

Record Number: 14010
Author, Monographic: Boucher, G.//Ouarda, T. B. M. J.//Bobée, B.
Author Role:
Title, Monographic: Étude de l'analyse des corrélations canoniques pour l'estimation des quantiles de crues
Translated Title:
Reprint Status:
Edition:
Author, Subsidiary:
Author Role:
Place of Publication: Québec
Publisher Name: INRS-Eau
Date of Publication: 1997
Original Publication Date: Août 1997
Volume Identification:
Extent of Work: viii, 73
Packaging Method: pages incluant 3 annexes
Series Editor:
Series Editor Role:
Series Title: INRS-Eau, rapport de recherche
Series Volume ID: 510
Location/URL:
ISBN: 2-89146-395-1
Notes: Rapport annuel 1997-1998
Abstract: chaire en hydrolgie statistique

15.00\$

Call Number:

R000510

Keywords:

rapport/ ok/ dl

**ÉTUDE DE L'ANALYSE DES
CORRÉLATIONS CANONIQUES
POUR L'ESTIMATION DES
QUANTILES DE CRUES**

ÉTUDE DE L'ANALYSE DES CORRÉLATIONS CANONIQUES POUR L'ESTIMATION DES QUANTILES DE CRUES

**Rapport préparé
par**

**Geneviève Boucher
Taha B.M.J. Ouarda
Bernard Bobée**

**Chaire industrielle en Hydrologie statistique
Institut national de la recherche scientifique, INRS-Eau
2800, rue Einstein, Case postale 7500, SAINTE-FOY (Québec), G1V 4C7**

Rapport de recherche No R-510

Août 1997



INRS-Eau, 1997
ISBN 2-89146-303-X

TABLE DES MATIÈRES

LISTE DES TABLEAUX	iv
LISTE DES FIGURES	v
1 Introduction	1
2 Analyse des corrélations canoniques	3
3 Détermination du voisinage	5
4 Présentation de la base de données	7
5 Application de l'analyse des corrélations canoniques	9
6 Étude du nombre optimal de stations à inclure dans le voisinage	13
7 Sensibilité au modèle d'estimation locale	17
8 Sensibilité de la méthode aux changements dans les réseaux hydrométriques	21
9 Étude des vecteurs d'erreurs	23
10 Conclusion	29
11 Références	31
APPENDICE: aspect théorique de l'analyse des corrélations canoniques	33
ANNEXE A: bases de données	35
ANNEXE B: programmes Matlab	45
ANNEXE C: figures	57

LISTE DES TABLEAUX

Tableau 5.1: Corrélation entre les variables originales	9
Tableau 5.2: Corrélations entre les variables canoniques et les variables originales	10
Tableau 7.1: Résultats des analyses des corrélations canoniques effectuées à partir des quantiles de crue estimés avec cinq distributions différentes	18
Tableau A.1: Base de données de la province de l'Ontario	36
Tableau A.2: Estimation des quantiles de crue suite à l'ajustement d'une loi GEV (PWM)	39
Tableau A.3: Estimation des quantiles de crue suite à l'ajustement d'une loi GEV (MV)	40
Tableau A.4: Estimation des quantiles de crue suite à l'ajustement d'une loi log-normal (MV) .	41
Tableau A.5: Estimation des quantiles de crue suite à l'ajustement d'une loi log-Pearson type III (BOB)	42
Tableau A.6: Estimation des quantiles de crue suite à l'ajustement d'un modèle non-paramétrique	43

LISTE DES FIGURES

Figure 5.1: Position des bassins dans l'espace canonique physiographique	11
Figure 5.2: Position des bassins dans l'espace canonique hydrographique	11
Figure 5.3: Estimation de Q_2 par la régression sur le voisinage	12
Figure 5.4: Estimation de Q_{100} par la régression sur le voisinage	12
Figure 6.1: Valeur optimale de α pour l'estimation de Q_2	15
Figure 6.2: Valeur optimale de α pour l'estimation de Q_{100}	15
Figure 6.3: Nombre moyen optimal de stations voisines pour l'estimation de Q_2	16
Figure 6.4: Nombre moyen optimal de stations voisines pour l'estimation de Q_{100}	16
Figure 7.1: Sensibilité de la méthode au modèle d'estimation locale	19
Figure 8.1: Sensibilité de la méthode à la réduction du réseau hydrométrique	22
Figure 9.1: Exemple de vecteurs d'erreurs	23
Figure 9.2: Effet de la pente du cours d'eau principal sur l'angle des vecteurs d'erreurs	25
Figure 9.3: Effet de la pente du cours d'eau principal sur la composante horizontale des vecteurs d'erreurs	26
Figure 9.4: Effet de la longueur du cours d'eau principal sur l'angle des vecteurs d'erreurs ...	26
Figure 9.5: Effet de la longueur du cours d'eau principal sur la composante verticale des vecteurs d'erreurs	27
Figure C.1: Sensibilité de la méthode au modèle d'estimation locale, Moyenne des erreurs relatives	58

Figure C.2: Sensibilité de la méthode au modèle d'estimation locale, Écart-type des erreurs relatives	58
Figure C.3: Sensibilité de la méthode au modèle d'estimation locale, Maximum des erreurs relatives	59
Figure C.4: Sensibilité de la méthode au modèle d'estimation locale, Biais des estimations régionales	59
Figure C.5: Effet de la superficie du bassin versant sur l'angle des vecteurs d'erreurs	60
Figure C.6: Effet de la superficie du bassin versant sur la norme des vecteurs d'erreurs	60
Figure C.7: Effet de la superficie du bassin versant sur la composante horizontale des vecteurs d'erreurs	61
Figure C.8: Effet de la superficie du bassin versant sur la composante verticale des vecteurs d'erreurs	61
Figure C.9: Effet de la superficie du bassin versant contrôlée par les lacs et marais sur l'angle des vecteurs d'erreurs	62
Figure C.10: Effet de la superficie du bassin versant contrôlée par les lacs et marais sur la norme des vecteurs d'erreurs	62
Figure C.11: Effet de la superficie du bassin versant contrôlée par les lacs et marais sur la composante horizontale des vecteurs d'erreurs	63
Figure C.12: Effet de la superficie du bassin versant contrôlée par les lacs et marais sur la composante verticale des vecteurs d'erreurs	63
Figure C.13: Effet de la précipitation totale moyenne annuelle sur l'angle des vecteurs d'erreurs	64
Figure C.14: Effet de la précipitation totale moyenne annuelle sur la norme des vecteurs d'erreurs	64

Figure C.15: Effet de la précipitation totale moyenne annuelle sur la composante horizontale des vecteurs d'erreurs	65
Figure C.16: Effet de la précipitation totale moyenne annuelle sur la composante verticale des vecteurs d'erreurs	65
Figure C.17: Effet de la pente du cours d'eau principal sur la norme des vecteurs d'erreurs ..	66
Figure C.18: Effet de la pente du cours d'eau principal sur la composante verticale des vecteurs d'erreurs	66
Figure C.19: Effet de la longueur du cours d'eau principal sur la norme des vecteurs d'erreurs	67
Figure C.20: Effet de la longueur du cours d'eau principal sur la composante horizontale des vecteurs d'erreurs	67
Figure C.21: Effet de la latitude sur l'angle des vecteurs d'erreurs	68
Figure C.22: Effet de la longitude sur l'angle des vecteurs d'erreurs	68
Figure C.23: Effet de la latitude sur la norme des vecteurs d'erreurs	69
Figure C.24: Effet de la longitude sur la norme des vecteurs d'erreurs	69
Figure C.25: Effet de la latitude sur la composante horizontale des vecteurs d'erreurs	70
Figure C.26: Effet de la longitude sur la composante horizontale des vecteurs d'erreurs	70
Figure C.27: Effet de la latitude sur la composante verticale des vecteurs d'erreurs	71
Figure C.28: Effet de la longitude sur la composante verticale des vecteurs d'erreurs	71

1 Introduction

Une bonne connaissance des débits de crue est importante afin de minimiser les risques d'inondations ainsi que les coûts de construction des structures. L'estimation des événements hydrologiques extrêmes (crues et étiages) dépend fortement de l'information disponible c'est-à-dire de la qualité et de la longueur des séries de débits. Cependant, il existe plusieurs sites d'intérêt où l'on dispose de peu ou d'aucune information hydrologique. Pour ces sites, il est possible d'estimer des débits de crue annuels Q_T , correspondant à une période de retour T , en se servant de l'information régionale provenant de sites jaugés considérés similaires aux sites d'intérêt selon des critères hydrologiques. L'approche générale consiste alors à :

- 1- déterminer des régions homogènes
- 2- utiliser une méthode d'estimation régionale

Il est reconnu que la proximité géographique de deux sites n'implique pas que ces sites soient similaires hydrologiquement. Les régions homogènes peuvent alors être déterminées à partir des caractéristiques physiographiques associées à chaque site. Ce rapport traite particulièrement d'une méthode de détermination des régions homogènes soit l'analyse des corrélations canoniques. GREHYS (1996) a montré que cette méthode produit des estimations adéquates des quantiles de crue à des sites non jaugés peu importe la méthode d'estimation régionale employée. Le but de ce rapport est de présenter une application de cette méthode et d'examiner différents aspects la concernant. La sensibilité de la méthode au modèle d'estimation locale ainsi qu'aux changements dans les réseaux hydrométriques ont été étudiés. Deux autres aspects ont aussi été considérés soient le nombre optimal de stations à inclure dans le voisinage et les vecteurs d'erreurs.

2 Analyse des corrélations canoniques

L'analyse des corrélations canoniques a pour but de représenter la relation de dépendance linéaire existant entre deux ensembles ou plus de variables aléatoires. Pour se faire, on recherche une combinaison linéaire de chaque ensemble de variables telle que la corrélation entre ces combinaisons linéaires soit maximum. On obtient ainsi un ensemble de variables canoniques indépendantes pour les deux ensembles de variables aléatoires associées à des coefficients de corrélations canoniques. Il est alors possible d'inférer sur les variables canoniques d'un ensemble connaissant les variables canoniques de l'autre ensemble. On note que l'analyse des corrélations canoniques peut entraîner des difficultés d'interprétation lors de l'application. Par la suite, on peut calculer une distance entre deux variables canoniques ce qui permettra de déterminer une région homogène. Cette approche considère que chaque station a sa propre région homogène (aussi appelé voisinage). La détermination du voisinage est présentée à la Section 3. Une explication plus détaillée de cette méthode est donnée dans les articles de [Torranin 1972] et de [Cavadias 1990]. L'aspect théorique de cette méthode est présenté en appendice.

Dans un contexte hydrologique, le vecteur X peut représenter un ensemble de caractéristiques physiques des bassins versants et le vecteur Y peut représenter un ensemble de caractéristiques de crue tels que les quantiles de débits maximum annuels. Le but cherché est d'inférer sur le vecteur Y connaissant le vecteur X . On définit les deux combinaisons linéaires suivantes :

$$W = \alpha_1 Y_1 + \dots + \alpha_p Y_p$$

$$V = \beta_1 X_1 + \dots + \beta_q X_q$$

où p est le nombre de variables hydrologiques utilisées et q est le nombre de variables physiographiques utilisées. L'analyse des corrélations canoniques permet d'optimiser la fonction de corrélation définie entre ces deux combinaisons linéaires. On obtient alors des coefficients de corrélations canoniques λ_i auxquels sont associés des variables canoniques $W_i = \alpha_{1i} Y_1 + \dots + \alpha_{pi} Y_p$ et $V_i = \beta_{1i} X_1 + \dots + \beta_{qi} X_q$ ($i \leq \text{rang de la matrice de covariance } \Sigma_{yx}$), le premier coefficient de corrélation étant le plus élevé et le dernier étant le plus faible. Sans perdre trop d'information, il est possible de considérer seulement les variables canoniques correspondant aux coefficients de corrélations canoniques les plus élevés. Les variables canoniques obtenues représentent l'espace canonique physiographique V_i et l'espace canonique hydrologique W_i . Il est possible d'estimer W_i par : $\hat{W}_i = \lambda_i V_i$. Donc, connaissant la position d'une station dans l'espace canonique physiographique V_i , il est possible d'évaluer sa position dans l'espace canonique hydrologique W_i .

3 Détermination du voisinage

L'analyse des corrélations canoniques permet, à partir des caractéristiques physiographiques d'une station non jaugée, de déterminer sa position dans l'espace canonique physiographique V_i et alors d'évaluer quelle est sa position dans l'espace canonique hydrologique W_i . Une fois que la station non jaugée est positionnée dans cet espace, il est possible de déterminer les stations qui lui ressemblent le plus en considérant celles qui sont les plus proches dans les espaces considérés. Graphiquement, la position de la station non jaugée dans l'espace canonique hydrologique est considérée comme le centre d'une ellipse. Les stations jaugées reconnues semblables sont alors celles qui sont positionnées dans les limites de cet ellipse. Pour se faire le calcul d'une distance est employée. Une explication sommaire sera présentée ici.

Soit W et V des vecteurs contenant les p variables canoniques W_i et V_i et soit Λ la matrice diagonale contenant les p coefficients de corrélations canoniques. La loi conditionnelle de W sachant V est une loi normale à p dimensions de moyenne ΛV et de variance $I_p - \Lambda \Lambda'$. Les stations jaugées voisines sont alors celles autour de la position moyenne ΛV , à l'intérieur d'une limite définie. La distance entre les positions des stations jaugées dans l'espace canonique hydrologique et la position moyenne dans le même espace peut être calculée par une équation quadratique (provenant de l'équation de la loi de $W|V$):

$$Q = (W - \Lambda V)' (I_p - \Lambda \Lambda')^{-1} (W - \Lambda V)$$

Q est une distance de type Mahalanobis et suit une loi χ^2 à p degrés de liberté pour Λ et V fixés.

Cependant, la position moyenne n'est pas connue et doit être estimée. La position moyenne de la station non jaugée est $\hat{\Lambda} V_0$ où $\hat{\Lambda}$ contient les coefficients de corrélations canoniques estimés et V_0 est la position de la station non jaugée dans l'espace canonique physiographique. La distance entre la position estimée de la station non jaugée dans l'espace canonique hydrologique et les positions des stations jaugées dans le même espace est alors:

$$Q = (W - \hat{\Lambda} V_0)' (I_p - \hat{\Lambda} \hat{\Lambda}')^{-1} (W - \hat{\Lambda} V_0).$$

Les stations voisines, correspondant à un niveau de confiance de $100(1-\alpha)\%$, sont celles pour lesquelles leur valeur de Q est inférieure à un certain seuil ($Q \leq \chi^2_{\alpha, p}$). L'ensemble des stations voisines est alors obtenu et l'estimation des caractéristiques hydrologiques peut être faite en utilisant ce voisinage avec diverses méthodes. Des applications de l'analyse des corrélations canoniques et du voisinage sont présentées dans les articles de [Cavadias 1995] et de [Ribeiro-Corréa et al. 1995].

4 Présentation de la base de données

La présente étude est basée sur les données de la province de l'Ontario (Tableau A.1, Annexe A). Cette base de données comporte 106 stations jaugées pour lesquelles les quantiles de crue de période de retour 2 ans et 100 ans ont été calculés. Pour se faire, plusieurs modèles, qui seront explicités subséquemment, ont été utilisés. Les variables hydrologiques servant aux analyses sont Q_2 et Q_{100}/Q_2 qui représentent respectivement la moyenne de crue annuelle et la dispersion de la distribution de crue. Cinq variables physiographiques et météorologiques sont disponibles soient:

- Aire, superficie du bassin versant (km^2)
- LCP, longueur du cours d'eau principal (km)
- PCP, pente du cours d'eau principal (m/km)
- SLM, superficie du bassin versant contrôlée par les lacs et marais (km^2)
- PTMA, précipitation totale moyenne annuelle (mm)

5 Application de l'analyse des corrélations canoniques

Considérons d'abord les corrélations existant entre chacune des variables. Ces corrélations sont présentées au Tableau 5.1. On constate que toutes les variables physiographiques sont assez fortement corrélées avec la superficie du bassin versant (Aire) sauf la variable PCP, qui n'est pas tellement corrélée avec les autres. On remarque aussi que les variables Aire et Q_2 sont fortement corrélées (corrélation de 0.86). L'Aire jouera un rôle prépondérant dans les analyses subséquentes.

Tableau 5.1 : Corrélations entre les variables originales.

	Aire	PCP	LCP	SLM	PTMA	Q_2	Q_{100}/Q_2
Aire	1.00	-0.28	0.79	0.96	-0.47	0.86	-0.05
PCP	-0.28	1.00	-0.41	-0.25	0.19	-0.32	0.29
LCP	0.79	-0.41	1.00	0.68	-0.50	0.72	-0.21
SLM	0.96	-0.25	0.68	1.00	-0.43	0.79	0.00
PTMA	-0.47	0.19	-0.50	-0.43	1.00	-0.36	-0.70
Q_2	0.86	-0.32	0.72	0.79	-0.36	1.00	-0.18
Q_{100}/Q_2	-0.05	0.29	-0.21	0.00	-0.07	-0.18	1.00

Les estimations de Q_2 et Q_{100} sont obtenues en ajustant les observations du débit maximum annuel à chaque site par une distribution GEV avec la méthode des moments pondérés (PWM) (Tableau A.2, Annexe A). Afin d'éliminer les problèmes d'échelle et d'unités de mesure, l'analyse a portée sur le logarithme naturel de toutes les variables considérées et celles-ci sont standardisées. Les valeurs nulles des variables sont remplacées par la valeur 1. La standardisation est obtenue en soustrayant la moyenne de l'attribut, calculée à partir de la série des valeurs observées à toutes les stations, de chaque valeur de cet attribut puis en divisant le tout par l'écart-type de ce même attribut.

L'analyse canonique est réalisée par programmation dans Matlab. Puisque l'on dispose de 2 variables hydrologiques et de 5 variables physiographiques, on obtient 2 coefficients de corrélations canoniques. Leur valeurs estimées sont $\lambda_1 = 0.96$ et $\lambda_2 = 0.31$. Le premier coefficient est très élevé tandis que le deuxième est faible. Ce dernier indique donc que la deuxième combinaison linéaire calculée a un faible pouvoir explicatif et que la première

combinaison pourrait seulement être utilisée. Les variables canoniques correspondantes sont les suivantes :

$$W_1 = 1.024(\ln(Q_2)) + 0.0765(\ln(Q_{100}/Q_2))$$

$$W_2 = 0.3093(\ln(Q_2)) + 1.06777(\ln(Q_{100}/Q_2))$$

$$V_1 = 1.2227(\ln(\text{Aire})) + 0.1221(\ln(\text{LCP})) + 0.1693(\ln(\text{PCP})) - 0.2614(\ln(\text{SLM})) \\ + 0.2688(\ln(\text{PTMA}))$$

$$V_2 = 1.9988(\ln(\text{Aire})) + 0.3050(\ln(\text{LCP})) - 1.8421(\ln(\text{PCP})) - 0.4856(\ln(\text{SLM})) \\ - 0.7615(\ln(\text{PTMA}))$$

Il est important de souligner que l'analyse est effectuée en considérant les variables transformées logarithmiquement. Leurs corrélations avec les variables canoniques calculées sont présentées au Tableau 5.2. On remarque que la variable canonique V_1 est fortement corrélée avec le logarithme de Aire, LCP et PTMA. V_1 étant un espace canonique physiographique, elle représente bien ces variables. La variable canonique W_1 est fortement corrélée avec ces mêmes variables physiographiques. En ce qui concerne les variables hydrologiques, W_1 est plus corrélée avec le logarithme de Q_2 . Ensuite, les corrélations pour la variable canonique V_2 sont plutôt faibles. La corrélation la plus forte se retrouve au niveau de la pente du cours d'eau principal. La variable canonique W_2 est fortement corrélée avec cette dernière (0.95) mais très peu avec les autres variables.

Tableau 5.2: Corrélations entre les variables canoniques et le logarithme des variables originales standardisé.

	V_1	V_2	W_1	W_2
$\ln(\text{Aire})$	0.96	-0.02	0.99	-0.07
$\ln(\text{PCP})$	-0.29	0.30	-0.31	0.95
$\ln(\text{LCP})$	0.94	0.11	0.91	0.03
$\ln(\text{SLM})$	-0.60	0.18	-0.58	0.06
$\ln(\text{PTMA})$	0.92	-0.14	0.88	-0.04
$\ln(Q_2)$	0.57	-0.10	0.54	-0.03
$\ln(Q_{100}/Q_2)$	-0.34	-0.56	-0.33	-0.18

La position des bassins dans les deux espaces canoniques sont présentés aux Figures 5.1 et 5.2. Compte tenu de la quantité de bassins considérée (106 stations), ils ne sont pas identifiés sur les graphiques. On constate une plus grande variabilité, entre les figures, sur l'axe verticale que sur l'axe horizontal. Cela est dû au fait que le coefficient de corrélation canonique entre la deuxième paire de variables canoniques est faible.

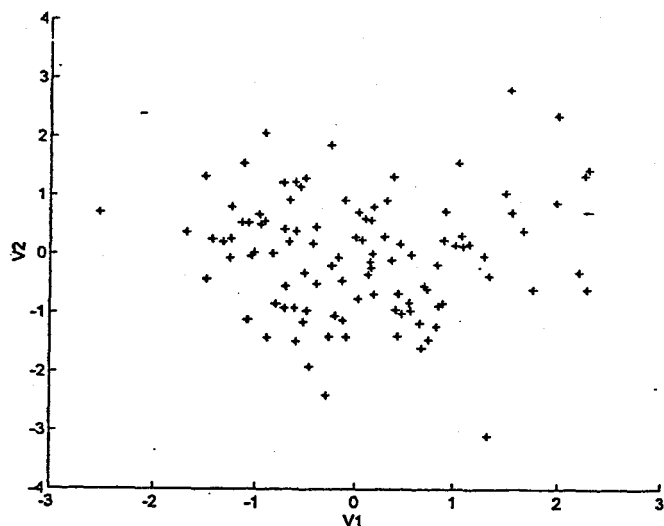


Figure 5.1: Position des bassins dans l'espace canonique physiographique

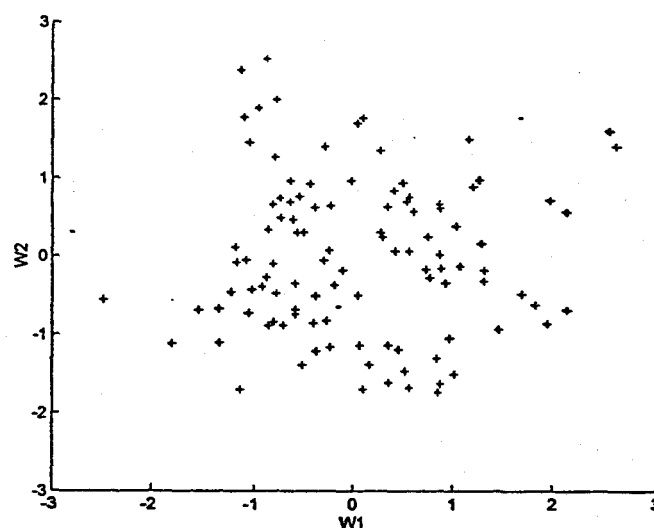


Figure 5.2: Position des bassins dans l'espace canonique hydrologique

Une méthode de rééchantillonnage de type jack-knife est ensuite utilisée. Chacune des 106 stations est tour à tour considérée comme non jaugée et les quantiles de crue y sont estimés en considérant les 105 stations restantes. Ces estimations au site cible sont faites en déterminant d'abord son voisinage par l'analyse des corrélations canoniques puis en effectuant une régression sur ce voisinage. L'équation de régression utilisée est $\text{Ln}(Q_2) = a \cdot \text{Ln}(\text{Aire}) + b$. Afin d'obtenir une régression acceptable, seule la variable Aire est employée car elle est fortement corrélée avec la majorité des variables. De plus, pour la même raison, les quantiles des stations non jaugées ayant moins de 5 stations jaugées voisines ne sont pas estimés. Neuf stations (numéros 21, 53, 67, 79, 85, 86, 89, 90 et 97) n'ont pas un voisinage de taille suffisante et ne sont pas représentées sur les graphiques. Le choix du voisinage est effectué à un niveau de confiance $100(1-\alpha)\%$. Ici nous considérons $\alpha=25\%$ afin d'avoir un bon compromis entre l'obtention de voisinages assez denses en moyenne et que ces voisinages ne soient pas trop larges. Ces calculs sont réalisés par programmation en langage Matlab. Le code du programme

utilisé est présenté à l'Annexe B. Les graphiques des estimations régionales des quantiles en fonction de ceux estimés aux sites sont présentés aux Figures 5.3 et 5.4. On constate, tant pour $T=2$ que $T=100$, une assez bonne adéquation entre ces deux estimations sauf pour les grandes valeurs des quantiles. L'utilisation d'un voisinage plus grand pourrait permettre de remédier à cette situation.

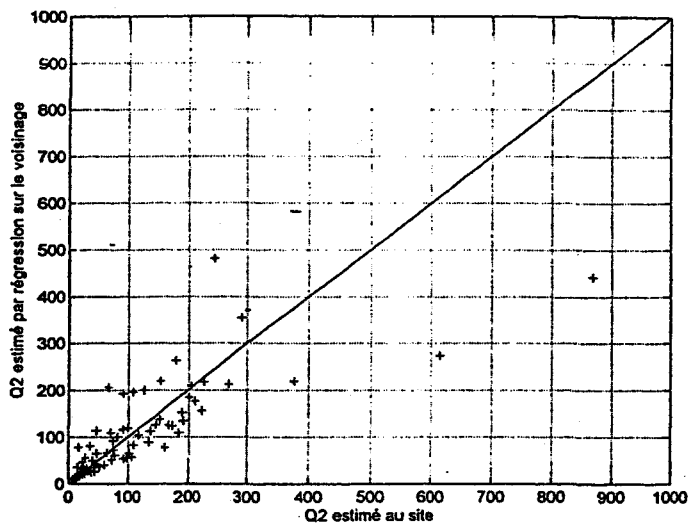


Figure 5.3: Estimation de Q_2 par la régression sur le voisinage

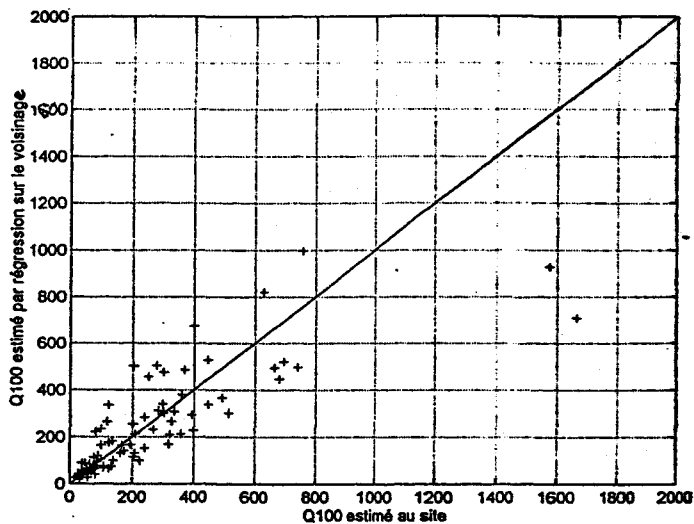


Figure 5.4: Estimation de Q_{100} par la régression sur le voisinage

6 Étude du nombre optimal de stations à inclure dans le voisinage

La détermination d'un voisinage implique le choix d'un seuil au-delà duquel les stations ne font pas partie de la région homogène. La détermination de ce seuil oblige donc l'utilisateur à effectuer un compromis. En effet, le nombre de stations appartenant au voisinage d'un site cible doit être petit afin d'assurer une bonne similarité mais doit être suffisamment grand pour pouvoir exécuter une estimation adéquate. Le but de cette étude est de déterminer le nombre optimal (compromis) de stations à inclure dans un voisinage pour la base de données de la province de l'Ontario.

Pour se faire, des indices de performance sont calculés pour plusieurs valeurs du seuil et pour différents nombres de stations dans le voisinage. L'étude porte sur les 106 stations jaugées de la base de données. Les variables et les transformations de ces variables sont les mêmes que dans la Section 5. Pour chaque valeur du seuil et pour chaque taille du voisinage fixée, chaque station, tour à tour, est considérée comme non jaugée et l'estimation des quantiles de crue est réalisée avec les 105 stations restantes. La procédure est la même que celle présentée dans la section précédente. Les calculs sont faits pour des tailles fixées de 6 à 105 stations voisines et pour des seuils $\alpha = 0.0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35$ et 0.40 . Des seuils où $\alpha > 0.40$ ne sont pas traités car le nombre de stations cibles où la taille du voisinage est insuffisante est supérieur à 10. Pour au moins un des seuils utilisés, dix stations (numéros 15, 21, 53, 67, 79, 85, 86, 89, 90 et 97) n'ont pas un voisinage suffisamment dense. Ces stations sont omises des calculs des indices de performance pour chaque seuil.

Les indices de performance mesurent la validité des estimations régionales en les comparant avec les estimations au site. Pour se faire, deux critères sont utilisés : le biais relatif d'estimation $([BIAIS]_r)$ et l'erreur quadratique moyenne relative $([EQM]_r)$. Ces deux critères de performance sont définis comme suit :

$$[BIAIS]_{r,\alpha} = \frac{1}{k} \sum_{i=1}^k \left[\frac{(\tilde{q}_{i,\alpha} - q_i)}{q_i} \right]$$

$$[EQM]_{r,\alpha} = \frac{1}{k} \sum_{i=1}^k \left[\frac{(\tilde{q}_{i,\alpha} - q_i)}{q_i} \right]^2$$

où k est le nombre de sites, q_i est l'estimation du quantile de crue au site pour la station i et $\tilde{q}_{i,\alpha}$ est l'estimation régionale du quantile de crue pour la station i avec un seuil α .

Pour mettre en évidence l'utilisation d'un voisinage, la comparaison avec le cas particulier $\alpha = 0$ est utile. Dans ce cas, les 105 stations restantes après avoir enlevé le site cible sont utilisées pour l'estimation des quantiles. L'objectif est de vérifier si les estimations des quantiles de crue pour un site cible s'améliorent lorsque le voisinage correspondant à une valeur de α non nulle est utilisée. Les indices de performance suivants ont alors été utilisés :

$$I_1(\alpha) = \frac{[\text{BIAIS}]_{r,\alpha}}{[\text{BIAIS}]_{r,0}} \quad (\text{indice de performance de biais})$$

$$I_2(\alpha) = \frac{[\text{EQM}]_{r,\alpha}}{[\text{EQM}]_{r,0}} \quad (\text{indice de performance de l'erreur moyenne au carré})$$

où $[\text{BIAIS}]_{r,0}$ et $[\text{EQM}]_{r,0}$ correspondent au cas $\alpha = 0$. Plus les indices de performance sont petits, meilleur est le gain de-performance.

Les Figures 6.1 à 6.4 présentent les indices de performance calculés pour chaque seuil et chaque taille fixée du voisinage, et ce pour les quantiles de crue Q_2 et Q_{100} . Sur les Figures 6.1 et 6.2, on remarque que les indices de performance tendent à diminuer lorsque α augmente donc lorsque le nombre de voisins diminue. Au delà du seuil $\alpha \approx 0.10$, les indices de performance sont presque stables. De même, sur les Figures 6.3 et 6.4, on voit que les indices de performance tendent à diminuer lorsque le nombre moyen de stations dans le voisinage diminue. Cependant, ce nombre augmente après une certaine limite (de l'ordre de 25 stations). On observe aussi un certain plateau pour les tailles du voisinage de 10 à 25 stations.

On observe que les résultats des Figures 6.3 et 6.4 semblent plus instables que ceux des Figures 6.1 et 6.2. On note que le nombre moyen de stations voisines correspondant à chacun des seuils utilisés est 105, 35, 30, 26, 23, 20, 18, 16 et 14. Si on tire une ligne sur les Figures 6.3 et 6.4 entre ces 9 points, on obtient sensiblement les mêmes courbes que sur les Figures 6.1 et 6.2. Donc la différence de stabilité observée est due au fait que ces derniers graphiques représentent un plus petit nombre de cas que les Figures 6.3 et 6.4.

Le nombre optimal de stations à inclure dans le voisinage correspond donc à un seuil α compris entre 0.20 et 0.40. Le nombre moyen de stations dans le voisinage varie alors de 14 à 23. Le seuil qui est choisi pour les analyses subséquentes est $\alpha = 0.25$, correspondant à un nombre moyen de stations voisines de 20.

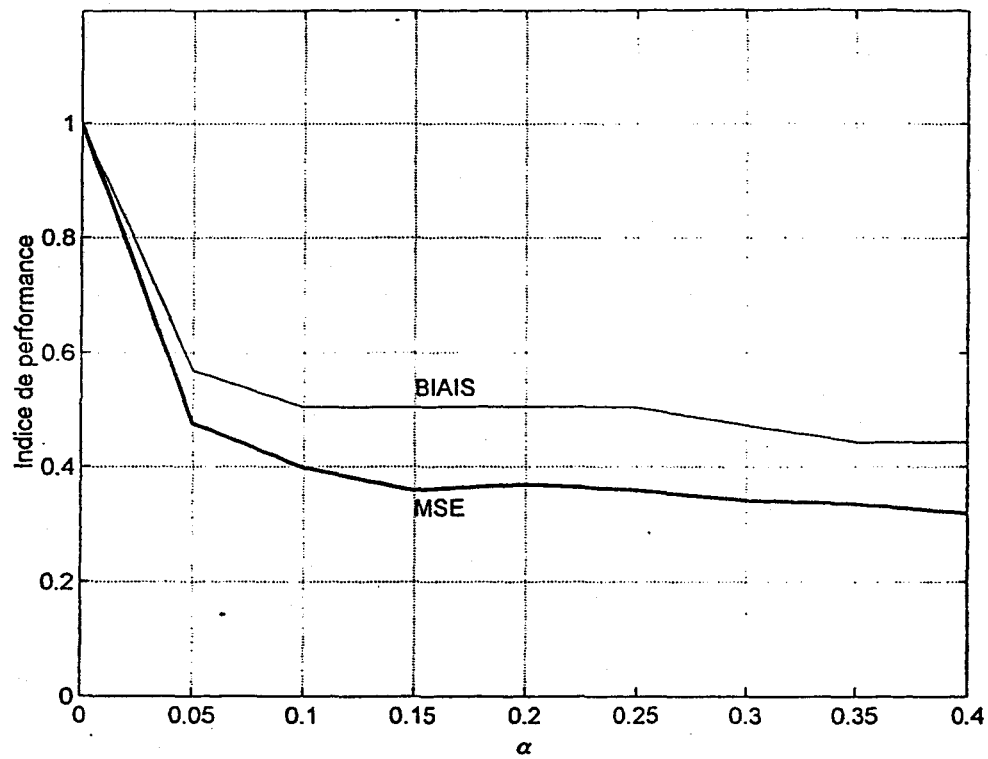


Figure 6.1: Valeur optimale de α pour l'estimation de Q_2

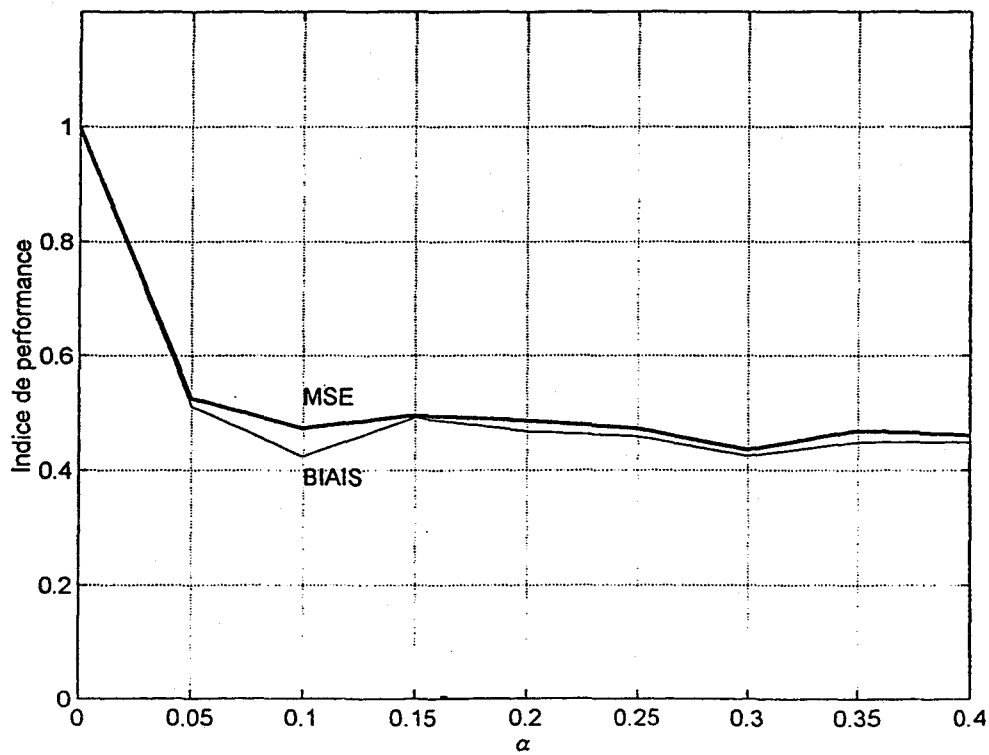


Figure 6.2: Valeur optimale de α pour l'estimation de Q_1

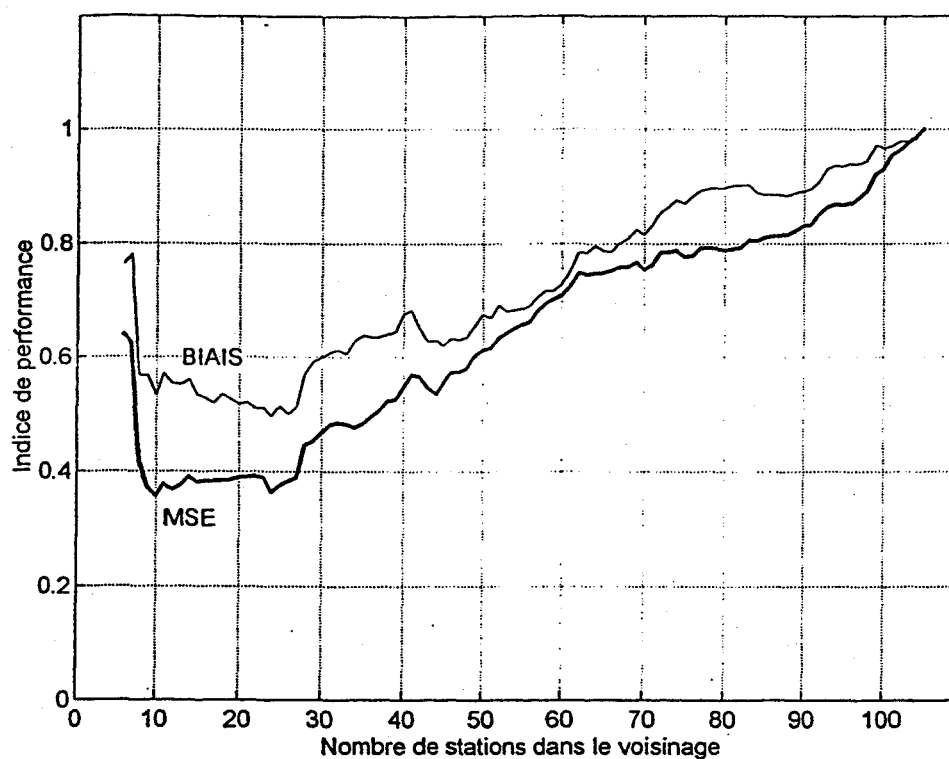


Figure 6.3: Nombre moyen optimal de stations voisines pour l'estimation de Q_2

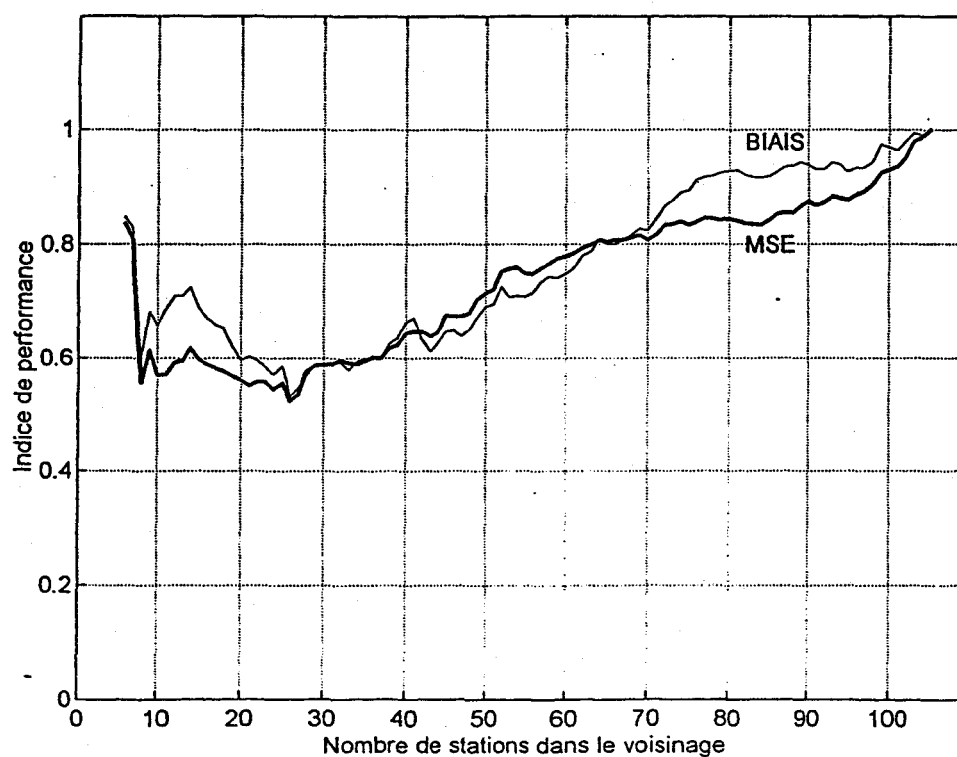


Figure 6.4: Nombre moyen optimal de stations voisines pour l'estimation de Q_{100}

7 Sensibilité au modèle d'estimation locale

Pour estimer les quantiles de crue aux sites non jaugés, des estimations de ces quantiles dans les sites jaugés sont nécessaires. Ceci est accompli en ajustant un modèle aux données recueillies pour chaque site. Le but principal de cette étude est de vérifier l'effet du modèle utilisé sur le calcul des coefficients de corrélations canoniques obtenus à partir des estimations aux sites jaugés et alors sur l'estimation régionale des quantiles de crue aux sites non jaugés. En deuxième lieu, l'objectif est de déterminer si une méthode non-paramétrique serait plus adéquate.

Pour se faire, les quantiles sont estimés aux sites par quatre méthodes paramétriques et une méthode non-paramétrique: GEV (probability weight moments), GEV (maximum de vraisemblance), log-normale (maximum de vraisemblance), log-Pearson type III (méthode directe des moments) et non-paramétrique avec noyau d'Epanechnikov. Ces méthodes seront notées GEV-PWM, GEV-MV, LN-MV, LP III-BOB et NP respectivement. Ces lois ont été choisies car elles sont couramment utilisées en hydrologie. Les estimations obtenues à partir de la loi GEV (PWM) et du modèle non-paramétrique sont calculés par programmation dans Matlab. Celles obtenues à partir de la GEV (MV) et de la log-normale sont calculées à l'aide du logiciel AJUST. Finalement, celles obtenues à partir de la log-Pearson type III sont tirées de [Ribeiro-Corréa et Rouselle 1993]. L'étude actuelle est faite avec les données de 104 sites jaugés car le rapport de [Ribeiro et Rousselle 1993] ne présente pas d'estimations pour les stations 28 et 74. Ces estimations sont présentées aux Tableaux A.2 à A.6 de l'Annexe A.

L'analyse des corrélations canoniques est d'abord faite pour les 5 méthodes d'estimations. Les variables et les transformations de ces variables sont les mêmes que dans la Section 5. Les résultats qui sont obtenus sont présentés au Tableau 7.1. On constate que les premiers coefficients de corrélations canoniques sont très semblables. Ils ne diffèrent qu'à partir de la troisième décimale. De plus, ils sont très élevés alors seulement la paire de variables canoniques correspondant à la première corrélation canonique pourrait être utilisée. Les deuxièmes coefficients de corrélations canoniques varient un peu plus mais restent dans le même ordre de grandeur et sont beaucoup plus petits que les premiers. Donc la loi utilisée pour effectuer les estimations aux sites ne semble pas avoir d'influence sur l'estimation des coefficients de corrélations canoniques et par conséquent sur l'estimation régionale des quantiles de crue.

Tableau 7.1: Résultats des analyses des corrélations canoniques effectuées à partir des quantiles de crue estimés avec cinq distributions différentes.

distribution	Vecteur i	coefficients de corrélations canoniques	Composantes des vecteurs canoniques						
			V _i					W _i	
			Aire	LCP	PCP	SLM	PTMA	Q ₂	Q ₁₀₀ /Q ₂
GEV-PWM	1	0.9592	1.2227	0.1221	0.1693	-0.2614	0.2688	1.0247	0.0765
GEV-MV		0.9590	1.2149	0.1207	0.1770	-0.2601	0.2728	1.0209	0.0709
LN-MV		0.9589	1.2183	0.1216	0.1749	-0.2617	0.2725	1.0169	0.0577
LP III-BOB		0.9594	1.2254	0.1230	0.1671	-0.2608	0.2698	1.0262	0.0800
NP		0.9596	1.2245	0.1215	0.1678	-0.2597	0.2737	1.0251	0.0832
GEV-PWM	2	0.3093	1.9988	0.3050	-1.8421	-0.4856	-0.7615	0.3093	1.0677
GEV-MV		0.3077	2.0443	0.1140	-1.8736	-0.6966	-0.7452	0.2777	1.0556
LN-MV		0.4222	2.2948	0.2691	-1.9121	-0.8032	-0.5946	0.2821	1.0538
LP III-BOB		0.2790	2.2563	0.3610	-2.0447	-0.4584	-0.6438	0.3129	1.0698
NP		0.2941	2.2999	0.4762	-2.0474	-0.3694	-0.5923	0.2798	1.0593

Ensuite, pour les 5 méthodes d'estimations, chaque station a été considérée tour à tour non jaugée et les quantiles de crue y sont estimée en considérant les 105 stations restantes. La procédure est la même que celle présentée dans l'exemple d'application. À partir des estimations régionales (notées $\tilde{q}$) obtenues pour une méthode, en considérant comme référence les estimations aux sites (notées q) correspondantes, la moyenne, l'écart-type et le maximum des erreurs relatives d'estimations ainsi que le biais relatif ($[BIAIS]_r$) et la racine de l'erreur quadratique moyenne relative ($[REQM]_r$) des estimations régionales sont calculés pour chaque distribution. Les formules utilisées pour ces calculs sont:

$$\text{erreur relative} = \frac{|\tilde{q} - q|}{q}$$

$$[BIAIS]_r = \frac{1}{m} \sum_{i=1}^m \frac{\tilde{q} - q(i)}{q(i)}$$

$$[\text{REQM}]_r = \sqrt{\frac{1}{m} \sum_{i=1}^m \frac{(\bar{q}(i) - q(i))^2}{q(i)^2}}$$

où $\bar{q}$ est la moyenne du quantile pour toutes les stations, i est le numéro de la station où le quantile est estimé et m est le nombre de quantiles estimés ($m=106$).

La Figure 7.1 présente la racine de l'erreur quadratique moyenne relative calculée pour chaque distribution et pour chaque quantile. Le calcul du biais relatif donne des résultats semblables (Figure C.4, Annexe C). La moyenne, l'écart-type et le maximum des erreurs relatives pour chaque distribution sont présentés aux Figures C.1 à C.3 de l'Annexe C. Les valeurs obtenues sont toutes similaires. Pour les deux quantiles considérés, aucune distribution ne s'est démarquée. Il semble donc que la méthode utilisée pour effectuer les estimations aux sites n'a pas d'influence sur l'estimation régionale des quantiles de crue.

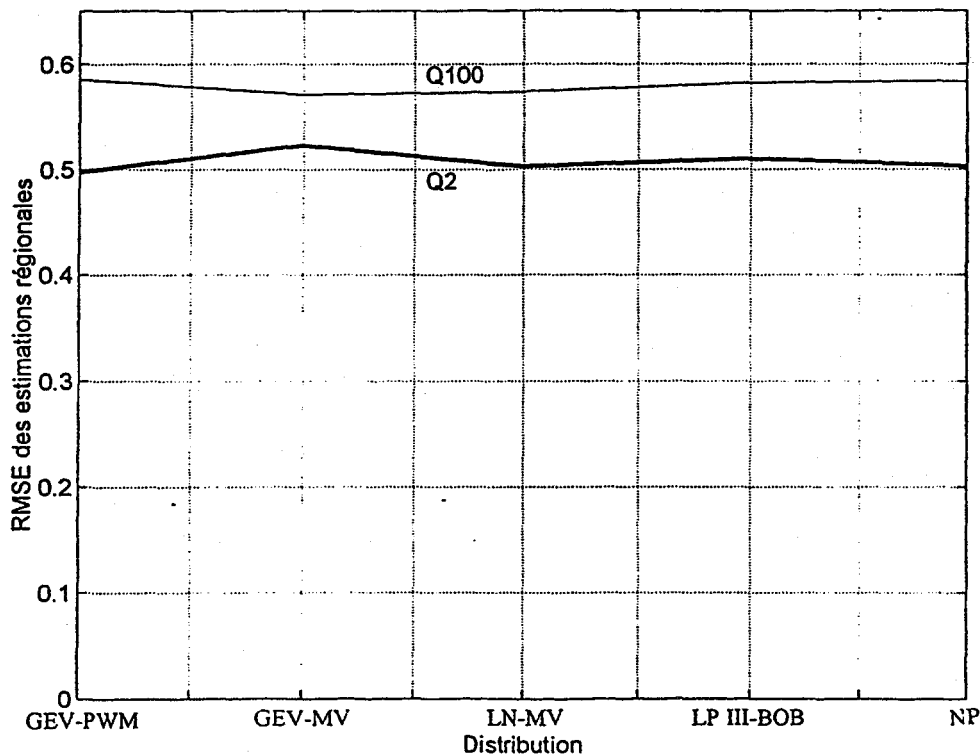


Figure 7.1: Sensibilité de la méthode au modèle d'estimation locale

8 Sensibilité de la méthode aux changements dans les réseaux hydrométriques

Il arrive de plus en plus fréquemment que l'on doive faire face à des réductions des réseaux de mesures hydrométriques. Le but de cette étude est de déterminer l'influence des changements dans les réseaux de mesures sur l'estimation des coefficients de corrélations canoniques et, par conséquent, l'estimation régionale des quantiles de crue.

Pour se faire, une méthode de ré-échantillonnage de type bootstrap est utilisée. Parmi les 106 stations disponibles, dix sont choisies aléatoirement et utilisées comme stations cibles tout au long de cette étude. Les numéros de ces stations sont 20, 24, 33, 42, 49, 66, 72, 78, 87 et 103. Chacune des stations cibles est consécutivement considérée comme étant non jaugée et éliminée de la base de données. Les estimations des quantiles de crue à cette station sont alors faites en enlevant aléatoirement des stations de la base de données. Une station est d'abord enlevée puis 2, puis 3 et ainsi de suite jusqu'à l'élimination de 53 stations, ce qui représente la moitié de la base de données. L'élimination d'un nombre plus grand de stations n'est pas considérée car elle conduit à des tailles de voisinage trop petites.

On note que le nombre de combinaisons possibles de x stations enlevées ($x = 2, 3, \dots, 53$) augmente rapidement et peut devenir considérable. Pour l'élimination de seulement 5 stations, près de 100 millions de combinaisons sont possibles ($C_{106}^{105} = 96\ 560\ 646$). Compte tenu des capacités de calcul des ordinateurs, l'estimation des quantiles pour toutes les combinaisons existantes est trop longue. C'est pourquoi, pour chaque station cible considérée (sauf pour l'élimination d'une station où seulement 105 combinaisons sont possibles) les estimations des quantiles sont effectuées en considérant 1000 combinaisons de x stations choisies aléatoirement. De plus, les quantiles sont estimés pour l'élimination de 1 à 5 stations puis pour les nombres impairs de stations ($x = 1, 2, 3, 4, 5, 7, 9, 11, \dots, 53$).

Les variables utilisées et les transformations de ces variables sont les mêmes que dans la Section 5. L'estimation des quantiles de crue est effectuée par régression sur les stations voisines comme dans l'exemple d'application présenté précédemment. À partir de ces estimations régionales en considérant les estimations aux sites comme celles de références, les erreurs quadratiques moyennes relatives des estimations en enlevant x stations sont calculées. La notation est la même que dans la section précédente. La formule utilisée pour ces calculs est la suivante:

$$[\text{REQM}]_r = 100 \left\{ \frac{1}{10} \sum_{j=1}^{10} \left[\sqrt{\frac{1}{1000} \sum_{i=1}^{1000} \frac{(\tilde{q}(i,j) - q(j))^2}{q(j)^2}} \right] \right\} \%$$

où i est l'indice du quantile estimé et j est l'indice de la station cible.

La Figure 8.1 présente le $[\text{REQM}]_r$ calculé en fonction du nombre de stations enlevées pour chaque quantile. On remarque d'abord que les $[\text{REQM}]_r$ relatifs pour l'estimation de Q_{100} sont moins élevés que ceux pour l'estimation de Q_2 . On constate aussi que les variations du $[\text{REQM}]_r$ semblent assez faibles. La méthode de l'analyse des corrélations canoniques semble donc robuste aux diminutions des réseaux hydrométriques. Toutefois, on observe une plus grande variation du $[\text{REQM}]_r$ pour le quantile Q_{100} qui semble plus sensible aux diminutions des réseaux hydrométriques.

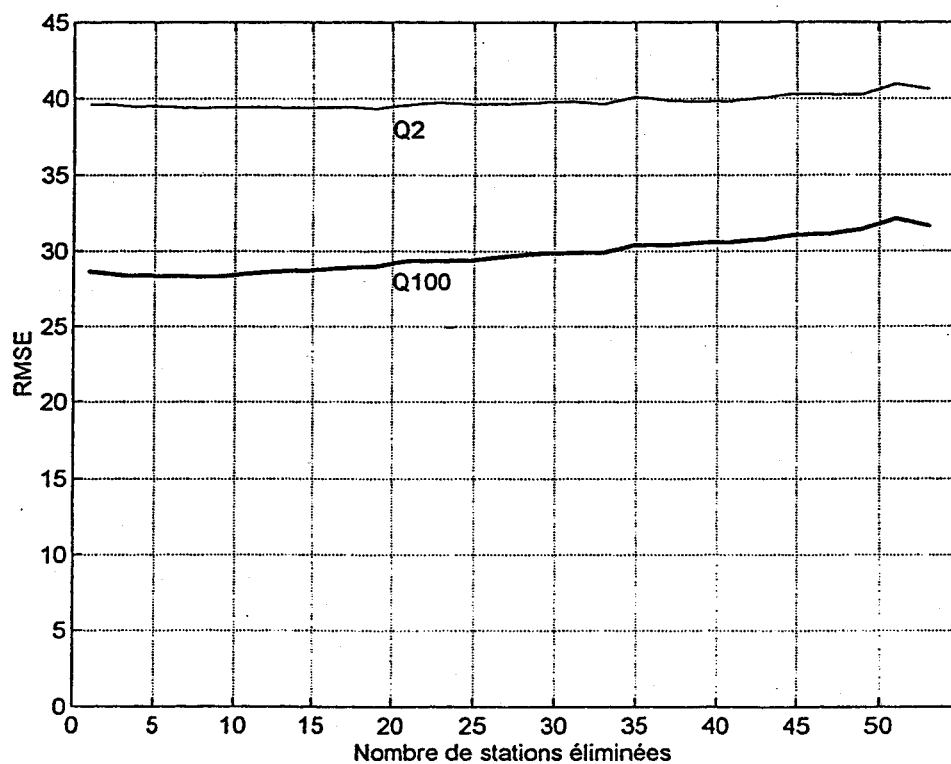


Figure 8.1: Sensibilité de la méthode à la réduction du réseau hydrométrique

9 Étude des vecteurs d'erreurs

L'étude des erreurs d'estimation des variables canoniques dans l'espace canonique hydrologique (W_1, W_2) est réalisée en faisant le graphique des points $w_i = (w_{1i}, w_{2i})$ et $\hat{w}_i = (\hat{w}_{1i}, \hat{w}_{2i})$, où i est le nombre d'observations et $\hat{w}_k$ est l'estimation de w_k ($k=1,2$), et suivant le vecteur correspondant $\overline{\hat{w}_i w_i}$. Le but de cette étude est de déterminer les variables qui influencent la longueur et la direction des vecteurs d'erreurs. L'étude de ces vecteurs d'erreurs peut être remplacée par l'étude des vecteurs d'erreurs des variables hydrologiques originales de coordonnées $(Q_{2i}, Q_{100i} / Q_{2i})$ et $(\hat{Q}_{2i}, \hat{Q}_{100i} / \hat{Q}_{2i})$ [Cavadias 1995]. L'avantage de ces derniers graphiques est qu'ils sont directement interprétables.

Ces vecteurs d'erreurs peuvent être déterminés pour chaque site de l'étude. La Figure 9.1 présente un exemple de 11 vecteurs d'erreurs obtenus à partir de la base de données étudiée (les lettres majuscules indiquent les points $(Q_{2i}, Q_{100i} / Q_{2i})$). Les valeurs de Q_2 et Q_{100} utilisées sont celles estimées au site avec une loi GEV (PWM). Les valeurs de $\hat{Q}_2$ et $\hat{Q}_{100}$ sont estimées de la même façon que celle présentée à la Section 5. Puisque l'on dispose de 106 stations jaugées, 106 vecteurs d'erreurs peuvent être calculés. Pour calculer les différentes caractéristiques des vecteurs d'erreurs, les valeurs de Q_{100}/Q_2 et de $\hat{Q}_{100}/\hat{Q}_2$ sont multipliées par 100 afin d'avoir des échelles similaires.

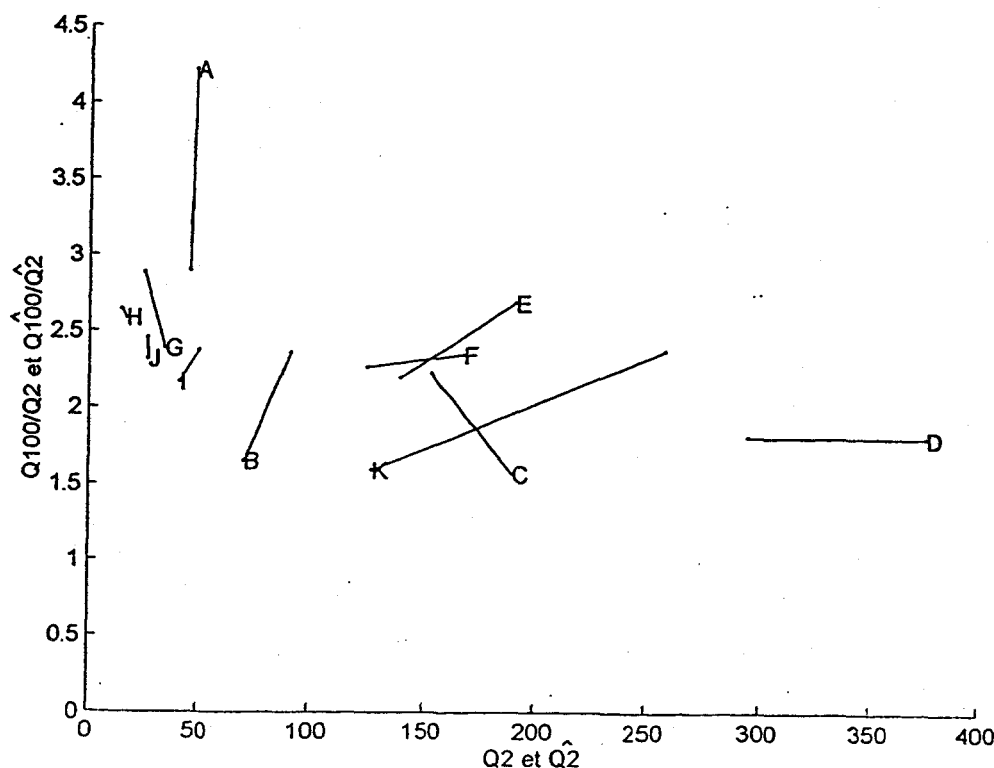


Figure 9.1: Exemple de vecteurs d'erreurs

Les différentes caractéristiques des vecteurs d'erreurs calculées sont: la norme, la composante verticale, la composante horizontale et l'angle (en degré) fait par le vecteur d'erreur avec l'axe horizontal. La valeur de l'angle utilisée est en fait une valeur transformée. En effet, les transformations suivantes sont appliquées:

$$\begin{aligned} 0^\circ \leq \text{angle} \leq 90^\circ & \rightarrow \text{angle utilisé} = \text{angle} \\ 90^\circ < \text{angle} \leq 180^\circ & \rightarrow \text{angle utilisé} = 180 - \text{angle} \\ 180^\circ < \text{angle} \leq 270^\circ & \rightarrow \text{angle utilisé} = \text{angle} - 180 \\ 270^\circ < \text{angle} < 360^\circ & \rightarrow \text{angle utilisé} = 360 - \text{angle} \end{aligned}$$

Ces transformations ramènent les angles à leur équivalent dans le premier quadrant (par symétrie). L'angle utilisé varie donc entre 0° et 90° et permet de considérer deux directions: les petites valeurs d'angles utilisés représentent un vecteur dans la direction horizontale tandis que les grandes valeurs représentent un vecteur dans la direction verticale.

En examinant les valeurs des composantes, on remarque que la composante verticale est en général plus grande que la composante horizontale. C'est-à-dire que l'erreur sur l'estimation de Q_{100}/Q_2 multipliée par 100 est en général plus grande que celle faite en estimant Q_2 . Les graphiques de chacune des caractéristiques des vecteurs d'erreurs en fonction de plusieurs variables sont produits. En premier lieu, on constate que la superficie du bassin versant, la surface contrôlée par les lacs et marais et la précipitation totale moyenne annuelle ne semblent pas avoir d'effet sur les vecteurs d'erreurs (Figures C.5 à C.16, Annexe C).

En second lieu, sur la Figure 9.2, on remarque qu'une grande valeur de la pente du cours d'eau principal est généralement associée à une grande valeur de l'angle fait par le vecteur avec l'horizontal (vecteurs d'erreurs verticaux). De plus, on voit que les angles petits sont associés aux petites valeurs de pente (vecteurs d'erreurs horizontaux). Finalement, sur la Figure 9.4, on constate que si les vecteurs d'erreurs sont verticaux on doit s'attendre à une petite valeur de la longueur du cours d'eau principal. Sur la Figure 9.3, on remarque qu'une grande valeur de la composante horizontale (ie une grande erreur sur l'estimation de Q_2) correspond à une petite valeur de la pente et les grandes valeurs de la pente sont associées à des composantes horizontales de petites valeurs.

Ensuite, avec la Figure 9.4, on constate que les grandes valeurs d'angles correspondent à des petites valeurs de la longueur du cours d'eau principal et que les grandes valeurs de la longueur sont généralement associées aux petites valeurs d'angles. De plus, sur la Figure 9.5, on remarque qu'une grande valeur de la composante verticale (ie une grande erreur sur l'estimation de Q_{100}/Q_2) est associée à une petite valeur de la longueur et une grande valeur

de la longueur correspond à une composante verticale de petite valeur. Finalement, les autres caractéristiques des vecteurs d'erreurs ne semblent pas être influencées par la pente et la longueur du cours d'eau principal (Figures C.17 à C.20, Annexe C).

En dernier lieu, l'influence de la position d'une station sur le vecteur d'erreur a été considérée. Pour se faire, les graphiques de chacune des caractéristiques des vecteurs d'erreurs en fonction de la latitude et de la longitude de chaque station ont été tracés. Ces graphiques sont présentés à l'Annexe C (Figures C.21 à C.28). On constate que la position de la station ne semble pas avoir d'effet sur les vecteurs d'erreurs.

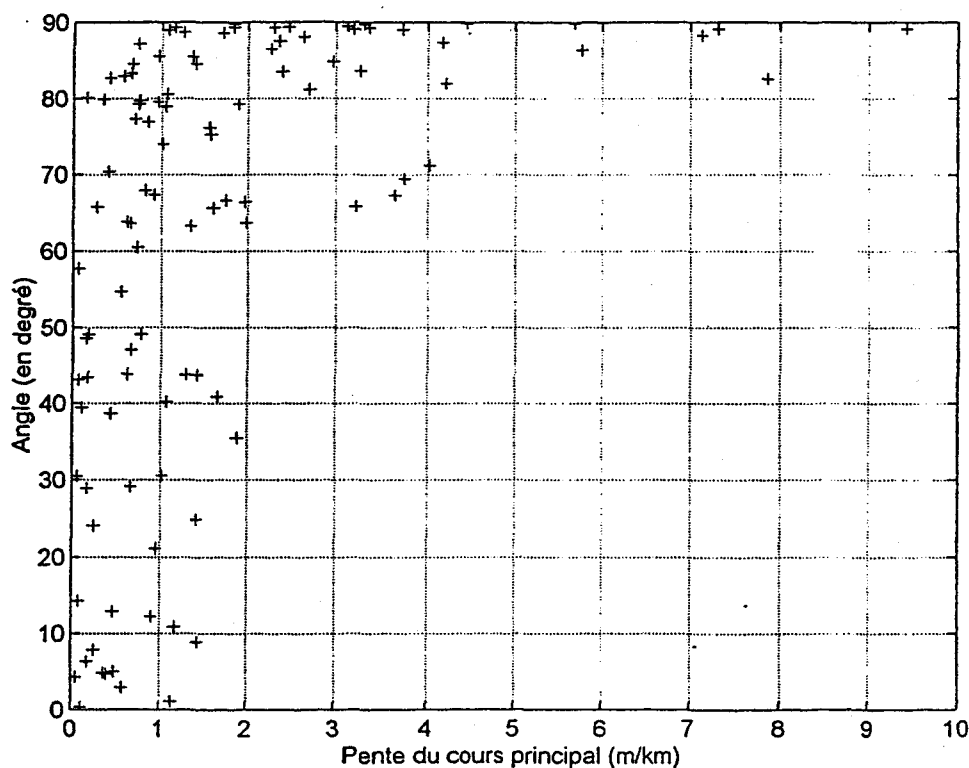


Figure 9.2: Effet de la pente du cours d'eau principal sur l'angle des vecteurs d'erreurs

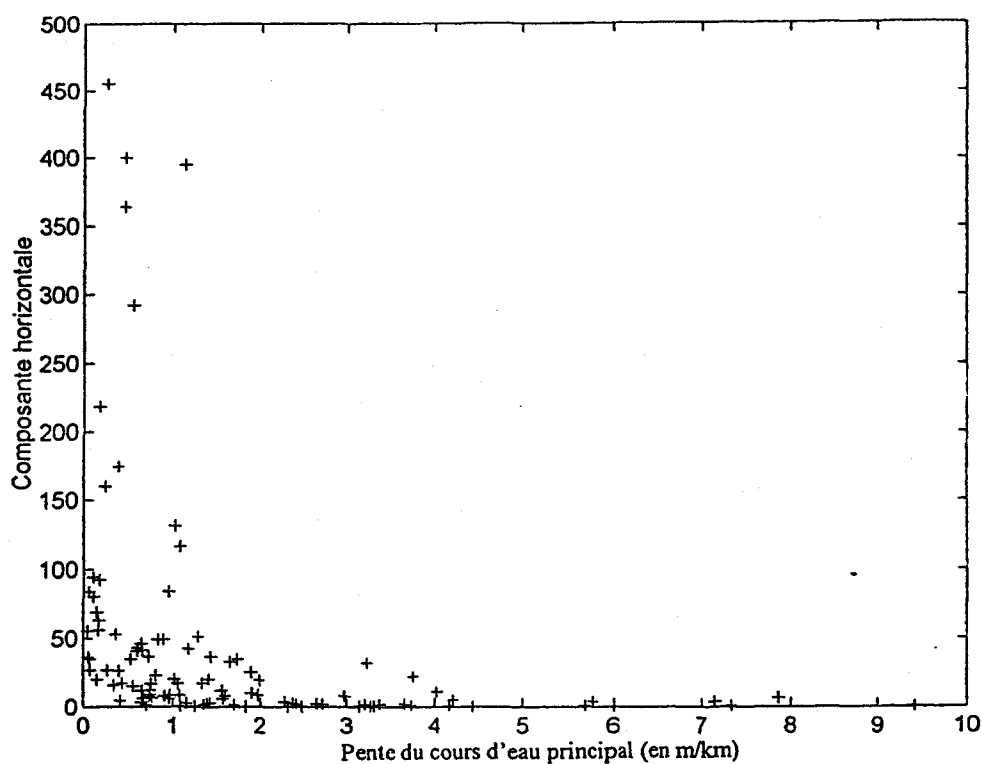


Figure 9.3: Effet de la pente du cours d'eau principal sur la composante horizontale des vecteurs d'erreurs

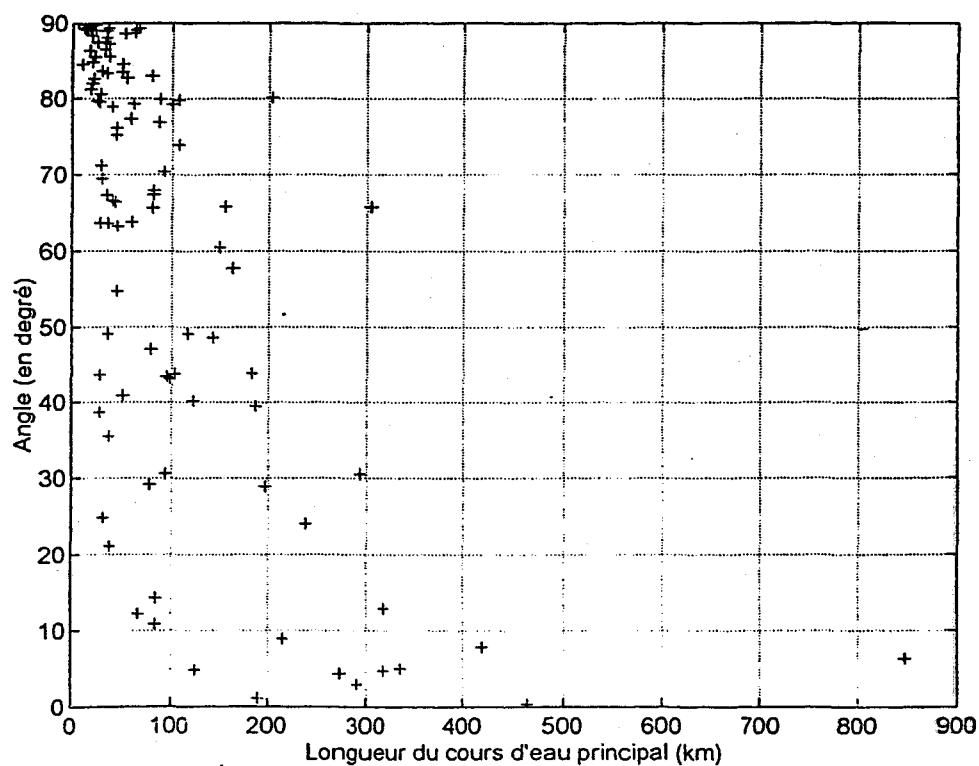


Figure 9.4: Effet de la longueur du cours d'eau principal sur l'angle des vecteurs d'erreurs.

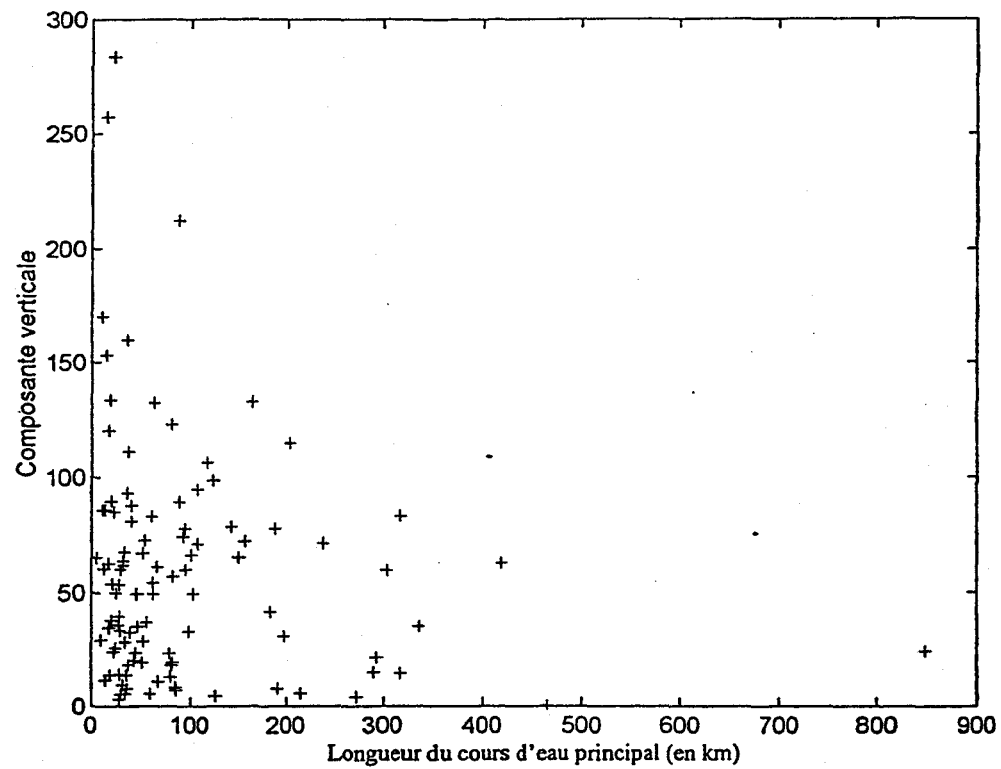


Figure 9.5: Effet de la longueur du cours d'eau principal sur la composante verticale des vecteurs d'erreurs

10 Conclusion

Certaines caractéristiques de la méthode de l'analyse des corrélations canoniques ont été étudiées à partir d'une base de données comprenant 106 stations jaugées de la province de l'Ontario. Cette méthode produit des estimations adéquates des quantiles de crue à des sites non jaugés. Pour ces sites, un voisinage leur était attribué puis les quantiles de crue y étaient estimés. Des procédures jackknif ou bootstrap ont montré que l'analyse des corrélations canoniques ne semble pas être influencée par le choix du modèle d'estimation locale ni par la diminution du nombre de stations. Cette méthode semble donc robuste. De même, pour cette base de données, un choix de seuil entre 0.20 et 0.40 est préconisé. Finalement, les vecteurs d'erreurs semblent reliés avec la pente et la longueur du cours d'eau principal.

11 Références

- Cavadias, G.S. (1990), The canonical correlation approach to regional flood estimation. *Regionalisation in hydrology, Proc. of the Ljubljana Symposium*, IAHS, Publ. no 191: 171-178, avril.
- Cavadias, G.S. (1995), Regionalization and multivariate analysis : the canonical correlation approach. *An international conference in honour of Jacques Bernier*, Unesco. Paris, 11-13 septembre. (Version préliminaire.)
- GREHYS (1996). Intercomparaison of flood frequency procedures for canadian rivers. *Journal of Hydrology*, 186(1-4): 85-103.
- Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28: 321-377.
- Kettenring, J.R. (1971). Canonical analysis for several sets of variables. *Biometrika*, 58: 433-451.
- Kshirsagar, A.M. (1972). *Multivariate Analysis*, New-York, Dekker, 534 p.
- Lebart, L., A. Morineau et N. Tabard (1977). *Techniques de la description statistique: Méthodes et logiciels pour l'analyse des grands tableaux*, Paris, Bordas, 351 p.
- Mathier, L., R. Roy, B. Bobée, H. Perron et V. Fortin (1993), *Estimation régionale des crues: description des banques de données hydrométriques, météorologiques et physiographiques pour le Québec et l'Ontario*, Ste-Foy, rapport interne INRS-Eau no I-123, 79 p.
- Ribeiro-Corréa, J., G.S. Cavadias, B. Clément, et J. Rouselle (1995), Identification of hydrological neighborhoods using canonical correlation analysis. *Journal of Hydrology*, 173: 71-89.
- Ribeiro-Corréa, J. et J. Rouselle (1993), *Détermination des régions homogènes par la méthode d'analyse des corrélations canoniques*, Montréal, rapport final Ecole polytechnique de Montréal, Département de génie civil, Section hydraulique, 49 p.
- Torranin, P. (1972). *Applicability of canonical correlation in hydrology*. Colorado State University, Hydrology Paper 58, 30p.

APPENDICE :
aspect théorique de l'analyse
des corrélations canoniques

La méthode de l'analyse des corrélations canoniques est une méthode statistique multivariée permettant de montrer la relation de dépendance linéaire entre deux ensembles de variables aléatoires. Lorsqu'un de ces ensembles contient une seule variable, on retrouve le cas particulier de la régression multivariée. La méthode de l'analyse des corrélations canoniques a été développée par [Hotelling 1936] et a été étendue pour l'analyse de plusieurs ensembles de variables aléatoires par [Kettenring 1971]. Une brève description de l'aspect théorique est présentée ici. Une présentation plus complète est faite par [Lebart et al. 1977] et [Kshirsagard 1972].

Soit $Y=(Y_1, \dots, Y_p)$ et $X=(X_1, \dots, X_q)$, deux ensembles de variables aléatoires de dimension p et q respectivement. Sans perte de généralité, on peut supposer que $p \leq q$. Posons $\text{Var}(X)=\Sigma_X$, $\text{Var}(Y)=\Sigma_Y$ et $\text{Cov}(X,Y)=\Sigma_{XY}$. La matrice de covariance entre X et Y est alors de rang p .

Soit les combinaisons linéaires des variables aléatoires suivantes: $W = \alpha_1 Y_1 + \dots + \alpha_p Y_p = \alpha'Y$ et $V = \beta_1 X_1 + \dots + \beta_q X_q = \beta'X$. De la théorie des combinaisons linéaires des variables aléatoires, on obtient $\text{Var}(W) = \alpha' \Sigma_Y \alpha$, $\text{Var}(V) = \beta' \Sigma_X \beta$ et $\text{Cov}(W,V) = \alpha' \Sigma_{YX} \beta$. Ce qui implique que le carré du coefficient de corrélation entre W et V est

$$\rho^2_{WV} = \frac{(\alpha' \Sigma_{YX} \beta)}{(\alpha' \Sigma_Y \alpha)(\beta' \Sigma_X \beta)}.$$

Donc, pour une matrice Σ_{YX} connue, ρ^2_{WV} est fonction des vecteurs α et β . Il est possible de démontrer que ρ^2_{WV} a p maximum locaux lorsqu'on fait varier α et β car Σ_{YX} est considéré de rang p . ρ^2_{WV} a alors p valeurs stationnaires qui sont les valeurs propres de la matrice $\Sigma_Y^{-1} \Sigma_{YX} \Sigma_X^{-1} \Sigma_{XY}$ c'est-à-dire qu'elles sont les p solutions, notées λ^2 , de l'équation caractéristique $|\Sigma_Y^{-1} \Sigma_{YX} \Sigma_X^{-1} \Sigma_{XY} - \lambda^2 I_p| = 0$. Il est possible d'ordonner les valeurs propres ainsi $0 \leq \lambda_p^2 \leq \lambda_{p-1}^2 \leq \dots \leq \lambda_1^2 \leq 1$ puisque $0 \leq \rho^2_{WV} \leq 1$. La racine carrée positive des valeurs de λ^2 obtenues sont les coefficients de corrélations canoniques. À chaque valeur propre λ_i^2 correspond la $i^{\text{ème}}$ corrélation canonique entre les variables canoniques W_i et V_i .

Chaque valeur propre λ_i^2 représente une direction dans l'espace Euclidien. Elle est donc associée à une infinité de vecteurs propres. Une contrainte doit alors être ajoutée afin de n'avoir qu'un seul vecteur propre associé. Les variances des vecteurs propres sont donc contraints à être égales à 1 c'est-à-dire $\text{Var}(\alpha'Y) = \text{Var}(\beta'X) = 1$. Les vecteurs propres α_i et β_i définissent alors une paire de variables canoniques W_i et V_i correspondant à un coefficient de corrélation canonique λ_i .

ANNEXE A :
bases de données

Tableau A.1: Base de données de la province de l'Ontario

No.	Code	Aire	PCP	LCP	SLM	PTMA
1	02AC001	736	3.189	63.50	257.6	687.4
2	02BA002	1190	1.015	107.50	1166.2	863.0
3	02BB002	1980	0.731	150.00	1405.8	863.0
4	02BB003	4270	0.111	465.00	2348.5	863.0
5	02BD003	1930	1.442	215.00	1775.6	691.7
6	02BF001	1190	1.289	103.00	690.2	989.6
7	02BF002	1160	1.173	84.50	962.8	989.6
8	02CA002	108	5.770	17.50	25.9	989.6
9	02CF007	272	0.746	24.50	51.7	868.0
10	02DD008	90	3.649	34.25	25.2	878.2
11	02EA005	321	1.970	43.60	292.1	1058.2
12	02EA010	149	1.838	36.00	41.7	1011.5
13	02EC002	1445	1.020	94.30	0.0	962.9
14	02EC009	181	2.281	32.50	7.2	817.2
15	02EC010	42.9	3.123	14.00	2.1	821.9
16	02EC011	282	0.415	55.25	208.7	948.2
17	02ED003	1180	1.900	61.75	601.8	916.7
18	02ED007	177	1.989	28.00	0.0	1006.6
19	02FA001	927	0.812	82.40	908.5	968.9
20	02FB007	181	1.430	27.60	104.9	1060.9
21	02FC001	3960	1.128	190.50	0.0	1060.9
22	02FC002	2150	1.600	81.00	236.5	1060.9
23	02FC004	249	1.892	37.75	204.2	1060.9
24	02FC011	163	3.750	29.50	0.0	1060.9
25	02FE008	648	0.908	81.71	12.9	1072.7
26	02FE009	376	1.340	46.46	0.0	1072.7
27	02FF002	865	0.740	100.85	0.0	956.2
28	02FF004	41.4	3.364	11.00	0.0	956.2
29	02FF007	466	1.656	52.50	0.0	876.4
30	02GA010	1008	0.950	39.40	181.4	866.4
31	02GA017	326	1.746	40.46	0.0	906.2
32	02GA018	552	0.900	67.90	0.0	864.8
33	02GB007	360	1.250	62.50	72.0	876.1
34	02GB009	92	1.360	22.30	0.0	864.8
35	02GC002	329	1.560	45.50	0.0	954.0
36	02GC010	342	1.056	40.50	0.0	967.1
37	02GC018	287	1.418	31.50	0.0	954.0
38	02GD010	150	1.070	28.10	0.0	949.7
39	02GD019	36	1.400	10.31	0.0	1046.2
40	02GD020	96	2.370	32.51	0.0	949.7
41	02GE005	146	0.966	37.50	0.0	954.0
42	02GG002	730	0.660	79.00	0.0	956.2
43	02GG004	609	0.341	88.75	0.0	912.6
44	02GG005	172	0.565	81.00	0.0	956.2
45	02GG006	267	0.735	36.50	0.0	912.6

Tableau A.1: (suite)

No.	Code	Aire	PCP	LCP	SLM	PTMA
46	02HA006	293	0.654	80.00	0.0	837.7
47	02HB004	199	2.650	34.05	0.0	836.5
48	02HB012	83	2.314	14.00	4.1	855.4
49	02HC009	197	2.409	51.00	27.6	780.6
50	02HC013	88	6.000	15.20	0.0	786.1
51	02HC018	106	7.870	21.50	2.1	908.8
52	02HC019	94	9.420	17.48	6.6	908.8
53	02HC023	62	4.438	11.75	3.1	780.6
54	02HC025	303	1.571	45.00	15.2	780.6
55	02HC027	58	3.729	22.50	0.0	769.9
56	02HC028	78	3.302	17.50	0.0	868.5
57	02HC029	130	4.212	20.00	0.0	874.5
58	02HC030	204	4.023	28.75	0.0	769.9
59	02HC031	148	4.168	24.00	0.0	780.6
60	02HC032	95	1.841	21.00	18.1	795.6
61	02HC033	71	3.267	29.50	0.0	769.9
62	02HD002	232	2.975	20.00	25.5	837.7
63	02HD006	82.9	5.680	22.00	5.8	883.2
64	02HD007	78	7.140	19.05	0.0	883.2
65	02HD008	125	7.319	13.25	2.5	802.1
66	02HD009	83	1.153	15.00	0.0	883.2
67	02HE001	13.9	2.487	5.50	0.0	855.3
68	02HJ001	110	2.714	18.50	97.9	838.8
69	02HL004	671	2.472	67.00	369.5	847.0
70	02HL005	308	1.710	53.25	52.4	820.7
71	02HM004	112	1.079	34.50	13.4	944.1
72	02HM005	155	0.957	27.00	54.3	954.8
73	02JC008	1780	0.174	95.25	1424.0	882.3
74	02LA007	559	0.608	60.50	95.0	886.8
75	02LB006	427	0.650	36.30	0.0	886.8
76	02LB007	246	0.436	28.00	236.2	981.6
77	02LB008	440	0.540	46.20	0.0	908.8
78	02MC001	358	0.670	51.20	0.0	973.7
79	04CA002	36500	0.177	847.72	21900.0	600.6
80	04CA003	619	0.706	59.06	123.8	600.6
81	04CB001	10800	0.054	294.01	6480.0	600.6
82	04CE002	4350	0.070	98.43	4350.0	600.6
83	04DA001	5960	0.153	203.20	2384.0	663.3
84	04DB001	7950	0.264	304.17	7950.0	600.6
85	04DC001	50000	0.254	419.73	40000.0	600.6
86	04DC002	4710	3.217	155.90	0.0	600.6
87	04EA001	10400	0.567	290.83	1040.0	600.6
88	04FA003	4900	0.051	273.05	1470.0	663.3
89	04FB001	24200	0.393	317.50	24200.0	663.3
90	04FC001	36000	0.464	317.50	34200.0	663.3

Tableau A.1: (suite)

No.	Code	Aire	PCP	LCP	SLM	PTMA
91	04GA002	5390	0.057	163.83	5390.0	701.1
92	04JA002	3780	0.108	187.96	2797.2	691.7
93	04JC003	3290	0.760	107.70	32.9	691.7
94	04JD005	2020	0.176	117.47	161.6	691.7
95	04JF001	5360	0.620	182.90	5360.0	691.7
96	04LJ001	8940	0.475	335.60	1609.2	764.7
97	04MF001	6680	1.076	123.82	0.0	752.3
98	05PA006	13400	0.169	196.90	13400.0	701.6
99	05PB014	4870	0.359	125.73	4870.0	701.6
100	05PC010	168	0.657	35.05	0.0	698.6
101	05PC011	461	0.770	36.19	0.0	689.6
102	05QA001	13900	0.240	238.13	13900.0	712.9
103	05QA004	4450	0.148	142.90	1468.5	712.9
104	05QC003	2370	0.075	-85.09	2133.0	661.4
105	05QE008	1690	0.838	87.62	1690.0	661.4
106	05QE009	1530	0.400	92.71	765.0	629.5

Tableau A.2: Estimation des quantiles de crue suite à l'ajustement d'une loi GEV (PWM)

No.	Q2	Q100	No.	Q2	Q100
=====	=====	=====	=====	=====	=====
1	48.44	204.37	54	25.58	77.54
2	71.51	118.02	55	10.86	23.99
3	190.07	298.72	56	14.92	27.47
4	375.87	680.13	57	18.06	61.71
5	145.47	334.36	58	39.48	109.17
6	191.36	514.83	59	22.56	76.56
7	167.75	393.97	60	10.37	25.79
8	23.86	53.03	61	12.30	35.44
9	34.83	83.17	62	33.00	132.04
10	17.30	44.77	63	14.07	80.59
11	41.52	90.01	64	12.57	54.57
12	27.50	63.75	65	17.64	71.09
13	127.22	202.83	66	10.37	58.43
14	19.78	68.82	67	2.39	6.80
15	6.03	13.85	68	13.83	39.83
16	37.49	73.82	69	64.69	125.12
17	98.91	303.39	70	32.16	61.10
18	21.02	47.60	71	19.72	44.27
19	118.29	200.98	72	27.66	63.10
20	26.52	67.38	73	174.25	300.86
21	481.71	963.62	74	97.66	161.23
22	269.07	665.67	75	105.99	195.15
23	15.16	39.23	76	43.90	107.22
24	23.84	75.85	77	102.23	317.95
25	137.37	326.20	78	75.73	135.60
26	75.95	210.01	79	573.34	1091.86
27	152.18	449.01	80	27.77	83.69
28	11.45	54.56	81	180.41	401.54
29	109.72	322.81	82	47.99	122.37
30	162.10	400.07	83	204.69	739.94
31	92.95	175.17	84	227.95	694.09
32	134.78	357.88	85	1298.93	4033.55
33	44.49	87.70	86	288.56	648.34
34	17.50	54.95	87	614.58	1664.45
35	78.42	239.81	88	211.96	493.71
36	47.99	166.53	89	765.24	1962.18
37	73.70	207.62	90	1171.26	3861.01
38	38.16	123.29	91	66.95	251.69
39	11.79	30.76	92	223.98	359.86
40	20.35	66.02	93	201.44	368.71
41	22.09	80.77	94	185.64	283.77
42	91.24	237.49	95	154.06	447.56
43	83.29	268.95	96	867.40	1577.68
44	25.35	87.08	97	689.77	1216.67
45	51.96	222.17	98	291.45	630.90
46	59.84	137.88	99	108.60	278.94
47	35.13	76.48	100	19.55	54.16
48	11.83	36.00	101	31.19	101.65
49	20.91	52.21	102	244.11	756.87
50	13.39	65.16	103	91.79	302.28
51	15.13	36.15	104	37.33	100.17
52	18.10	48.19	105	16.85	82.99
53	8.05	20.31	106	28.44	101.21

Tableau A.3: Estimation des quantiles de crue suite à l'ajustement d'une loi GEV (MV)

No.	Q2	Q100	No.	Q2	Q100
=====	=====	=====	=====	=====	=====
1	50.62	180.57	54	25.61	76.88
2	72.71	109.28	55	10.97	22.40
3	189.65	297.42	56	15.08	25.94
4	378.23	652.61	57	18.06	62.18
5	146.88	328.11	58	40.28	100.84
6	192.86	494.64	59	22.73	75.59
7	170.97	366.06	60	10.44	24.44
8	24.26	48.05	61	12.27	35.52
9	34.78	83.13	62	32.00	151.91
10	17.68	41.04	63	14.23	81.28
11	41.56	88.81	64	12.56	54.58
12	27.91	58.70	65	17.72	70.27
13	126.63	208.85	66	10.59	56.18
14	18.80	142.67	67	2.36	7.39
15	6.13	12.57	68	14.22	36.16
16	37.62	72.23	69	64.49	125.65
17	99.02	302.21	70	32.57	55.66
18	21.15	45.87	71	19.87	43.07
19	118.50	195.18	72	28.01	59.88
20	28.68	65.45	73	178.40	266.96
21	480.85	962.23	74	---	---
22	270.67	648.43	75	106.91	185.72
23	15.73	34.13	76	43.74	106.71
24	24.15	72.62	77	102.79	310.55
25	140.17	297.47	78	76.03	130.12
26	77.87	190.36	79	581.23	1006.34
27	151.60	447.17	80	28.42	76.55
28	---	---	81	180.59	386.56
29	111.94	301.44	82	48.18	117.83
30	164.53	369.47	83	207.46	715.85
31	95.49	158.09	84	230.04	663.12
32	132.95	369.54	85	1273.00	4452.29
33	45.57	79.13	86	291.33	606.21
34	17.61	53.18	87	609.32	1693.16
35	80.08	224.14	88	213.69	464.58
36	47.56	171.23	89	771.43	1858.55
37	74.61	198.07	90	1192.38	3618.07
38	37.89	125.12	91	67.20	250.57
39	11.87	29.80	92	225.42	341.70
40	20.51	64.24	93	204.24	342.99
41	21.82	90.34	94	186.08	279.71
42	92.84	222.63	95	159.25	396.17
43	81.04	320.37	96	865.67	1588.94
44	26.22	78.08	97	699.61	1139.86
45	49.34	295.40	98	284.54	680.38
46	59.87	134.21	99	107.45	279.51
47	35.00	78.28	100	19.47	54.00
48	12.02	33.91	101	30.30	107.98
49	21.06	49.37	102	244.93	746.84
50	12.90	80.99	103	92.86	288.23
51	15.54	32.12	104	37.32	99.57
52	18.14	47.62	105	16.97	84.54
53	8.18	18.43	106	28.65	98.37

Tableau A.4: Estimation des quantiles de crue suite à l'ajustement d'une loi log-normale (MV)

No.	Q2	Q100	No.	Q2	Q100
1	47.33	245.11	54	25.72	75.59
2	65.54	168.63	55	10.59	26.73
3	181.39	369.84	56	13.53	40.63
4	353.59	861.96	57	18.30	58.54
5	145.57	338.14	58	39.32	113.37
6	196.01	465.64	59	23.66	63.14
7	165.53	430.73	60	10.03	29.38
8	23.33	58.48	61	12.44	33.86
9	34.28	88.82	62	32.13	142.36
10	17.83	39.65	63	14.95	63.65
11	41.65	88.72	64	11.78	67.49
12	28.49	54.71	65	16.46	96.11
13	123.89	231.18	66	11.04	45.70
14	20.69	57.21	67	2.44	6.28
15	5.87	15.53	68	13.79	41.72
16	35.40	95.78	69	63.21	138.53
17	100.01	287.62	70	31.94	62.61
18	20.88	49.14	71	19.62	45.85
19	111.04	260.51	72	27.86	62.06
20	25.99	73.30	73	170.94	327.98
21	460.39	1163.44	74	---	---
22	266.29	695.86	75	99.52	253.35
23	15.81	33.25	76	42.79	117.70
24	23.89	75.39	77	101.45	326.63
25	134.28	367.02	78	71.51	171.18
26	74.55	234.83	79	549.34	1293.32
27	148.10	493.59	80	28.28	79.04
28	---	---	81	181.90	386.00
29	108.10	353.50	82	48.47	117.68
30	150.82	542.63	83	214.54	617.55
31	82.80	272.12	84	228.57	692.31
32	132.99	371.67	85	1311.65	3857.34
33	42.14	111.33	86	289.12	646.75
34	17.35	57.02	87	616.46	1636.72
35	76.58	272.53	88	208.54	528.30
36	46.52	184.61	89	746.47	2169.17
37	71.74	236.67	90	1217.27	3331.39
38	36.10	149.40	91	68.05	235.50
39	11.80	30.81	92	221.57	376.50
40	20.78	60.73	93	200.55	379.61
41	23.27	64.97	94	176.13	368.42
42	89.44	263.00	95	158.92	403.44
43	85.47	242.16	96	842.30	1803.97
44	25.67	85.63	97	630.10	1822.82
45	51.76	220.79	98	260.32	977.36
46	57.09	165.23	99	103.02	331.45
47	33.25	97.08	100	17.66	79.33
48	11.97	34.62	101	27.23	158.66
49	19.20	71.66	102	243.91	751.95
50	13.53	60.38	103	90.45	323.94
51	13.64	52.10	104	35.78	119.36
52	16.40	75.91	105	16.99	80.63
53	7.55	25.34	106	27.00	125.59

Tableau A.5: Estimation des quantiles de crue suite à l'ajustement d'une loi log-Pearson type III (BOB)

No.	Q2	Q100	No.	Q2	Q100
=====	=====	=====	=====	=====	=====
1	49.48	182.10	54	26.09	70.67
2	72.51	108.09	55	11.04	22.02
3	191.22	286.28	56	15.08	25.30
4	378.6	640.96	57	18.73	53.05
5	144.19	342.92	58	40.31	99.18
6	194.52	477.79	59	23.39	66.30
7	169.74	370.06	60	10.57	23.33
8	24.37	47.65	61	12.84	29.86
9	34.56	83.64	62	34.13	110.41
10	17.50	42.28	63	14.77	67.11
11	41.85	86.93	64	12.06	56.67
12	28.51	54.59	65	17.97	61.28
13	126.82	205.36	66	10.74	51.01
14	20.87	56.38	67	2.51	5.68
15	6.17	12.37	68	14.17	35.89
16	37.58	70.81	69	64.33	125.92
17	101.18	275.04	70	32.74	55.99
18	21.15	45.67	71	19.70	43.85
19	118.85	191.92	72	27.99	59.70
20	26.72	64.21	73	177.09	276.12
21	480.71	952.38	74	---	---
22	271.13	639.22	75	106.66	185.45
23	15.40	36.46	76	44.04	103.21
24	24.17	70.81	77	103.92	291.98
25	140.23	295.01	78	76.11	129.05
26	77.93	186.57	79	579.58	1020.41
27	156.71	391.09	80	28.54	74.68
28	---	---	81	185.09	357.61
29	110.84	302.86	82	49.60	106.41
30	164.87	359.60	83	215.89	608.20
31	94.24	157.61	84	236.88	590.91
32	138.18	319.22	85	1348.78	3446.81
33	45.23	80.26	86	295.34	585.01
34	18.13	47.14	87	636.66	1430.98
35	79.02	225.49	88	215.62	451.64
36	49.62	140.75	89	778.89	1782.68
37	74.00	197.60	90	1222.33	3278.16
38	37.84	118.90	91	68.17	229.73
39	11.85	29.62	92	225.56	346.03
40	20.84	59.71	93	204.29	346.44
41	22.96	69.75	94	188.33	266.29
42	92.68	220.34	95	158.24	401.68
43	86.79	227.83	96	865.59	1577.59
44	25.94	78.89	97	702.32	1103.97
45	54.62	178.16	98	282.01	659.47
46	60.09	130.93	99	109.99	254.98
47	34.56	78.72	100	19.28	52.21
48	12.04	33.25	101	30.60	94.50
49	21.10	47.43	102	248.43	696.00
50	13.60	58.82	103	95.05	258.60
51	15.40	31.66	104	36.84	100.13
52	17.82	47.26	105	17.28	72.76
53	8.27	17.62	106	28.49	94.02

Tableau A.6: Estimation des quantiles de crue suite à l'ajustement d'un modèle non-paramétrique

No.	Q2	Q100	No.	Q2	Q100
=====	=====	=====	=====	=====	=====
1	50.50	179.10	54	25.70	71.70
2	72.60	111.60	55	10.80	22.00
3	193.60	305.90	56	15.10	26.40
4	387.70	657.00	57	117.90	51.80
5	147.60	371.50	58	40.30	96.60
6	196.40	442.90	59	22.10	66.20
7	172.90	359.50	60	10.90	21.50
8	24.10	44.40	61	11.60	26.10
9	37.20	89.10	62	31.10	98.90
10	17.60	44.20	63	14.10	64.60
11	41.80	91.70	64	14.20	60.60
12	25.90	48.80	65	17.70	61.90
13	128.80	225.10	66	10.90	53.70
14	19.50	54.30	67	2.30	5.00
15	6.00	11.50	68	13.80	35.30
16	38.20	74.40	69	67.80	131.10
17	99.40	267.70	70	31.60	52.50
18	21.30	46.60	71	20.60	44.90
19	120.00	199.90	72	29.20	58.60
20	25.70	67.20	73	169.60	265.40
21	493.10	1007.20	74	---	---
22	265.70	637.90	75	106.70	186.20
23	15.00	38.10	76	45.10	102.90
24	24.70	68.50	77	108.00	274.30
25	137.70	287.70	78	77.40	132.60
26	76.10	172.60	79	637.80	992.50
27	140.40	358.70	80	28.20	73.40
28	---	---	81	181.20	334.10
29	113.70	285.20	82	46.70	101.60
30	159.70	330.60	83	192.70	566.20
31	93.10	161.00	84	222.80	568.90
32	140.00	304.90	85	1305.70	3306.30
33	44.50	78.10	86	269.00	542.20
34	16.70	43.20	87	586.30	1349.10
35	79.60	228.90	88	218.60	445.00
36	47.30	125.80	89	780.10	1751.20
37	72.30	204.10	90	1128.50	3167.10
38	39.70	128.10	91	71.50	229.70
39	12.20	30.30	92	227.40	341.90
40	21.20	56.40	93	203.10	341.90
41	23.30	65.20	94	186.30	285.60
42	91.20	207.60	95	155.50	394.80
43	85.30	219.00	96	885.60	1712.10
44	26.00	78.00	97	705.30	1162.90
45	52.60	159.50	98	309.00	786.60
46	63.30	131.20	99	109.60	250.60
47	36.60	83.40	100	20.70	53.50
48	12.10	33.50	101	32.40	105.70
49	20.60	46.50	102	243.50	703.70
50	13.80	63.10	103	90.00	230.30
51	15.70	30.70	104	39.80	102.00
52	18.80	50.50	105	16.70	72.00
53	8.50	16.20	106	28.30	94.30

ANNEXE B :

programmes Matlab

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Ce programme calcul les estimations régionales des quantiles de crue Q2 et %
% Q100. En effectuant une procédure jack-knife, les voisinages sont déterminés %
% avec l'analyse canonique des corrélations et les estimations sont faites par %
% régression sur le voisinage obtenu. %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

% Lecture des variables

```

```

load physiog.txt -ascii; % données physiographiques
load Meteo.txt -ascii; % données météorologiques
load Quantile.txt -ascii; % quantiles de crue

```

```

physio(:,1:4)=physiog(:,2:5);
physio(:,5)=Meteo;
hydro(:,1)=Quantile(:,1);
hydro(:,2)=Quantile(:,2)./Quantile(:,1);

```

```

clear physiog
clear Meteo
clear Quantile

```

```

[m,p]=size(physio);
[m,r]=size(hydro);

```

```

% Remplacement des 0 par des 1

```

```

for i=1:m
    if (physio(i,4)==0)
        physio(i,4)=1;
    end
end

```

```

% Initialisation des variables

```

```

comp=0; % Nombre de station où il y a estimations
num=0; % Numéro de la station
compteur=0; % Nombre de stations sans voisinage

```



```
% Analyse pour la première station
```

```
num=num+1;
```

```
% Formation des matrices nécessaires à l'analyse
```

```
XX=log(physio(2:m,:));
```

```
XX1=physio(2:m,:);
```

```
YY=log(hydro(2:m,:));
```

```
YY1=hydro(2:m,:);
```

```
cible=log(physio(1,:));
```

```
% Standardisation
```

```
moy=ones(m-1,1)*mean(XX);
```

```
ecart=ones(m-1,1)*std(XX);
```

```
X=(XX-moy)./ecart;
```

```
clear moy
```

```
clear ecart
```

```
moy=ones(m-1,1)*mean(YY);
```

```
ecart=ones(m-1,1)*std(YY);
```

```
Y=(YY-moy)./ecart;
```

```
clear moy
```

```
clear ecart
```

```
% Calcul des variables canoniques
```

```
[V,W,Vu,Wu,lambda2,a]=acc(X,Y,m-1,p,r);
```

```
% Calcul du voisinage
```

```
% Standardisation
```

```
cible_st=(cible-mean(XX))./std(XX);
```

```
[c,d]=size(cible_st);
```

```
[voisin]=voisinag(cible_st,Vu,W,lambda2,m-1,a,c);
```

```
t=length(voisin);
```

```
% Calcul des coefficients de régression
```

```
if (t>5)
```

```
comp=comp+1;
```

```
YY1(:,2)=YY1(:,2).*YY1(:,1); % Calcul de Q2
```

```
for i=1:t
```

```
    Xr(i,1)=1;
```

```
    Xr(i,2)=log(XX1(voisin(1,i),1));
```

```
    Yr2(i,1)=log(YY1(voisin(1,i),1));
```

```
    Yr100(i,1)=log(YY1(voisin(1,i),2));
```

```
end
```

```
b2=inv(Xr'*Xr)*Xr'*Yr2;
```

```
b100=inv(Xr'*Xr)*Xr'*Yr100;
```

```
% Estimation de Q2 et Q100 par la régression
```

```
q=[1 cible(1,1)]*b2;
```

```
estq2(comp)=exp(q); % Estimations de Q2
```

```
q=[1 cible(1,1)]*b100;
```

```
estq100(comp)=exp(q); % Estimations de Q100
```

```
else
```

```
    compteur=compteur+1;
```

```
    station(compteur)=num;
```

```
end
```

```
clear voisin
```

```
clear Xr
```

```
clear Yr2
```

```
clear Yr100
```

```
% Analyse pour les stations 2 à m-1
```

```
for l=2:m-1
```

```
    num=num+1;
```

```
% Formation des matrices nécessaires à l'analyse
```

```
XX(1:l-1,:)=log(physio(1:l-1,:));
XX(l:m-1,:)=log(physio(l+1:m,:));
XX1(1:l-1,:)=physio(1:l-1,:);
XX1(l:m-1,:)=physio(l+1:m,:);
YY(1:l-1,:)=log(hydro(1:l-1,:));
YY(l:m-1,:)=log(hydro(l+1:m,:));
YY1(1:l-1,:)=hydro(1:l-1,:);
YY1(l:m-1,:)=hydro(l+1:m,:);
cible=log(physio(l,:));
```

```
% Standardisation
```

```
moy=ones(m-1,1)*mean(XX);
ecart=ones(m-1,1)*std(XX);
X=(XX-moy)/ecart;
clear moy
clear ecart
```

```
moy=ones(m-1,1)*mean(YY);
ecart=ones(m-1,1)*std(YY);
Y=(YY-moy)/ecart;
clear moy
clear ecart
```

```
% Calcul des variables canoniques
```

```
[V,W,Vu,Wu,lambda2,a]=acc(X,Y,m-1,p,r);
```

```
% Calcul du voisinage
```

```
% Standardisation
```

```
cible_st=(cible-mean(XX))/std(XX);
[c,d]=size(cible_st);
```

```
[voisir]=voisinag(cible_st,Vu,W,lambda2,m-1,a,c);
t=length(voisin);
```

```

% Calcul des coefficients de régression
if (t>5)
    comp=comp+1;

    YY1(:,2)=YY1(:,2).*YY1(:,1); % Calcul de Q2

    for i=1:t
        Xr(i,1)=1;
        Xr(i,2)=log(XX1(voisin(1,i),1));
        Yr2(i,1)=log(YY1(voisin(1,i),1));
        Yr100(i,1)=log(YY1(voisin(1,i),2));
    end

    b2=inv(Xr'*Xr)*Xr'*Yr2;
    b100=inv(Xr'*Xr)*Xr'*Yr100;

    % Estimation de Q2 et Q100 par la régression
    q=[1 cible(1,1)]*b2;
    estq2(comp)=exp(q); % Estimations de Q2

    q=[1 cible(1,1)]*b100;
    estq100(comp)=exp(q); % Estimations de Q100

else
    compteur=compteur+1;
    station(compteur)=num;

end

clear voisin
clear Xr
clear Yr2
clear Yr100

end

```

```
% Analyse pour la dernière station
num=num+1;

% Formation des matrices nécessaires à l'analyse
XX=log(physio(1:m-1,:));
XX1=physio(1:m-1,:);
YY=log(hydro(1:m-1,:));
YY1=hydro(1:m-1,:);
cible=log(physio(m,:));

% Standardisation
moy=ones(m-1,1)*mean(XX);
ecart=ones(m-1,1)*std(XX);
X=(XX-moy)./ecart;
clear moy
clear ecart

moy=ones(m-1,1)*mean(YY);
ecart=ones(m-1,1)*std(YY);
Y=(YY-moy)./ecart;
clear moy
clear ecart

% Calcul des variables canoniques
[V,W,Vu,Wu,lambda2,a]=acc(X,Y,m-1,p,r);

% Calcul du voisinage

% Standardisation
cible_st=(cible-mean(XX))./std(XX);
[c,d]=size(cible_st);

[voisin]=voisinag(cible_st,Vu,W,lambda2,m-1,a,c);
t=length(voisin);

% Calcul des coefficients de régression
if (t>5)
    comp=comp+1;
```

```
YY1(:,2)=YY1(:,2).*YY1(:,1); % Calcul de Q2
```

```
for i=1:t
```

```
    Xr(i,1)=1;
```

```
    Xr(i,2)=log(XX1(voisin(1,i),1));
```

```
    Yr2(i,1)=log(YY1(voisin(1,i),1));
```

```
    Yr100(i,1)=log(YY1(voisin(1,i),2));
```

```
end
```

```
b2=inv(Xr'*Xr)*Xr'*Yr2;
```

```
b100=inv(Xr'*Xr)*Xr'*Yr100;
```

```
% Estimation de Q2 et Q100 par la régression
```

```
q=[1 cible(1,1)]*b2;
```

```
estq2(comp)=exp(q); % Estimation de Q2
```

```
q=[1 cible(1,1)]*b100;
```

```
estq100(comp)=exp(q); % Estimation de Q100
```

```
else
```

```
    compteur=compteur+1;
```

```
    station(compteur)=num;
```

```
end
```

```
clear voisin
```

```
clear Xr
```

```
clear Yr2
```

```
clear Yr100
```

% Fonction qui calcule les variables canoniques. Les arguments fournis sont: les matrices de
 % variables physiographiques (X) et des variables hydrologiques (Y), le nombre de sites au
 % départ (m), de variables physiographiques(p) et de variables hydrologiques (r). Elle
 % retourne les valeurs des variables canoniques pour chaque observation et pour chaque espace
 % (V et W), les composantes des combinaisons linéaires des variables canoniques (Vu et Wu),
 % les coefficients de corrélations canoniques (lambda2) et le nombre de variables canoniques
 % (a).

```
function[V,W,Vu,Wu,lambda2,a] = acc(X,Y,m,p,r)
```

% Calcul de la matrice des corrélations des variables initiales transformées

```
Z(1:m,1:r) = Y;
```

```
Z(1:m,r+1:r+p) = X;
```

```
C = cov(Z);
```

```
CY = C(1:r,1:r);
```

```
CYX = C(1:r,r+1:r+p);
```

```
CXY = C(r+1:r+p,1:r);
```

```
CX = C(r+1:r+p,r+1:r+p);
```

```
clear C
```

% Calcul des coefficients de corrélations canoniques (valeurs propres) et des vecteurs propres
 % associés pour l'espace hydrologique (W)

```
mat = inv(CY)*CYX*inv(CX)*CXY;
```

```
[Wu,lambda] = eig(mat);
```

```
lambda2 = sqrt(lambda); % Coefficients de corrélations
```

```
clear mat
```

% Tri des valeurs et des vecteurs propres calculés

```
[a,a] = size(Wu);
```

```
for i = 1:a-1
```

```
    for j = i+1:a
```

```
        if (lambda2(i,i) <= lambda2(j,j))
```

```
            tempo = lambda2(i,i);
```

```
            lambda2(i,i) = lambda2(j,j);
```

```
            lambda2(j,j) = tempo;
```

```
            tempo = Wu(:,i);
```

```
            Wu(:,i) = Wu(:,j);
```

```
    Wu(:,j) = tempo;  
end  
end  
end
```

```
% Normalisation des variables canoniques
```

```
for j = 1:a  
    coef(1,j) = 1;  
    for i = 2:a  
        coef(i,j) = Wu(i,j)/Wu(1,j);  
    end  
end
```

```
elem = sqrt(inv(coef*CY*coef));
```

```
for j = 1:a  
    Wu(1,j) = elem(j,j);  
    for i = 2:a  
        Wu(i,j) = coef(i,j)*elem(j,j);  
    end  
end
```

```
% Calcul des vecteurs propres pour l'espace physiographique (V)
```

```
for i = 1:a  
    Vu(:,i) = (inv(CX)*CYX*Wu(:,i))./lambda2(i,i);  
end
```

```
% Calcul des variables canoniques pour chaque observation
```

```
for j = 1:a  
    for i = 1:m  
        W(i,j) = Y(i,:)*Wu(:,j);  
        V(i,j) = X(i,:)*Vu(:,j);  
    end  
end
```


% Fonction qui détermine les stations voisines d'une station cible. Les arguments fournis sont:
 % la matrice des sites cibles (cible), les composantes des combinaisons linéaires dans l'espace
 % physiographique (Vu), les valeurs des variables canoniques de chaque observation dans
 % l'espace hydrologique (W), les coefficients de corrélations canoniques (lambda2), le nombre
 % de sites au départ (m), le nombre de variables canoniques (a) et le nombre de sites cibles (c).
 % Elle retourne le numéro des stations voisines.

```
function[voisin] = voisinag(cible,Vu,W,lambda2,m,a,c);
```

```
% Estimation des variables canoniques pour chaque nouvelle observation
```

```
for j = 1:a
```

```
    for i = 1:c
```

```
        VV(i,j) = cible(i,:)*Vu(:,j);
```

```
    end
```

```
end
```

```
% Calcul de la région voisine
```

```
alpha = 0.25;    % Seuil désiré
```

```
chi2 = chi2inv(1-alpha,a); % Valeur critique
```

```
for i = 1:c
```

```
    k = 1;
```

```
    for j = 1:m
```

```
        % Calcul de la distance
```

```
        Q = (W(j,:)-lambda2*VV(i,:))'*inv(eye(a)-lambda2*lambda2')*(W(j,:)-lambda2*VV(i,:));
```

```
        if (Q <= chi2)
```

```
            voisin(i,k) = j; % Indique le numéro de la station cible et ses stations voisines
```

```
            k=k+1;
```

```
        end
```

```
    end
```

```
end
```

```
clear Q
```


ANNEXE C :
figures

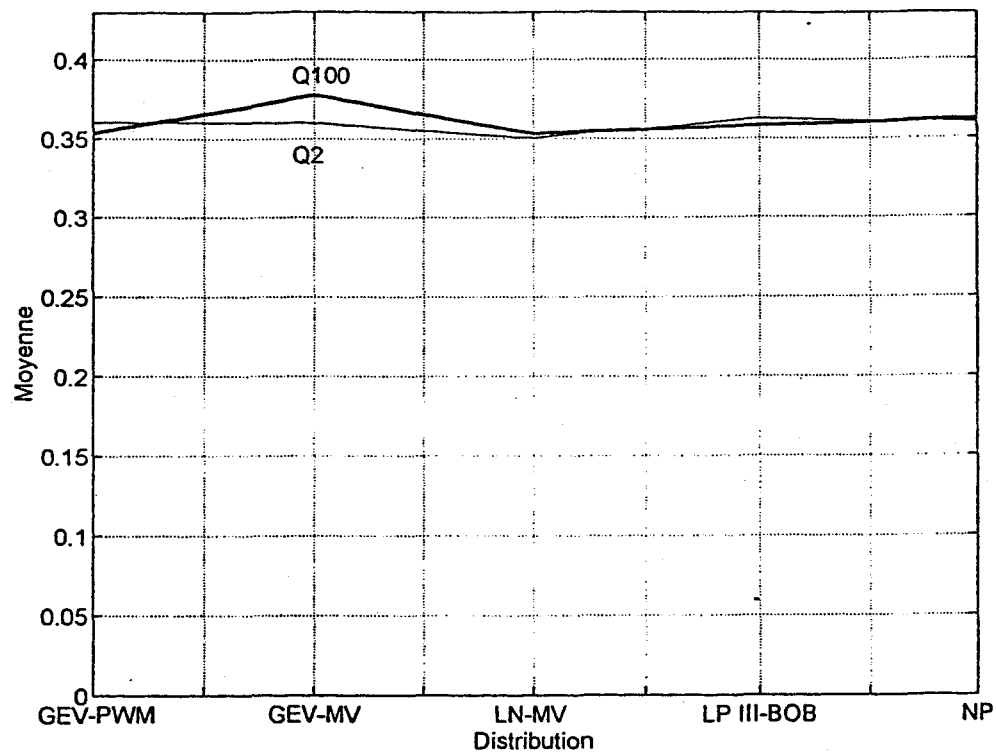


Figure C.1: Sensibilité de la méthode au modèle d'estimation locale, Moyenne des erreurs relatives

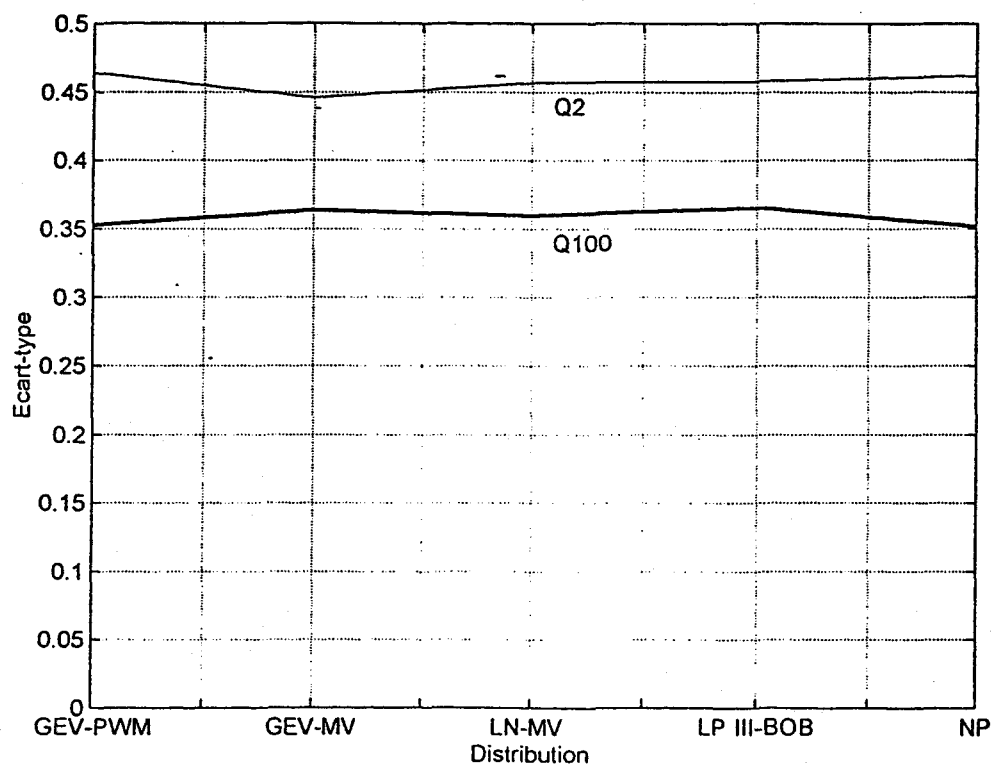


Figure C.2: Sensibilité de la méthode au modèle d'estimation locale, Écart-type des erreurs relatives

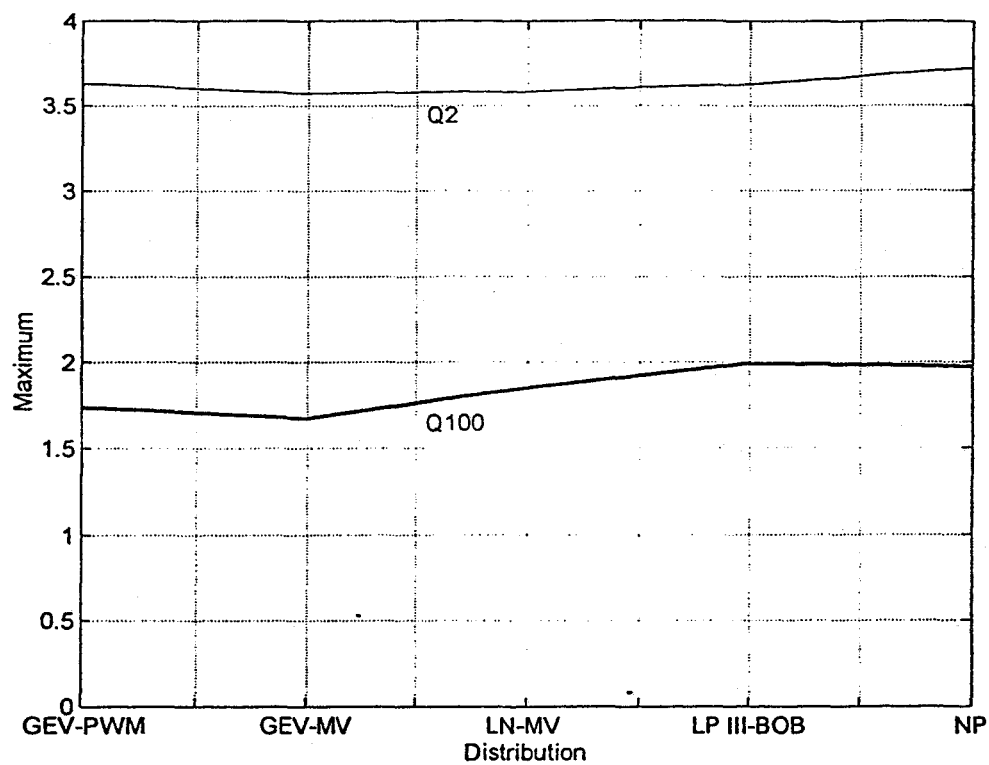


Figure C.3: Sensibilité de la méthode au modèle d'estimation locale, Maximum des erreurs relatives

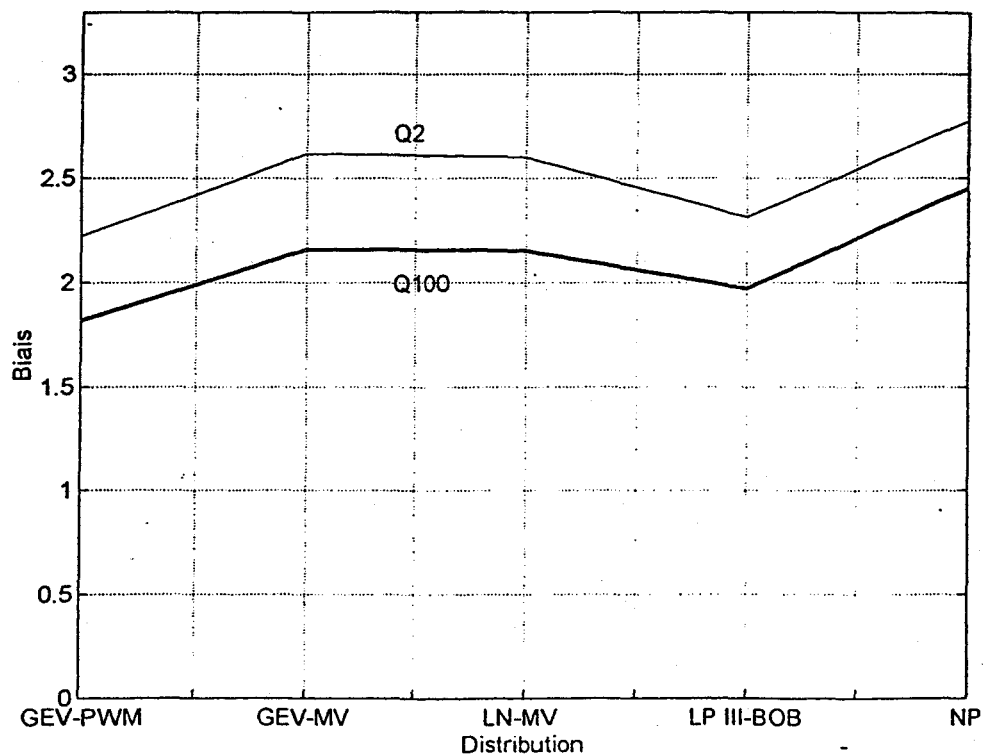


Figure C.4: Sensibilité de la méthode au modèle d'estimation locale, Biais des estimations régionales

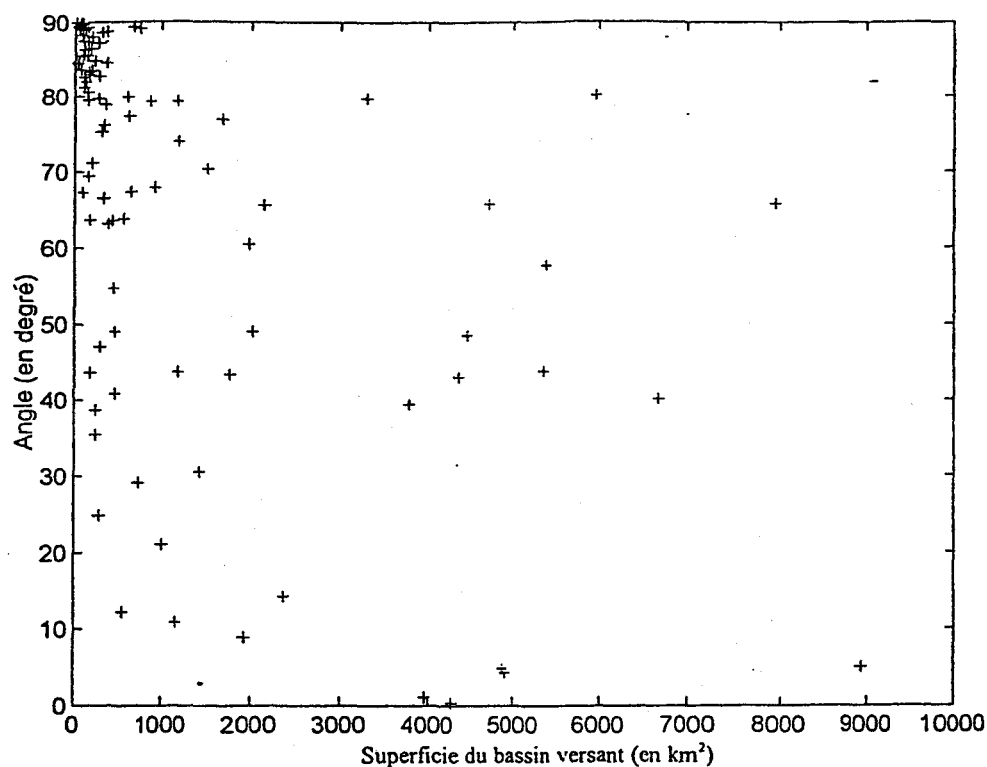


Figure C.5: Effet de la superficie du bassin versant sur l'angle des vecteurs d'erreurs

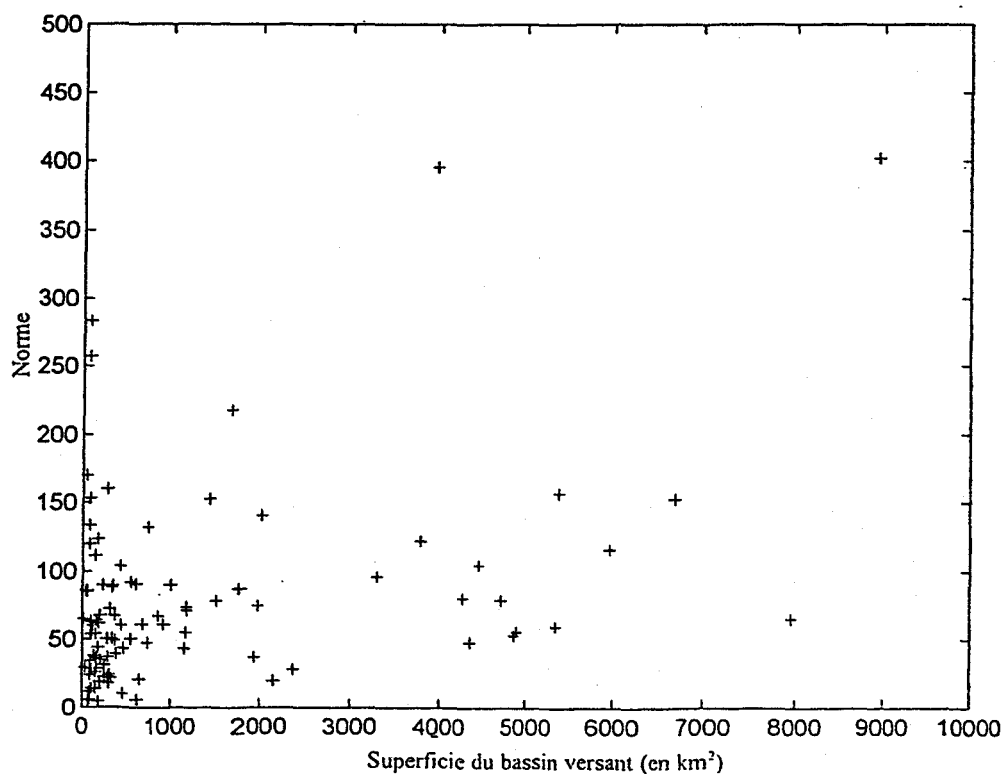


Figure C.6: Effet de la superficie du bassin versant sur la norme des vecteurs d'erreurs

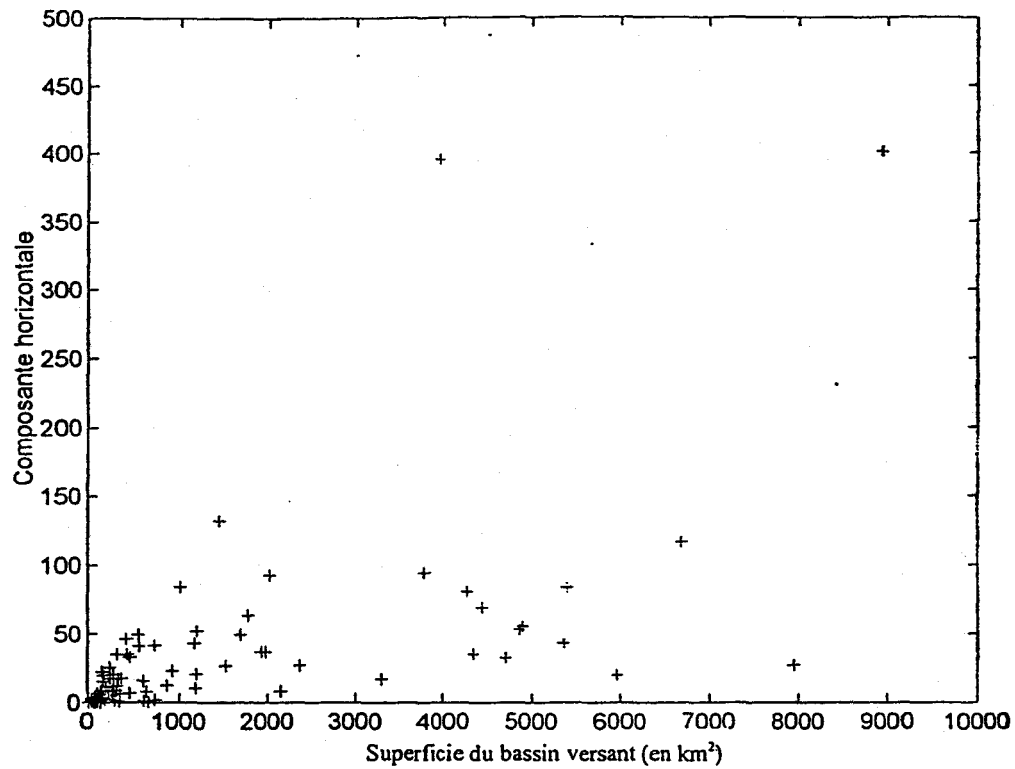


Figure C.7: Effet de la superficie du bassin versant sur la composante horizontale des vecteurs d'erreurs

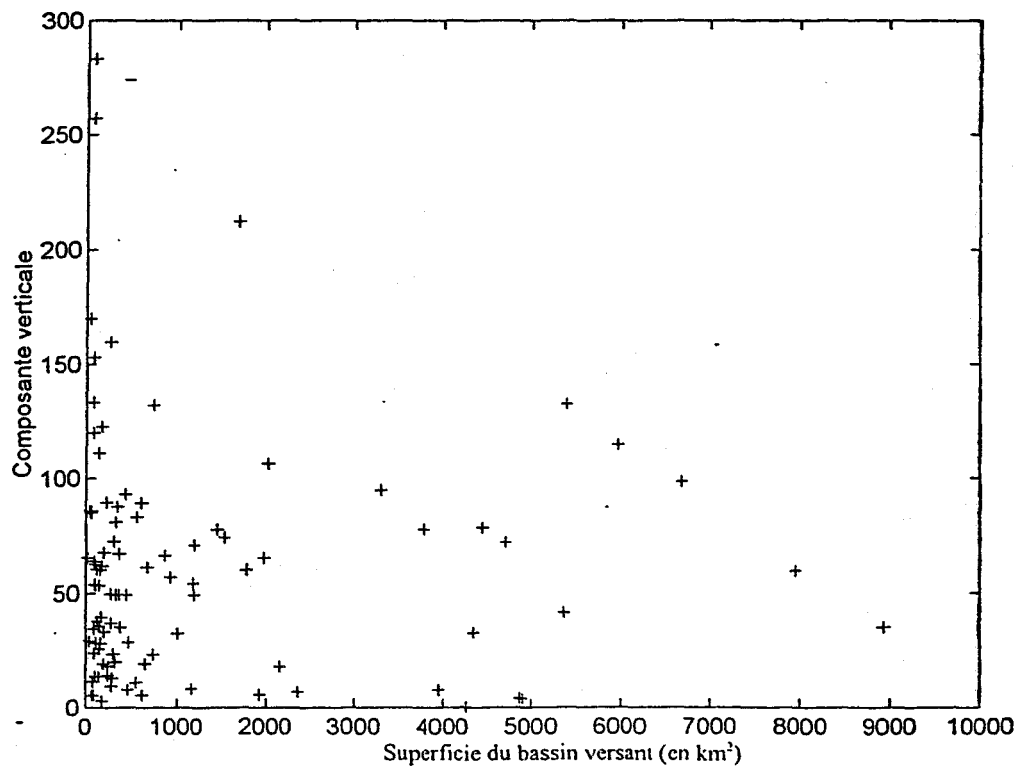


Figure C.8: Effet de la superficie du bassin versant sur la composante verticale des vecteurs d'erreurs

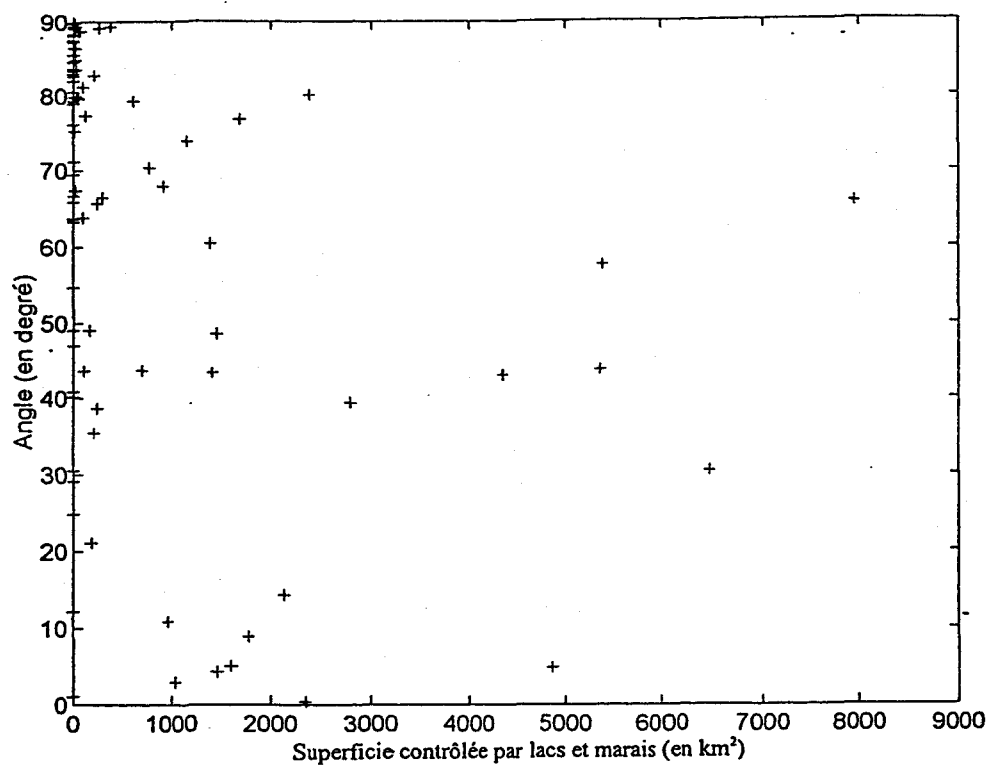


Figure C.9: Effet de la superficie du bassin versant contrôlée par les lacs et marais sur l'angle des vecteurs d'erreurs

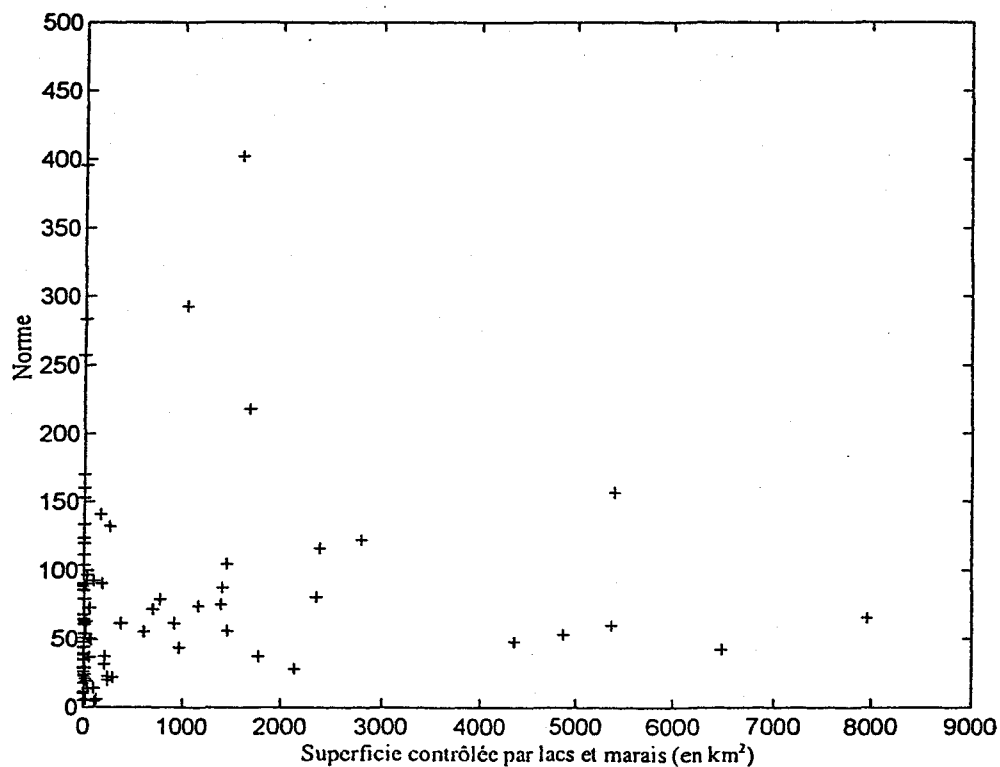


Figure C.10: Effet de la superficie du bassin versant contrôlée par les lacs et marais sur la norme des vecteurs d'erreurs

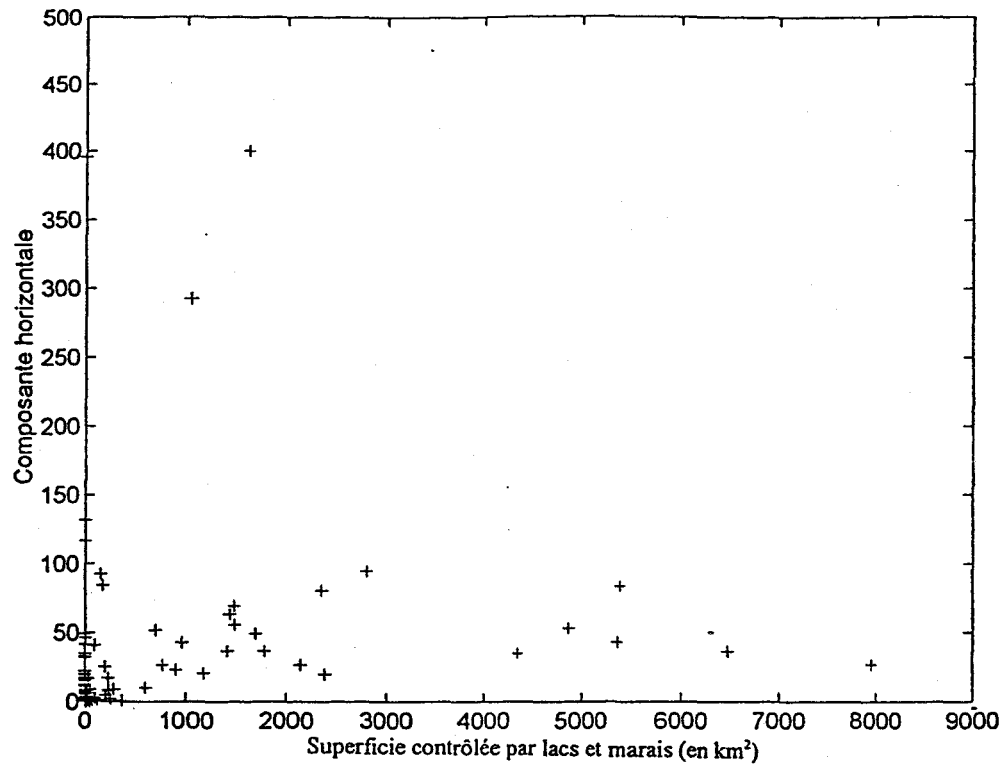


Figure C.11: Effet de la superficie du bassin versant contrôlée par les lacs et marais sur la composante horizontale des vecteurs d'erreurs

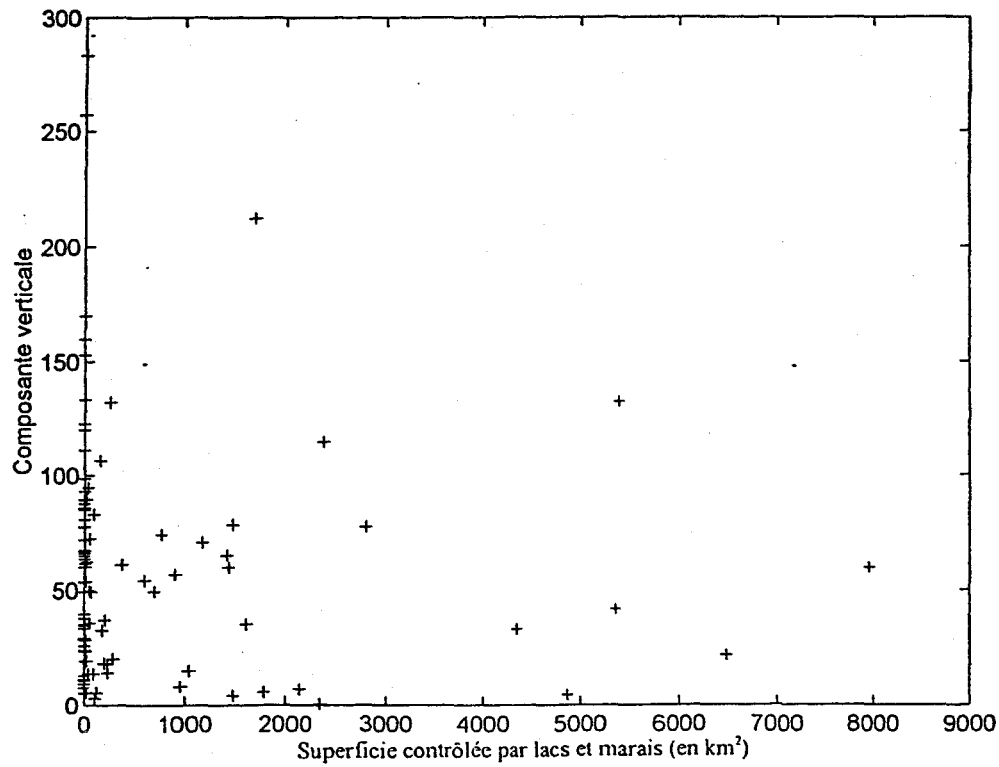


Figure C.12: Effet de la superficie du bassin versant contrôlée par les lacs et marais sur la composante verticale des vecteurs d'erreurs

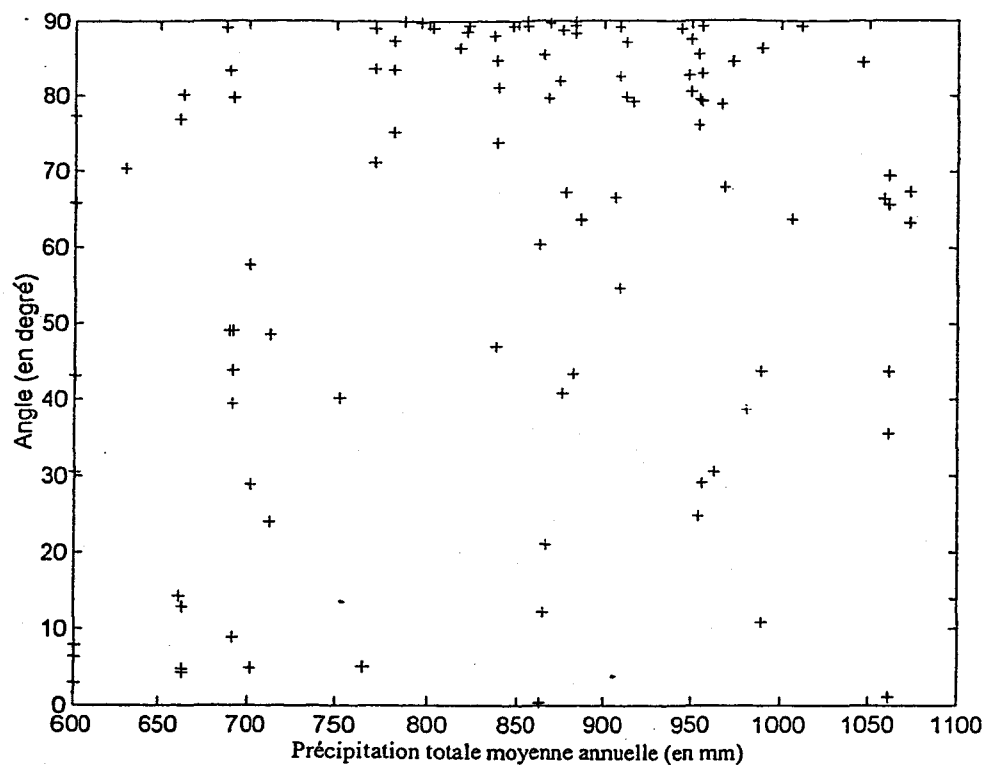


Figure C.13: Effet de la précipitation totale moyenne annuelle sur l'angle des vecteurs d'erreurs

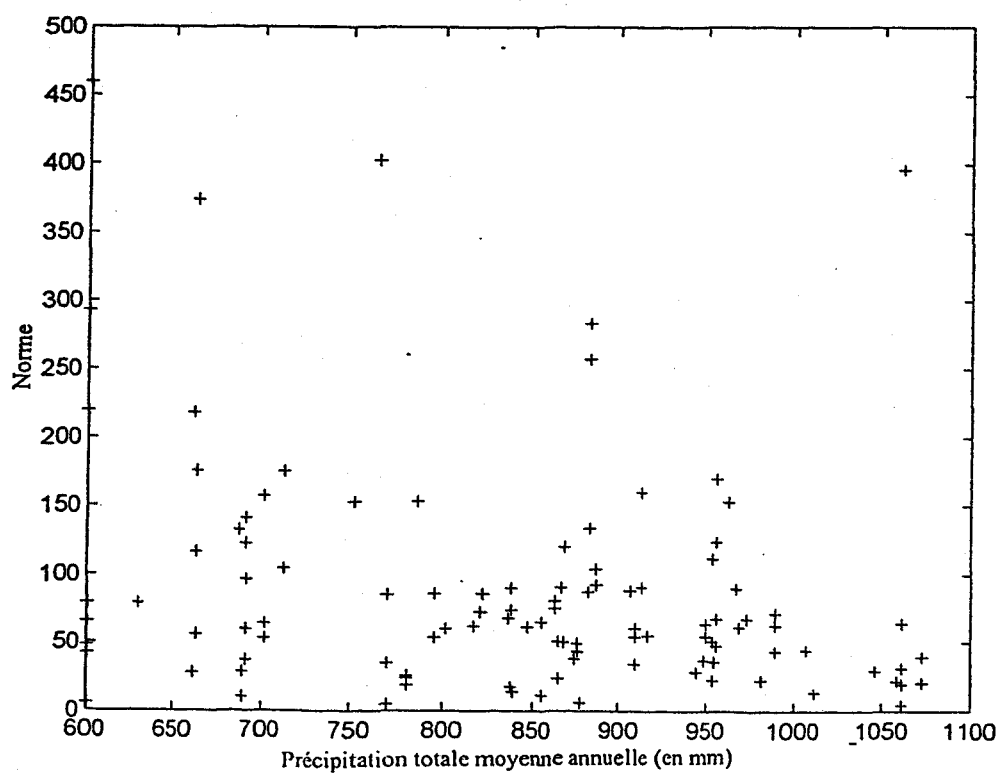


Figure C.14: Effet de la précipitation totale moyenne annuelle sur la norme des vecteurs d'erreurs

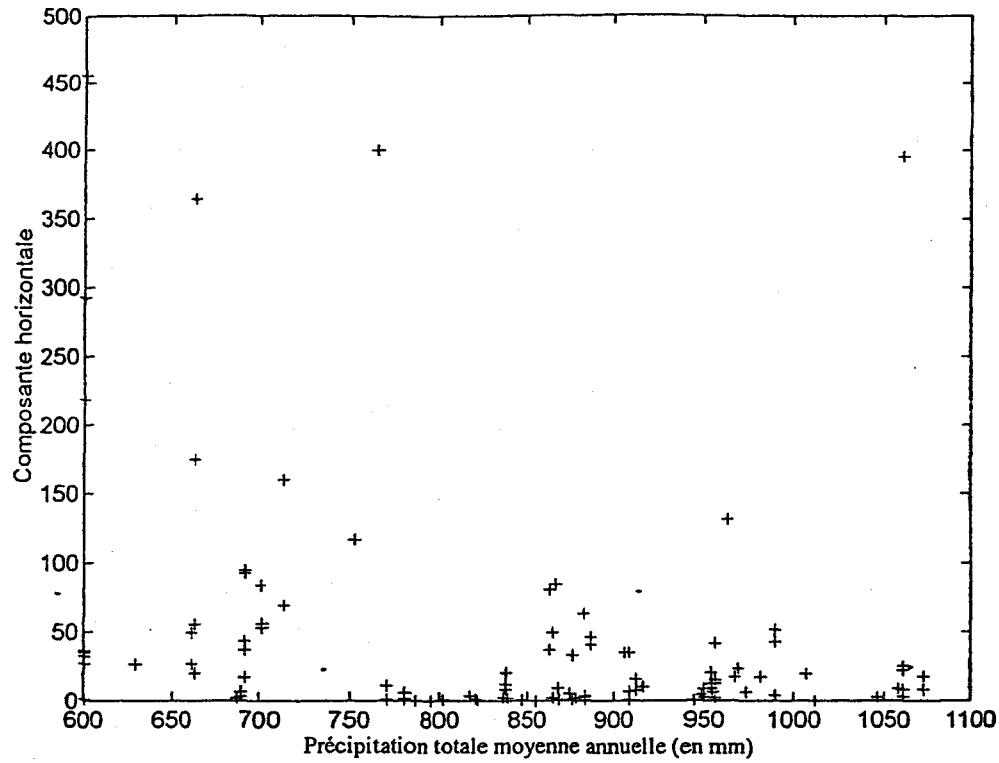


Figure C.15: Effet de la précipitation totale moyenne annuelle sur la composante horizontale des vecteurs d'erreurs

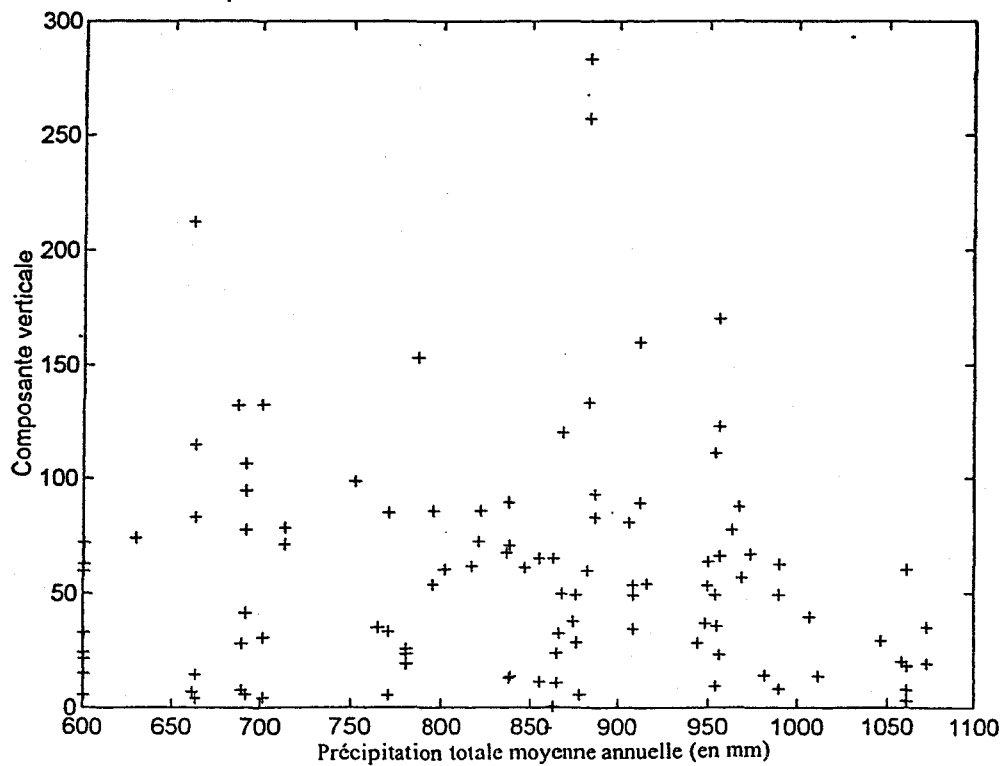


Figure C.16: Effet de la précipitation totale moyenne annuelle sur la composante verticale des vecteurs d'erreurs

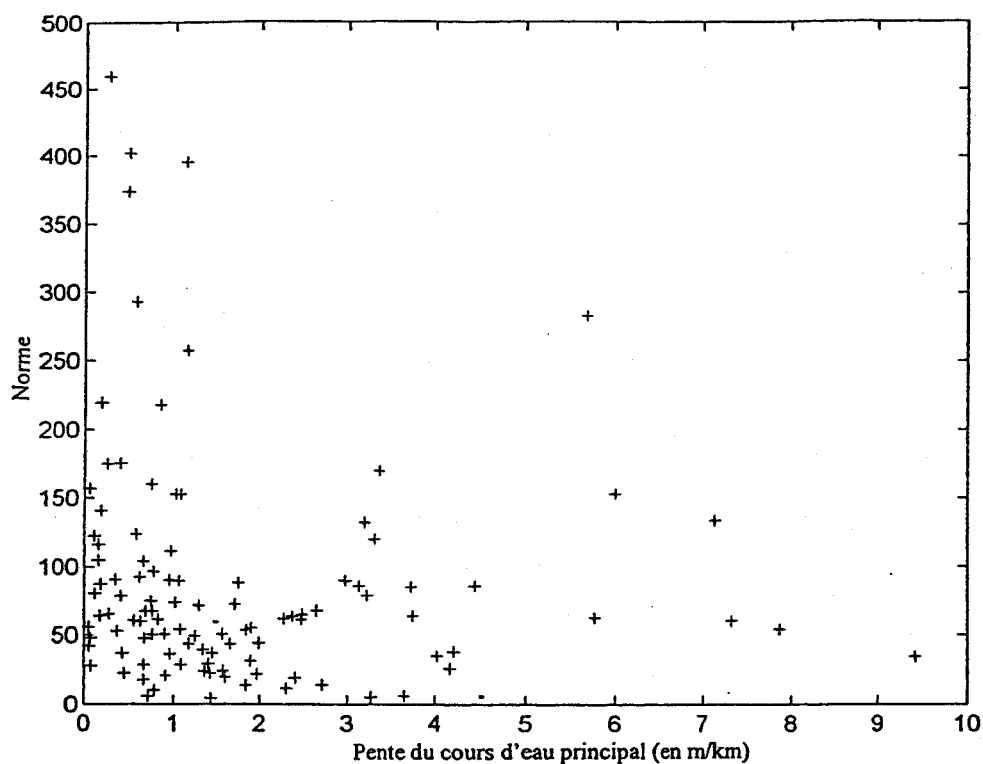


Figure C.17: Effet de la pente du cours d'eau principal sur la norme des vecteurs d'erreurs

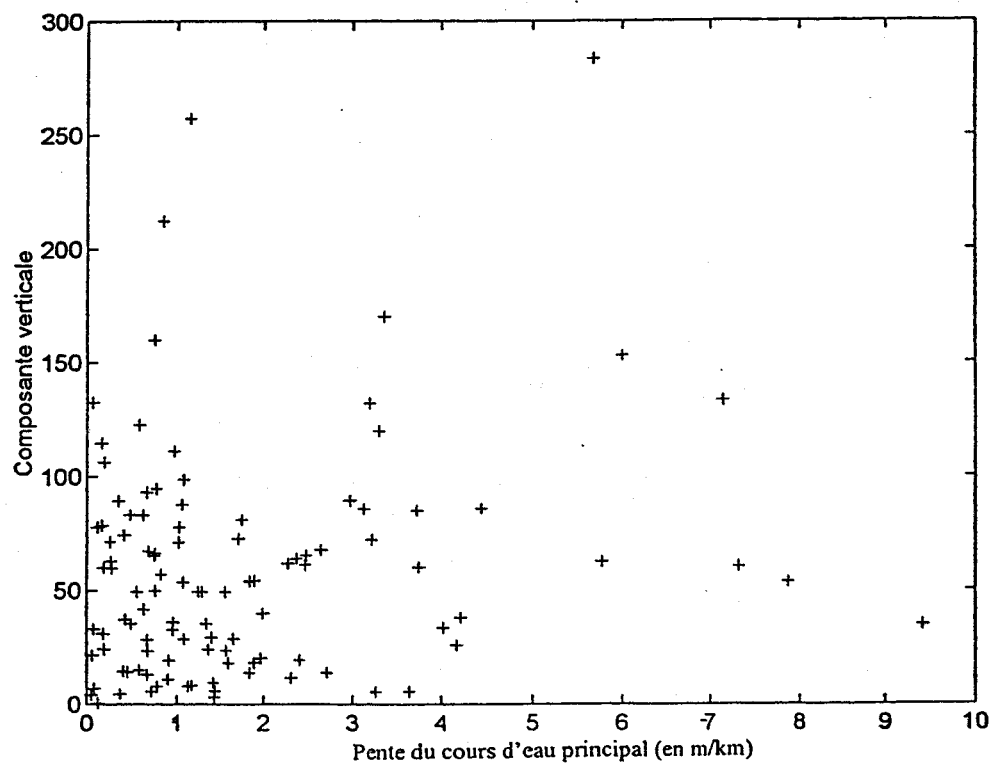


Figure C.18: Effet de la pente du cours d'eau principal sur la composante verticale des vecteurs d'erreurs.

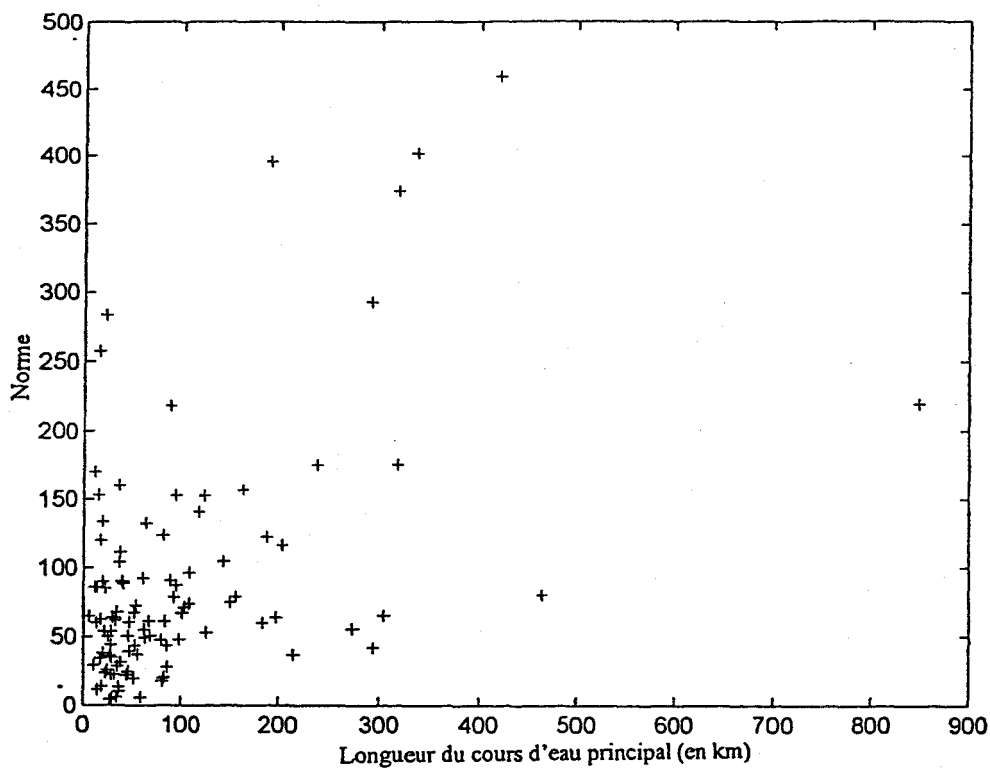


Figure C.19: Effet de la longueur du cours d'eau principal sur la norme des vecteurs d'erreurs

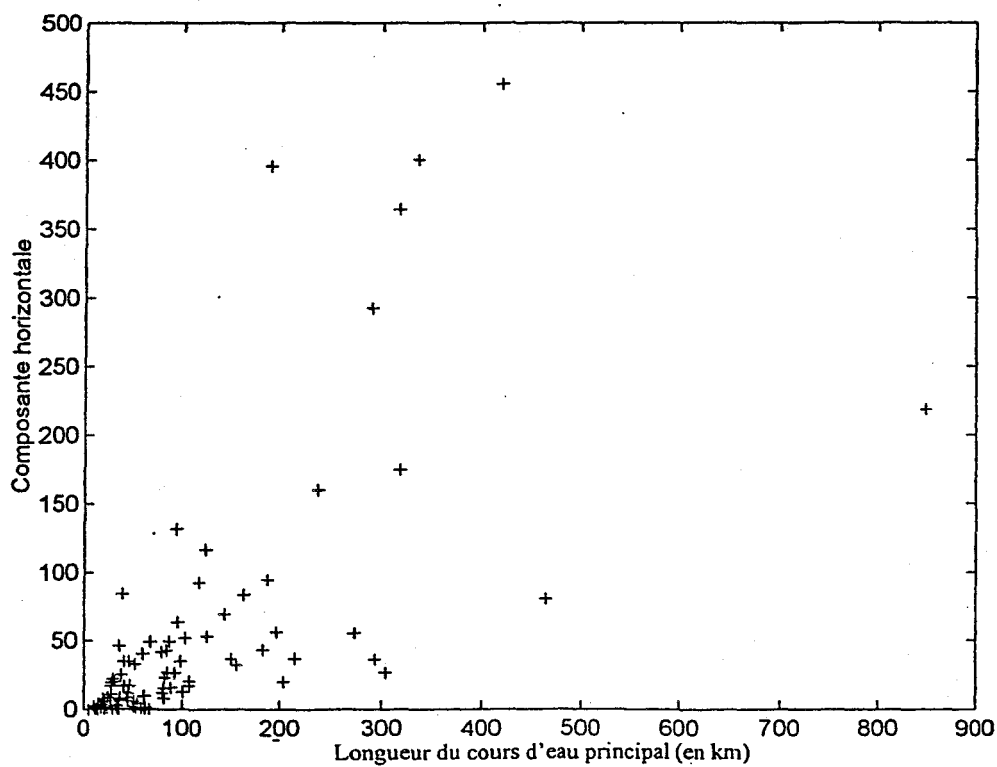


Figure C.20: Effet de la longueur du cours d'eau principal sur la composante horizontale des vecteurs d'erreurs

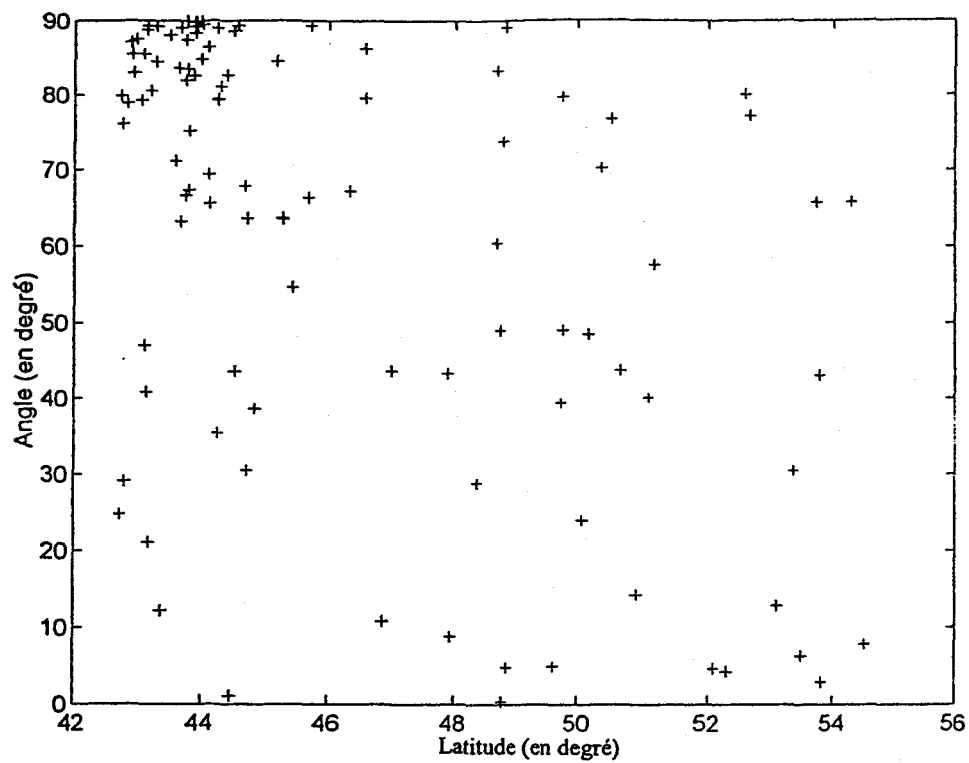


Figure C.21: Effet de la latitude sur l'angle des vecteurs d'erreurs

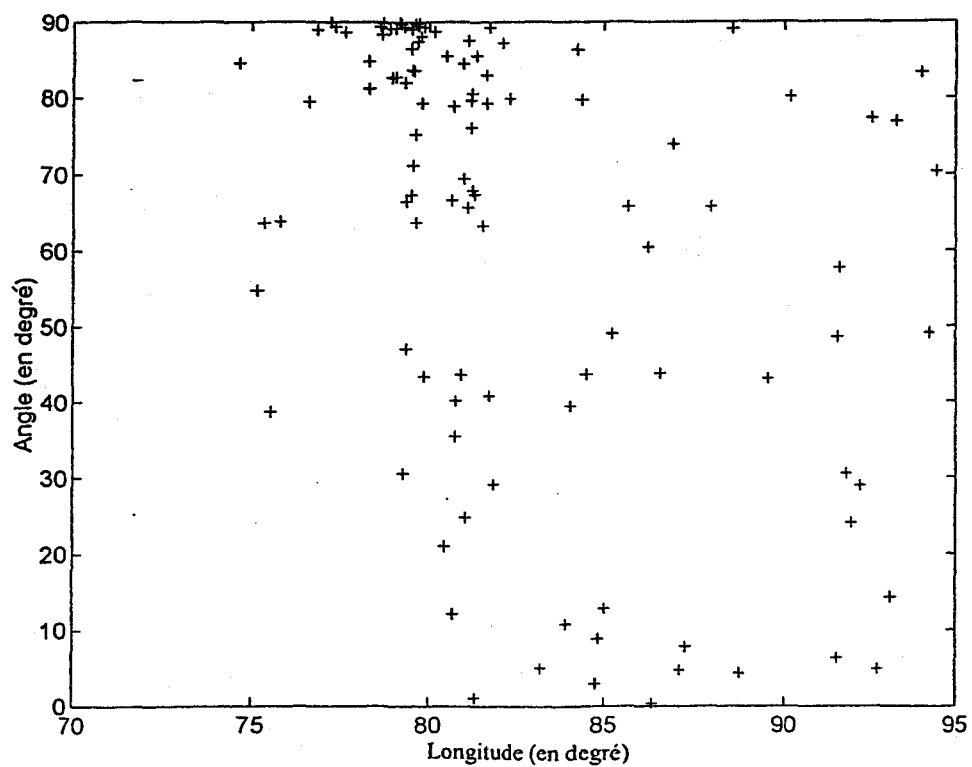


Figure C.22: Effet de la longitude sur l'angle des vecteurs d'erreurs

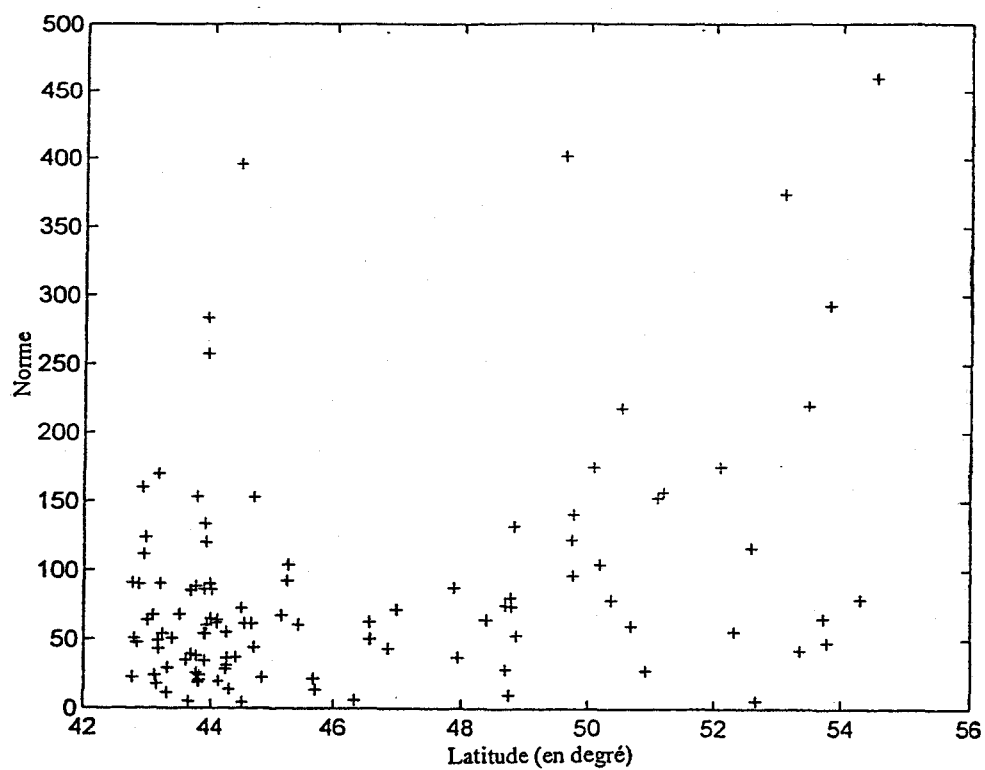


Figure C.23: Effet de la latitude sur la norme des vecteurs d'erreurs

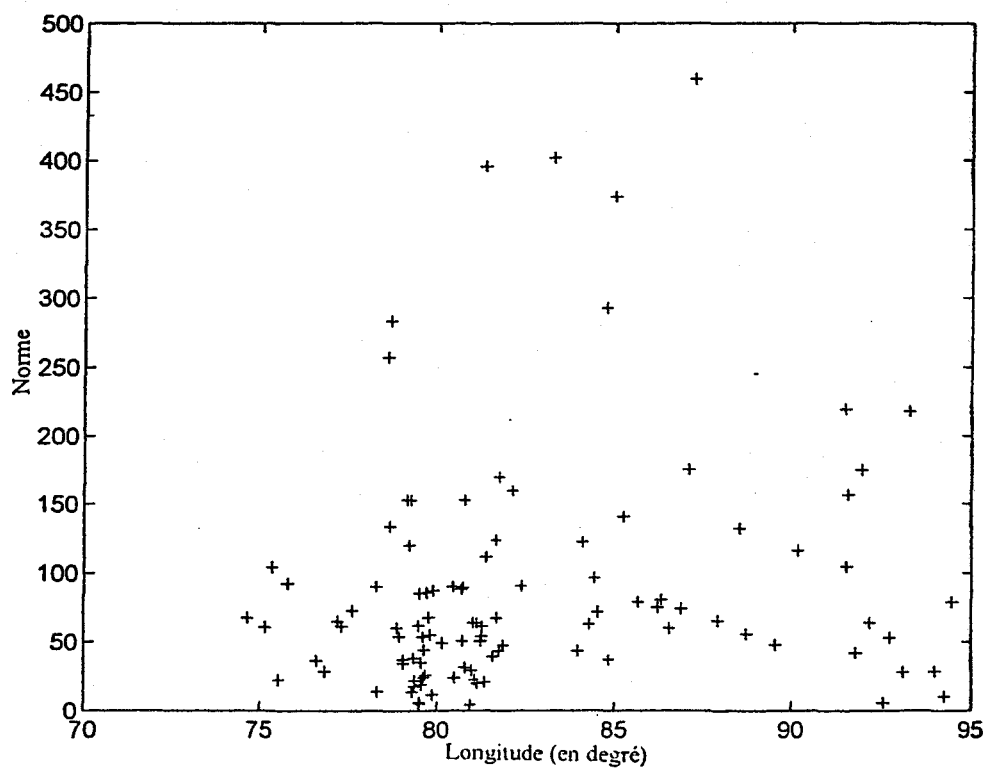


Figure C.24: Effet de la longitude sur la norme des vecteurs d'erreurs

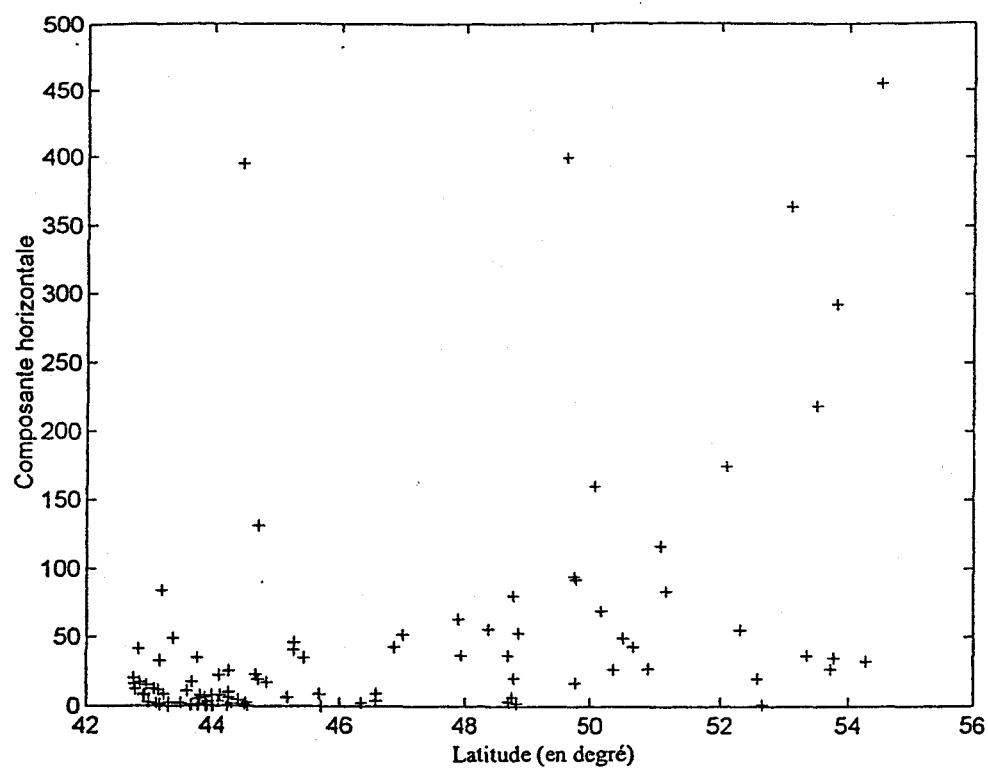


Figure C.25: Effet de la latitude sur la composante horizontale des vecteurs d'erreurs

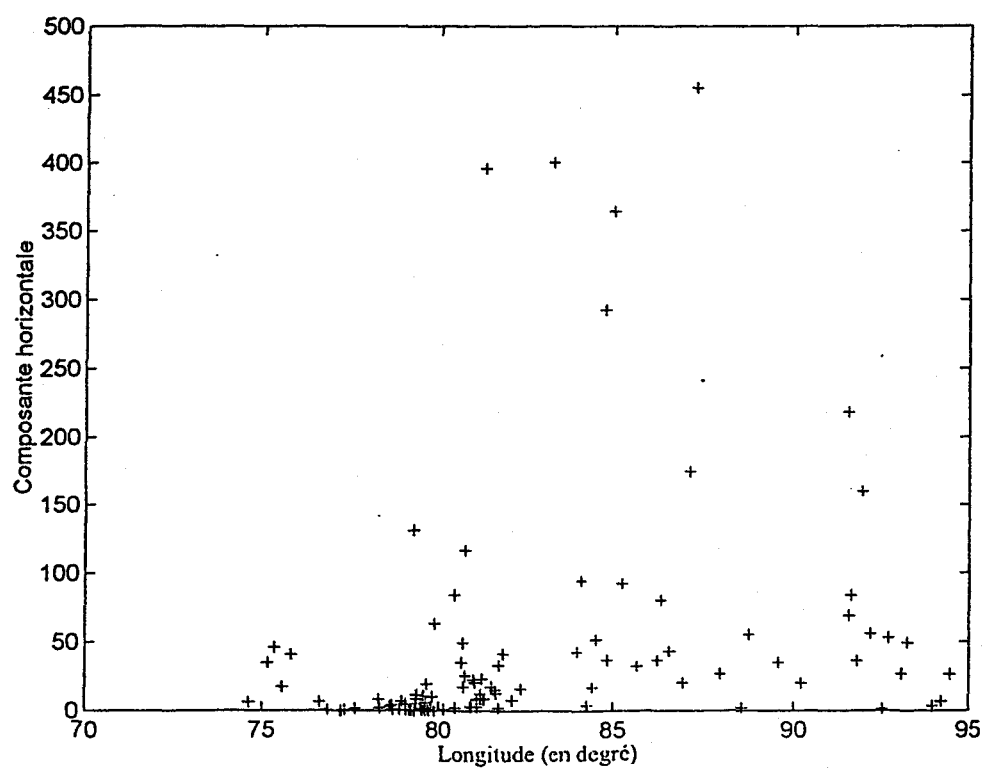


Figure C.26: Effet de la longitude sur la composante horizontale des vecteurs d'erreurs

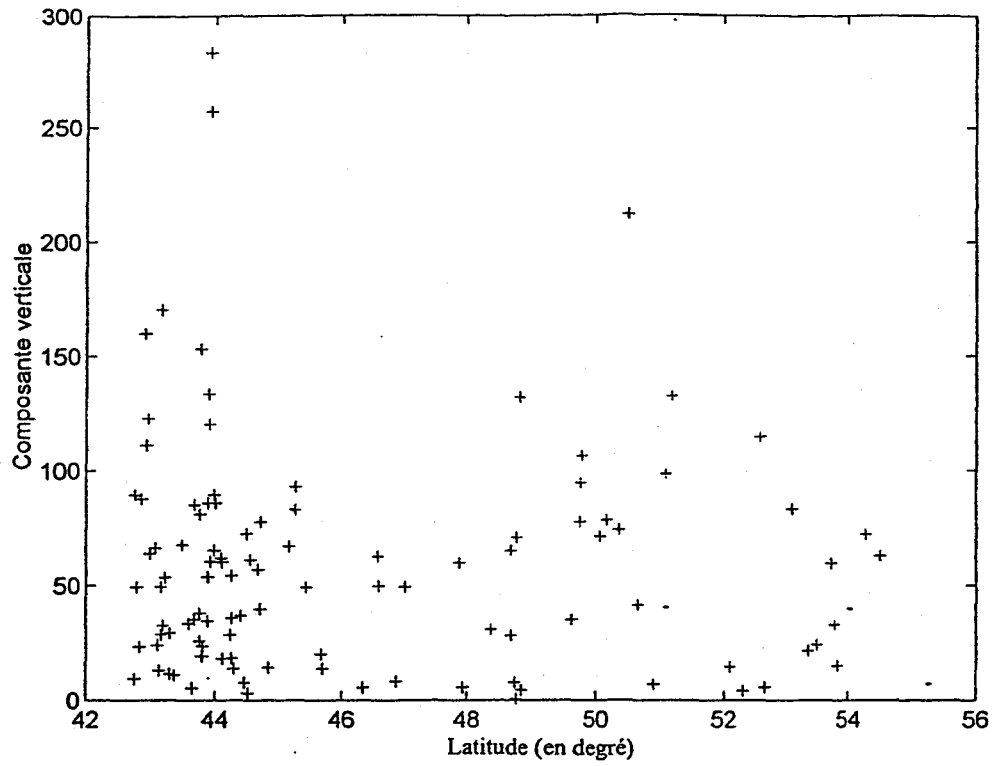


Figure C.27: Effet de la latitude sur la composante verticale des vecteurs d'erreurs

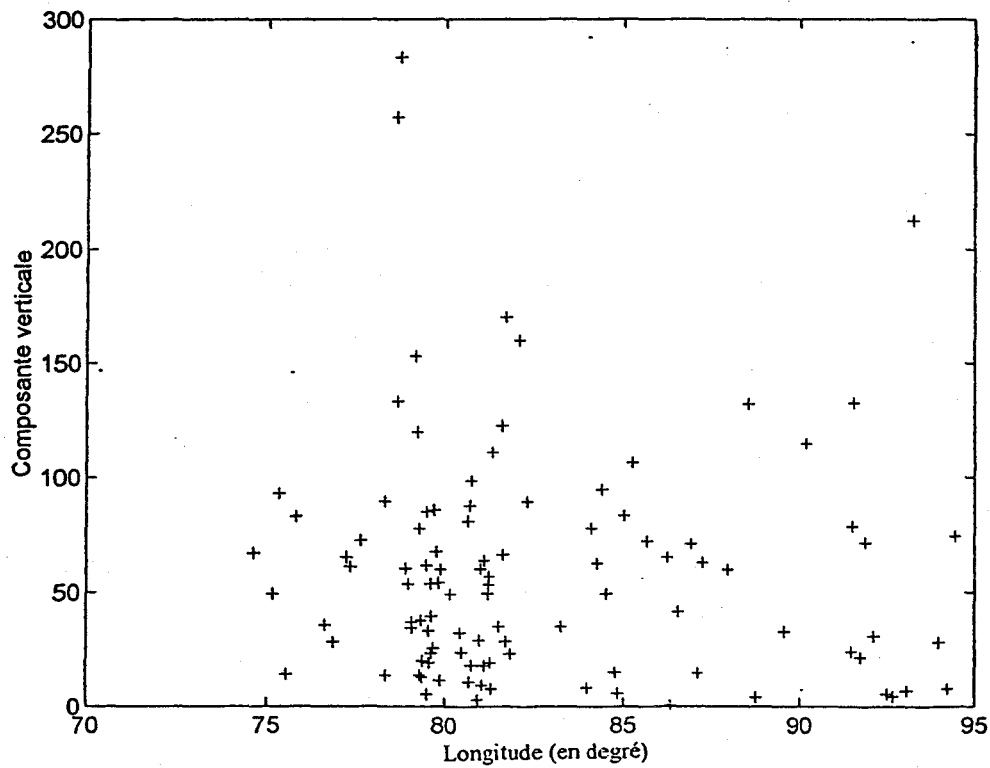


Figure C.28: Effet de la longitude sur la composante verticale des vecteurs d'erreurs