

Rapport scientifique No 305

par

Luc Perreault
René Roy
Bernard Bobée
Louis Mathier

**ESTIMATION RÉGIONALE DES DÉBITS
MAXIMUMS ANNUELS X_T DE
PÉRIODE DE RETOUR T**

Juin 1991

**INRS-Eau
Université du Québec
C.P. 7500
Sainte-Foy, Québec
G1V 4C7**

TABLE DES MATIERES

Liste des figures.....	iii
Liste des tableaux.....	v
1. INTRODUCTION ET PROBLÉMATIQUE.....	1
2. DESCRIPTION DES DONNÉES	3
3. DESCRIPTION STATISTIQUE DES COMPORTEMENTS HYDROLOGIQUES ET FORMATION DE GROUPES HOMOGENES	5
3.1 Étude pilote	5
3.2 Formation de groupes de stations homogènes.....	10
3.2.1 Analyse des points-variables.....	11
3.2.2 Analyse des points-stations	14
3.3. Affectation d'une rivière non-jaugée à un des groupes.....	20
4. ESTIMATION RÉGIONALE NE FAISANT INTERVENIR QUE LA VARIABLE SUBV.....	20
4.1. Lois de probabilité.....	22
4.2. Comparaison des relations X_T -SUBV	24
5. ESTIMATION RÉGIONALE ET ÉTUDE EMPIRIQUE DE L'EFFICACITÉ.....	27
6. CONCLUSIONS ET RECOMMANDATIONS.....	33
BIBLIOGRAPHIE.....	34
ANNEXE A: Contributions relatives des variables aux 7 premiers facteurs de l'analyse factorielle des correspondances	36
ANNEXE B: Estimation de X_T pour 8 périodes de retour selon la loi Pearson Type 3	38
ANNEXE C: Résultats des comparaisons des trois relations X_T -SUBV	42

ANNEXE D: Biais brut et biais unitaire pour $T = 2, 5, 10, 20, 50, 100, 200$ ans.....45

LISTE DES FIGURES

Figure 2.1 Stations hydrométriques du MENVIQ	4
Figure 3.1 Représentation des variables dans le plan F_1 - F_2 de l'ACP	7
Figure 3.2 Représentation des variables dans le plan F_1 - F_3 de l'ACP	8
Figure 3.3 Représentation des variables dans le plan F_2 - F_3 de l'ACP	8
Figure 3.4 Représentation des variables dans le plan F_1 - F_2 de l'AFC	13
Figure 3.5 Représentation des variables dans le plan F_1 - F_3 de l'AFC	13
Figure 3.6 Représentation des variables dans le plan F_2 - F_3 de l'AFC	14
Figure 3.7 Hiérarchie supérieure du classement.	15
Figure 3.8. Représentation des centres de gravité dans le plan F_1 - F_2 de l'AFC	16
Figure 3.9. Représentation des centres de gravité dans le plan F_1 - F_3 de l'AFC	16
Figure 3.10. Représentation des centres de gravité dans le plan F_2 - F_3 de l'AFC	17
Figure 3.11. Représentation simultanée dans le plan F_1 - F_2	17
Figure 3.12. Représentation simultanée dans le plan F_1 - F_3	18
Figure 3.13. Représentation simultanée dans le plan F_2 - F_3	18
Figure 3.14. Représentation géographique des 113 stations selon les quatre groupes	19
Figure 4.1. Les différentes avenues possibles de régionalisation	22
Figure 4.2. Graphique de X_T en fonction de SUBV ($T = 1000$)	24

Figure 4.3. Graphique de $\text{Log}(X_T)$ en fonction de $\text{Log}(\text{SUBV})$ ($T = 1000$)	25
Figure 5.1. Plan de la simulation.....	29
Figure 5.2. Biais brut pour $T = 1000$ ans.....	31
Figure 5.3. Biais unitaire pour $T = 1000$ ans.	32

LISTE DES TABLEAUX

Tableau 3.1. Corrélations entre les variables mesurées sur les 53 stations.....	5
Tableau 3.2. Résultat du codage disjonctif complet	11
Tableau 4.1. Seuils observés P des tests de Kolmogorov-Smirnov pour la loi Pearson Type 3	23
Tableau 5.1. Résultats des régressions entre $\text{Log}(X_T)$ et SUBV pour les huit périodes de retour T	28
Tableau 5.2. Nombre de fois que les intervalles contiennent X_T	30

1. INTRODUCTION ET PROBLÉMATIQUE

Ce projet conjoint Université / Industrie est une collaboration entre l'INRS-Eau et Hydro-Québec d'une durée de deux ans (janvier 1989 - janvier 1991) et subventionné par le Conseil National de Recherches en Sciences Naturelles et en Génie du Canada. Ce projet de recherche a pour thème principal "la validation des données d'apports naturels: modélisation des volumes et des débits, à un site et régionalement". Il se divise en trois sous-projets principaux:

- 1- Validation et estimation des apports
- 2- Modélisation des pointes et des volumes
- 3- Régionalisation

Ce document vise à présenter les résultats des travaux de recherches concernant le sous-projet 3 "Régionalisation". Ceux des deux autres sous-projets sont présentés dans Perreault *et al* (1991a, b).

L'estimation des débits X_T de période de retour T est importante pour Hydro-Québec dans le cas du dimensionnement d'ouvrages et pour la gestion d'ouvrages existants. Pour les sites où l'on dispose de séries courtes ou d'aucune mesure de débits, cette estimation de X_T est problématique. On souhaite donc déterminer X_T avec le plus de précision possible afin de limiter les risques associés au sous-dimensionnement ou les coûts superflus encourus par la construction d'un ouvrage sur-dimensionné en utilisant l'information régionale. La régionalisation a pour but non seulement d'améliorer l'estimation de X_T lorsque peu de données sont disponibles, mais aussi d'estimer cette valeur lorsqu'aucune mesure de débit n'est connue (rivière non-jaugée).

Donc, conformément à l'objectif énoncé dans la demande de subvention initiale du projet de recherche, le travail concernant ce sous-projet consiste à développer une méthode permettant le transfert de l'information disponible à des sites jaugés à des sites non-jaugés.

L'une des techniques de régionalisation les plus répandues dans la littérature a été proposée par Benson (1962). Cette méthode, dont nous nous sommes inspirés pour estimer le débit maximum annuel de période de retour donnée d'une rivière non-jaugée, est basée sur la relation régressive existant entre X_T et certaines caractéristiques physiographiques des bassins versants. Les différentes étapes de la méthode de régionalisation que nous proposons sont les suivantes :

- 1] choix de variables permettant de former des groupes de stations ayant un comportement homogène;
- 2] formation de groupes de stations ayant un comportement hydrologique similaire (groupes homogènes);
- 3] validation des groupes (les stations d'un même groupe ont-elles un comportement similaire? Les stations appartenant à des groupes différents ont-elles bel et bien un comportement différent?);
- 4] choix d'une loi de probabilité pour chaque groupe représentant bien les débits maximums annuels;
- 5] estimation de X_T pour toutes les stations de chaque groupe homogène;
- 6] estimation des paramètres des modèles de régression, mettant en relation X_T et les variables sélectionnées à l'étape 1, pour toutes les périodes de retours considérées.

Une fois toutes ces étapes effectuées, il est possible en affectant une rivière non-jaugée à un des groupes homogènes de prédire, à l'aide du modèle régressif approprié, le débit X_T de période de retour T .

Dans un premier temps, nous présentons, dans la section 2, les données utilisées dans l'élaboration de la méthode de régionalisation. Nous y décrivons brièvement les variables physiographiques et hydrologiques que nous avons considérées. A la section 3, nous donnons les résultats d'une analyse pilote qui nous a permis de déterminer les variables pertinentes à la formation de groupes homogènes. Nous appliquons ensuite l'analyse factorielle des correspondances et l'analyse de classification ascendante hiérarchique afin de former des groupes de stations homogènes. Les groupes n'étant pas délimités géographiquement, et certaines variables utilisées dans ces analyses ne pouvant être mesurées ou même être estimées sur une rivière non-jaugée, nous proposons, dans la section 4, une démarche d'estimation régionale avec comme seule variable la superficie du bassin versant. Nous présentons aussi dans cette section les résultats d'une simulation effectuée afin de vérifier l'efficacité de cette méthode. Enfin, à la section 5, nous donnons quelques recommandations quant à l'application de la méthode de régionalisation telle que proposée dans ce rapport et nous discutons brièvement des nouvelles approches de régionalisation proposées dans la littérature récente.

2. DESCRIPTION DES DONNÉES

L'échantillon utilisé provient du Ministère de l'Environnement du Québec (*MENVIQ*). Il contient 113 stations de jaugeages situées au nord du fleuve St-Laurent pour lesquelles des mesures de débits maximums annuels ainsi que la superficie du bassin versant sont disponibles. De plus, pour 53 de ces stations, une série de variables physiographiques autres que la superficie de bassin ont été mesurées. Principalement, deux raisons ont motivé l'utilisation des données du *MENVIQ* plutôt que celles des bassins du réseau d'Hydro-Québec :

- les débits pour les bassins d'Hydro-Québec sont calculés par bilan hydrique alors que ceux du *MENVIQ* sont obtenus par jaugeage en rivières naturelles. Ces derniers sont donc mesurés avec plus de précision.
- Hydro-Québec ne possède pas un inventaire aussi complet des caractéristiques physiographiques que le *MENVIQ*.

L'analyse pilote, dont les résultats sont présentés à la section 3, a été effectuée sur le sous-échantillon des 53 stations pour lesquelles toutes les variables suivantes étaient disponibles :

variables physiographiques:

- la latitude (*LATI*),
- la longitude (*LONG*),
- la superficie du bassin versant (*SUBV*),
- la superficie des lacs et des marais (*SULM*),
- la superficie des forêts (*SUFO*),
- l'élévation (*ELEV*),
- la pente du bassin versant (*PEBV*),

variables hydrologiques:

- le débit maximum annuel moyen (*QMAM*),
- le coefficient d'asymétrie de la série des débits maximums annuels (*COAS*),
- le coefficient d'aplatissement de la série des débits maximums annuels (*COAP*).

Ces 53 stations ainsi que les 60 autres constituant notre ensemble de données pour l'étude de régionalisation sont situées sur la carte du Québec présentée à la figure 2.1.

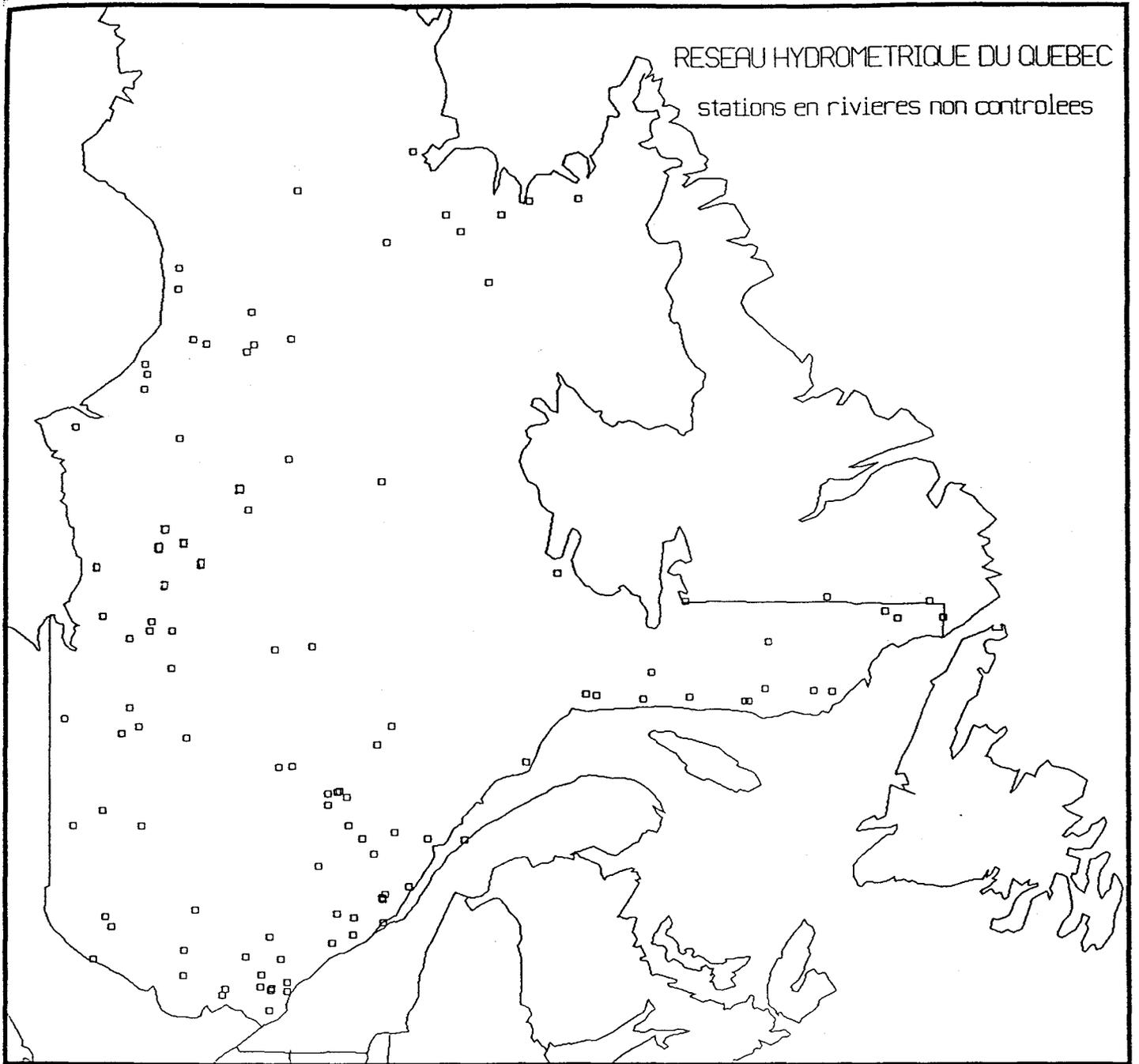


Figure 2.1 Stations hydrométriques du MENVIQ

On tire principalement du Tableau 3.1 les remarques suivantes :

- 1) les variables de superficie (SUBV, SUFO et SULM) sont très corrélées entre elles;
- 2) les variables de superficie sont très corrélées avec la variable débit maximum annuel moyen (QMAM);
- 3) les variables pente du bassin versant (PEBV) et élévation (ELEV) sont très peu corrélées avec le débit maximum annuel moyen (QMAM);
- 4) les coefficients d'asymétrie (COAS) et d'aplatissement (COAP) sont corrélés significativement entre eux.

Pour mieux interpréter les différents liens existants entre ces variables, nous avons effectué une analyse en composantes principales sur ces 53 stations. L'objet de cette analyse de données est d'extraire l'essentiel de l'information contenue dans un tableau de données, et d'en fournir une représentation graphique se prêtant plus aisément à l'interprétation. L'examen des résultats nous permettra entre autres d'éliminer quelques variables moins importantes. On retrouve une description détaillée de l'analyse en composante principale dans Lebart, Morineau et Tabard (1977). Nous donnons ici une brève interprétation géométrique de la méthode.

Pour comprendre les opérations de réduction (élimination de variables) pratiquées par ce type de technique, il est utile de représenter géométriquement les stations et les variables de notre tableau de données par des points d'un espace à p ou n dimensions (ici, $p = 10$ variables et $n = 53$ stations), points dont les coordonnées sont précisément les éléments du tableau. Les 53 stations peuvent donc être considérées comme 53 points d'un espace à 10 dimensions, alors que les 10 variables sont des points d'un espace à 53 dimensions.

Si les valeurs prises par deux variables particulières sont très voisines pour toutes les stations, ces variables seront représentées par deux points très proche dans l'espace des stations (à 53 dimensions). Cela peut vouloir dire que ces variables mesurent la même chose, ou encore qu'elles sont liées par une relation particulière. Toutefois, l'interprétation de ces proximités dans les deux espaces est très lourde à faire. L'analyse en composante principale consiste à représenter graphiquement les proximités entre les points au terme d'une série d'ajustements : afin de simplifier le problème, on cherche, dans l'espace des stations (à 53 dimensions) des sous-espaces (facteurs) de faibles dimensions (une, deux ou trois) s'ajustant le mieux possible aux nuages de points-variables, c'est-à-dire de façon à ce que les proximités, mesurées dans les sous-espaces, reflètent autant que

possible les proximités réelles.

Pour ce faire, on construit des axes factoriels, expliquant chacun un pourcentage de la variance totale, sur lesquels on projette les points-variables. Ces facteurs sont par la suite interprétés. Par exemple, en ne retenant que les 3 axes expliquant le plus de variabilité, nous obtenons une représentation graphique des points-variables sur un sous-espace à trois dimensions. L'interprétation est donc plus facile, cependant la qualité de la représentation dépend de la variance expliquée cumulée des trois axes factoriels. Nous notons ces axes F_1 , F_2 et F_3 .

Les Figures 3.1, 3.2 et 3.3 donnent respectivement la représentation des points-variables dans les plans factoriels F_1 - F_2 , F_1 - F_3 et F_2 - F_3 . Le pourcentage de variance totale expliquée par les trois axes factoriels (représentation à trois dimensions) est de 77.3%.

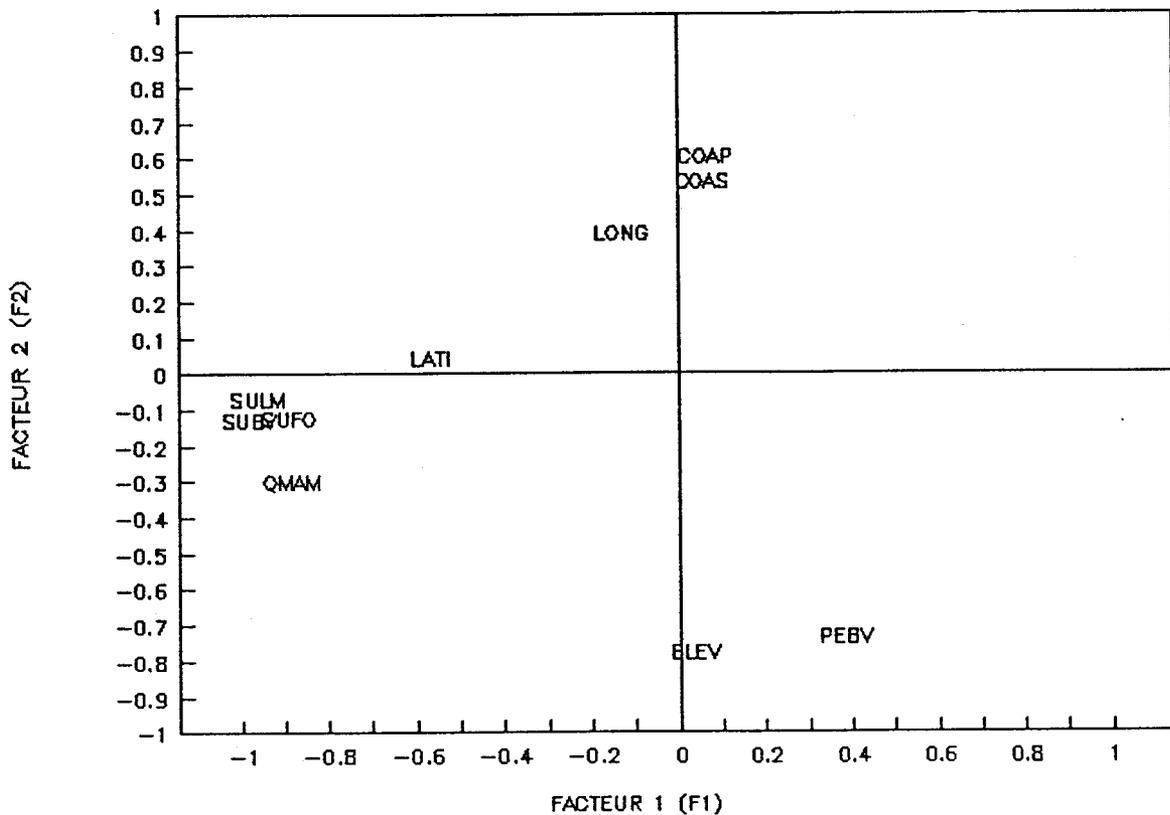


Figure 3.1 Représentation des variables dans le plan factoriel F_1 - F_2 de l'ACP

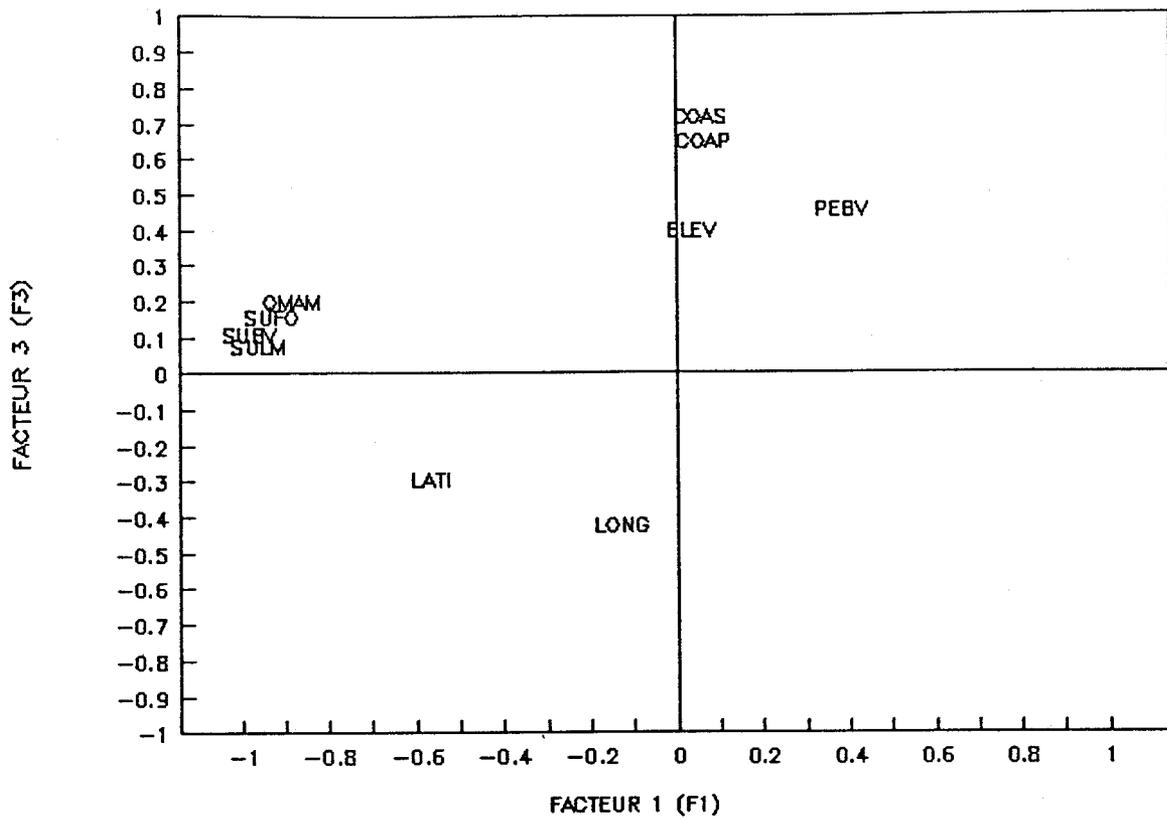


Figure 3.2 Représentation des variables dans le plan factoriel F_1 - F_3 de l'ACP

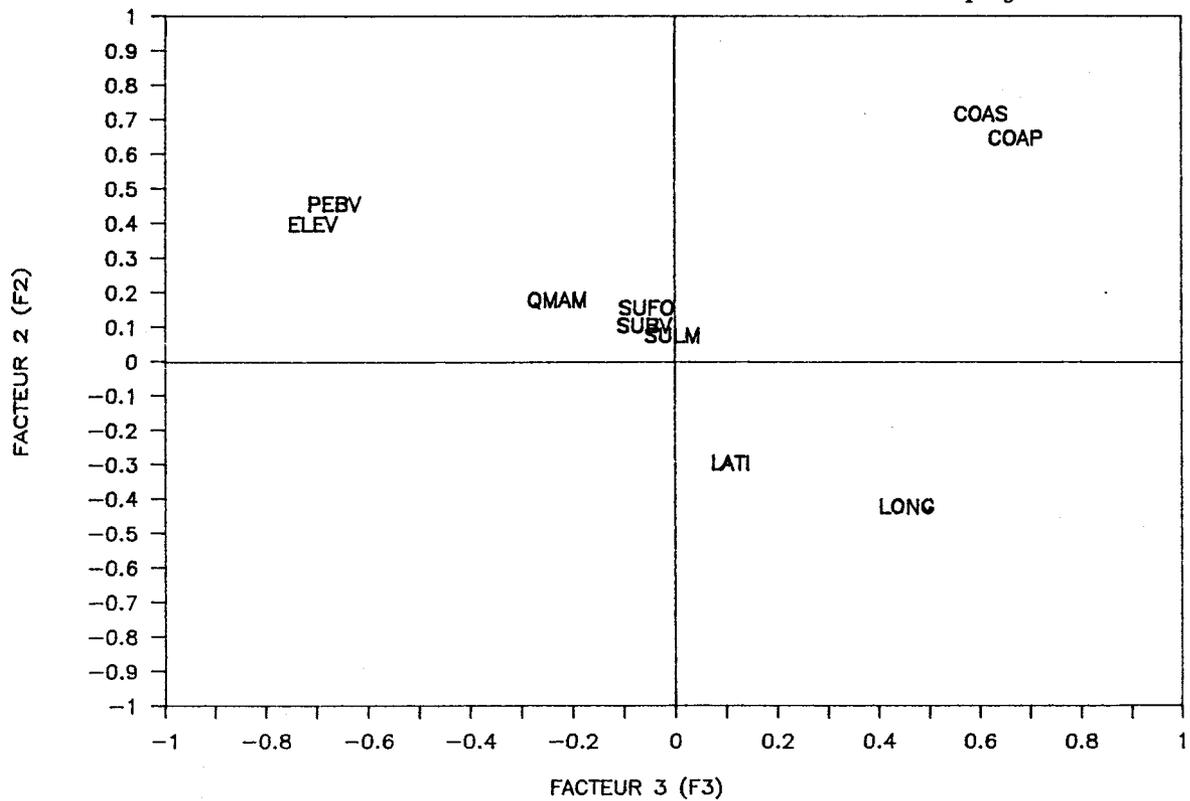


Figure 3.3 Représentation des variables dans le plan factoriel F_2 - F_3 de l'ACP

Sur les trois plans factoriels présentés aux Figures 3.1, 3.2 et 3.3, on constate premièrement que le premier facteur, prédominant (qui explique 40.4% de la variabilité), peut être qualifié d'axe de taille de bassin puisque les variables SUBV, SUFO et SULM y sont très bien représentées (coordonnées ≈ -1 sur le facteur 1, Fig. 3.1 et 3.2). On remarque, de plus, que ces variables sont très proches l'une de l'autre, ce qui signifie qu'elles apportent probablement la même information, et que la variable QMAM est fortement reliée aux variables de superficie (QMAM est située à proximité des variables de taille).

On observe ensuite que les variables COAS et COAP s'opposent sur le facteur 2, de façon plus ou moins marquée, aux variables ELEV et PEBV (Fig. 3.2). Ceci signifie qu'en général les stations ayant de fortes élévations ainsi que de fortes pentes du bassin versant auront des coefficients d'asymétrie et d'aplatissement faibles. L'interprétation de cet axe reste tout de même délicate, d'une part à cause de son faible pourcentage de variance expliquée (20.4%), et d'autre part parce qu'une telle opposition est difficile à expliquer physiquement.

Finalement, on constate que les variables LATI et LONG se situent à proximité de l'origine dans les trois plans factoriels (Fig. 3.1, 3.2 et 3.3). Elles sont donc mal représentées (faibles coordonnées) dans l'espace des trois facteurs expliquant le plus de variabilité. Ceci signifie que la latitude et la longitude sont des variables de moindre importance pour expliquer les différents comportements hydrologiques des stations considérées dans cette étude.

L'examen des résultats de l'analyse en composantes principales dans l'espace des trois premiers facteurs, qui donne une bonne représentation de la variabilité totale (77.284%), nous permet de tirer les conclusions suivantes :

- 1] il n'est pas nécessaire d'utiliser les trois variables de superficie pour la régionalisation puisqu'elles semblent expliquer la même chose. L'utilisation de la variable SUBV devrait être suffisante.
- 2] les variables LATI et LONG sont mal représentées dans l'espace des trois premiers facteurs. Elles ne sont donc pas utiles pour discriminer les stations hétérogènes.
- 3] les variables PEBV et ELEV ne sont aucunement liées à la variable QMAM. Celles-ci sont donc peu pertinentes à la formation de groupes de stations homogènes pour la région à l'étude.

Ainsi, nous retenons, pour former des groupes de stations homogènes, les

variables SUBV, QMAM, COAS et COAP. On remarque que trois variables hydrologiques sur quatre ont été considérées. Selon nous, il est important de considérer des variables hydrologiques dans l'analyse parce que l'objectif principal de la régionalisation est l'estimation du débit X_T sur une rivière non-jaugée : il faut entre autres s'assurer que les débits des stations formant un groupe homogène suivent le même type de loi de probabilité. En particulier, les coefficients d'asymétrie (COAS) et d'aplatissement (COAP) sont des variables nous informant sur la forme de la loi d'où proviennent les mesures de débits. C'est pourquoi, même si elles sont peu corrélées à QMAM, nous considérons qu'il est important de les inclure dans l'analyse de régionalisation. Certains auteurs ont utilisé ces deux coefficients pour construire des diagrammes servant à identifier la loi d'où proviennent les observations (voir Bobée, Perreault et Ashkar, 1991). Plus précisément, ils utilisent les statistiques $\beta_1 = \text{COAS}^2$ et $\beta_2 = \text{COAP}$. Par convention, nous employons aussi cette notation pour les analyses subséquentes.

3.2 Formation de groupes de stations homogènes

Pour former des groupes homogènes à l'aide des 113 stations du MENVIQ et des 4 variables retenues, nous avons combiné l'analyse factorielle des correspondances, l'AFC (Lebart, Morineau et Tabard, 1977), et l'analyse de la classification ascendante hiérarchique, la CAH (Jambu, 1976).

L'AFC cherche principalement la meilleure représentation simultanée de deux ensembles constituant les lignes (les stations) et les colonnes (les variables) d'un tableau de données. Cette technique, qui est une ACP particulière (métrique du khi-deux), est donc très semblable à l'analyse en composantes principales. L'AFC a de plus l'avantage de permettre l'interprétation des proximités entre les points-variables et les points-stations (propriété barycentrique de l'AFC), ce que l'analyse en composantes principales ne permet pas. Cette propriété est intéressante puisqu'elle permet de caractériser les stations selon leur proximité aux différentes variables.

L'analyse de la classification ascendante hiérarchique, pour sa part, est une technique permettant de regrouper des éléments d'un tableau (par exemple les stations) ayant des caractéristiques semblables. Ces regroupements se font par étapes en calculant certaines distances (Jambu, 1976) entre les éléments. On produit ensuite ce que l'on appelle un arbre hiérarchique des classifications qui nous permet de suivre et d'interpréter toutes les étapes de cette procédure de regroupement, et de sélectionner un nombre adéquat de groupes.

Dans ce qui suit, nous présentons l'application des deux analyses au fichier des 113 stations mesurées sur les 4 variables (SUBV, QMAM, β_1 et β_2) fournissant, selon les résultats de la section 3.1, le plus d'information pertinente sur le processus hydrologique

qui nous intéresse.

3.2.1 Analyse des points-variables

Alors que l'analyse en composantes principales s'applique aux tableaux de mesures continues, l'AFC est particulièrement utile dans le cas des tableaux de contingence et par extension pour des tableaux de codages discontinus. Pour appliquer l'analyse factorielle des correspondances à nos mesures, on doit donc procéder à une transformation des données : il est nécessaire que les variables considérées soient transformées en variables qualitatives. Nous avons alors effectué un codage dijontif complet (Jambu, 1976) sur les données en considérant trois classes quasi-équiprobables par variable (contenant approximativement le même nombre d'observations par classe). La table 3.2 donne pour chacune des 4 variables les bornes des trois classes ainsi formées.

Tableau 3.2. Résultat du codage dijontif complet

Variable	Classe	Bornes		Nombre d'observations
		inférieure	supérieure	
SUBV	1	0.00	1410.00	38
	2	1410.00	7280.00	38
	3	7280.00	57500.00	37
QMAM	1	0.00	136.34	38
	2	136.34	609.72	38
	3	609.72	4829.48	37
β_1	1	0.00	0.06	38
	2	0.06	0.35	38
	3	0.35	10.42	37
β_2	1	0.00	2.67	38
	2	2.67	3.65	38
	3	3.65	17.04	37

Le tableau analysé au moyen de l'AFC a donc comme dimensions 113 lignes correspondant aux 113 stations et 12 colonnes correspondant aux quatre variables codées selon trois modalités (SUBV1, SUBV2, SUBV3, QMAM1, etc ...).

En analyse factorielle des correspondances on peut calculer, pour chaque variable, un indice nous permettant de juger de la qualité de leur représentation sur chacun des facteurs (113 facteurs dans l'espace des stations). Cette mesure, la contribution relative des variables aux facteurs, est donnée pour les 7 premiers facteurs à l'annexe A. Le tableau A.1 donne les contributions relatives individuelles aux facteurs de chaque modalité des quatre variables. Le tableau A.2 donne les contributions relatives moyennes cumulées des variables dans les espace à 1, 2, 3, 4, 5, 6 et 7 dimensions. On remarque en particulier, dans ce dernier tableau, que le pourcentage de variance expliquée pour les variables SUBV et QMAM est de 78% dans l'espace à trois dimension tandis que celui des variables β^1 et β^2 est 42% dans ce même espace. Ainsi, cette application de l'AFC permet d'expliquer sur les trois premiers facteurs 60.6% de la variabilité totale (moyenne des contributions relatives cumulées sur les 3 premiers facteurs). Ces trois premiers facteurs permettent donc une bonne représentation de l'ensemble des points-variables.

L'examen de la position des points-variables dans le plan des axes F1-F2, F1-F3 et F2-F3 (Figures 3.4, 3.5 et 3.6) montre, d'une part, que les variables SUBV et QMAM ont une variation identique le long du facteur 1 (Figures 3.4 et 3.5). On obtient en effet, une similitude de parcours pour ces deux variables en reliant par des droites les points représentant des modalités voisines. L'axe 1 peut donc être qualifié, ici aussi, d'axe de taille de bassin opposant les petites superficies aux grandes. D'autre part, on constate, de la même façon, que β_1 et β_2 ont un comportement très semblable mais le long de l'axe 2 (Figures 3.4 et 3.6). Ainsi, le facteur 2 est un axe qui oppose les fortes valeurs aux faibles valeurs de coefficients d'asymétrie et d'aplatissement. Ce facteur est un axe caractérisant le type de loi des débits maximums. Enfin, le facteur 3 oppose les superficies de bassin extrêmes (grandes et petites) et les superficies de bassin moyennes (Figures 3.5 et 3.6).

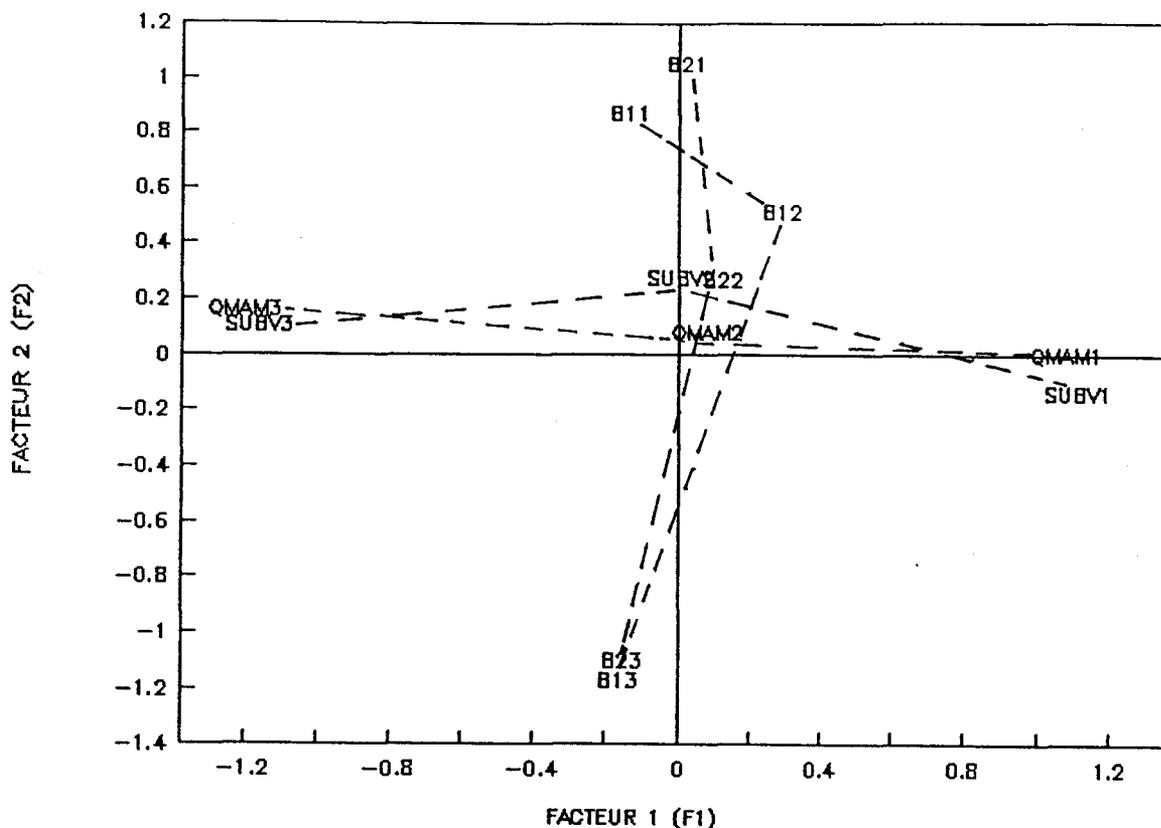


Figure 3.4 Représentation des variables dans le plan factoriel F_1 - F_2 de l'AFC

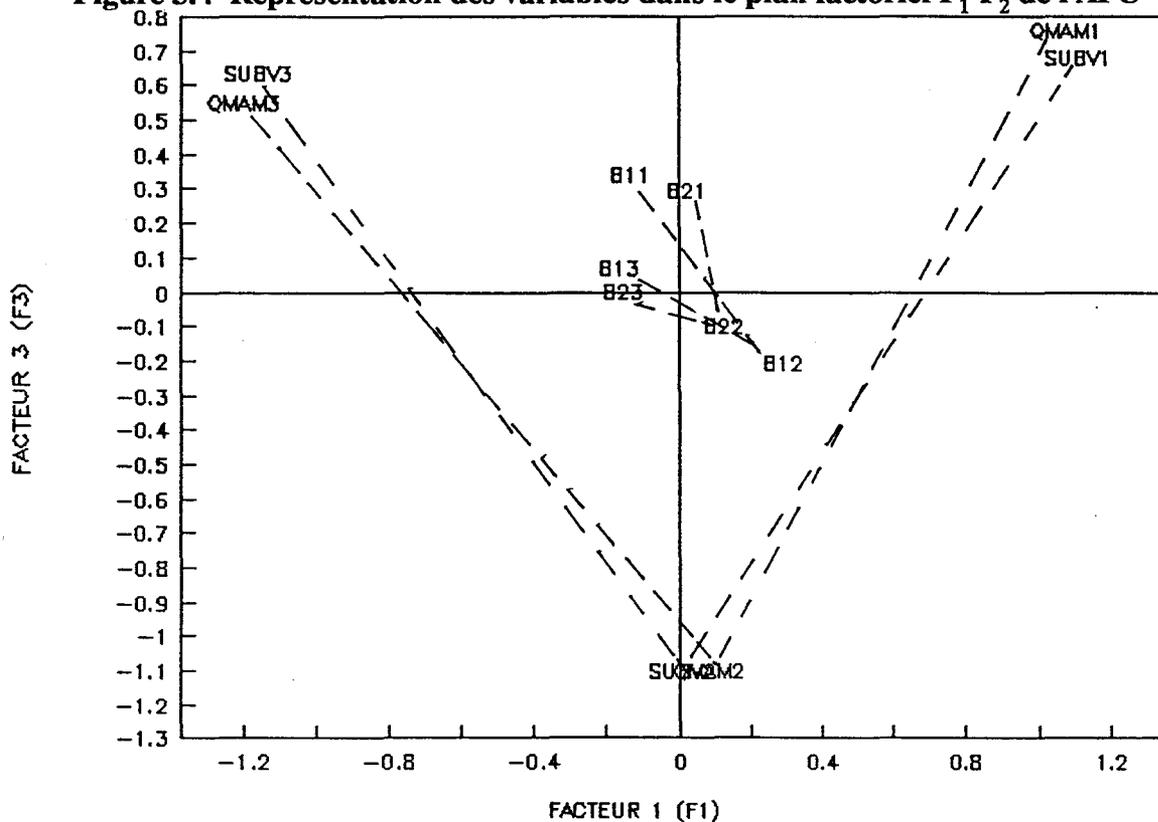


Figure 3.5 Représentation des variables dans le plan factoriel F_1 - F_3 de l'AFC

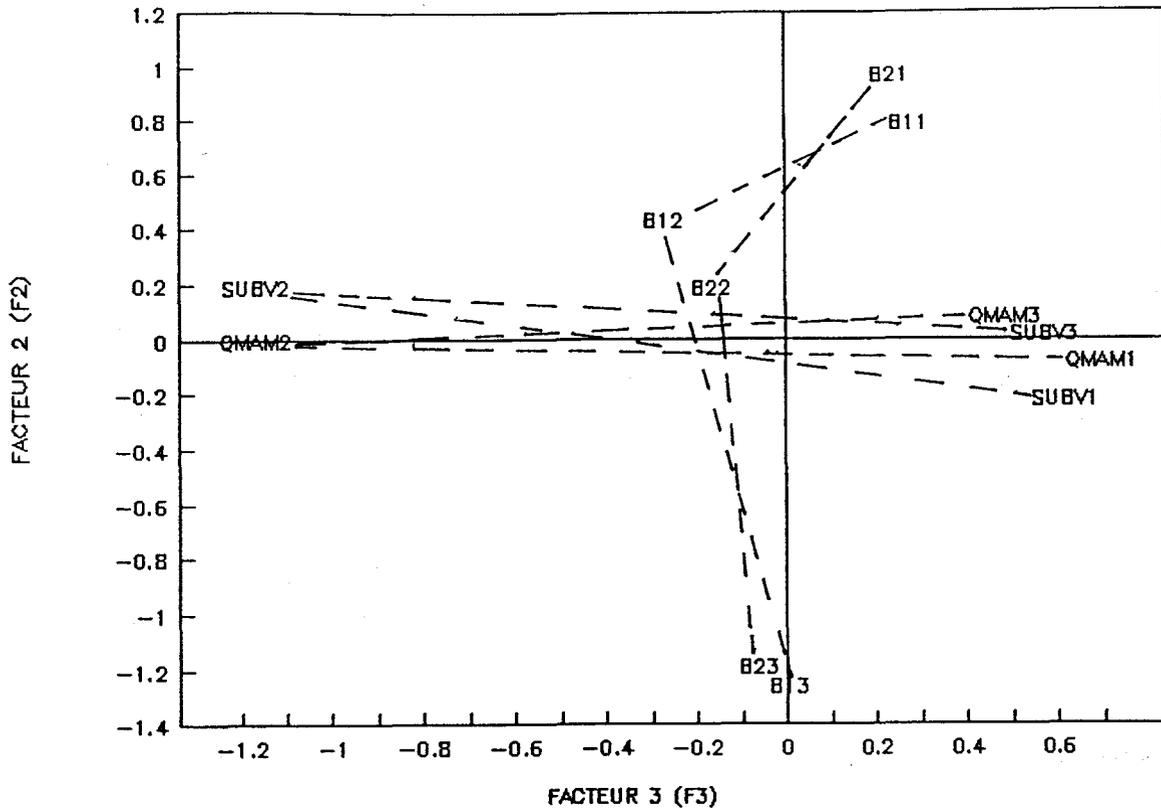


Figure 3.6 Représentation des variables dans le plan factoriel F_2 - F_3 de l'AFC

Comme l'AFC permet d'interpréter simultanément les proximités entre les éléments des deux espaces (variables et stations), on peut conclure que le premier facteur discrimine les stations selon leur superficie de bassin et leur débit maximum moyen, le second facteur selon le type de loi des débits maximums, et le troisième facteur distingue les stations aux superficies et débits extrêmes de celles ayant des superficies et débits moyens.

3.2.2 Analyse des points-stations

Pour l'analyse et l'interprétation des points-stations dans l'espace factoriel, ainsi que la formation de groupes homogènes, on utilise la classification ascendante hiérarchique (CAH) selon l'approche décrite par Jambu (1976). Cette méthode peut être appliquée, soit sur un tableau de mesures brutes, soit sur un tableau de facteurs. Dans ce cas-ci, on a choisi d'appliquer la CAH sur le tableau des facteurs, car en procédant ainsi, on élimine du tableau de données brutes un bruit de fond attribuable aux fluctuations d'échantillonnage ou à l'imprécision des mesures. Nous avons donc convenu d'appliquer la CAH sur le tableau des six premiers facteurs puisqu'ils permettent une très bonne représentation des quatre variables (contribution relative cumulée supérieure à 90% pour toutes les variables). La représentation de la hiérarchie supérieure du classement (arbre hiérarchique) est fournie à la Figure 3.7.

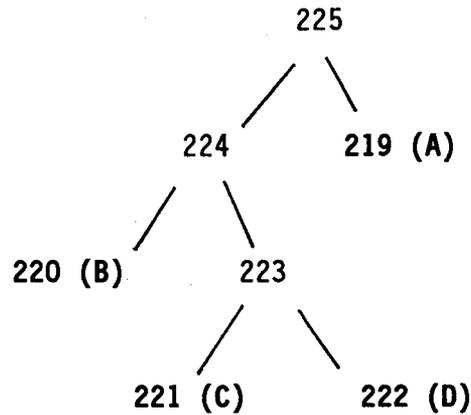


Figure 3.7 Hiérarchie supérieure du classement.

La classe correspondant au noeud 225 représente le nuage de l'ensemble des 113 points-stations et le premier niveau de la hiérarchie. Les deuxième et troisième niveaux sont constitués des classes correspondant respectivement aux noeuds 224 et 223. En coupant l'arbre hiérarchique au quatrième niveau, on obtient une partition des stations en quatre classes, soit les classes de stations associées aux noeuds 219, 220, 221 et 222 que l'on note respectivement groupes A, B, C et D.

La représentation des centres de gravité des classes de la hiérarchie dans l'espace factoriel (Figures 3.8, 3.9 et 3.10) nous permet d'interpréter les différentes étapes de la CAH en examinant leur position dans l'espace. On voit que la première division des stations (séparation de la classe 225 en 224 et 219) se fait selon les axes 1 et 3 (Fig. 3.9), la seconde division (séparation de la classe 224 en 220 et 223), principalement selon l'axe 3 (Fig. 3.9 et 3.10), et la dernière, obtenue en considérant la séparation de la classe 223 en ses deux constituants 221 et 222, est plutôt dans la direction de l'axe 2 (Fig. 3.8 et 3.10).

En vertu de la propriété barycentrique de l'AFC, on peut caractériser les classes de la partition en examinant les proximités entre les points-variables et les centres de gravité des classes de points-stations dans l'espace factoriel (Figures 3.11, 3.12 et 3.13). On peut ainsi distinguer que la classe 220 est caractérisée par de fortes valeurs de débits maximums moyens et de superficies de bassin alors que la classe 219 est caractérisée par de faibles valeurs de ces deux variables. On constate aussi que la classe 221 est associée à de faibles valeurs de coefficients d'asymétrie et d'aplatissement. Enfin, la classe 222 semble, à un degré moindre, caractérisée par des valeurs moyennes de superficie et de débits maximums moyens. De l'examen de ces représentations simultanées, on voit que ce sont les modalités des variables superficie de bassin et débit maximum moyen qui caractérisent principalement les quatre groupes de stations (la classe 221 à un degré moindre).

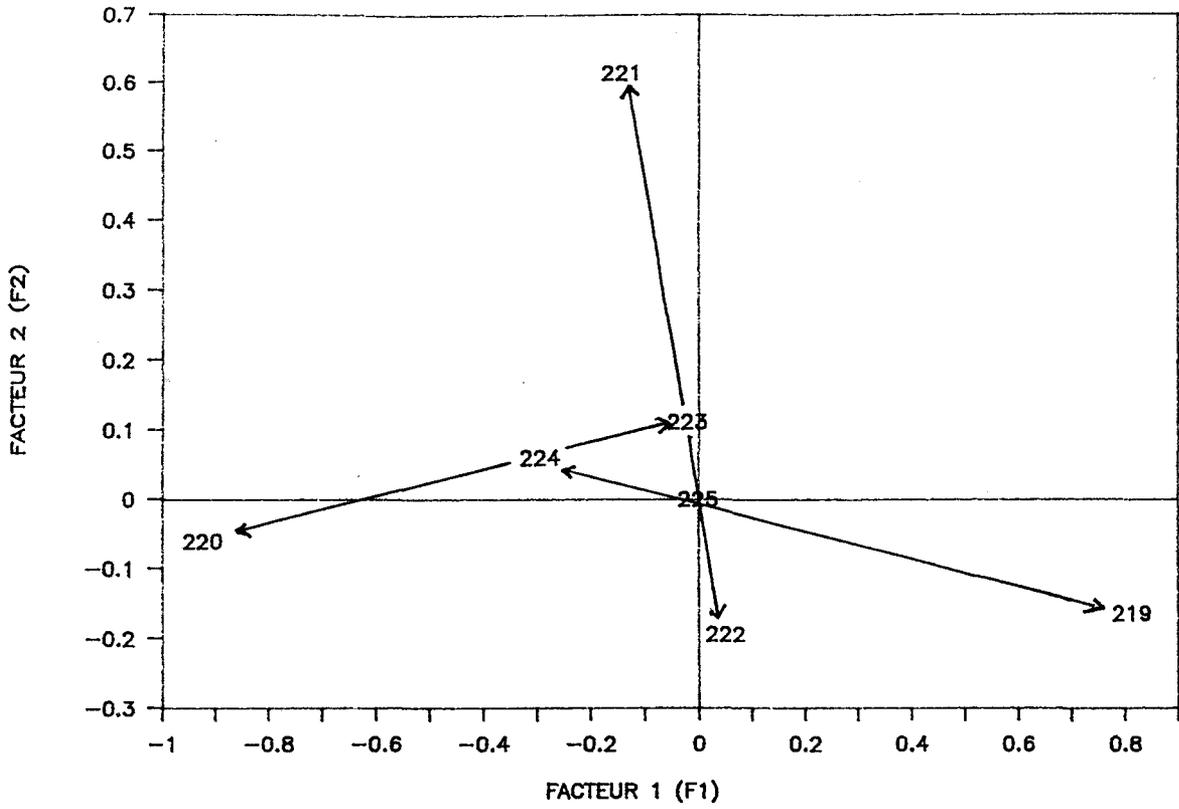


Figure 3.8. Représentation des centres de gravité dans le plan F_1 - F_2 de l'AFC

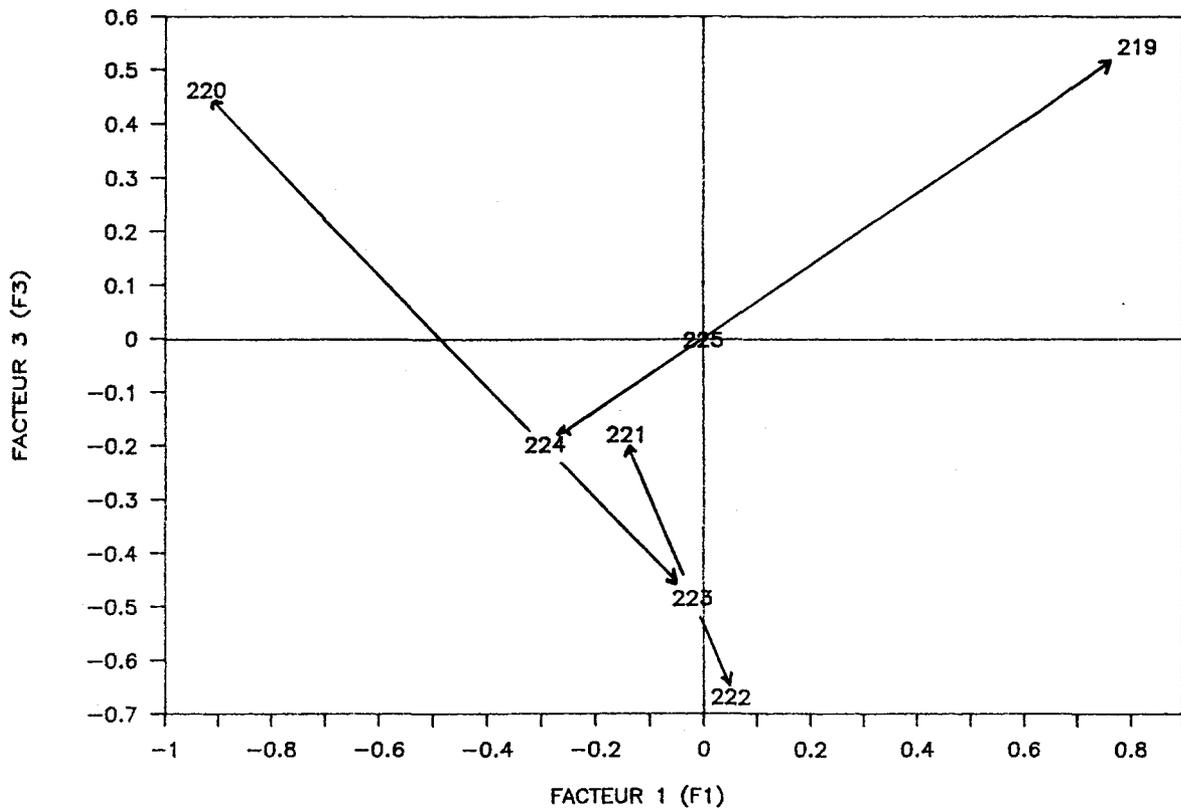


Figure 3.9. Représentation des centres de gravité dans le plan F_1 - F_3 de l'AFC

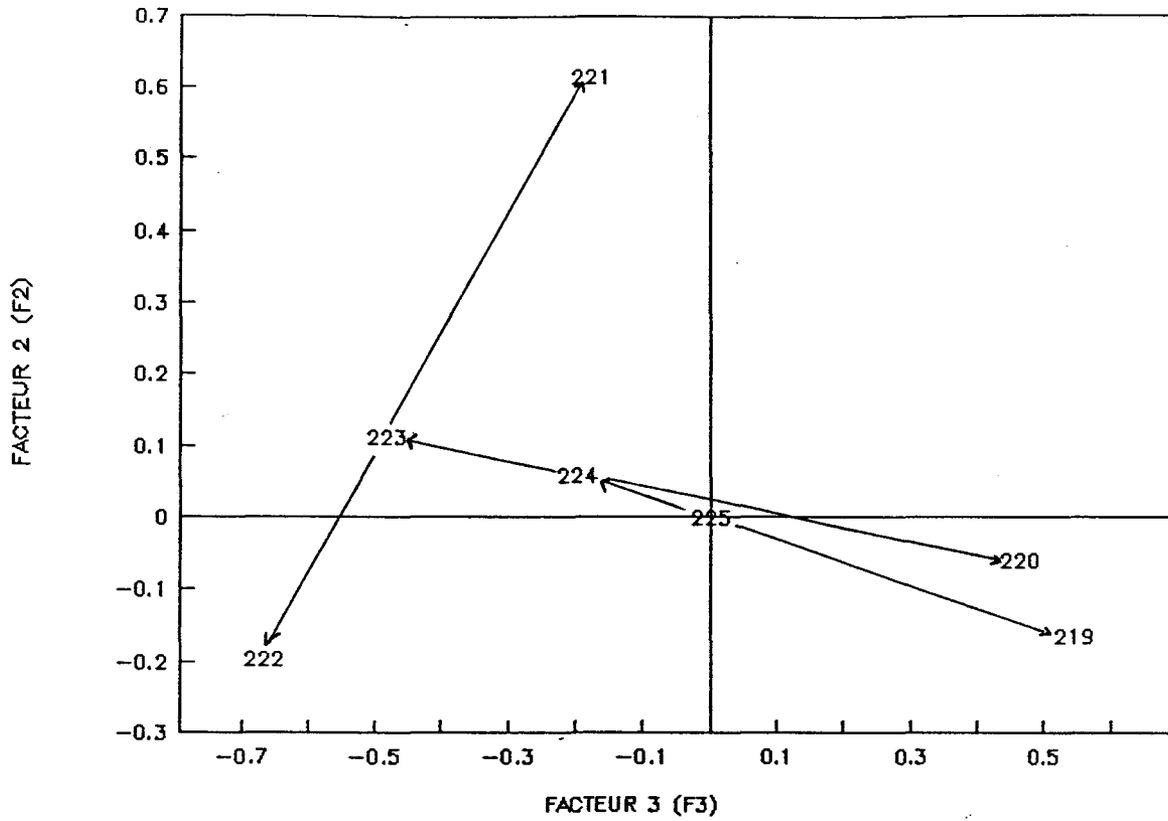


Figure 3.10. Représentation des centres de gravité dans le plan F_2 - F_3 de l'AFC

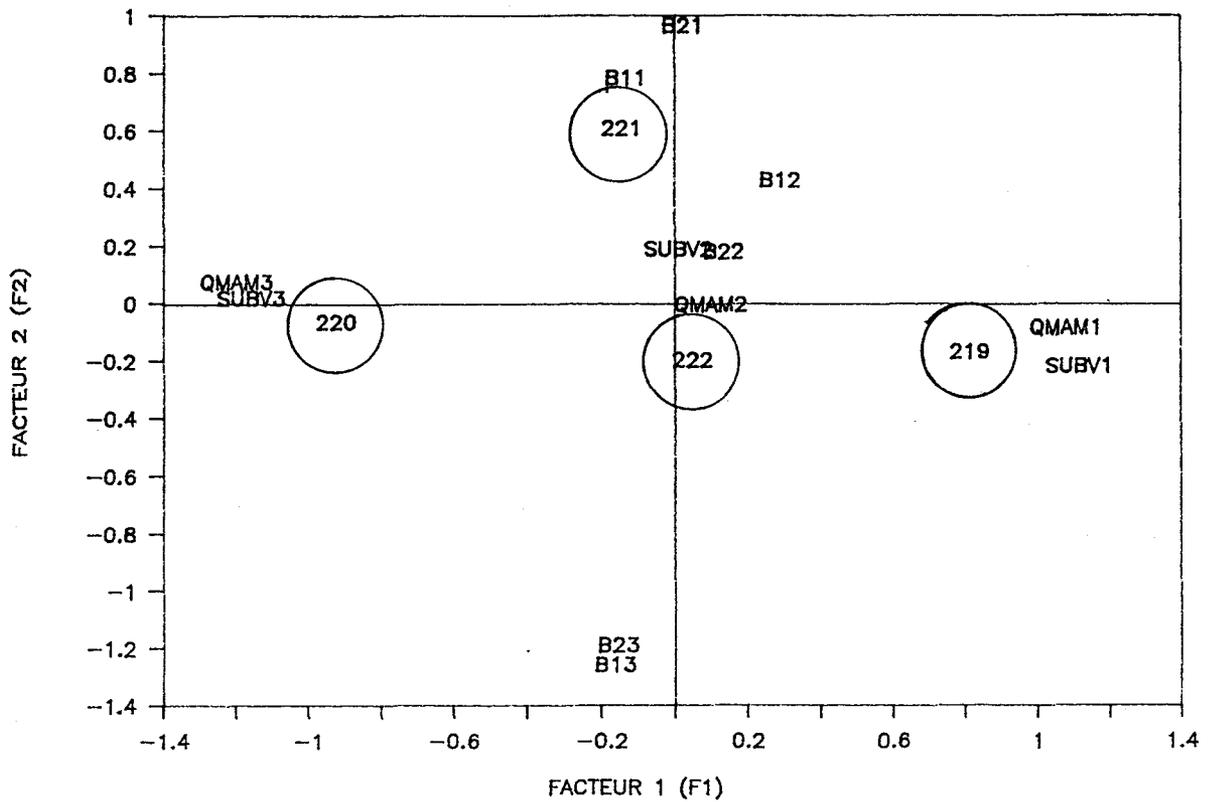


Figure 3.11. Représentation simultanée dans le plan F_1 - F_2

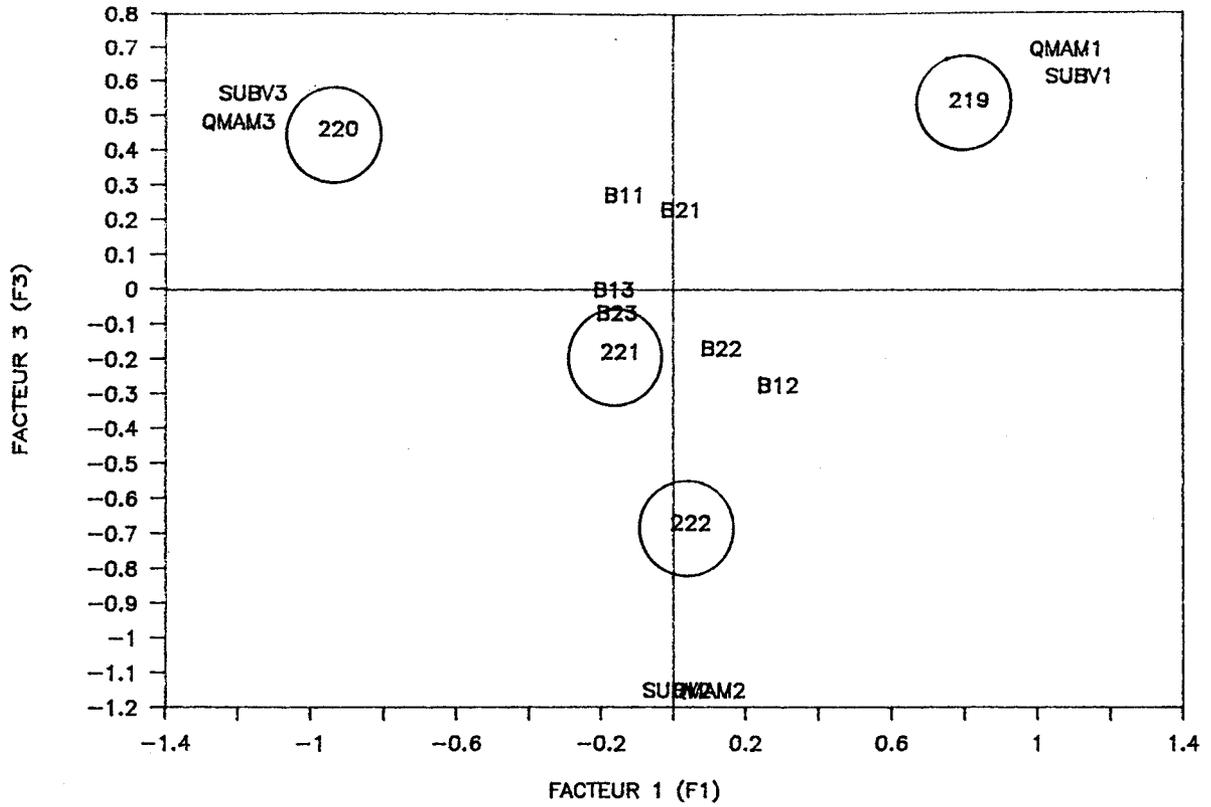


Figure 3.12. Représentation simultanée dans le plan F_1 - F_3

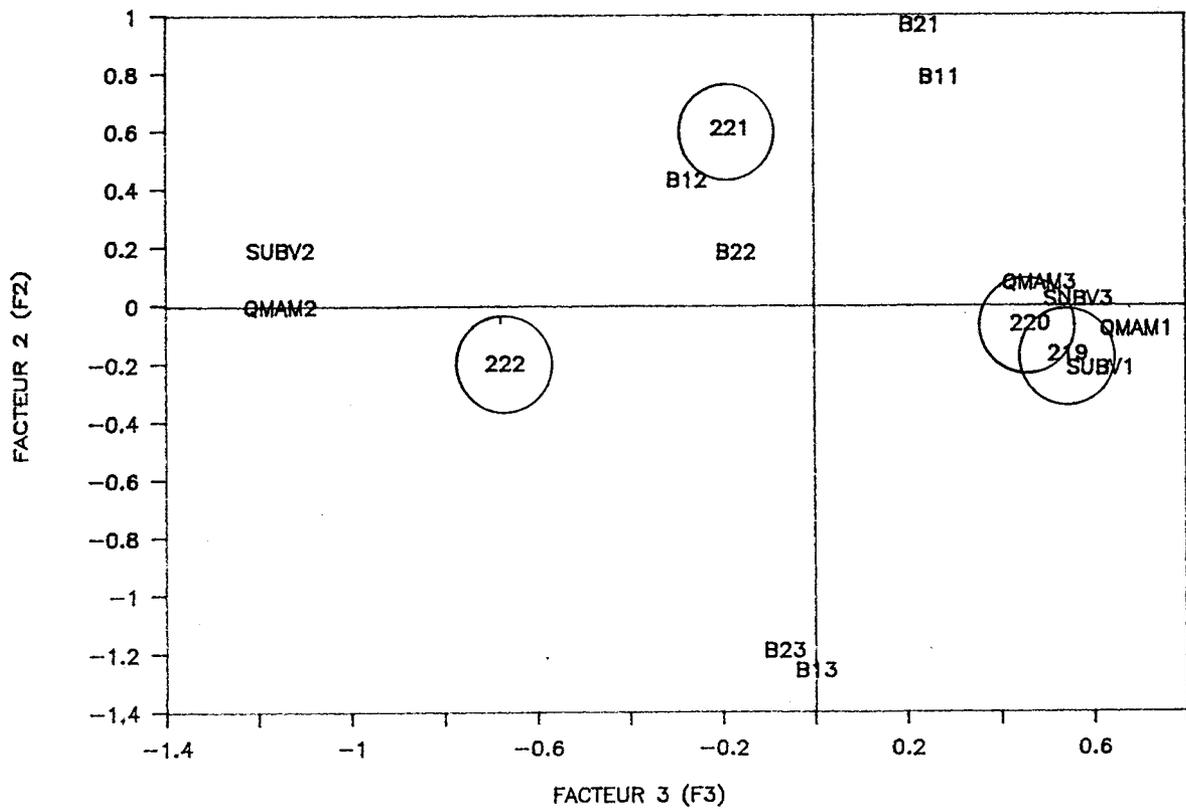


Figure 3.13. Représentation simultanée dans le plan F_2 - F_3

Pour mieux interpréter ces résultats, nous avons reporté sur la carte du Québec les 113 stations en les identifiant d'un symbole correspondant à la classe dont elles font partie (Figure 3.14). L'examen de la répartition géographique des quatre groupes de stations montre, malgré quelques régions où l'on observe de fortes densités de stations appartenant à la même classe (dans la partie sud-ouest), qu'il n'est pas possible de délimiter géographiquement les quatre groupes de stations homogènes. Ceci complique l'affectation, à un des groupes, d'une rivière non-jaugée pour laquelle on voudrait estimer le débit X_T de période de retour T à l'aide de stations ayant un comportement hydrologique similaire.

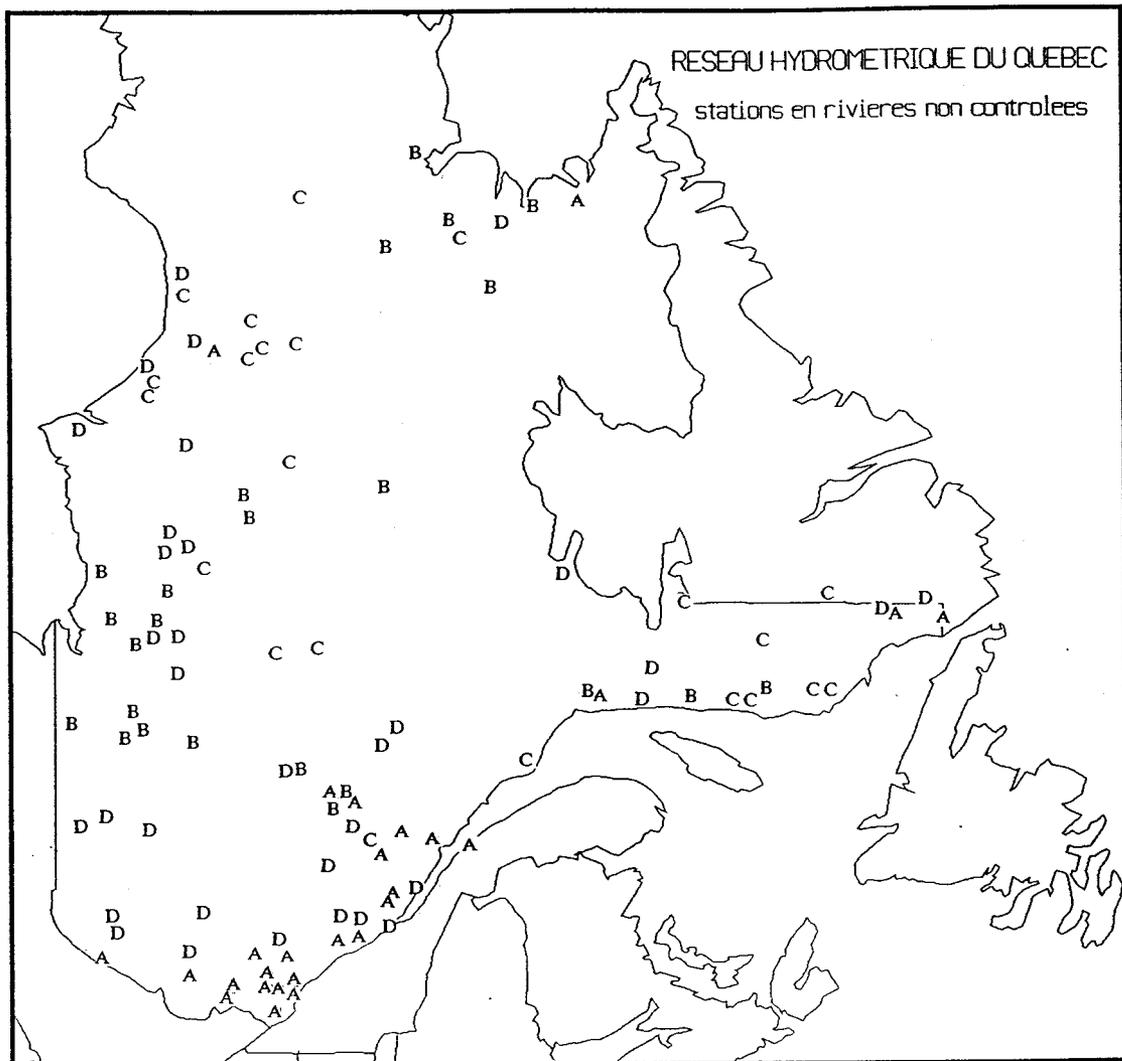


Figure 3.14. Représentation géographique des 113 stations selon les quatre groupes

3.3. Affectation d'une rivière non-jaugée à un des groupes

Lorsque l'on veut estimer le débit X_T régionalement pour une station non-jaugée, nous devons l'affecter à un des quatre groupes de stations homogènes pour lesquelles nous disposons d'une série de débits maximums annuels mesurés. Si les classes obtenues à la section précédente étaient délimitées géographiquement, il suffirait de situer la rivière non-jaugée sur la carte de la figure 3.14 et de l'affecter à son groupe. Puisque les résultats ne permettent pas une telle délimitation, nous devons envisager une autre approche.

Une façon d'affecter une rivière non-jaugée à un des groupes pourrait consister, dans un premier temps, à développer des modèles de prédiction à l'aide des stations jaugées (regressions simples) mettant en relation la superficie de bassin (SUBV) et chacune des variables hydrologiques (QMAM, β_1 et β_2). Ensuite, connaissant la superficie de bassin d'une rivière non-jaugée, il serait possible d'estimer ces variables hydrologiques non-mesurées pour enfin entrer la rivière non-jaugée dans l'analyse factorielle et dans la classification hiérarchique comme station supplémentaire (voir Lebart, Morineau et Tabard, 1977; Jambu, 1978). De cette façon, les deux analyses permettraient d'affecter automatiquement la rivière non-jaugée à un des quatre groupes homogènes.

Toutefois, cette méthode s'avère infructueuse puisque les relations entre la superficie de bassin (SUBV) et les coefficients β_1 et β_2 ne sont pas significatives. Nous avons donc remplacé ces deux variables par de nouveaux coefficients, notés δ_1 et δ_2 , définis par Bobée *et al* 1991. Ces nouveaux coefficients ont l'avantage d'être moins variables puisqu'ils dépendent de moments d'ordre inférieurs. Les relations entre la superficie de bassin et ces deux nouvelles variables, quoique significatives, ne permettent pas d'obtenir des estimations efficaces de δ_1 et δ_2 pour une rivière non-jaugée. C'est pourquoi, nous proposons dans la section qui suit d'effectuer la régionalisation en considérant seulement la variable SUBV.

4. ESTIMATION RÉGIONALE NE FAISANT INTERVENIR QUE LA VARIABLE SUBV

Puisque les classes obtenues à la section 3.2 ne sont pas géographiques et puisqu'on ne peut estimer les variables hydrologiques de façon efficace, il est préférable de nous restreindre, en pratique, à régionaliser à l'aide de la seule variable disponible sur une rivière non-jaugée, la superficie de bassin. Nous diminuons ainsi les sources d'erreurs engendrées, soit par une affectation géographique, soit par une estimation imprécise des variables hydrologiques. Cette alternative pratique est aussi justifiée par les résultats

obtenus à l'aide de l'analyse en composantes principales, de l'analyse factorielle des correspondances et de la classification hiérarchique effectuées aux sections précédentes. En effet, nous avons constaté que :

- SUBV est très corrélée avec le débit maximum annuel moyen (section 3.1)
- SUBV est très bien représentée sur le premier facteur de l'analyse factorielle des correspondances (section 3.2.1 et annexe A)
- SUBV discrimine bien les stations aux comportements hydrologiques différents (les principales divisions obtenues avec la CAH sont dues à cette variable, section 3.2.2)

Les 113 stations ont donc été divisées en 3 groupes suivant la taille de leur bassin versant : petite, moyenne ou grande. Cette classification des stations est la même que celle obtenue lors du codage dijontif des variables utilisées pour l'AFC (voir Tableau 3.1).

Toutefois, avant d'effectuer l'estimation régionale de X_T en considérant ces trois classes de stations, on doit s'assurer que les stations d'un même groupe soient homogènes et que celles appartenant à des groupes différents soient hétérogènes. En effet, la partition des 113 stations en trois classes permet une meilleure estimation régionale de X_T sur une rivière non-jaugée si les stations de différents groupes ont bel et bien des comportements hydrologiques différents, c'est-à-dire :

- 1) si les débits maximums annuels de stations appartenant à différents groupes sont distribués selon des lois de probabilité différentes;
- 2) si les trois relations régressives entre les superficies de bassin et les débits X_T diffèrent.

Dans le cas où ces deux conditions ne sont pas vérifiées, il est préférable de considérer une seule loi pour les 113 stations et d'effectuer la régression à l'aide de l'échantillon tout entier. Les 113 stations sont alors jugées homogènes.

Les deux sections qui suivent sont consacrées à la vérification de ces deux conditions. Le schéma de la Figure 4.1 résume les étapes de cette étude et les différentes avenues possibles advenant le rejet ou l'acceptation des deux hypothèses.

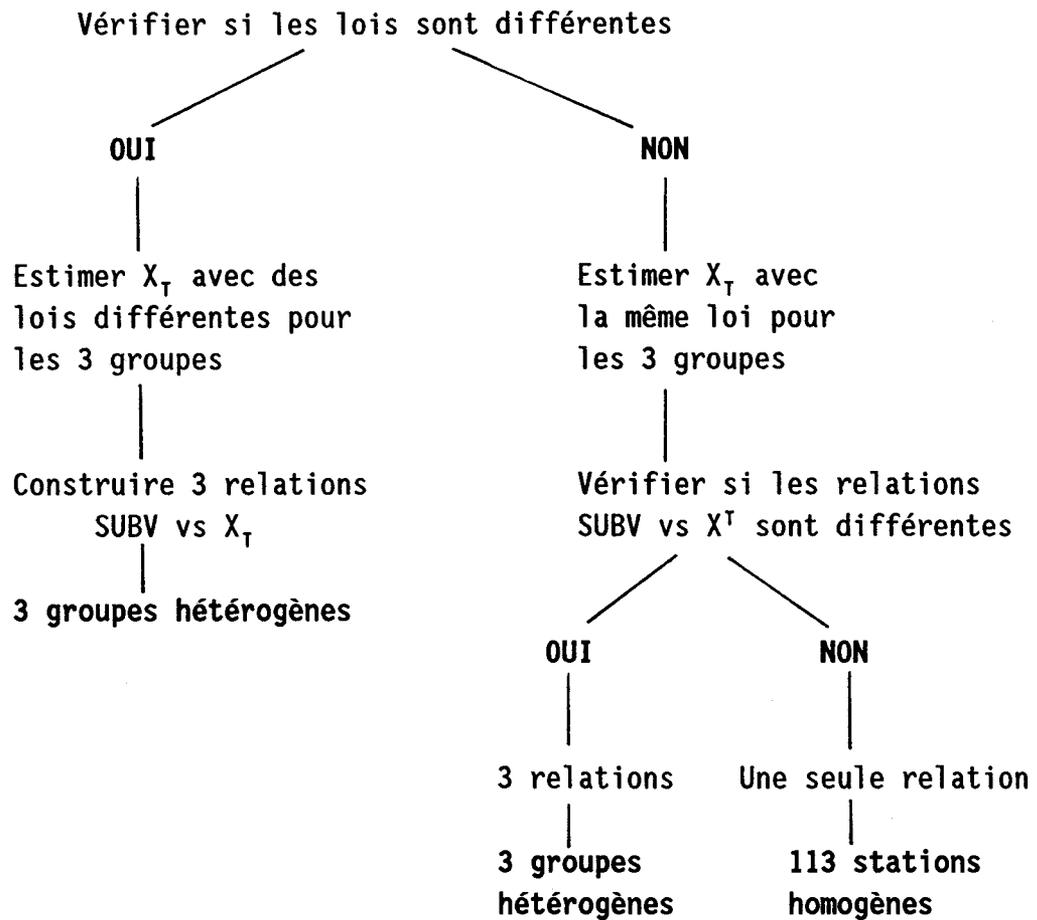


Figure 4.1. Les différentes avenues possibles de régionalisation

4.1. Lois de probabilité

Nous avons vérifié, dans un premier temps, si les débits des trois groupes de stations de superficies différentes suivaient des lois différentes. Des tests d'ajustements de Kolmogorov-Smirnov (D'Agostino et Stephens, 1986) ainsi que des ajustements de lois à l'aide du logiciel AJUST (Bobée et Ashkar, 1991) ont été effectués. On conclut que d'un groupe à l'autre les débits sont bien représentés par une même loi, la loi Pearson Type 3. Les résultats des tests de Kolmogorov-Smirnov, à savoir si les débits sont distribués selon une loi Pearson Type 3, sont donnés à la Table 4.1.

Dans ce tableau, nous donnons, pour 30 stations (dix de chaque groupe), le seuil observé P obtenu. Le seuil observé du test de Kolmogorov-Smirnov est la probabilité, sous l'hypothèse que les observations sont bel et bien distribuées selon une loi Pearson Type 3 (hypothèse H_0), que la statistique du test soit plus extrême que ce que l'on a calculé avec

l'échantillon. Ainsi, un grand seuil observé supporte l'hypothèse H_0 alors qu'un faible seuil observé favorise le rejet de cette hypothèse. Cette mesure permet de quantifier la décision de rejeter ou d'accepter la loi Pearson Type 3. On peut montrer que pour effectuer le test, comparer la statistique calculée au quantile de la loi théorique de la statistique de Kolmogorov-Smirnov (méthode conventionnelle) est équivalent à comparer la valeur de P au niveau de signification qu'on doit fixer a priori. On dira donc que l'hypothèse H_0 est rejetée, pour un niveau de signification 5%, si le seuil observé est inférieur à cette valeur. Dans le tableau 4.1, nous avons identifié par une étoile en indice supérieur les cas où nous rejetons la loi Pearson Type 3 à un niveau de signification de 5%.

Tableau 4.1. Seuils observés P des tests de Kolmogorov-Smirnov pour la loi Pearson Type 3

PETITS		MOYENS		GRANDS	
Stations	P	Stations	P	Stations	P
40212	0.633	40810	0.342	61901	0.912
50119	0.949	42607	0.881	61905	0.886
50409	0.868	43012	0.637	62101	0.443
50701	0.661	50135	0.999	62102	0.973
51001	0.846	50304	0.997	73801	0.716
51002	0.277	61906	0.331	92715	0.844
51003	0.659	74601	0.893	93804	0.683
51004	0.418	76601	0.922	102701	0.272
51301	0.558	80101	0.162	103605	0.158
73301	0.312	93801	0.663	104001	0.014*

* rejet de H_0 à un niveau de signification de 5%.

On remarque que la loi Pearson Type 3 n'est pas rejetée à un niveau de signification de 5% pour 29 des 30 stations présentées dans la Table 4.1. On rejette cette loi seulement pour la station 104001. On constate aussi que la valeur de P est généralement élevée, ce qui justifie l'utilisation de la loi Pearson Type 3. Cette conclusion peut être généralisée à l'ensemble des 113 stations étudiées d'autant plus que cette loi est reconnue en hydrologie pour bien représenter les débits maximums annuels (Bobée et Robitaille, 1977).

L'ajustement de la loi Pearson Type 3 a donc été effectué pour les 113 échantillons de débits maximums annuels à l'aide du logiciel AJUST (Bobée et Ashkar, 1991). Les paramètres de la loi ont été estimés par la méthode des moments en utilisant les coefficients d'asymétrie Cs_2 corrigés pour le biais (Bobée et Robitaille, 1977). Les débits X_T sont ensuite estimés pour les périodes de retour T de 2, 5, 10, 20, 50, 100, 200 et 1000 ans. Les estimations de ces débits sont présentés à l'annexe B. On remarque que ces tableaux donnent les résultats que pour 104 stations. En effet, 9 séries de débits ont dû être rejetées à cause de problèmes de convergence des méthodes d'estimation. Nous considérons donc, pour la suite, seulement 104 stations.

4.2. Comparaison des relations X_T -SUBV

Pour étudier les relations X_T -SUBV pour les huit périodes de retour considérées, nous avons premièrement effectué une transformation logarithmique sur les données originales (pour les deux variables, SUBV et X_T). Cette transformation a pour but de linéariser (hypothèse de base en régression) la relation entre ces deux variables.

Les Figures 4.1 et 4.2 illustre, pour $T = 1000$ ans et pour les 104 stations, la relation X_T -SUBV avant et après la transformation.

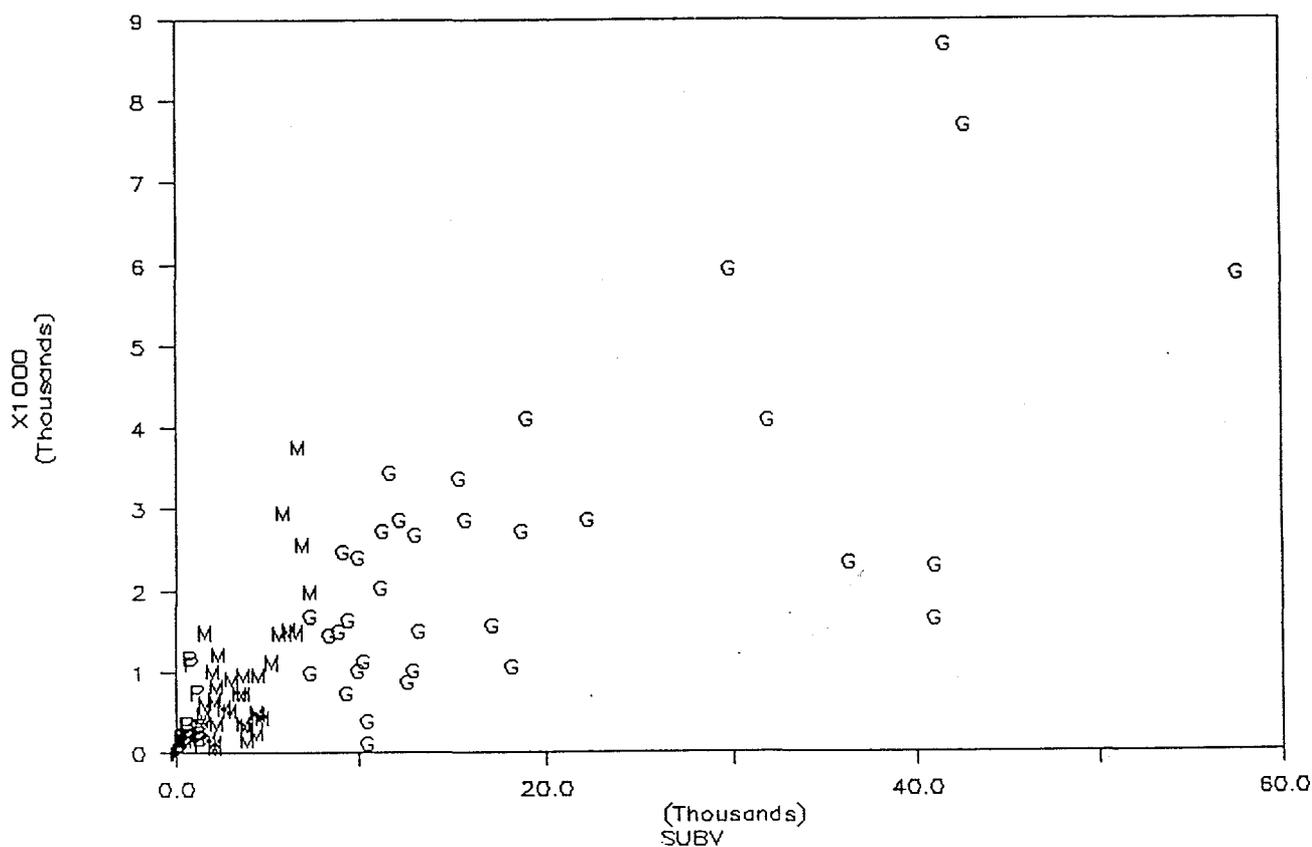


Figure 4.1. Graphique de X_T en fonction de SUBV ($T = 1000$)

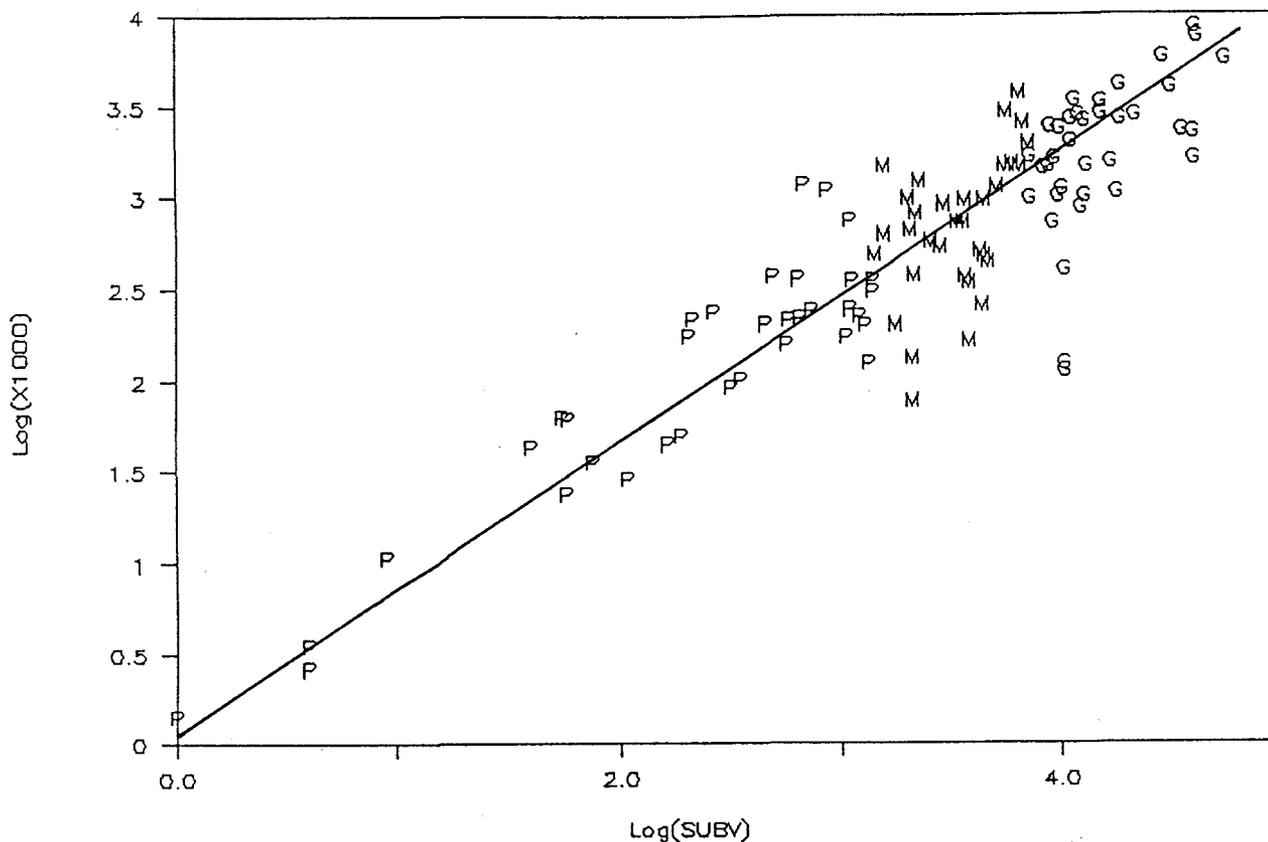


Figure 4.2. Graphique de $\text{Log}(X_T)$ en fonction de $\text{Log}(\text{SUBV})$ ($T = 1000$)

Pour comparer les trois droites de régression (petites, moyennes et grandes superficies) pour chaque période de retour, nous utilisons la méthode des variables indicatrices dichotomiques telle que décrite dans Neter, Wasserman et Kutner (1985). Nous décrivons ici cette méthode pour le cas de la comparaison de deux droites de régression, pour ensuite généralisée à trois droites.

Supposons que l'on veuille comparer les régressions simples $\text{Log}(X_T)$ - $\text{Log}(\text{SUBV})$ pour deux groupes de stations. Exprimons les deux relations linéaires comme suit :

$$\text{Groupe 1 : } \quad \text{Log}(X_{T1}) = \gamma_0 + \gamma_1 \text{Log}(\text{SUBV}_1)$$

$$\text{Groupe 2 : } \quad \text{Log}(X_{T2}) = \delta_0 + \delta_1 \text{Log}(\text{SUBV}_2)$$

En fait, nous désirons savoir si les paramètres de ces deux modèles sont significativement différents. Plus précisément, nous voulons tester les hypothèses suivantes :

$$H_1 : \quad \gamma_0 = \delta_0 \quad \text{et} \quad H_2 : \quad \gamma_1 = \delta_1$$

Pour ce faire, on crée une variable indicatrice dichotomique X identifiant le groupe dans lequel est classée chacune des observations. Cette variable se définit comme suit :

$$\begin{aligned} X &= 0 \text{ si la station appartient au Groupe 1 et} \\ &= 1 \text{ sinon} \end{aligned}$$

On estime ensuite par régression multiple les paramètres β_0 , β_1 , β_2 et β_3 du modèle suivant à l'aide de toutes les observations (les deux groupes) :

$$\text{Log}(X_T) = \beta_0 + \beta_1 \text{Log}(\text{SUBV}) + \beta_2 X + \beta_3 X * \text{Log}(\text{SUBV}) \quad (4.1)$$

On remarque que ce modèle peut se réécrire de la façon suivante :

Si $X = 0$ (la station appartient au Groupe 1), alors

$$\text{Log}(X_T) = \beta_0 + \beta_1 \text{Log}(\text{SUBV})$$

Si $X = 1$ (la station appartient au Groupe 2), alors

$$\text{Log}(X_T) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) \text{Log}(\text{SUBV})$$

On constate aisément que les tests conventionnels de régressions linéaires appliqués au modèle (4.1), à savoir si les paramètres β_2 et β_3 sont significatifs, sont équivalents à tester respectivement si les ordonnées à l'origine (γ_0 et δ_0) et les pentes (γ_1 et δ_1) des deux relations sont différentes.

On peut généraliser cette méthode pour le cas qui nous intéresse, soit la comparaison de trois relations régressives. Il suffit cette fois-ci de créer 2 variables indicatrices X_1 et X_2 définies comme suit :

$$\begin{aligned} X_1 &= 0 \text{ si la station possède une superficie de bassin moyenne} \\ &= 1 \text{ sinon} \end{aligned}$$

et

$$\begin{aligned} X_2 &= 0 \text{ si la station possède une grande superficie de bassin} \\ &= 1 \text{ sinon} \end{aligned}$$

et d'utiliser le modèle suivant :

$$\begin{aligned} \text{Log}(X_T) = & \beta_0 + \beta_1 \text{Log}(\text{SUBV}) + \beta_2 X_1 + \beta_3 + \beta_4 X_1 * \text{Log}(\text{SUBV}) \\ & + \beta_5 X_2 * \text{Log}(\text{SUBV}) \end{aligned} \quad (4.2)$$

On estime les six paramètres de ce modèle puis on vérifie, à l'aide des tests usuels de la régression multiple, si les paramètres β_2 , β_3 , β_4 et β_5 sont significatifs. Les tests effectués sur ces quatre paramètres correspondent respectivement à :

- comparer les ordonnées à l'origine des régressions pour les stations de petites superficies à celles de superficies moyennes;
- comparer les ordonnées à l'origine des régressions pour les stations de petites superficies à celles de grandes superficies;
- comparer les pentes des régressions pour les stations de petites superficies à celles de superficies moyennes;
- comparer les pentes des régressions pour les stations de petites superficies à celles de grandes superficies.

Les résultats des tests (estimations des paramètres, statistiques des tests et seuils observés) sur ces quatre paramètres, pour les huit périodes de retour considérées, sont donnés à l'annexe C. On constate systématiquement pour toutes les périodes de retour que β_2 , β_3 , β_4 et β_5 ne sont pas significativement différents de zéro. En effet, tous les seuils observés sont élevés (> 5%, en particulier). Les régressions $\text{Log}(X_T)$ - $\text{Log}(\text{SUBV})$ ne diffèrent donc pas d'un groupe à l'autre : il est préférable d'utiliser une seule relation globale $\text{Log}(X_T)$ - $\text{Log}(\text{SUBV})$.

Les résultats obtenus dans cette section ainsi qu'à la précédente montrent qu'il n'est pas nécessaire de partitionner les 113 stations en différents groupes. On considère donc l'ensemble de ces stations comme homogènes.

5. ESTIMATION RÉGIONALE ET ÉTUDE EMPIRIQUE DE L'EFFICACITÉ

Comme nous considérons l'ensemble des stations comme homogènes, le modèle utilisé pour estimer le débit X_T de période de retour T est

$$\text{Log}(X_T) = \beta_0 + \beta_1 \text{Log}(\text{SUBV})$$

et la régression est effectuée sur les 104 stations pour lesquelles nous avons obtenu une

estimation de X_T . L'estimation des paramètres du modèle ainsi que le R^2 sont donnés, pour les huit périodes de retour, à la Table 5.1.

Tableau 5.1. Résultats des régressions entre $\text{Log}(X^T)$ et SUBV pour les huit périodes de retour T

T	β_0	β_1	R^2
1000	0.3159*	0.7123**	0.8086
200	0.2558*	0.7169**	0.8232
100	0.2249*	0.7193**	0.8299
50	0.1891	0.7223**	0.8371
20	0.1321	0.7270**	0.8472
10	0.0760	0.7322**	0.8555
5	0.0019	0.7392**	0.8645
2	-0.1621	0.7562**	0.8794

* significatif à 5%, ** significatif à 1%

La prévision X_{TN} de X_T pour une rivière non-jaugée ayant une superficie de bassin SUBV_N s'exprime, en appliquant la transformation inverse, comme suit :

$$X_{TN} = 10^{b_0} * \text{SUBV}_N^{b_1}$$

où b_0 et b_1 sont les estimations de β_0 et β_1 respectivement. On peut aussi construire un intervalle de prévision à un niveau de confiance donné. La construction d'un tel intervalle est décrite dans Perreault *et al* (1991b).

Une simulation a été effectuée pour vérifier l'efficacité de l'estimation régionale de X_T par cette approche. Tour à tour, nous avons considéré chacune des stations comme non-jaugée, nous avons estimé X_T à l'aide du modèle de régression correspondant à la période de retour T et nous avons construit les intervalles de prévision à 95% et 99% pour les huit périodes de retour. Le schéma de la Figure 5.1 donne plus précisément les différentes étapes de la simulation pour une période de retour T donnée.

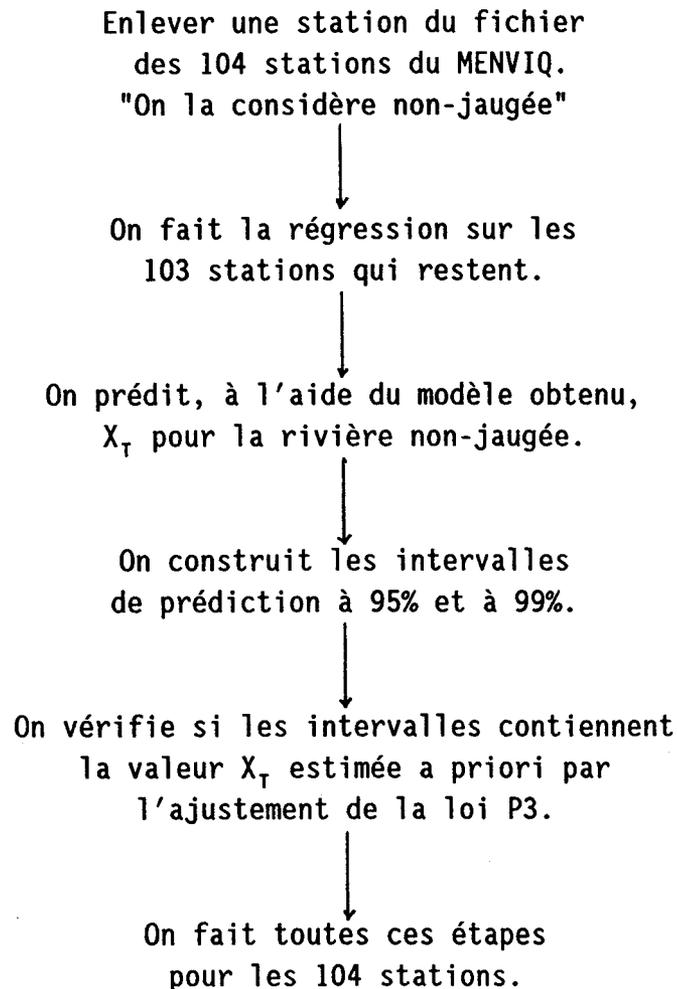


Figure 5.1. Plan de la simulation.

Nous avons dénombré, parmi les 104 stations considérées tour à tour comme non-jaugées, celles dont les intervalles à 95% et 99% ne contiennent pas la valeur de X_T estimée a priori par l'ajustement de la loi Pearson Type 3 (voir section 4.1). Les résultats sont présentés au Tableau de fréquences 5.2.

On constate que la majorité des intervalles de prévision obtenus contiennent les valeurs de X_T connues a priori (94% des intervalles à 95% et 97% des intervalles à 99%). La méthode proposée semble donc adéquate pour estimer les débits de périodes de retour d'une rivière non-jaugée. Toutefois, il est important de noter que ces intervalles sont assez larges, ce qui fait foi de la grande variabilité dans la prévision de X_T .

Tableau 5.2. Nombre de fois que les intervalles contiennent X_T

T	95%		99%	
	OUI	NON	OUI	NON
1000	98	6	102	2
200	99	5	101	3
100	99	5	101	3
50	100	4	101	3
20	100	4	101	3
10	100	4	101	3
5	100	4	101	3
2	100	4	101	3

Nous avons, de plus, étudié le biais de la prévision de X_T pour l'ensemble des 104 stations (considérées comme non-jaugées) et pour les huit périodes de retour. Nous avons considéré le biais brut qui se définit comme suit :

$$B = X_{TN} - X_T$$

où X_{TN} est la prévision du débit de période de retour T pour la rivière non-jaugée et X_T le débit de période de retour T estimé par l'ajustement de la loi Pearson Type 3.

L'interprétation de cette mesure est toutefois trompeuse puisqu'elle ne tient pas compte de la taille du bassin versant de la rivière. En effet, le biais brut dans la prévision de X_T est beaucoup plus grand pour une rivière possédant une grande superficie de bassin. Il est donc préférable, pour que les biais de l'ensemble des stations soient comparables, de considérer le biais unitaire que l'on définit de la façon suivante :

$$BU = \frac{X_{TN} - X_T}{SUBV}$$

Cette mesure standardisée nous permet de relativiser l'erreur de la prévision en

tenant compte de la taille du bassin versant de la rivière. De plus, l'interprétation de BU est pratique puisqu'il donne le biais de la prévision X_{TN} par unité de superficie de bassin (km^2).

Le biais brut (B) ainsi que le biais unitaire (BU) calculés à l'aide des résultats de la simulation pour chacune des 104 stations et pour la période de retour $T = 1000$ ans sont présentés aux Figures 5.2 et 5.3. On retrouve en ordonnée les biais (B et BU) et en abscisse les stations classées en ordre croissant de leur superficie de bassin. Ces mêmes graphiques sont donnés à l'annexe D pour les sept autres périodes de retour.

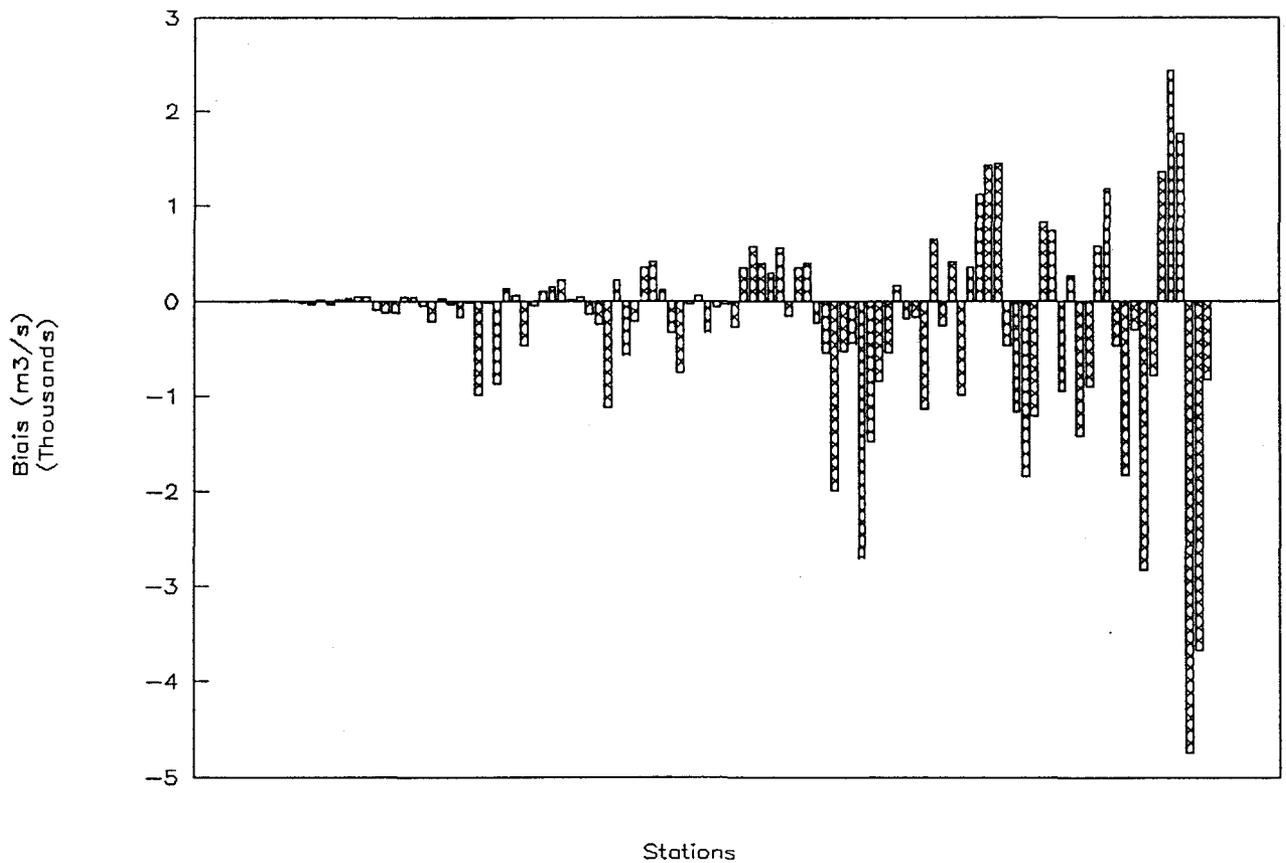


Figure 5.2. Biais brut pour $T = 1000$ ans.

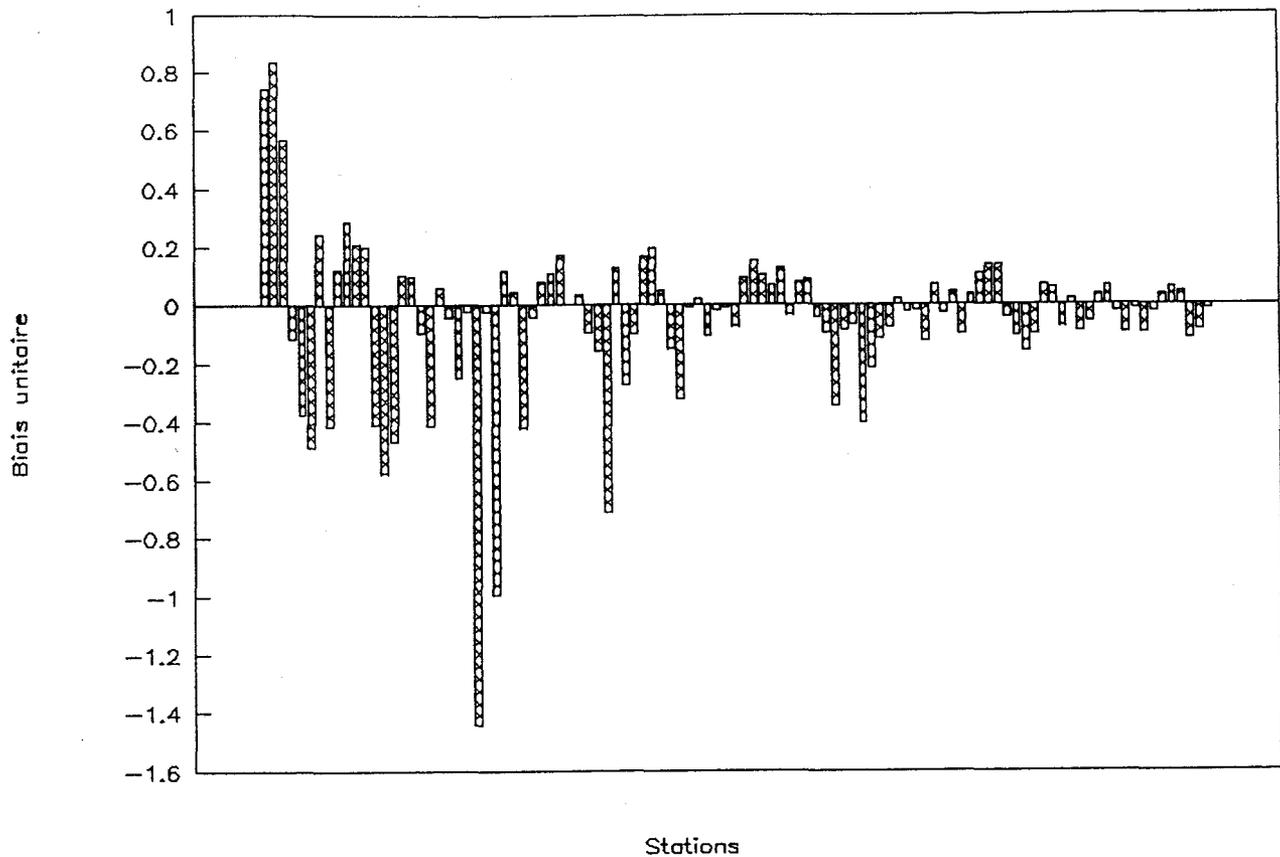


Figure 5.3. Biais unitaire pour T = 1000 ans.

Après l'examen de ces graphiques, on constate qu'il n'y a aucune sous-estimation ou sur-estimation systématique de X_T , et ce pour toutes les valeurs de T. On remarque aussi que l'étude exclusive du biais brut nous aurait amené à conclure que la prévision de X_T est moins bonne pour les rivières ayant de grands bassins que pour celles possédant de petits bassins alors que l'étude du biais unitaire nous indique le contraire. En effet, l'erreur de prévision par km^2 est plus importante pour les petites superficies de bassin. Ceci, peut s'expliquer en partie, par le fait que le débit de ces rivières réagit plus rapidement aux précipitations et à la fonte des neiges, ce qui engendre de plus grandes fluctuations dans les mesures. Enfin, on observe une augmentation du biais à mesure que la période de retour T augmente. Cette constatation n'est pas surprenante puisque l'adéquation du modèle régressif diminue avec T (voir le Tableau 5.1).

6. CONCLUSIONS ET RECOMMANDATIONS

Dans un premier temps, l'analyse factorielle nous a permis d'effectuer une typologie des comportements hydrologiques sur le territoire québécois. Cette approche a montré en particulier que la variable "Superficie du bassin versant" est très corrélée au débit maximum annuel et qu'elle discrimine bien les stations ayant des comportements hydrologiques différents. De plus, l'analyse de classification hiérarchique a permis de former des groupes homogènes. On remarque principalement que ces groupes ne peuvent être délimités géographiquement. Ceci rejoint les conclusions obtenues par plusieurs auteurs (Linsley (1982), Wiltshire (1985), USWRC (1988), Cunnane (1989) et Cavadias (1990)) dans des recherches récentes en régionalisation. Ces auteurs ont montré que la proximité géographique de bassins versants n'est pas nécessairement une garantie de l'homogénéité des caractéristiques de crue, puisque des bassins voisins peuvent être hydrologiquement différents.

Ainsi, suite à ces résultats, nous avons effectué la régionalisation en considérant seulement la variable "Superficie du bassin versant". Cette approche donne de bons résultats. Une simulation montre que la prévision de X_T pour une rivière non-jaugée, est efficace mais variable (intervalles de prévision larges). De plus, aucun biais systématique n'a été observé.

Même si les résultats obtenus avec la variable "Superficie du bassin" seulement sont bons, il serait intéressant de raffiner la méthodologie en considérant de nouvelles variables physiographiques (mesurables sur une rivière non-jaugée) qui, selon nous, influencent grandement le comportement hydrologique d'une rivière. Ces variables peuvent représenter des indices de forme des bassins (allongés, circulaires, ...) ou des indices de compacité (Chow (1964)). Finalement, nous recommandons d'exploiter les nouvelles avenues qui émergent de plus en plus dans la littérature sur la régionalisation, comme par exemple l'approche de voisinage (Burn (1989), Cavadias (1990)).

BIBLIOGRAPHIE

BENSON, M.A. (1962). Factors influencing the occurrence of floods in a humid region of diverse terrain. U.S. Geol. Survey Water Supply Paper, 1580-b.

BOBÉE, B. and R. ROBITAILLE (1977). The use of the Pearson Type 3 and Log Pearson Type 3 distributions revisited. *Water Resources Research*, No 2, pp. 427-443.

BOBÉE, B. and F. ASHKAR (1991). The gamma family and derived distributions applied in hydrology. *Water Resources Publications*. Yevjevich Edt. 204 p.

BOBÉE, B., PERREAULT, L. and F. ASHKAR (1991). Two kinds of moment ratio diagrams and their applications in hydrology. *Journal of Hydrology* (en préparation).

BURN, D.M. (1989). An appraisal of the regions of influence approach to flood frequency analysis. *Journal des Sciences Hydrologiques*, Vol. 35, No 2, pp 149-165.

CAVADIAS, G. (1990). The canonical correlation approach to regional flood estimation. *Regionalisation in Hydrology, Proc. of the Ljubiana Symposium IAHS*, Publ. No 191, pp 171-178.

CHOW, V.T. (1964). *Handbook of applied hydrology*. McGraw Hill.

CUNNANE (1989). *Statistical distributions for flood frequency analysis operational hydrology*. Report No 33, OMM.

D'AGOSTINO, R.B. and M.A. STEPHENS (1986). *Goodness-of-fit techniques*. Marcel Dekker, New York.

JAMBU, M. (1976). Sur l'interprétation mutuelle d'une classification hiérarchique et d'une analyse des correspondances. *Revue de Statistique Appliquée*, No 24, pp. 45-73.

JAMBU, M. (1978). *Classification automatique pour l'analyse des données, Tome I : Méthodes et algorithmes*. Dunod.

LEBART, L., MORINEAU, A. et N. TABARD (1977). *Techniques de la description statistique*. Dunod.

LINSLEY, R.K. (1982). Flood estimate. How good are they? *Wat. Resour. Res.*, No 22, Vol 9.

NETER, J., WASSERMAN, W. and M.H. KUTNER (1985). Applied linear statistical models. Irwin.

PERREAULT, L., ROY, R., BOBÉE, B. et L. MATHIER (1991a). Validation et estimation des apports journaliers. INRS-Eau, rapport scientifique no 271.

PERREAULT, L., ROY, R., BOBÉE, B. et D. TREMBLAY (1991b). Modélisation des pointes et des volumes. INRS-Eau, rapport scientifique no 304.

USWRC (Water Resources Council, 1988). Guidelines for determining flood flow frequency. 2120 Long Island, NW, Washington, DC.

WILTSHIRE, S.E. (1985). Grouping basins for regional flood frequency analysis. Hydrological Sciences Journal, No 30, Vol 1, pp. 151-159.

A N N E X E A

**Contributions relatives des variables
aux 7 premiers facteurs de l'analyse
factorielle des correspondances**

Tableau A.1. Contributions relatives des variables aux facteurs

Variabes	F ₁	F ₂	F ₃	F ₄	F ₅	F ₆	F ₇
SUBV1	0.632	0.023	0.192	0.001	0.002	0.055	0.001
SUBV2	0.000	0.019	0.692	0.010	0.001	0.268	0.008
SUBV3	0.653	0.000	0.157	0.016	0.007	0.081	0.004
QMAM1	0.592	0.003	0.245	0.001	0.000	0.070	0.003
QMAM2	0.004	0.000	0.690	0.038	0.005	0.248	0.016
QMAM3	0.698	0.003	0.115	0.029	0.003	0.055	0.006
β_{11}	0.009	0.315	0.038	0.219	0.343	0.006	0.070
β_{12}	0.043	0.095	0.038	0.304	0.490	0.009	0.021
β_{13}	0.013	0.767	0.000	0.007	0.013	0.030	0.170
β_{21}	0.000	0.475	0.027	0.233	0.145	0.008	0.111
β_{22}	0.008	0.018	0.014	0.605	0.351	0.000	0.004
β_{23}	0.011	0.685	0.002	0.088	0.046	0.006	0.162

Tableau A.2. Contributions relatives moyennes cumulées

Variabes	Dimension de l'espace (Nbre de Facteurs)						
	1	2	3	4	5	6	7
SUBV	0.428	0.442	0.789	0.798	0.802	0.936	0.941
QMAM	0.431	0.433	0.783	0.806	0.809	0.933	0.941
β_1	0.022	0.414	0.439	0.616	0.898	0.913	1.000
β_2	0.006	0.399	0.413	0.722	0.903	0.907	0.999
Totale	0.222	0.422	0.606	0.736	0.853	0.922	0.970

ANNEXE B

**Estimation de X_T pour 8 périodes de retour
selon la loi Pearson Type 3**

Tableau B.1. Estimation des X_T pour les stations de petites superficies

Station	Périodes de retour (années)							
	1000	200	100	50	20	10	5	2
510041	1.45	1.24	1.15	1.05	0.92	0.80	0.67	0.46
510021	2.67	2.39	2.26	2.12	1.92	1.74	1.53	1.14
510071	3.56	3.13	2.93	2.72	2.42	2.16	1.86	1.33
510031	10.86	9.06	8.27	7.47	6.37	5.52	4.62	3.28
402121	43.18	36.47	33.46	30.34	25.99	22.46	18.58	12.34
404091	63.02	57.80	55.15	52.16	47.50	43.16	37.66	26.39
522321	23.90	22.23	21.37	20.39	18.83	17.37	15.50	11.58
510061	61.52	53.44	49.79	46.00	40.69	36.35	31.56	23.74
522311	36.33	33.11	31.51	29.74	27.03	24.58	21.53	15.50
401291	28.73	28.72	28.72	28.70	28.60	28.38	27.78	24.63
401271	45.34	45.23	45.10	44.87	44.23	43.26	41.33	34.42
522281	50.27	50.20	50.11	49.94	49.43	48.59	46.86	40.30
768011	175.79	153.06	142.64	131.70	116.15	103.18	88.53	63.60
467081	217.20	173.29	154.36	135.40	110.30	91.27	72.17	46.72
510051	235.48	206.09	192.75	178.85	159.28	143.17	125.23	95.51
522301	93.44	91.84	90.72	89.18	86.10	82.46	76.71	60.91
507011	102.14	98.68	96.75	94.45	90.60	86.75	81.54	69.73
725021	205.76	178.51	166.21	153.44	135.58	120.99	104.89	78.64
610221	374.82	307.74	278.27	248.31	207.69	175.84	142.40	93.14
606011	157.48	142.37	134.96	126.77	114.34	103.13	89.34	62.32
421031	216.20	172.73	154.18	135.78	111.78	93.97	76.67	55.32
504091	367.27	325.99	307.05	287.17	258.90	235.31	208.63	163.18
522331	222.88	201.64	191.58	180.78	164.93	151.22	135.05	105.48
733011	1199.35	862.07	722.53	587.97	420.98	307.02	210.21	128.34
601011	245.58	229.48	221.43	212.44	198.56	185.83	169.90	137.84
513011	1112.10	871.04	766.89	662.46	523.80	418.25	311.76	168.07
762011	172.71	166.88	163.68	159.87	153.55	147.31	138.91	120.10
618011	251.66	224.45	211.80	198.38	179.03	162.63	143.73	110.44
510011	762.33	681.51	643.78	603.62	545.48	495.90	438.44	336.23
610201	353.17	324.75	311.01	296.02	273.59	253.73	229.75	184.10
942041	228.99	207.75	197.51	186.35	169.69	154.96	137.22	103.59
408141	200.39	184.10	176.09	167.27	153.85	141.75	126.85	97.68
404061	126.63	126.37	126.08	125.55	124.15	122.02	117.83	103.01
942011	355.73	324.37	309.23	292.73	268.08	246.27	219.99	170.08
501191	312.81	276.94	260.46	243.13	218.42	197.76	174.32	134.17

Tableau B.2. Estimation des X_T pour les stations de superficies moyennes

Station	Périodes de retour (années)							
	1000	200	100	50	20	10	5	2
947011	498.09	432.57	403.15	372.72	330.39	296.05	258.46	198.17
501351	637.63	549.71	509.98	468.70	410.90	363.59	311.28	225.74
714011	1500.41	1178.74	1040.89	903.62	723.38	588.41	455.36	285.44
408101	204.62	192.28	185.84	178.44	166.64	155.42	140.90	110.10
704011	1014.37	928.47	886.09	839.23	767.75	703.00	623.00	465.22
746011	681.77	625.14	598.23	569.25	526.59	489.50	445.59	364.66
419031	135.71	127.97	124.13	119.86	113.31	107.33	99.90	85.08
426071	77.42	70.93	67.82	64.46	59.46	55.07	49.83	40.01
103801	382.08	362.82	352.24	339.72	318.99	298.52	271.05	209.72
622101	833.32	675.11	608.08	542.00	456.69	394.43	335.37	266.96
615021	1243.04	995.61	888.27	780.26	636.10	525.5	412.74	256.84
430121	582.95	513.84	482.22	449.06	402.01	362.87	318.75	244.01
503051	539.90	535.73	532.43	527.51	516.82	503.10	479.89	410.58
756011	928.78	852.72	815.66	775.03	713.77	659.03	592.32	463.51
723021	743.98	702.01	680.45	655.99	617.48	581.39	535.27	439.40
622091	740.56	724.49	714.51	701.74	678.48	653.26	616.37	524.21
104401	987.82	927.66	897.12	862.71	809.00	759.11	695.91	598.22
801011	375.18	336.35	318.52	299.76	273.03	250.69	225.35	181.98
945021	165.99	156.25	151.88	147.38	141.13	136.08	130.57	121.82
102704	345.67	309.07	291.87	273.50	246.75	223.79	196.98	148.69
619061	523.12	494.82	480.22	463.60	437.36	412.69	381.04	314.93
933021	259.53	259.54	259.54	259.54	259.52	259.42	258.89	252.76
503041	975.97	912.13	881.23	847.52	797.06	752.32	698.26	595.24
810081	483.31	410.89	379.03	346.61	302.56	267.94	231.49	177.42
938041	454.18	425.30	410.71	394.33	368.87	345.32	315.62	255.12
754011	1147.56	1124.87	1110.54	1091.99	1057.80	1020.23	964.68	823.90
747011	1496.63	1400.94	1354.56	1303.90	1227.95	1160.50	1078.89	922.90
761011	2970.35	2569.17	2388.61	2201.58	1940.82	1728.61	1495.48	1118.99
811011	1537.65	1286.10	1175.93	1064.21	913.30	795.54	672.77	494.42
738021	1508.91	1358.89	1291.16	1220.84	1122.45	1042.04	953.24	808.41
766011	3767.45	3100.26	2808.04	2511.75	2111.48	1799.15	1473.50	1000.44
408301	2579.22	2183.32	2005.57	1821.82	1566.32	1359.13	1132.48	769.41
735031	1986.34	1796.72	1705.15	1605.39	1456.32	1324.45	1165.49	863.60

Tableau B.3. Estimation des X_T pour les stations de grandes superficies

Station	Périodes de retour (années)							
	1000	200	100	50	20	10	5	2
810061	1704.27	1550.64	1480.03	1405.77	1299.93	1211.39	1110.98	938.94
938051	1008.76	915.99	871.81	824.13	753.78	692.44	619.64	484.77
807171	1459.19	1179.97	1062.47	947.34	800.25	694.73	597.23	492.13
103603	1510.42	1367.80	1298.91	1223.84	1111.63	1012.35	892.63	665.16
103715	2480.09	2100.27	1931.35	1757.98	1519.46	1328.70	1123.50	805.51
927231	743.62	741.24	739.07	735.60	727.38	716.00	695.50	629.49
621011	1647.98	1565.04	1522.05	1472.96	1395.12	1321.62	1227.00	1028.19
808091	1031.05	910.44	855.15	797.07	714.50	645.62	567.75	435.09
621021	2419.96	2185.15	2074.94	1957.23	1785.92	1638.96	1467.52	1159.08
927171	1137.37	1038.98	992.64	943.04	870.64	808.29	735.27	603.00
942061	406.22	375.73	361.71	346.95	325.89	308.25	288.22	253.81
950011	126.08	121.84	119.52	116.77	112.22	107.73	101.71	88.28
942071	114.50	102.40	96.85	91.04	82.79	75.92	68.17	55.04
619051	2037.78	1864.19	1780.98	1690.79	1556.92	1439.41	1298.91	1035.54
801041	2738.69	2523.27	2417.75	2301.61	2125.59	1967.27	1773.12	1394.54
749021	3447.42	3116.58	2962.58	2799.05	2562.99	2362.40	2130.85	1721.70
757021	2866.45	2749.90	2683.27	2602.32	2464.07	2323.33	2128.86	1677.52
950021	895.71	824.55	790.57	753.84	699.52	652.04	595.51	490.31
927221	1021.69	973.34	948.58	920.51	876.40	835.14	782.49	673.38
738011	2693.56	2580.14	2517.53	2443.17	2319.54	2197.04	2031.96	1661.60
927151	1516.23	1412.58	1361.98	1306.40	1222.43	1147.18	1055.23	876.94
619011	3370.56	2936.86	2741.14	2538.03	2254.02	2022.06	1766.13	1349.42
749031	2884.33	2683.02	2582.89	2471.55	2300.55	2144.44	1950.08	1562.10
808011	1585.67	1369.52	1273.49	1175.01	1039.68	931.63	815.62	636.54
810071	1082.79	998.35	959.59	918.86	860.89	812.48	757.67	664.07
807041	2739.91	2311.48	2125.89	1939.40	1690.95	1500.88	1307.83	1043.25
723011	4119.14	3875.01	3750.49	3609.77	3389.35	3183.87	2922.60	2383.89
807071	2877.10	2536.52	2384.84	2229.01	2014.27	1842.19	1656.61	1367.55
104001	5958.13	5677.58	5520.67	5332.72	5017.10	4701.21	4271.49	3294.90
807181	4091.86	3353.96	3040.64	2731.14	2330.17	2036.01	1754.65	1421.76
938011	2349.40	2157.03	2065.88	1967.90	1824.03	1699.31	1552.17	1282.36
810021	1651.24	1651.00	1650.88	1650.73	1649.97	1647.16	1635.70	1544.95
810011	2284.25	2095.31	2007.15	1913.40	1777.74	1662.16	1528.35	1290.69
102701	8695.83	7804.54	7369.92	6893.08	6173.60	5529.80	4744.47	3225.46
103605	7702.34	7490.04	7361.41	7199.22	6909.47	6600.94	6157.13	5071.23
807011	5871.73	5150.69	4826.36	4490.60	4022.78	3642.41	3224.99	2552.28

ANNEXE C

**Résultats des comparaisons des
trois relations X_1 -SUBV**

T = 1000

Paramètre	Estimation	Statistique	P
β_2	-0.490453	-0.505	0.6150
β_3	-0.663782	-0.722	0.4723
β_4	0.112292	0.402	0.6885
β_5	0.112098	0.496	0.6210

T = 200

Paramètre	Estimation	Statistique	P
β_2	-0.524336	-0.563	0.5746
β_3	-0.662965	-0.752	0.4536
β_4	0.122154	0.457	0.6489
β_5	0.112668	0.520	0.6039

T = 100

Paramètre	Estimation	Statistique	P
β_2	-0.544107	-0.597	0.5521
β_3	-0.662783	-0.768	0.4443
β_4	0.128068	0.489	0.6261
β_5	0.113259	0.534	0.5945

T = 50

Paramètre	Estimation	Statistique	P
β_2	-0.567157	-0.636	0.5263
β_3	-0.661586	-0.784	0.4350
β_4	0.135001	0.527	0.5995
β_5	0.113741	0.548	0.5846

T = 20

Paramètre	Estimation	Statistique	P
β_2	-0.606918	-0.703	0.4839
β_3	-0.660053	-0.808	0.4213
β_4	0.147284	0.594	0.5542
β_5	0.115089	0.573	0.5679

T = 10

Paramètre	Estimation	Statistique	P
β_2	-0.642969	-0.764	0.4466
β_3	-0.654442	-0.822	0.4131
β_4	0.158475	0.656	0.5136
β_5	0.115321	0.589	0.5569

T = 5

Paramètre	Estimation	Statistique	P
β_2	-0.687675	-0.841	0.4026
β_3	-0.642572	-0.830	0.4085
β_4	0.172636	0.735	0.4644
β_5	0.114756	0.603	0.5477

T = 2

Paramètre	Estimation	Statistique	P
β_2	-0.755365	-0.965	0.3369
β_3	-0.581023	-0.784	0.4347
β_4	0.194536	0.865	0.3891
β_5	0.103871	0.571	0.5695

ANNEXE D

**Biais brut et biais unitaire pour
T = 2, 5, 10, 20, 50, 100, 200 ans**

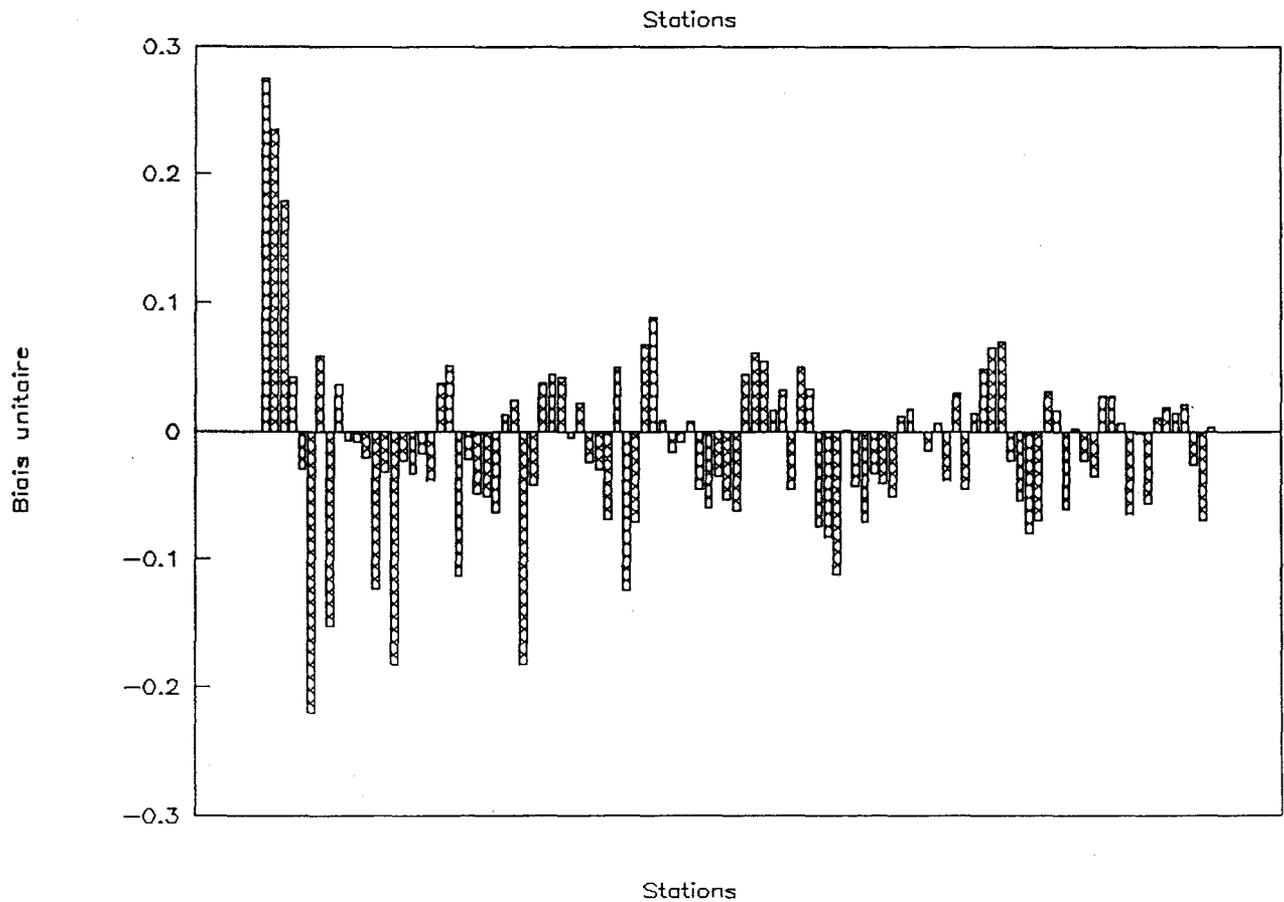
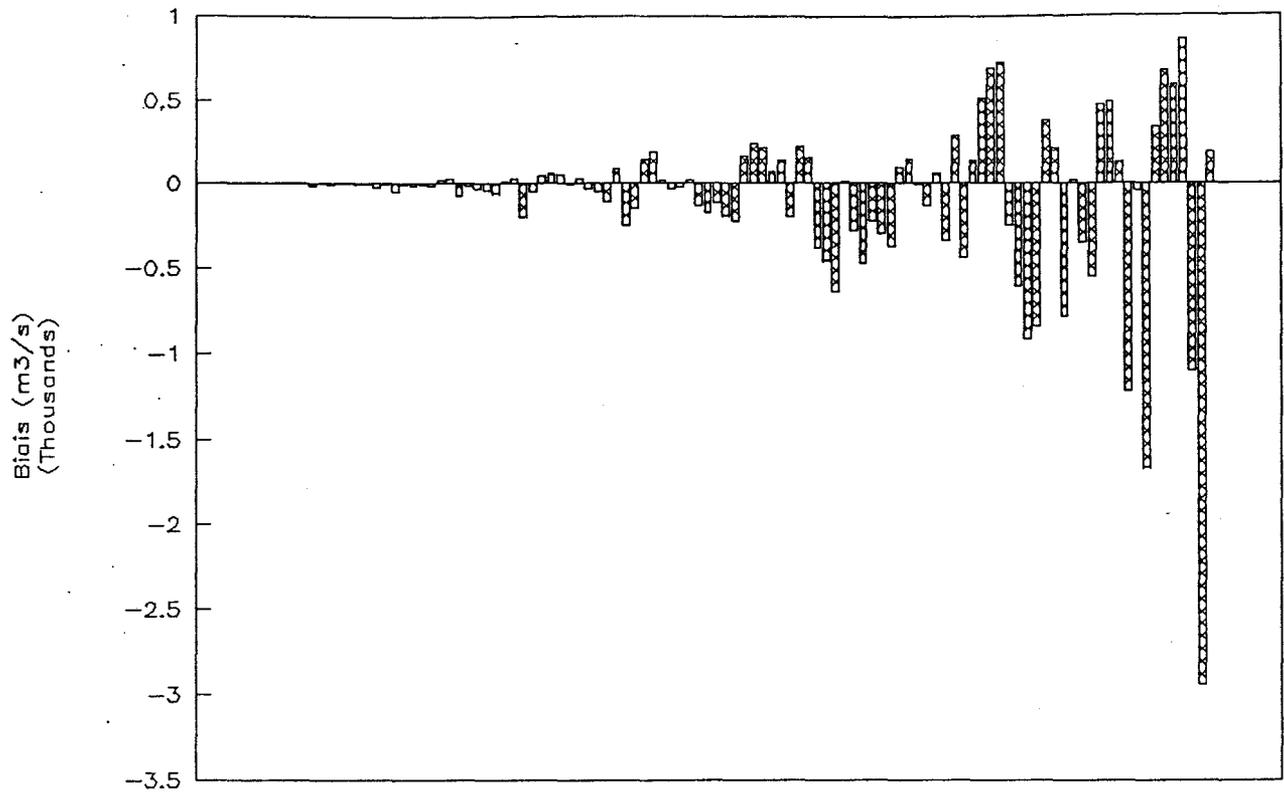
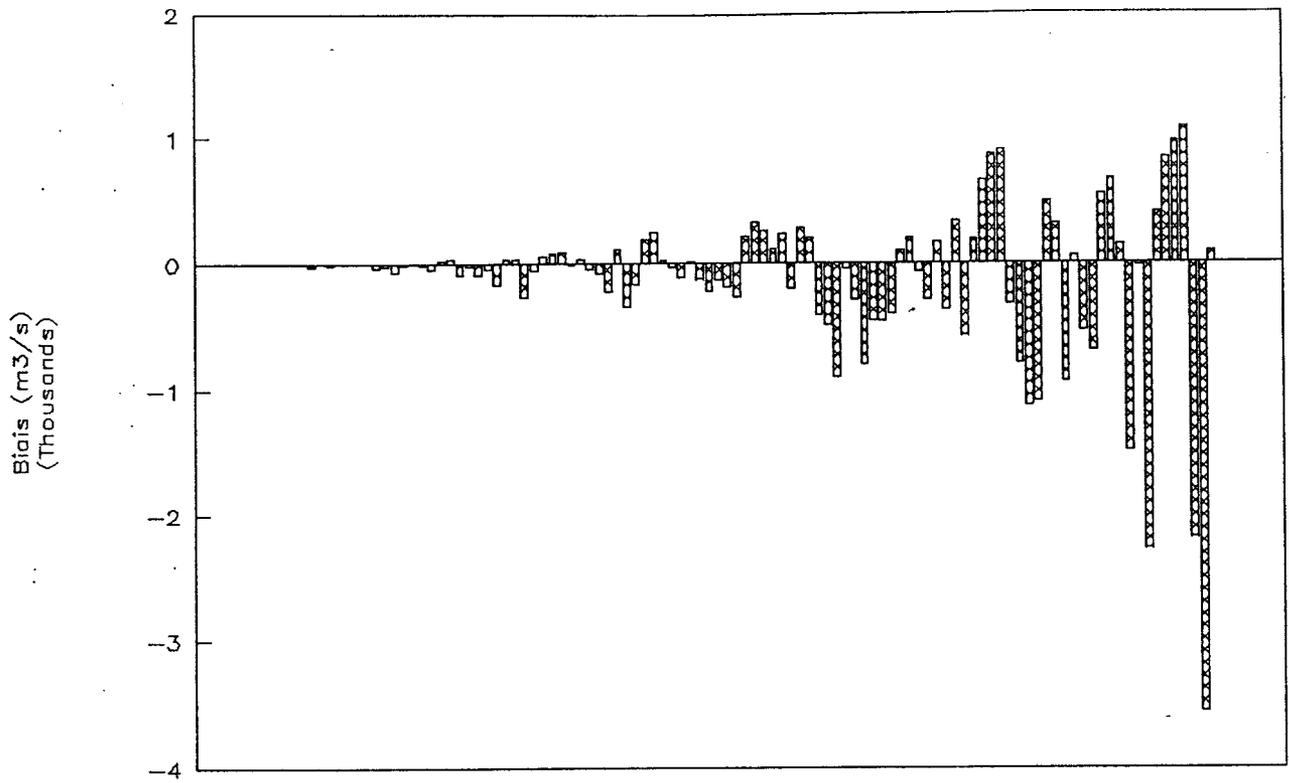
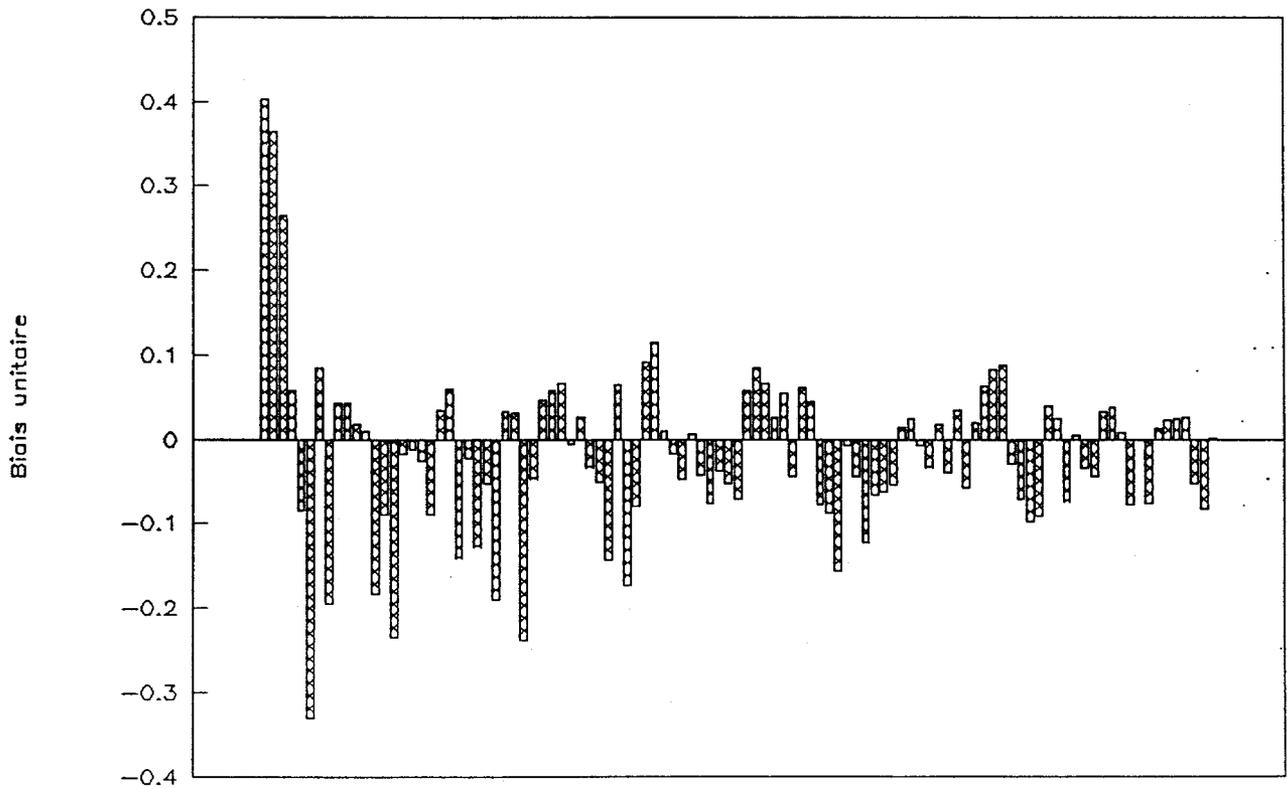


Figure D.1. Biais brut et biais unitaire, T = 2 ans



Stations



Stations

Figure D.2. Biais brut et biais unitaire, T = 5 ans

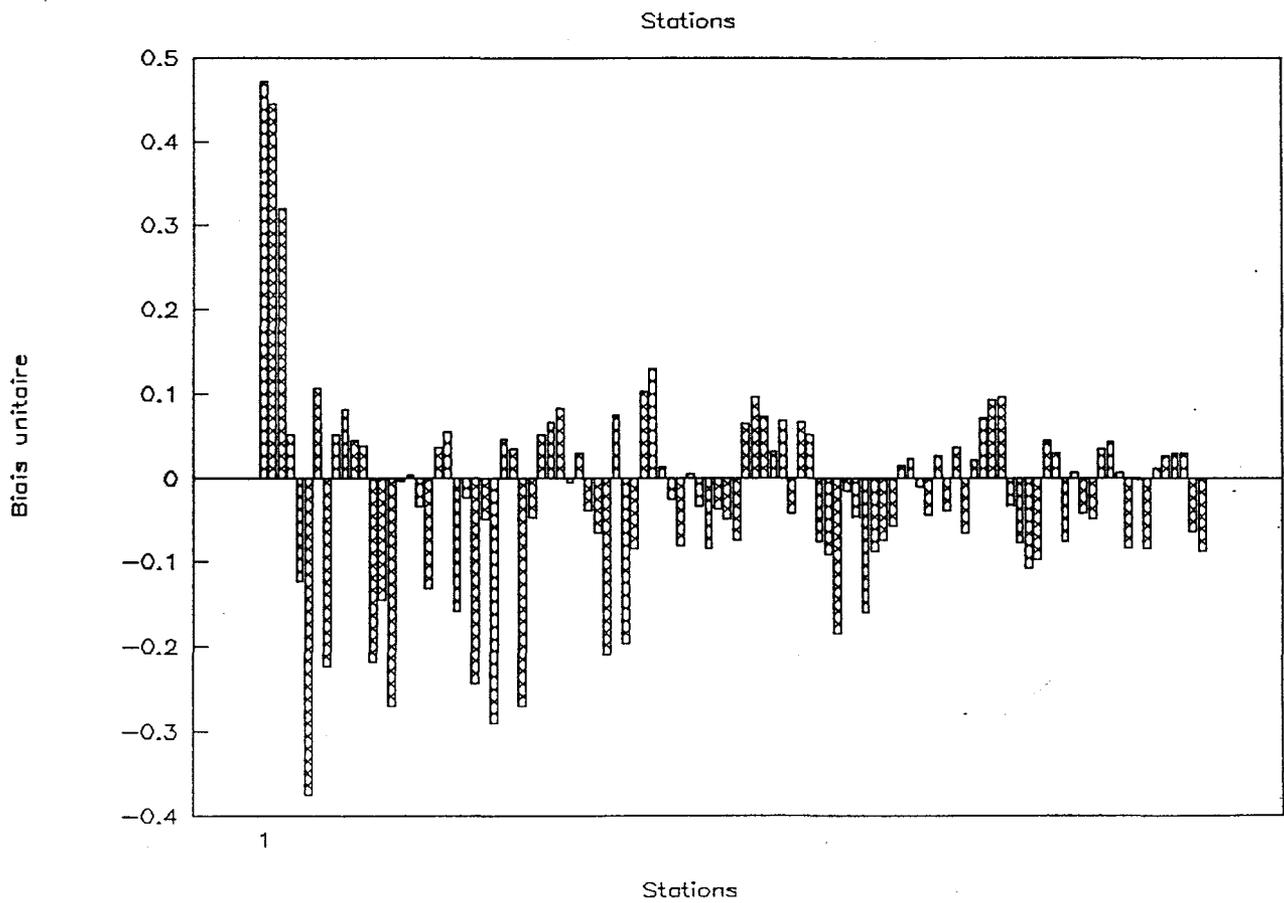
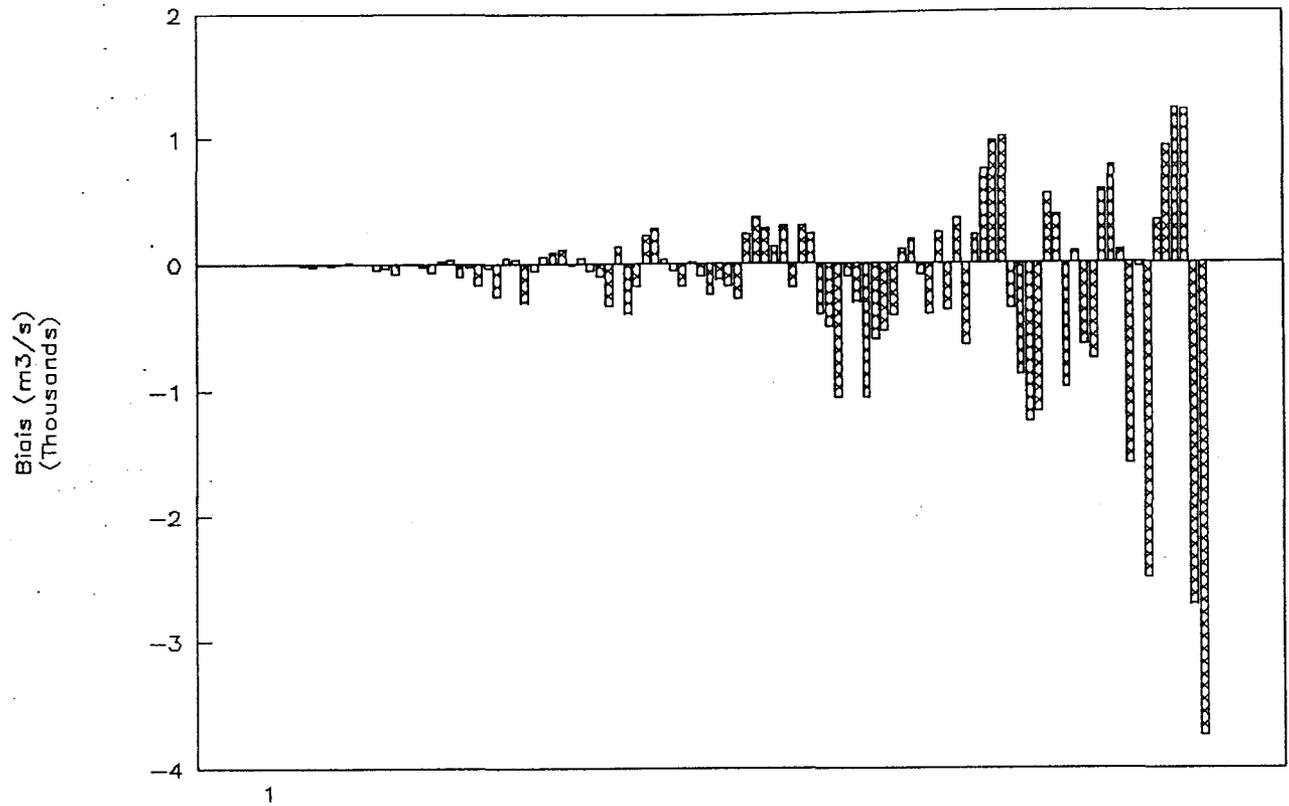
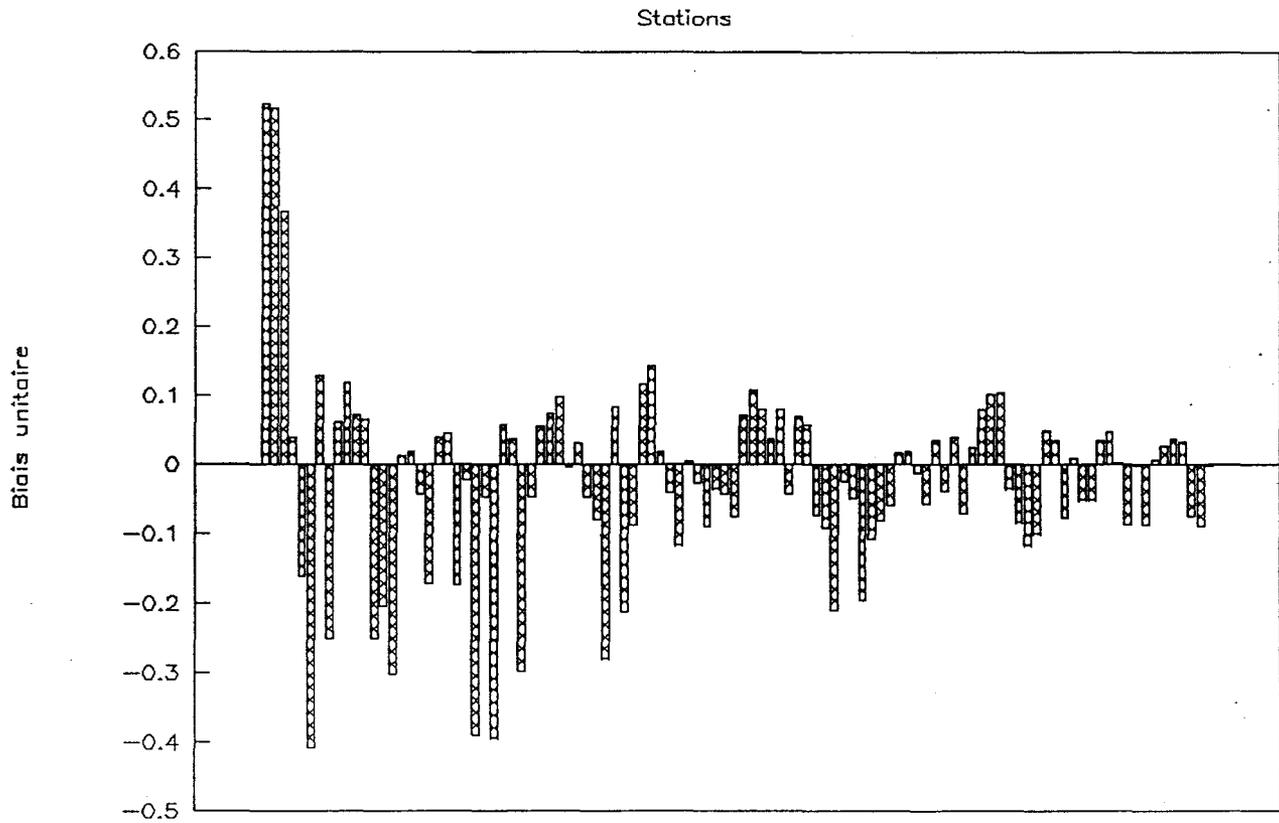
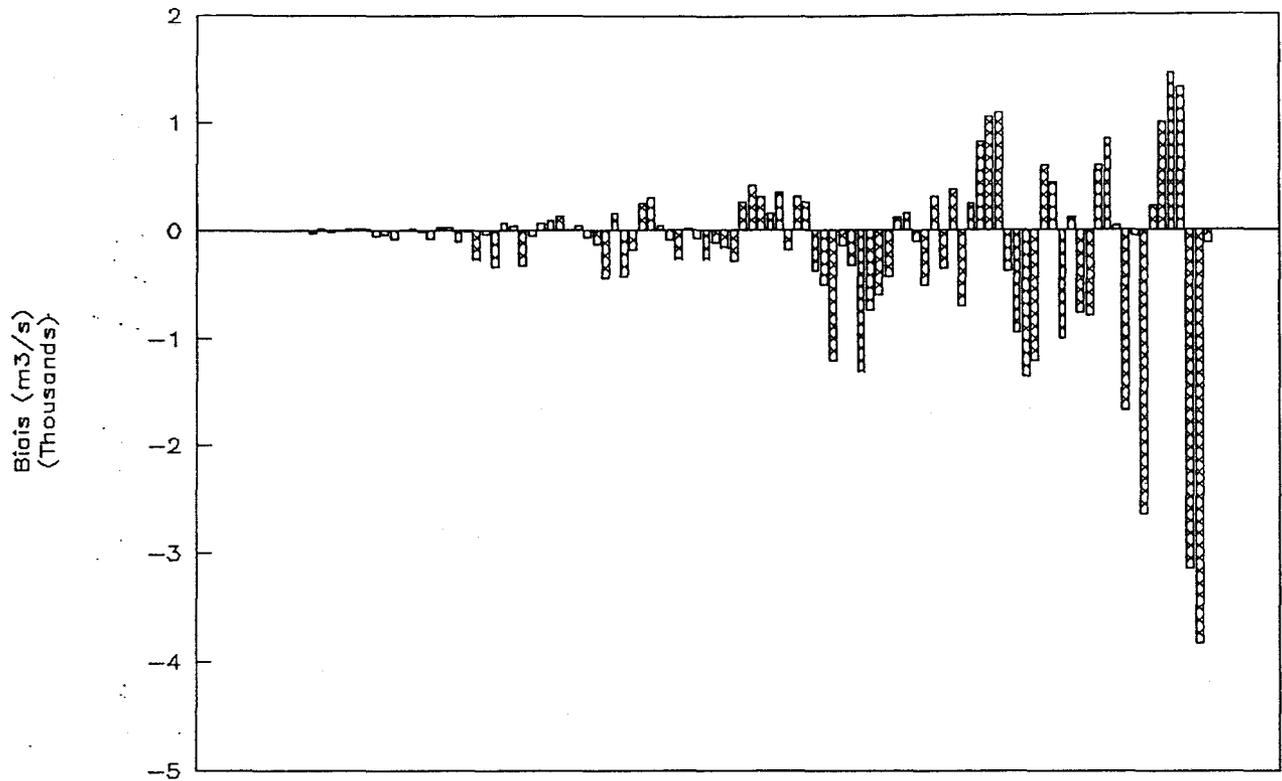
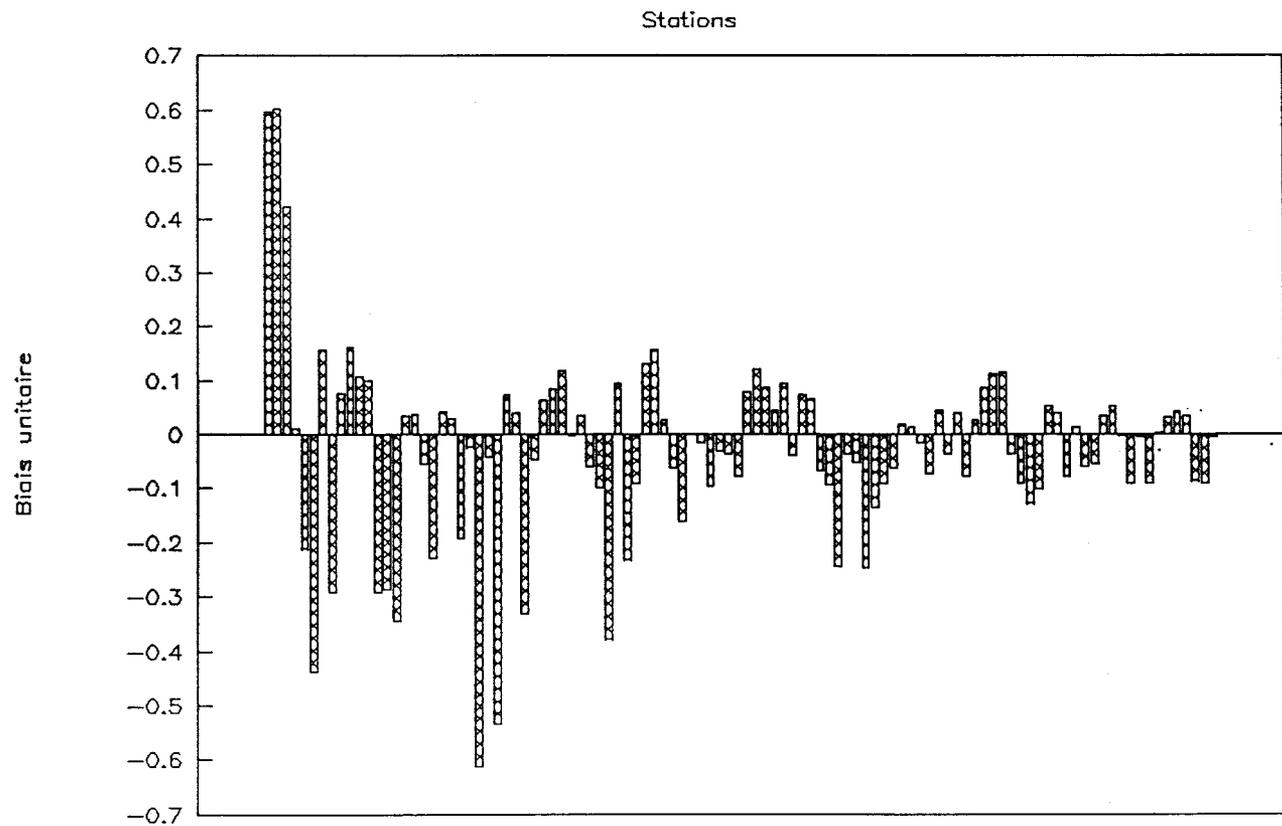
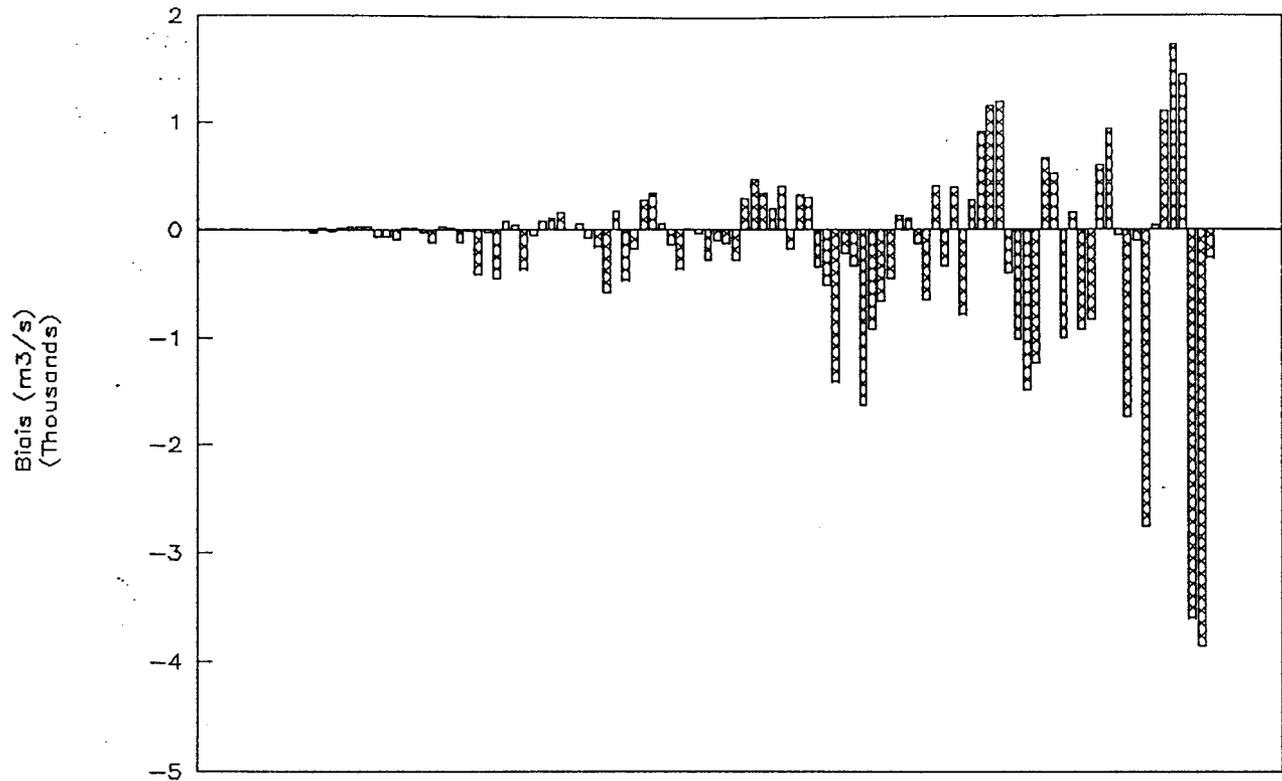


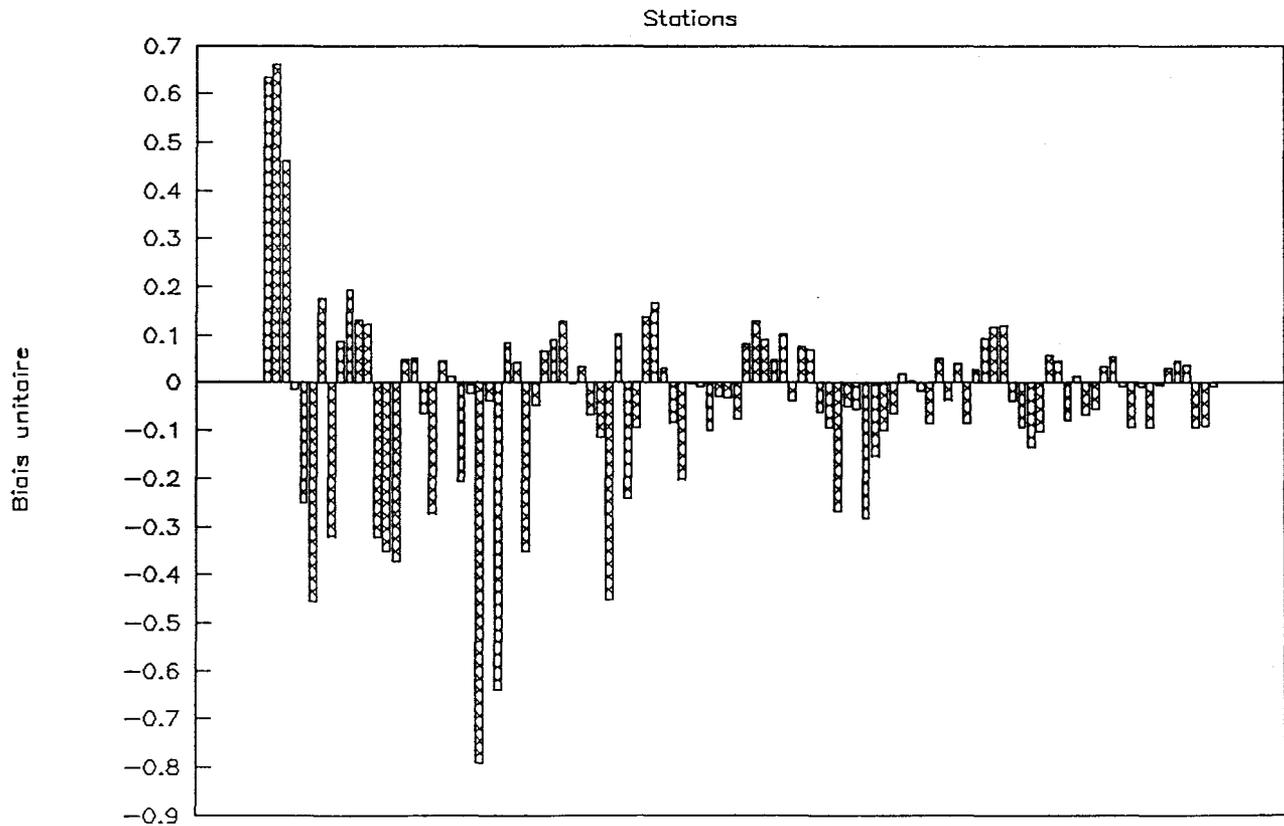
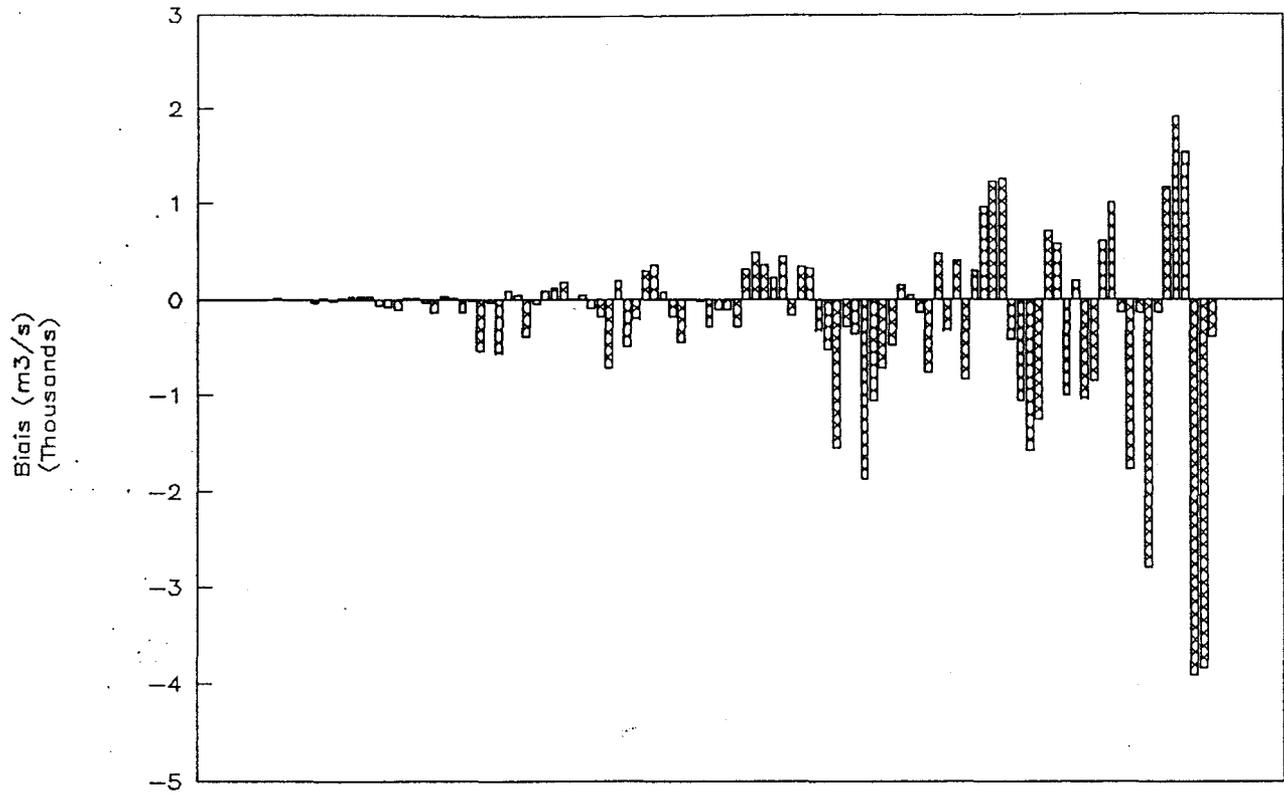
Figure D.3. Biais brut et biais unitaire, T = 10 ans



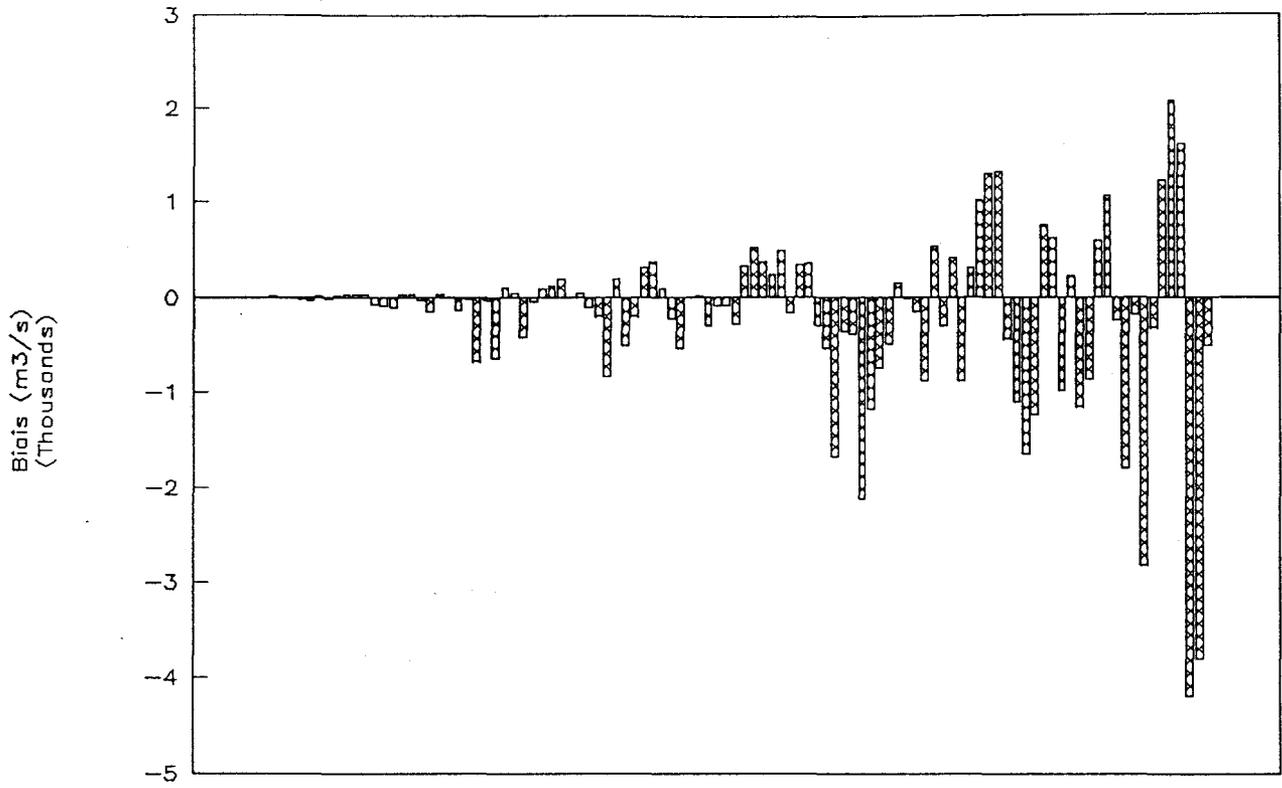
Stations
Figure D.4. Biais brut et biais unitaire, T = 20 ans



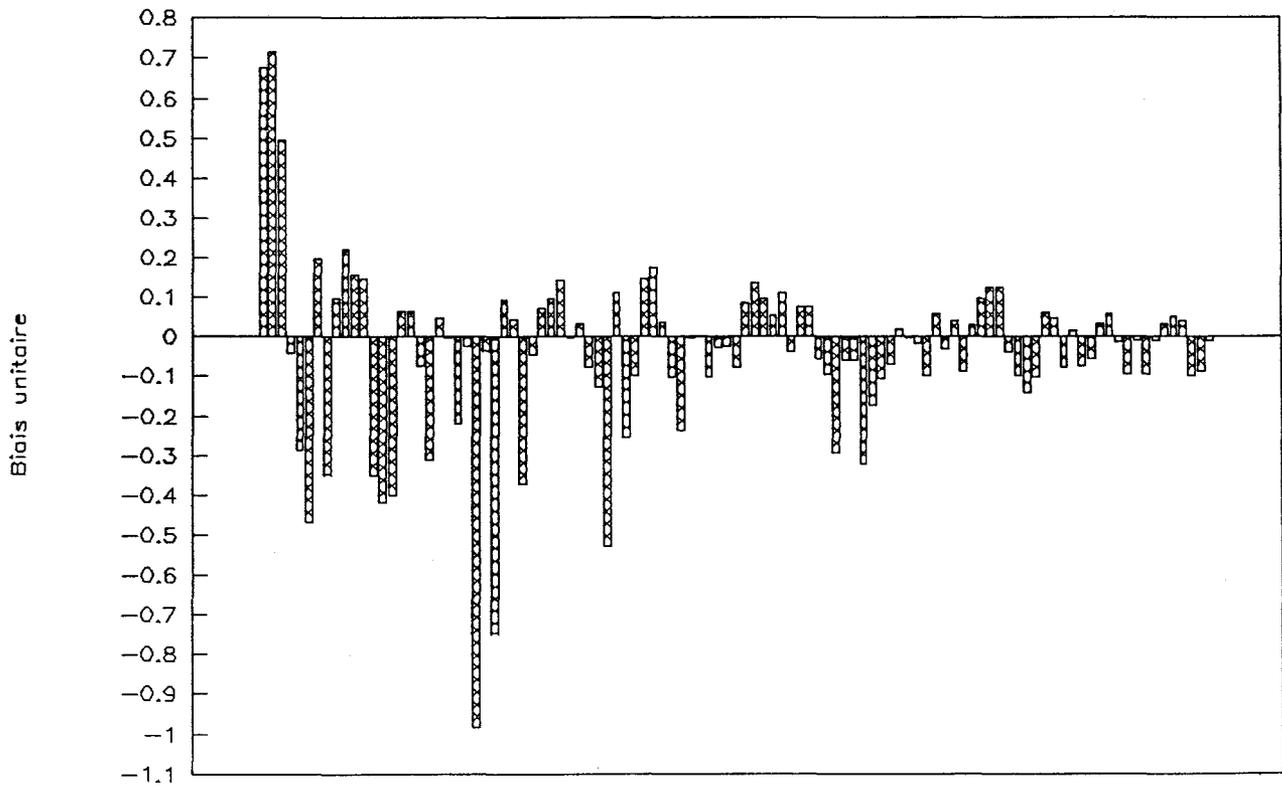
Stations
Figure D.5. Biais brut et biais unitaire, T = 50 ans



Stations
Figure D.6. Biais brut et biais unitaire, T = 100 ans



Stations



Stations

Figure D.7. Biais brut et biais unitaire, T = 200 ans