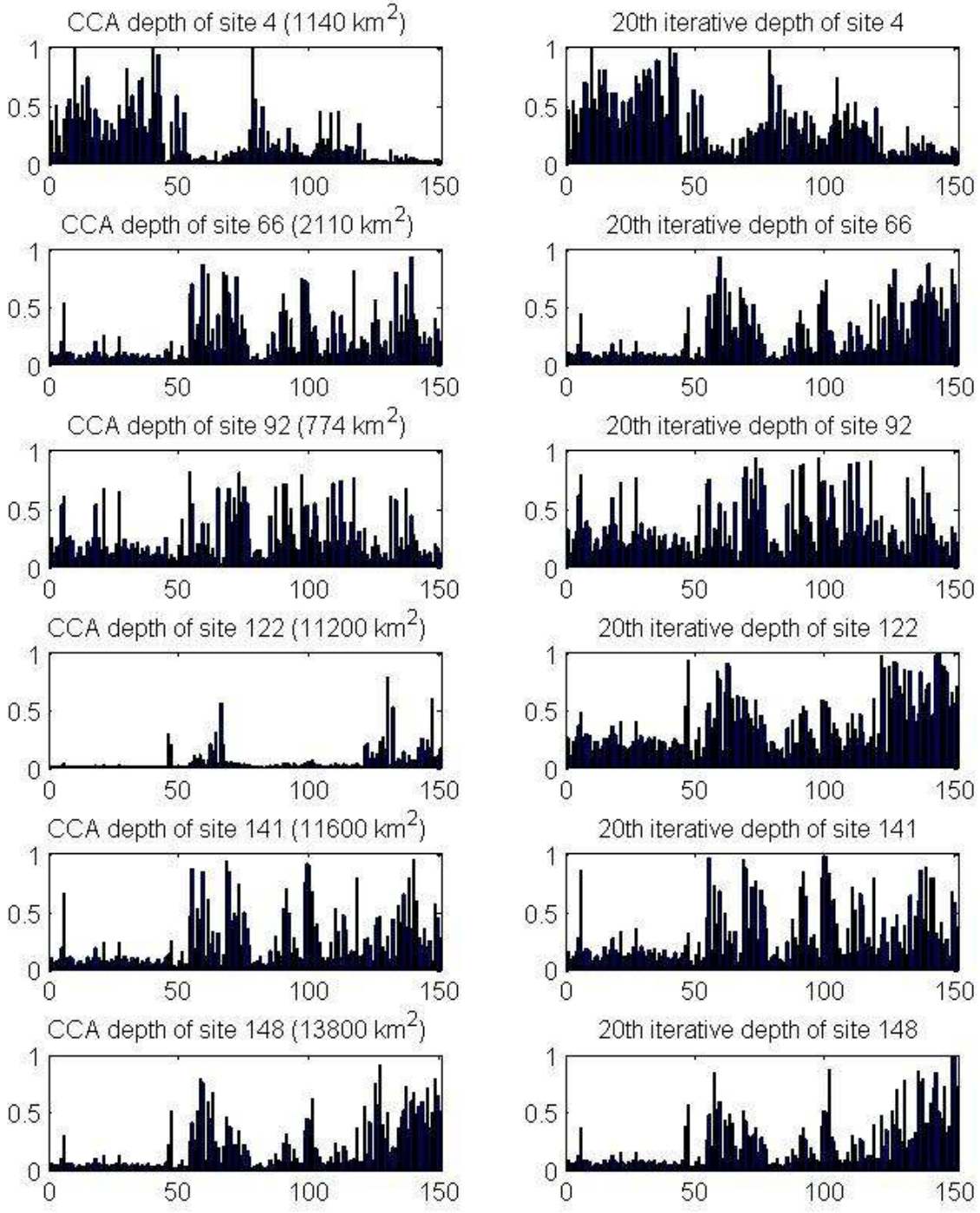

Figure 8. Depth values from the traditional CCA (II) and from the 20th iteration of the method (III.c) for a selection of sites. Values between parentheses represent the area of the corresponding site