Optimal depth-based regional frequency analysis

H. Wazneh, F. Chebana and T.B.M.J. Ouarda

MS No.: hess-2012-537

Dear Editor,

Please find enclosed the response to all comments made by the two Referees to the manuscript "Optimal depth-based regional frequency analysis" that is being considered for publication in Hydrology and Earth System Science.

The authors gratefully acknowledge the helpful comments that have contributed to the improvement of the paper. Detailed replies to these comments are provided below**.**

Should you have any questions or require further information concerning this revision please contact me at the address listed below.

Sincerely,

Hussein Wazneh
INRS-ETE, University of Quebec
490, de la Couronne, Québec (Québec)
G1K 9A9, CANADA
Email: hussein.wazneh@ete.inrs.ca

Reply to Referee 1

The authors are grateful to the referee for his/her comments which contributed to improve the quality of the paper. The authors provide hereafter the answers to the referee's comments.

Comment 1: The paper should clearly outline that the criteria already introduced in section 3.2 can also be used as KPI to compare different RFA methodologies.

Answer: The authors agree with the referee that the criteria defined in section 3.2 Eq. (17) and (18) can be used as key performance indicators (KPI- quantitative indicators) to compare different RFA approaches. This is indicated in the revised version by adding the following lines:

"In the hydrological framework, the previously defined criteria are used as key performance indicators (KPI) to compare different RFA approaches (see for instance, Gaal et al., 2008, Shu, and Ouarda 2007)."

References:

L. Gaal, J. Kysely, and J. Szolga (2008): Region-of-influence approach to a frequency analysis of heavy precipitation in Slovakia, Hydrol. Earth Syst. Sci., 12, 825–839, 2008,

C. Shu, and T. B. M. J. Ouarda (2007): Flood frequency analysis at ungauged sites using artificial neural networks in canonical correlation analysis physiographic space, Water Resour. Res., 43, W07438, doi:10.1029/2006WR005142.

Comment 2: Thereafter typical ranges for a mean, good and very good estimations should be defined. This would create the basis for a better assessment of the results of the case study.

Answer: The authors agree with the referee's comment about the eventual usefulness of defining typical ranges to classify the performance of the considered models. However, to the knowledge of the authors there is no objective way to define these ranges. In addition, this issue is general and goes beyond the scope of this paper which aims to optimize the performance of the DBRFA approach and compare it with available models.

Comment 3: The case study compares the new approach with only one other classical RFA method. A comparison with further approaches would increase the credibility of the case study.

Answer: The authors thank the referee for raising this point. In this paper the authors compared the results of the optimal DBRFA approach with the basic formulation of one of the most popular RFA approaches (that is linear regression model with canonical correlation analysis LR-CCA). Indeed, comparing the optimal DBRFA with further approaches increases the credibility of its performance. However, to perform such a comparison, a number of competitive approaches must be introduced. Therefore, the paper becomes too long (it is already long). On the other hand, such comparison requires either at-site datasets or existed results associated to those approaches for regions to be compared. However, this is not the case for the Texas and Arkansas regions. The authors suggest adding a discussion in the revised version including the following table which summarizes the obtained results for the Quebec region with different available approaches.

Results of RFA approaches applied in the southern of Quebec

Method	QS10 RB (%)	QS10 RRMSE (%)	QS100 RB (%)	QS100 RRMSE (%)
Linear regression[1] (LR)	-9	55	-11	64
Nonlinear regression (NLR)[3]	-9	61	-12	70
Nonlinear regression with regionalisation approach (NLR-R)[3]	-19	67	-24	79
LR-Canonical correlation Analysis (LR-CCA)[1]	-7	44	-8	52
Kriging in the CCA Physiographical Space[2]	-20	66	-27	86
Kriging in the PCA Physiographical Space[2]	-16	51	-23	70
Adaptive Neuro-Fuzzy Inference Systems (ANFIS)[3]	-8	57	-14	64
Artificial Neural Networks (ANN)[3]	-8	53	-10	60
Single ANN-CCA space (SANN-CCA)[4]	-5	38	-4	46
Ensemble ANN –(EANN)[4]	-7	44	-10	60
Ensemble ANN-CCA space (EANN-CCA)	-5	**37**	-6	45
Optimal DBRFA[1]	**-3**	38	**-2**	**44**

Best results are in bold character

[1] H. Wazneh, F. Chebana and T.B.J.M. Ouarda, (2013): Optimal depth-based regional frequency analysis, HESSD, 10, 519-555, 2013, doi:10.5194/hessd-10-519-2013.

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

[3] C. Shu, and T.B.M.J. Ouarda, (2008): Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system, J. Hydrol., 349, 31-43, 2008.

[4] C. Shu, and T. B. M. J. Ouarda, (2007): Flood frequency analysis at ungauged sites using artificial neural networks in canonical correlation analysis physiographic space, Water Resour. Res., 43, W07438, doi:10.1029/2006WR005142.

Comment 4: The results of the case study should be discussed in a more comprehensive way.

Answer: The revised version includes a more extensive discussion of the results of the case study.

<u>**Reply to Referee 2**</u>

The authors are grateful to the referee for his/her comments which contributed to improve the quality of the paper. The authors provide hereafter the answers to the referee's comments.

Comment 1: The method is poorly explained. It took me quite some time to understand a bit what the authors have done. My difficulties are explained below. The presentation needs much improvement.

Answer: The authors believe that the corrections and modifications in the revised version clarify and improve the readability of the method (please refer to the answers of specific comments below). For instance, the background section is modified to become clear, short, specific and easily understood (see answers to comments 3, 6 and 9). In the methodology section, the optimization procedure is rewritten and discussed in a more comprehensive way (see answer to comment 2) and a diagram is added to provide an overview of the procedure (see also answer to comment 5).

Major comments

Comment 2: In the introduction the authors state that they are interested in design event quantiles at ungauged sites (p. 520, l. 23, 24). In line with this, the quantiles at the start of their procedure are unknown (p. 531, l. 17). Then an iterative procedure is described on p. 532, which ends with an estimate of the quantiles at the target site on top of p. 533. However, on the same page two performance indices are given, in which there are N sites for which a local quantile estimate is available. This suggests that the authors have applied the iterative procedure N times to the quantiles of gauged sites rather than an ungauged target site.

Answer: The referee raises an important point which requires more description. Indeed, the iterative procedure described on p. 531-532 aims to find the optimal weight function in DBRFA approach with respect to predefined criteria (l. 6 p. 531). The authors recognize that the way to use the optimal DBRFA to estimate the quantile of target sites was not clearly presented in the original manuscript. For this reason and in order to clarify this concept, in the revision version, the general procedure (p. 531 l.5 through p534 l.7) is rewritten as follows:

“ **3.1 General procedure**

In order to find the optimal weight function $\varphi_{Optimal}$ in the DBRFA approach, the procedure is composed of three main steps. They are summarized as follows:

i. For a given class of weight function φ and a set of gauged sites (region), use a jackknife procedure to assess the regional flood quantile estimators (Eq. 8) for the sites of the region using the DBRFA approach. These estimators depend on the weight function φ through its coefficient;

ii. For a pre-selected criterion, calculate its value to quantify the performance of the estimates obtained from step i;

iii. Using an optimization algorithm, optimize the criterion (objective function) calculated in step ii. The outputs of this step are $\varphi_{Optimal}$ and the value of the selected criterion.

3.2 Description of the procedure

In the first step of the procedure, we use a jackknife resampling procedure to assess the regional flood quantile estimators for the sites of the region. This jackknife procedure consists in considering each site l ($l = 1, \ldots N$) in the region as an ungauged one by removing it temporarily from the region (i.e. we assume that the hydrological variable Y_l of site l is unknown and the physio-meteorological variable X_l is known since it can be easily estimated from existing physiographic maps and climatic data). Then we calculate the regional estimator $\left(\hat{Y}_l\right)_\varphi$ of site l by the iterative DBRFA approach, using the $N-1$ remaining sites, which is related to the given weight function φ. The parameters of the starting estimator (initial point) of DBRFA, denoted by $\hat{\beta}_{1,l}$ and $\hat{\Gamma}_{1,l}$, are calculated by assuming that $X = X^{<-l>}$, $Y = Y^{<-l>}$ and $\Omega = I_{N-1}$ in (10) and (11), where $X^{<-l>}$ represents the matrix of physio-meteorological variables excluding site l, $Y^{<-l>}$ is the matrix of hydrological variables excluding site l and I_{N-1} is the identity matrix of dimension $(N-1)\times(N-1)$. The starting estimator $\left(\hat{Y}_{1,l}\right)_\varphi$ is obtained by replacing β with $\hat{\beta}_{1,l}$ in (8). Then for each DBRFA iteration k, $k = 2, 3, .., k_{iter}$, we calculate the Mahalanobis depth (2) of the gauged site i, $i = 1, ..., N-1$, with respect to the ungauged site l denoted by

$\left(D_{k,(i,l)}\right)_\varphi = MHD_{\left(\hat{\Gamma}_{k-1,l}\right)_\varphi}\left(\log Y_i ; \left(\log \hat{Y}_{k-1,l}\right)_\varphi\right)$. The number of iterations k_{iter} is fixed to ensure the convergence of the depth function (generally k_{iter} = 25 can be appropriate). The weight matrix at iteration k is defined by applying the function φ to the depth calculated at this iteration. The parameters of the MR model at the k^{th} iteration are estimated by:

$$\left(\hat{\beta}_{k,l}\right)_\varphi = \left(\left(\log X^{<-l>}\right)' \left(\Omega_{k,l}\right)_\varphi \left(\log X^{<-l>}\right)\right)^{-1} \left(\log X^{<-l>}\right)' \left(\Omega_{k,l}\right)_\varphi \log Y^{<-l>} \tag{13}$$

$$\left(\hat{\Gamma}_{k,l}\right)_\varphi = \frac{\left(\log Y^{<-l>} - \left(\log X^{<-l>}\right)\left(\hat{\beta}_{k,l}\right)_\varphi\right)' \left(\log Y^{<-l>} - \left(\log X^{<-l>}\right)\left(\hat{\beta}_{k,l}\right)_\varphi\right)}{\left(N-1\right) - r - 1} \tag{14}$$

where $\left(\Omega_{k,l}\right)_\varphi$ is a N-1 diagonal matrix with elements:

$$\varphi\left[\left(D_{k,(1,l)}\right)_\varphi\right], \ldots\ldots, \varphi\left[\left(D_{k,(N-1,l)}\right)_\varphi\right] \tag{15}$$

Note that all these parameters depend on φ. Then, the regional quantile estimator for the site l in this iteration is:

$$\left(\hat{Y}_{k,l}\right)_\varphi = \exp\left[\left(\log X_l\right)\left(\hat{\beta}_{k,l}\right)_\varphi\right] \tag{16}$$

In the second step of the procedure, we use the regional estimators at the last iteration since their associated estimation errors are the minimum possible by construction. Consequently, in order to simplify the notations in the rest of this paper, we denote $\left(\hat{Y}_1\right)_\varphi = \left(\hat{Y}_{k_{iter},1}\right)_\varphi, \ldots, \left(\hat{Y}_l\right)_\varphi = \left(\hat{Y}_{k_{iter},l}\right)_\varphi, \ldots, \left(\hat{Y}_N\right)_\varphi = \left(\hat{Y}_{k_{iter},N}\right)_\varphi$.

After calculating $\left(\hat{Y}_l\right)_\varphi$, $l = 1,..,N$ in step i, we consider and evaluate one or several performance criteria in step ii. The considered criteria are employed as objective functions in the optimization step iii.

The relative bias (RB) and the relative root mean square error (RRMSE) are widely used in hydrology, particularly in RFA, as criteria to evaluate model performances. These two criteria are defined by:

$$RB_{\varphi} = 100 \times \frac{1}{N} \sum_{l=1}^{N} \left(\frac{Y_l - \left(\hat{Y}_l \right)_{\varphi}}{Y_l} \right) \tag{17}$$

$$RRMSE_{\varphi} = 100 \times \sqrt{\frac{1}{N-1} \sum_{l=1}^{N} \left(\frac{Y_l - \left(\hat{Y}_l \right)_{\varphi}}{Y_l} \right)^2} \tag{18}$$

where Y_l is the local quantile estimation for the l^{th} site, $\left(\hat{Y}_l \right)_{\varphi}$ is the regional estimation by DBRFA approach according to φ and excluding site l, and N is the number of sites in the region. The RB_{φ} measures the tendency of quantile estimates to be uniformly too high or too low across the whole region and the RRMSE_{φ} measures the overall deviation of estimated quantiles from true quantiles (e.g., Hosking and Wallis, 1997). Note that other criteria can also be considered such as the Nash criterion (NASH) and the coefficient of determination (R^2).

Finally in step iii, we apply an optimization algorithm on the selected and evaluated criterion in step ii. The algorithms to be considered are presented in the introduction section. The formulation of the criteria to be optimized, generally complex and non-explicit, suggests the use of zero-order algorithms. The application of these algorithms allows to find the optimal function $\varphi_{Optimal}$ with respect to selected criteria."

In order to explain how the user can use the optimal DBRFA to estimate the quantile at the target site (ungauged) the following lines are added to the revised version:

"The procedure described above aims to calculate $\varphi_{Optimal}$ according to the desired criterion. In order to estimate the quantile Y_u of an ungauged site u using the optimal DBRFA approach, the user simply repeats step i of the procedure without excluding any site and while fixing the weight function, i.e. step i with $\varphi = \varphi_{Optimal}$."

Comment 3: The definition of vectors and matrices in the description of the weighted least squares procedure needs more care. From p. 528, l. 4-9, I understand that Y and log Y are row vectors with s elements (a 1 * s matrix) and log X is a row vector with r+1 elements [a 1*(r+1) matrix]. Then, if beta is an (r+1)*s matrix, the product log X * beta is indeed a 1*s matrix. However, it is not possible to write beta * log X as is done on p. 528, line 15, unless s = 1. Furthermore, log X and log Y on p. 528, l. 16 and p. 528, l. 20 [Eq. (11)] seem to be no longer row vectors but matrices with N rows (l.16) or N columns (l. 20)

Answer: The authors thank the referee for pointing out the need for clear definitions of vectors and matrices on p. 528. In fact, *X* and *Y* are two matrices with dimensions $N \times (r+1)$ and $N \times s$ respectively, where *N* is the number of sites in the region, *r* is the number of physio-meteorological variables and *s* is the number of quantiles (hydrological variables). To be accurate and clear about this point, lines 4 through 22 (p.528) of the original manuscript are replaced by the following in the revised version:

"Let *s* be the number of quantiles *QT* corresponding to *s* return periods and *N* be the total number of sites in the region. A matrix of hydrological variables $Y = (QT_1, QT_2, ..., QT_s)$ of dimension $N \times s$ is then constructed. With a log-transformation in (7) we obtain the multivariate log-linear model in the following form:

$$\log Y = (\log X)\beta + \varepsilon \qquad (8)$$

where $\log X = (1, \log A_1, \log A_2, ..., \log A_r)$ is the $N \times (r+1)$ matrix formed by (*r*) physio-meteorological variables series, β is the $(r+1) \times s$ matrix of parameters and $\varepsilon = (\varepsilon^1, ..., \varepsilon^s)$ is the $N \times s$ matrix that represents the model error (residual) with null mean vectors and variance-covariance matrix Γ:

$$E(\varepsilon) = (0,..,0) \quad \text{and} \quad Var(\varepsilon) = \Gamma = \begin{pmatrix} Var(\varepsilon^1) & \dots & Cov(\varepsilon^1, \varepsilon^s) \\ \vdots & \ddots & \vdots \\ Cov(\varepsilon^s, \varepsilon^1) & \cdots & Var(\varepsilon^s) \end{pmatrix} \qquad (9)$$

The parameter β can be estimated, using the WLS estimation, by:

$$\begin{aligned} \hat{\beta}_w &= \arg\min_{\beta} (logY - logX\beta)' \, \Omega (logY - logX\beta) \\ &= ((\log X)'\Omega \log X)^{-1} (\log X)'\Omega \log Y \end{aligned} \qquad (10)$$

where $\Omega = \text{diag}(w_1,...,w_N)$ is the diagonal matrix with diagonal elements w_i which is the weight for site i. The matrix Γ is estimated by:

$$\hat{\Gamma}_w = \frac{\left(\log Y - \log X\hat{\beta}_w\right)'\left(\log Y - \log X\hat{\beta}_w\right)}{N - r - 1} \qquad (11)$$ ”

Comment 4: Moreover, if Yi is an 1*s matrix in Eqs. (17) and (18), the question arises how the ratio of two such matrices is defined. I had the same difficulties with the Chebana and Ouarda (2008) paper on depth-based regional frequency estimation.

Answer: The authors agree with the referee concerning this comment and thank him for pointing it out. Indeed, the ratio of these vectors is defined by using an element-by-element division. This precision is indicated in the revised version.

Comment 5: Section 2 is a rather long section on background material which is sometimes difficult to read (in particular sections 2.1 and 2.4) and it leads to questions about the connection between the various topics. For instance, the sentence “The subset A represents the neighborhood or the region in the classical RFA approaches” (p. 527, l. 6,7) would be immediately clear to the reader if he/she realizes that the argument x of the weight function is the Mahalanobis depth. Weight functions are defined in section 2.2 and weights wi appear in the next section 2.3. The reader may ask at that stage how these two are related. It is not explained in section 2.4 that the optimization refers to the coefficients of the chosen weight function. This is even not clear in the later section 3.1 where the general procedure is outlined. A flow chart may be useful. I have the impression that step i is an iteratively weighted least-squares procedure within the optimization step iii.

Answer: First, regarding the comment on the length of the background material (section 2), the authors would like to mention that it is treated in comments 6 and 9 below.

Second, the authors agree with the referee that a diagram (flow chart) would be useful to show the connection between the various topics defined in section 2, and to summarize the steps of the optimization of the DBRFA approach. Moreover, a diagram would also be helpful to show that the argument of the weight function φ is the Mahalanobis depth (step i) and that the objective function (criterion to be optimized) is parameterized by the coefficient of φ (step iii). Therefore, the following diagram is included in the revised version.

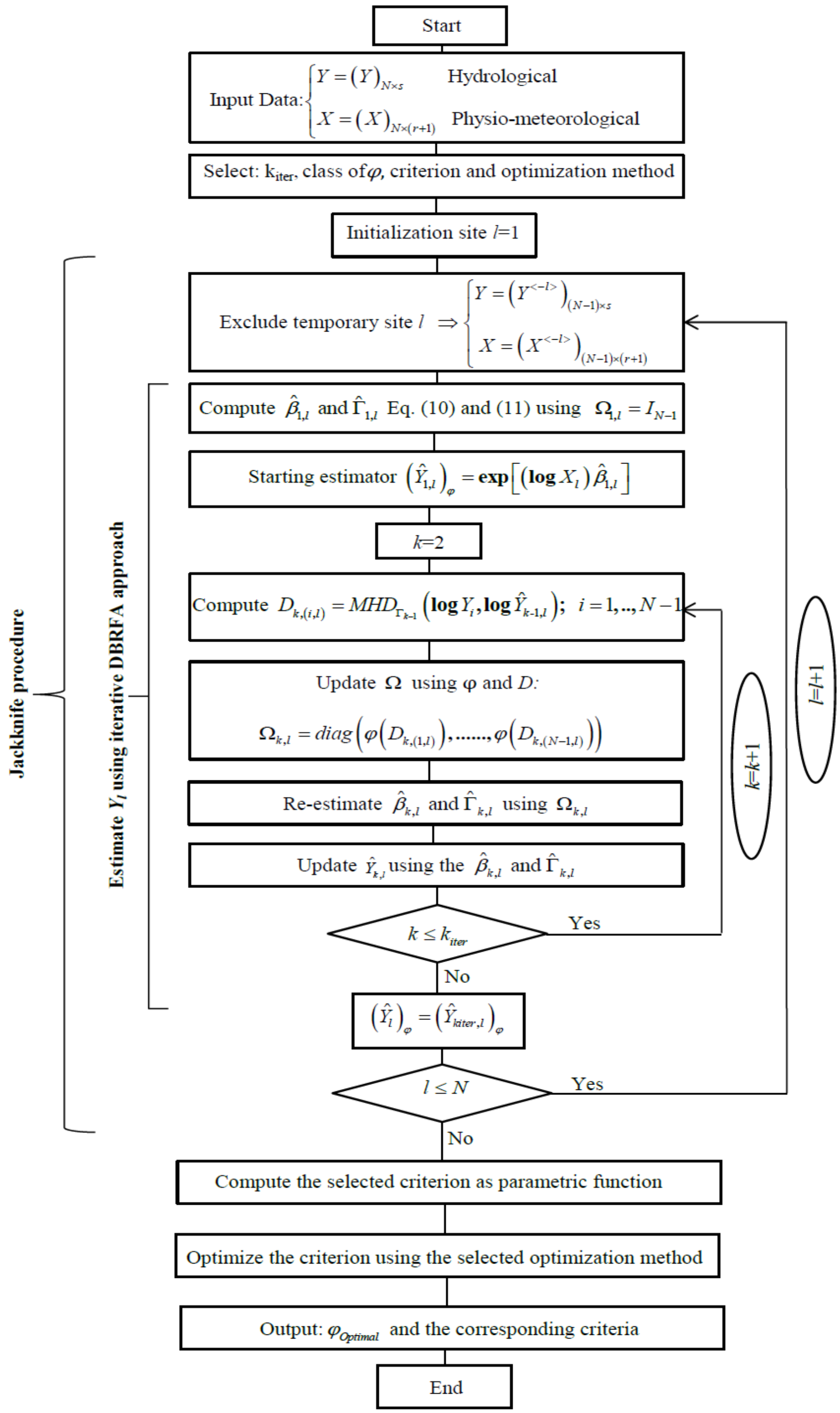

Figure 1. An overview diagram summarizing the optimization procedure of the DBRFA approach

Comment 6: To be more specific on the readableness of section 2.1, why should the reader be bothered with terms like affine invariance, simplicial volume, halfspace? (p. 524).

Answer: In the revised version, the authors focus on the definition of the Mahalanobis depth which is the one used in the methodology part (section 3). The corresponding text in the original version (line 6 p. 524 to 1ine 10 p.525) is replaced by the following paragraph:

"**2.1 Mahalanobis depth function**

The absence of a natural order to classify multivariate data led to the introduction of the depth functions (Tukey, 1975). They are used in many research fields, and were introduced in water science by Chebana and Ouarda (2008). Several depth functions were introduced in the literature (Zuo and Serfling, 2000). Depth functions have a number of features that fit well with the constraint of RFA (Chebana and Ouarda, 2008).

In this study, the Mahalanobis depth function is used to sort sites where the deeper the site is the more it is hydrologically similar to the target site. This function is used for its simplicity, value interpretability, and for the relationship with the CCA approach used in RFA. The Mahalanobis depth function is defined on the basis of the Mahalanobis distance given by $d_A^2(x,y)=(x-y)' A^{-1}(x-y)$ between two points $x,y \in R^d$ $(d \geq 1)$ where A is a positive definite matrix (Mahalanobis, 1936). This distance was used by Ouarda et al. (2001) in the development of the CCA approach. The Mahalanobis depth is given by:

$$MHD(x;F)=\frac{1}{1+d_A^2(x,\mu)}; \qquad x \text{ in } \mathrm{R}^d \tag{1}$$

for a cumulative distribution function F characterized by a location parameter μ and a matrix of covariance A. Note that the Mahalanobis depth function has values in the interval $[0,1]$.

An empirical version of the Mahalanobis depth of x with respect to μ is defined by replacing F by a suitable empirical function $\hat{F}_N$ for a sample of size N (Liu and Singh, 1993). In the context of the present paper, the notation in (1) is replaced by:

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

"

Comment 7: It is curious that if F is replaced by the empirical distribution function, there appears a mu in the left-hand side of Eq.(2) instead of the empirical distribution function, and that the right-hand sides of Eqs. (1) and (2) are identical.

Answer: In Eq.(2) (p. 525 l.10), the authors substitute the function F by its characteristics μ and A in the definition of Mahalanobis depth in order to indicate how it is used in RFA (p.532 l.4). Note that this equation was used in Chebana and Ouarda (2008).

Comment 8: The discussion of the depth function should be reduced. The need for a separate subsection should be questioned. It may be better to introduce the depth function in the discussion of the weights.

Answer: After modifying section 2.1 of the original manuscript by focusing on the definition of Mahalanobis depth function (comment 6), the discussion of the depth function is reduced.

Comment 9: It is not necessary to explain the Nelder-Mead and pattern search optimization methods (sections 2.4.1 and 2.4.2). A reference to the relevant literature suffices. The discussion on p. 523, l. 3-11 on optimization algorithms should also be removed.

Answer: In the revised version, the description of the optimization methods is removed. However, the short discussion on p. 523 (l3-11) is necessary to indicate the employed optimization methods (with the references) since the optimization is a key element of the paper.

Minor comments: answers (where required) are given in italic

- The authors should bring their reference list in agreement with the main text. For instance, the often cited paper Chebana and Ouarda (2008) is not found in the reference list. This is also the case for Singh and Bardossy (2008), p. 521, l. 2
- p. 520, l. 3 Please replace “correspond to” by “lead to” or “result in”.
- p. 521, l. 16 “method” should read “methods
- p. 522, l. 2 “depth function” should read “a depth function” or “depth functions”.
- p. 525, l.1 “Mahalanobis depth” should it not be “Mahalanobis depth of x with respect to mu”? This is more in line with p. 532, l. 2, 3.
- p. 528, l.9 Please reformulates “matrix formed by “r” vector
- p. 528, l. 11 “error of model” should read “model error”
- p. 528, l. 17 Please change “wi and wi” into “wi where wi”.

- p. 531, l.23 Please change “identity matrix of dimension N” into “N*N identity matrix”.
- In contrast to Eq. (2), the two arguments of the depth function are separated by a comma instead of a semicolon.
- p. 532, l.4 I understand why the quantity D has been enclosed by large round brackets. A consequence of this is that more round brackets are needed in the righthand side of Eq. (15) to indicate that the depth is the argument of the function phi.
- p. 534, l. 17 It is uncommon to refer to section 3 within section 3.
- p. 535, l.4 “wil” should read “will”
- p. 535, l. 14 “performed”. Should it not be “used”?
- p. 535, l. 17 “precipitations” should read “precipitation”.
- p. 536, l. 21 The fact that alpha refers to the neighborhood coefficient should be mentioned on p. 536, l. 20 and not on
- p. 536, l. 21. p. 537, l. 9 The word “consequently” could be omitted.
- p. 537, l. 3,4 the sentence has no verb.
- p. 538, l.13 Please delete “in terms of values”.
- p. 552 Please change “with respect RRMSE” into “with respect to RMSE” and “with respect RB” into “with respect to RB”.
- p. 555 please use the same vertical scales in the top and bottom rows of Fig. 8.

The authors thank the referee for all these careful comments which are considered in the revised version.

- p.522, l.29 “three families of weight functions”. However, four different weight functions are discussed in section 2.2 and p.525,l. 12.

The sentence on p. 522, l. 29: “In the present context, three families of φ are considered: Gompertz (φ_G) [Gompertz, 1825], logistic ($\varphi_{\text{logistic}}$) [Verhulst, 1838] and linear (φ_{Linear})” becomes in the revised version:

“In the present context, four families of φ are considered: Gompertz (φ_G) [Gompertz, 1825], logistic ($\varphi_{\text{logistic}}$) [Verhulst, 1838], linear (φ_{Linear}) and indicator (φ_I)”.

The sentence on p. 525, l. 12 “Below are the definitions of the three families of weight functions φ_G, $\varphi_{\text{logistic}}$ and φ_{Linear}” becomes:

"Below are the definitions of the four families of weight functions $\varphi_G, \varphi_{\text{logistic}}$, φ_{Linear} *and* φ_I *considered in…"*

- p. 525, l. 21 and p. 526, l. 14 What is a "derivable" function?

The term "derivable" is removed because it is no longer necessary after the previous modifications.

- p. 526, l. 4 The coefficient b in Eqs. (3) and (4) is a scale parameter. It determines the spread and not the shape of the curve. For the Gompertz function this is demonstrated in Fig. 1c.

The sentence on p. 526, l. 3-4: "where c is its upper limit, a and b are two coefficients which respectively allow to translate and change the shape of the curve" becomes:

"where c is its upper limit, a and b are two coefficients which respectively allow to translate and determine the spread of the curve".

- Note that the caption of Fig. 1c is not correct. It should read b varies with fixed a and c (and the caption of Fig. 1b should read a varies with fixed b and c).

The caption of Figure 1: "Illustration of Gompertz function: (a) c varies with fixed a and b, (b) b varies with fixed a and c and (c) a varies with fixed b and c." becomes:

" Illustration of Gompertz function: (a) c varies with fixed a and b, (b) a varies with fixed b and c and (c) b varies with fixed a and c."

- p. 527, l. 10 What is "the quantile of order alpha for p degrees of freedom?". Does it relate to some quantile of the chi-square distribution?

This is replaced by: $\chi^2_{\alpha,p}$ *is the quantile of order* α *associated to the chi-squared distribution with p degrees of freedom.*

- p. 528, l.3 The variable epsilon in Eq. (7) is not defined. Note that in contrast to Eq. (8), epsilon is not a vector in Eq. (7).

Eq. (7): $QT = \beta_0 A_1^{\beta_1} A_2^{\beta_2} \ldots A_r^{\beta_r} e^{\varepsilon}$ *becomes:* $QT = \beta_0 A_1^{\beta_1} A_2^{\beta_2} \ldots A_r^{\beta_r} e$ *, where e is the model error*

.

- p. 528, l. 11 The mean of epsilon is the null vector. It is not clear what the authors mean by "null mean vectors".

Please see answer of comment 3.

- p. 528, l.12 and p. 528, l. 18-20.The authors assume that the covariance matrix of the quantile estimates is the same for all sites. How valid is this assumption? Is it better fulfilled after the log-transformation?

The authors do not assume that the covariance matrix of quantile is the same for all sites. In fact, the matrix variance-covariance Γ *of dimension* $s \times s$ *contains the inter-quantiles correlation for all sites within the region. The sites are seen as individuals where the series for which we evaluate the variance and covariance are constituted by the quantile over sites.*

- p. 528, l. 21, 22 the fact that the regional estimate contains a transformation bias may question the use of RB in Eq. (17) as a criterion for optimizing the coefficients in the weight function. RB will differ from zero if the local quantile estimate is unbiased.

The criteria used in this paper (RB and RRMSE) are most frequently used in hydrology (see e.g., Hosking and Wallis, 1997). They are used in this paper for comparison with other studies where they are already used. In addition, they do not affect the validity of the proposed optimization procedure since it is general and any other criterion can be considered.

<u>References:</u>

Hosking and wallis, (1997): Regional Frequency Analysis: An Approach Based on L-Moments.

- p. 532, l.4 The second argument in the depth function seems to be no longer a location parameter mu.

Generally, in the RFA approaches and especially in the ROI and the CCA neighborhood approach, the target site represents the center of its hydrological region (see for instance, Tasket et al., 1996, Ouarda et al., 2001). As a result, the quantile of the target site can be used as location parameter for the empirical distribution of the quantiles of gauged sites.

<u>References:</u>

G.D. Tasker, S.A. Hodge, C.S. Barks, 1996: Region of influence regression for estimating the 50-year flood at ungaged sites. Journal of the American Water Resources Association, 32(1): 163-170.

T.B.M.J. Ouarda, C. Girard, G.S. Cavadias, B. Bobée, 2001. Regional flood frequency estimation with canonical correlation analysis. Journal of Hydrology, 254(1-4): 157-173.

- p. 537, l.19, 20 what is R^2?

R^2 is defined on p.534 l.2 (the coefficient of determination).

- p. 538, l.9 In statistical theory the term "efficient" refers to the variance of an estimator. Therefore an estimate cannot be efficient in terms of the relative bias RB

The term efficient is replaced by "accurate"

- p. 538, l. 21 what is a high S-curve?

It is an S-curve with short upper extremity. This is clarified in the revised version.

- p. 539, l. 9-11. How can we see from the results in Table 2 that the optimal function keeps the S shape?

Indeed, it is not possible to see directly from Table 2 the shape of the weight function. However, the figure which allows to show this is not incorporated in the paper for space limitation. This is mentioned in the revised version.

- p. 540, l. 1,3 Do the results in Fig. 8 really imply that there is an underestimation by a factor 2-5 for some sites? For Southern Quebec and Arkansas the largest underestimation is strongly reduced by the depth-based approach. How much contributes this to the reduction in the average relative bias?

In figure 8, we observe large negative errors (underestimation) for some sites, in particular site number 66 in Quebec and site 175 in Arkansas. Note that, for Texas, there are no sites with such high underestimation. The authors agree with the referee that the largest underestimation is strongly reduced by the DBRFA approach. To quantify the contribution of this reduction in the average of RB and RRMSE, the authors apply the CCA and the optimal DBRFA approaches on these regions after removing the largest underestimated sites (site 66 in Quebec and site 175 in Arkansas). The following table summarizes the obtained results:

Method		Southern of Quebec				Arkansas			
		QS10		QS100		Q10		Q100	
		RB (%)	RRMSE (%)	RB (%)	RRMSE (%)	RB (%)	RRMSE (%)	RB (%)	RRMSE (%)
CCA	All sites	-7.5	44.6	-8.1	51.8	-7.8	48.1	-9.3	59.5
	Without: site 66 Quebec, site 175 Arkansas	-5.2	34.6	-5.5	40.8	-6	42.3	-6.9	49.0
Optimal DBRFA	All sites	-3.8	38.7	-2.2	44.5	-6.0	41.5	-6.3	47.7
	Without: site 66 Quebec, site 175 Arkansas	-2.3	32.2	-1.6	37.9	-5.7	41.3	-5.5	46.5

We observe that removing the largest underestimated sites provides an improvement in the performance of both approaches (optimal DBRFA and CCA). This improvement is more significant for the traditional CCA approach than optimal DBRFA. Note that in the both cases (all sites and without the underestimated sites), the optimal DBRFA leads to more accurate estimates in term of RB and RRMSE than those obtained using the CCA approach.

- p. 550 Fig. 3 is unnecessary.

In the revised version Figure 3 are deleted.