

GUIDE DE L'USAGER DU LOGICIEL *REGIONS*
(PHASE II)

Rapport de recherche No R-821

Octobre 2005

GUIDE DE L'USAGER DU LOGICIEL *REGIONS* (PHASE II)

Rapport préparé à l'attention de:

Monsieur Marc Barbet, ing. M.Sc.A.

Hydro-Québec

855 rue Ste-Catherine E. (19^{ème} étage)

Montréal (Québec), Canada, H2L 4P5

par:

Taha B.M.J. Ouarda

Véronique Jourdain

Ousmane Seidou

Hugo Gingras

Bernard Bobée

Chaire de recherche du Canada en estimation des variables hydrologiques

Institut National de la Recherche Scientifique, INRS-ETE

490, de la Couronne, Québec (Québec), G1K 9A9

Rapport de recherche No R-821

ISBN : 2-89146-312-9

Octobre 2005

ÉQUIPE DE RECHERCHE

Ont participé à la réalisation de cette étude:

Chaire en Hydrologie Statistique

Institut National de la Recherche Scientifique, INRS-ETE

Hugo Gingras

Taha B.M.J. Ouarda

Zeljka Ristic-Rudolf

Aziz Kouider

Mario Haché

Bernard Bobée

Hydro-Québec

Marc Barbet

Pierre Bruneau

TABLE DES MATIÈRES

TABLE DES MATIÈRES.....	iii
LISTE DES TABLEAUX.....	v
LISTE DES FIGURES.....	vii
1 INTRODUCTION.....	1
2 DESCRIPTION DU FONCTIONNEMENT DES DIFFÉRENTS MODULES DE REGIONS.....	3
2.1 MENU PRINCIPAL.....	3
2.2 MODULE DE CRÉATION D'UN NOUVEAU PROJET.....	3
2.3 MODULE DE STATISTIQUES DESCRIPTIVES.....	7
2.4 MODULE DE CLASSIFICATION HIÉRARCHIQUE.....	12
2.4.1 Description générale du module.....	12
2.4.2 Choix des paramètres de la classification.....	13
2.4.3 Obtention et réarrangement des classes.....	17
2.5 MODULE DE DÉFINITION MANUELLE DES CLASSES.....	19
2.6 MODULE D'ANALYSE CANONIQUE DES CORRÉLATIONS (APPROCHE PAR OPTIMISATION DU PARAMÈTRE ALPHA).....	21
2.6.1 Description générale du module.....	21
2.6.2 Choix des paramètres de l'analyse des corrélations canoniques.....	22
2.6.3 Obtention des voisinages et optimisation du paramètre "alpha".....	25
2.7 MODULE D'ANALYSE CANONIQUE DES CORRÉLATIONS (APPROCHE PAR CLASSIFICATION).....	29
2.7.1 Description générale du module.....	29
2.7.2 Choix des paramètres de l'analyse des corrélations canoniques.....	30
2.7.3 Obtention des voisinages.....	32
2.8 MODULE DE RÉGRESSION MULTIPLE.....	34
2.8.1 Choix du type de régions homogènes.....	34
2.8.2 Sélection du modèle régressif.....	36
2.8.3 Description des résultats de la régionalisation.....	38
2.9 MODULE DE L'INDICE DE CRUE.....	41
2.9.1 Choix du type de régions homogènes.....	42

2.9.2 <i>Sélection du modèle régressif</i>	43
2.9.3 <i>Description des résultats de la régionalisation</i>	45
2.10 MODULE DE COMBINAISON BAYÉSIENNE DE L'INFORMATION RÉGIONALE ET LOCALE.....	48
2.11 FENÊTRE DE COMMENTAIRES.....	55
RÉFÉRENCES	57

LISTE DES TABLEAUX

Tableau 2.1 Description des différents types de distances	16
Tableau 2.2 Description des différents algorithmes de regroupement	16



LISTE DES FIGURES

Figure 2.1	Fenêtre du menu principal	4
Figure 2.2	Fenêtre de création d'un nouveau projet	5
Figure 2.3	Exemple d'un fichier de données	6
Figure 2.4	Fenêtre de sélection du répertoire de projet.....	7
Figure 2.5	Fenêtre de statistiques descriptives.....	8
Figure 2.6	Fenêtre de visualisation des données.....	10
Figure 2.7	Fenêtre d'affichage des caractéristiques de base.....	10
Figure 2.8	Fenêtre d'affichage de la matrice de corrélation	11
Figure 2.9	Fenêtre de validation graphique de l'hypothèse de normalité.....	12
Figure 2.10	Fenêtre de classification hiérarchique	14
Figure 2.11	Fenêtre de sélection des paramètres de la classification.....	15
Figure 2.12	Fenêtre de représentation de l'arbre de classification	17
Figure 2.13	Fenêtre de définition manuelle des classes.....	19
Figure 2.14	Fenêtre d'analyse canonique des corrélations	23
Figure 2.15	Fenêtre de position des stations dans l'espace canonique physiographique	25
Figure 2.16	Fenêtre de position des stations dans l'espace canonique hydrologique.....	26
Figure 2.17	Fenêtre de représentation graphique des indices de performance en fonction du paramètre "alpha"	27

Figure 2.18	Fenêtre d'analyse canonique des corrélations (approche par classification) ...	31
Figure 2.19	Fenêtre de position des stations dans l'espace canonique physiographique....	32
Figure 2.20	Fenêtre de position des stations dans l'espace canonique hydrologique	33
Figure 2.21	Fenêtre de spécification du type de régions homogènes à considérer au moment de l'exécution du module de régression multiple	35
Figure 2.22	Fenêtre de régression multiple (option voisinages – ACC/approche par classification)	37
Figure 2.23	Fenêtre de régression multiple (options de validation du modèle régressif)...	39
Figure 2.24	Graphiques de validation du modèle régressif.....	40
Figure 2.25	Fenêtre de spécification du type de régions homogènes à considérer au moment de l'exécution du module de l'indice de crue	43
Figure 2.26	Fenêtre de l'indice de crue (option voisinages – ACC/approche par classification)	44
Figure 2.27	Fenêtre de répartition des coefficients régionaux (module d'indice de crue)..	47
Figure 2.28	Spécification des trois quantiles et de leurs probabilités respectives.....	49
Figure 2.29	Affichage des résultats de l'analyse régionale.....	50
Figure 2.30	Méthode bayésienne empirique de combinaison de l'information locale et régionale	51
Figure 2.31	Combinaison de l'information locale et régionale par MCMC.....	52
Figure 2.32	Histogrammes des paramètres et des quantiles.....	53
Figure 2.33	Fenêtre de commentaires.....	55

1 INTRODUCTION

Le logiciel *REGIONS* a été développé dans le but d'automatiser les opérations à effectuer au cours d'une procédure de régionalisation des événements hydrologiques extrêmes (e.g. débits et volumes de crues). Essentiellement, une telle procédure se divise en deux grandes étapes, soit:

- a) la définition et la détermination de régions hydrologiquement homogènes, i.e. le regroupement des stations ayant un comportement hydrologique semblable ;
- b) l'estimation régionale, i.e. le transfert (à l'intérieur d'une même région homogène) de l'information disponible aux sites jaugés à un site non jaugé ou encore partiellement jaugé.

L'équipe de la Chaire Hydro-Québec/CRSNG/Alcan en Hydrologie statistique, établie à l'INRS-ETE, a donc été sollicitée par Hydro-Québec afin de développer le logiciel *REGIONS* dans le but de répondre aux besoins de régionalisation de l'entreprise, notamment en ce qui touche la question du dimensionnement des ouvrages hydrauliques. Cette demande constitue en fait une suite logique à l'adoption par Hydro-Québec de la méthodologie mise au point par la Chaire (suite aux travaux effectués par l'INRS-Eau dans le cadre d'un projet CRSNG Stratégique au cours de la période 1991-94, puis dans le cadre du projet C5 de la Chaire) et faisant intervenir la méthode d'analyse des corrélations canoniques pour la détermination des voisinages hydrologiques ainsi que la régression multiple pour l'estimation régionale.

En effet, le but premier poursuivi avec *REGIONS* était d'implanter cette approche à l'intérieur d'une application présentant à l'utilisateur une interface conviviale et flexible lui permettant de passer aisément au travers des étapes requises par la procédure de régionalisation. Le logiciel a été créé à l'aide de l'environnement de travail *MATLAB* 6.1, qui grâce à ses nombreuses fonctions mathématiques et statistiques de même qu'à ses capacités de développement d'interfaces graphiques, se trouve à offrir un cadre efficace à la

réalisation d'un tel projet. L'utilisateur doit donc posséder la version 6 de *MATLAB* pour être en mesure d'exécuter *REGIONS*, ce qui ne présente toutefois pas de problème pour Hydro-Québec puisque *MATLAB* y est déjà bien implanté.

Le présent rapport propose donc un survol des différentes fonctionnalités offertes par *REGIONS* tout en décrivant les diverses options qui se présentent à l'utilisateur au moyen des boutons et menus définis pour chaque module. Sa structure est donc similaire à celle du précédent rapport (Gingras *et al.*, 2002) couvrant les travaux réalisés dans la première phase du projet. Ce nouveau rapport reprend donc pour l'essentiel le contenu de l'ancien rapport en plus de décrire les derniers ajouts et développements réalisés dans la phase II du projet.

2 DESCRIPTION DU FONCTIONNEMENT DES DIFFÉRENTS MODULES DE *REGIONS*

2.1 Menu principal

Le menu principal du logiciel *REGIONS* est présenté à la Figure 2.1. Les noms des boutons qui le composent correspondent aux modules qu'ils permettent d'exécuter après les avoir appuyés. On y retrouve également un menu défilant intitulé "Projet" au haut de la fenêtre. Au moment de l'ouverture de *REGIONS*, les boutons du menu principal demeurent inactifs jusqu'à ce que l'utilisateur se serve du menu "Projet" afin de créer un nouveau projet ou encore pour ouvrir un projet déjà existant (options "Nouveau" et "Ouvrir" respectivement). Si les régions homogènes n'ont pas été définies pour le projet en cours, les boutons de l'encadré "Méthodes d'Estimation Régionale" restent inactifs. Le bouton "Quitter" permet de mettre fin à l'exécution du logiciel tout en sauvegardant les modifications apportées au projet (l'option "Quitter" du menu "Projet" peut aussi être utilisée de façon équivalente).

2.2 Module de création d'un nouveau projet

Pour amorcer la création d'un nouveau projet, il faut sélectionner l'option "Nouveau" du menu "Projet" dans le menu principal de *REGIONS*. Cette action fera apparaître la fenêtre de création d'un nouveau projet, telle qu'illustrée à la Figure 2.2. Ce module permet d'importer les données nécessaires (i.e. noms/numéros des stations + noms, unités et valeurs des variables hydrologiques et physiographiques/météorologiques) pour appliquer la procédure de régionalisation. Pour l'instant, celles-ci doivent être rassemblées dans un fichier texte où les deux premières lignes sont réservées pour l'identification des variables et des unités, tandis que pour les lignes subséquentes, on réserve la première colonne pour les noms/numéros des stations étudiées (sites jaugés et non jaugés rassemblés) et les autres colonnes pour les valeurs des différentes variables. Un exemple d'un tel fichier est présenté à la Figure 2.3. Les colonnes n'ont pas besoin d'être alignées mais doivent être séparées par

au moins un espace (i.e. que chaque ligne doit contenir le même nombre d'éléments, les espaces servant de délimiteur entre ceux-ci).

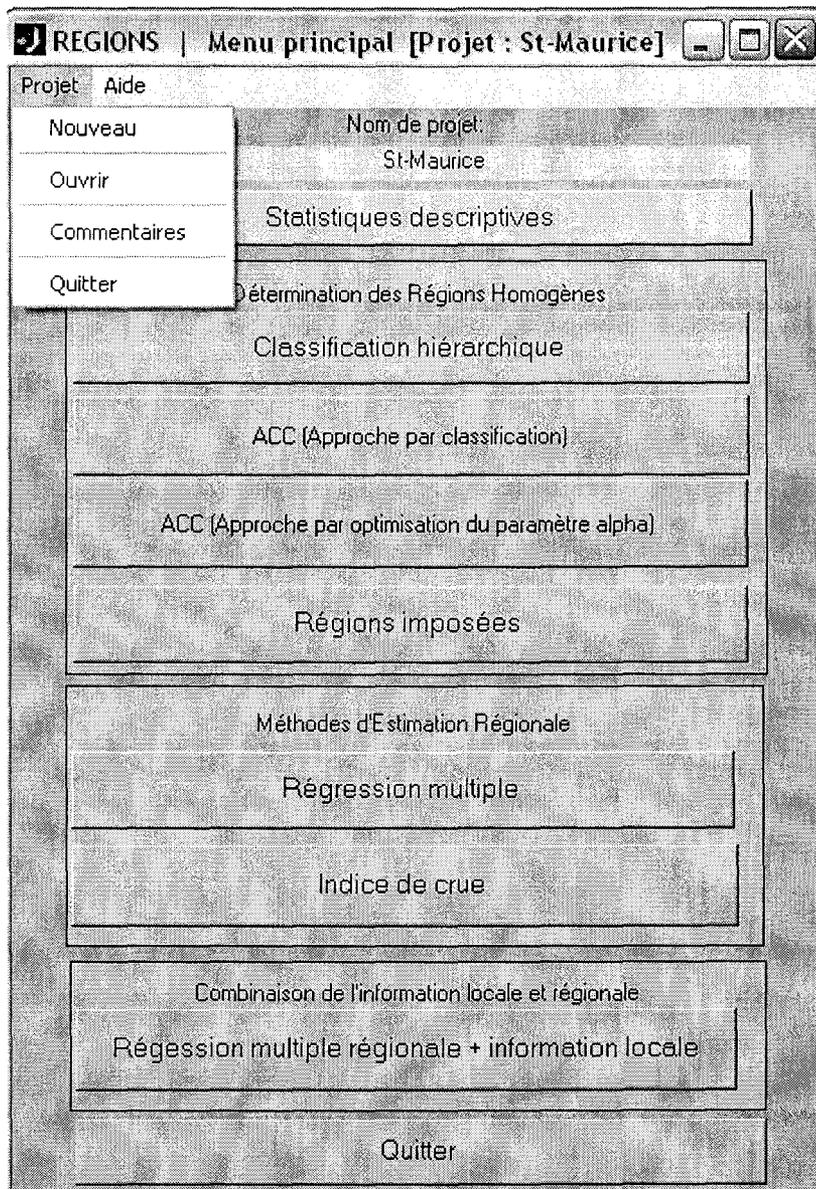


Figure 2.1 Fenêtre du menu principal

La première étape dans la création d'un nouveau projet consiste donc à importer le fichier de données préalablement préparé. Le bouton "Parcourir" situé dans l'encadré au haut à gauche de la fenêtre de création d'un nouveau projet (voir Figure 2.2) permet de sélectionner le fichier en question. On peut ensuite importer les données qu'il contient en

appuyant sur le bouton "Lire les données", situé dans le même encadré. Il est à noter que ces deux opérations (i.e sélection + lecture du fichier) peuvent être condensées en une seule en utilisant l'option "Ouvrir" du menu "Fichier". On peut s'assurer que les données ont été lues correctement à l'aide du bouton "Visualisation des données".

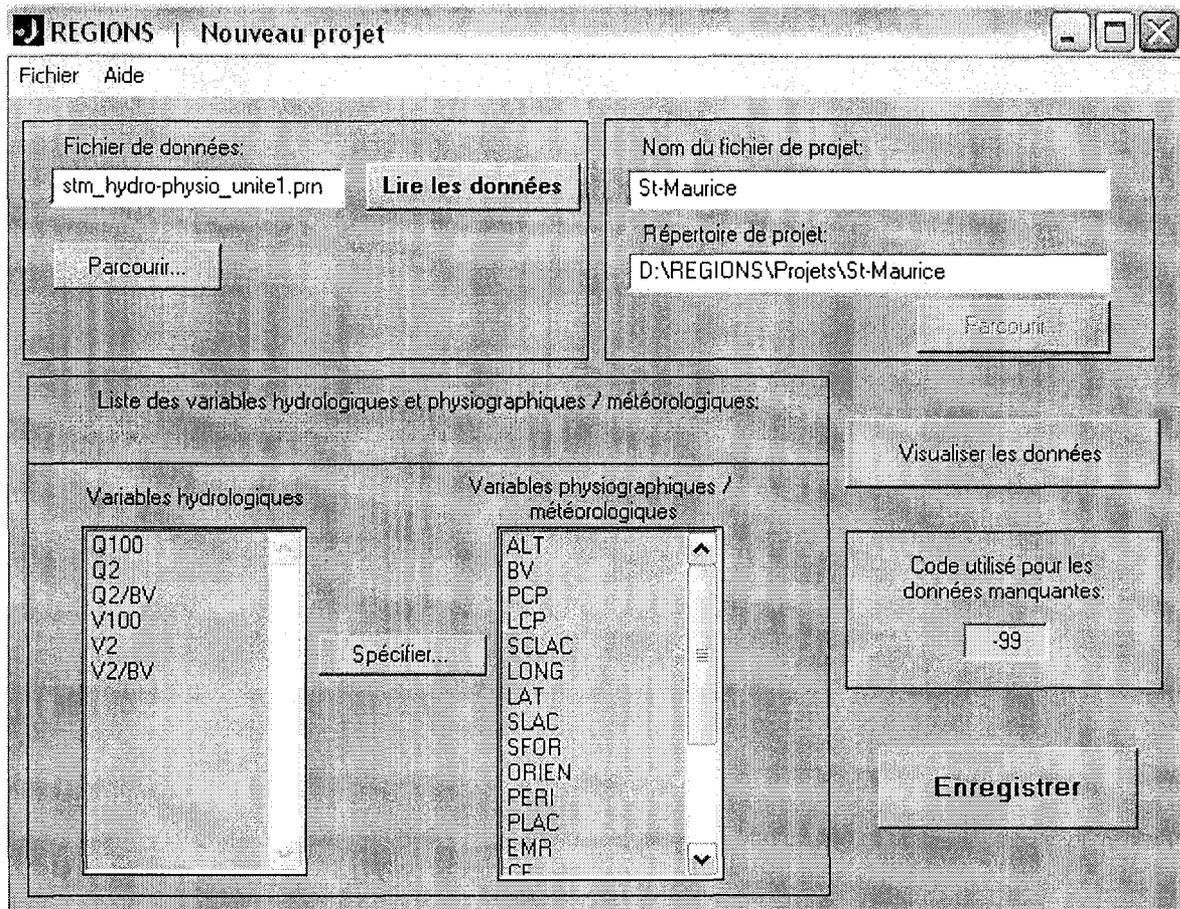


Figure 2.2 Fenêtre de création d'un nouveau projet

Dans l'encadré situé au haut à droite de la fenêtre, on demande d'entrer un nom pour le nouveau fichier de projet de même que le répertoire dans lequel on souhaite l'enregistrer. Il est important de bien spécifier le chemin complet du répertoire. On peut aussi le spécifier en utilisant le bouton "Parcourir" situé sous la boîte d'édition du répertoire, ce qui a pour effet de lancer la fenêtre de sélection du répertoire de projet (telle qu'illustré à la Figure 2.4) qui permet de naviguer entre les divers répertoires définis sur la machine de l'utilisateur (il n'y a qu'à suivre les instructions déjà inscrites sur la fenêtre).

La prochaine étape consiste à préciser quelles sont les variables hydrologiques présentes dans le jeu de données (ce qui est nécessaire pour que *REGIONS* puisse faire la distinction entre les variables hydro et les variables physio/météo) à l'aide du bouton "Spécifier", situé dans l'encadré intitulé "Liste des variables hydrologiques et physiographiques / météorologiques". Les variables ainsi discriminées sont alors reportées dans les deux boîtes d'affichage situées de chaque côté du bouton "Spécifier". Si le code utilisé pour indiquer les données manquantes dans le fichier de données (e.g. les valeurs des variables hydrologiques dans le cas des sites non jaugés) est différent de -99 (valeur par défaut), il faut alors entrer le nouveau code dans l'encadré prévu à cet effet.

	Q1.00 m ³ /s	Q2 m ³ /s	Q2/BV m ³ /s/km ²	V100 m ³	V2 m ³	V2/BV m ³ /km ²	ALT m	BV km ²	PCP m/km
040401	580.90	272.20	0.1056	1357.00	719.00	0.2790	229	2577	2.42
040202	804.40	518.10	0.0947	2433.30	1418.20	0.2593	122	5470	1.83
040203	151.00	65.50	0.0504	459.50	253.00	0.1946	183	1300	2.07
040405	248.40	123.90	0.0623	817.90	431.70	0.2169	137	1990	1.85
040406	138.90	95.50	0.0729	514.90	337.10	0.2573	162	1310	2.16
040814	185.60	96.00	0.0733	497.00	259.20	0.1979	168	1310	1.87
041903	182.80	121.40	0.0575	630.10	447.20	0.2119	320	2110	0.53
043008	86.50	38.80	0.0388	349.40	168.50	0.1685	290	1000	0.24
043012	500.30	256.20	0.0989	968.70	577.70	0.2231	274	2590	0.26
050117	295.20	149.10	0.0898	610.00	352.90	0.2126	289	1660	0.22
050119	345.10	216.60	0.1165	922.90	491.70	0.2644	168	1860	2.07
050125	216.40	108.30	0.1031	488.00	245.50	0.2338	122	1050	3.98
050126	257.80	135.30	0.0973	615.30	347.30	0.2499	366	1390	2.30
050135	516.00	223.80	0.1435	807.90	440.30	0.2822	183	1560	2.66
050301	453.20	207.90	0.0790	1236.00	570.40	0.2169	243	2630	1.60
050304	891.20	551.80	0.1201	2201.10	1344.70	0.2926	15	4596	2.37
050409	311.00	168.40	0.2656	415.40	274.10	0.4323	162	634	6.57
050701	95.60	69.40	0.1891	206.90	117.40	0.3199	27	367	4.81
052201	265.50	86.80	0.1581	347.00	171.50	0.3124	244	549	3.71
052203	307.40	213.10	0.1678	715.20	394.60	0.3107	152	1270	4.47
052202	355.10	178.00	0.1328	690.80	368.40	0.2749	53	1340	2.92
052601	199.90	118.30	0.1149	499.60	300.00	0.2913	76	1030	4.21
052801	256.00	151.40	0.1059	696.10	401.40	0.2807	15	1430	3.12
061901	2650.60	1398.60	0.0914	6633.70	4088.50	0.2672	111	15300	1.36
061905	1707.10	1056.10	0.0935	5256.50	2971.80	0.2630	305	11300	1.38
062101	1559.10	1065.50	0.1226	4252.30	2766.40	0.3183	116	8690	1.62
062102	2020.90	1241.70	0.1291	4554.30	2837.60	0.2950	117	9620	2.04
080101	309.80	180.90	0.0492	1265.30	765.80	0.2081	305	3680	0.36
080705	1099.10	527.50	0.0624	3579.80	1942.30	0.2296	305	8460	0.38
080707	2522.70	1333.00	0.0604	8535.90	5061.80	0.2295	250	22060	0.32
cible	-99	-99	-99	-99	-99	-99	185	4000	2.19

Figure 2.3 Exemple d'un fichier de données

Une fois que toutes les informations ont été entrées, il ne reste plus qu'à appuyer sur le bouton "Enregistrer" afin de créer le nouveau fichier de projet. Si, pour une raison quelconque, l'utilisateur veut interrompre le processus de création du fichier de projet, il peut utiliser l'option "Quitter" du menu "Fichier" ; un message apparaîtra alors demandant à l'utilisateur de bien confirmer ses intentions.

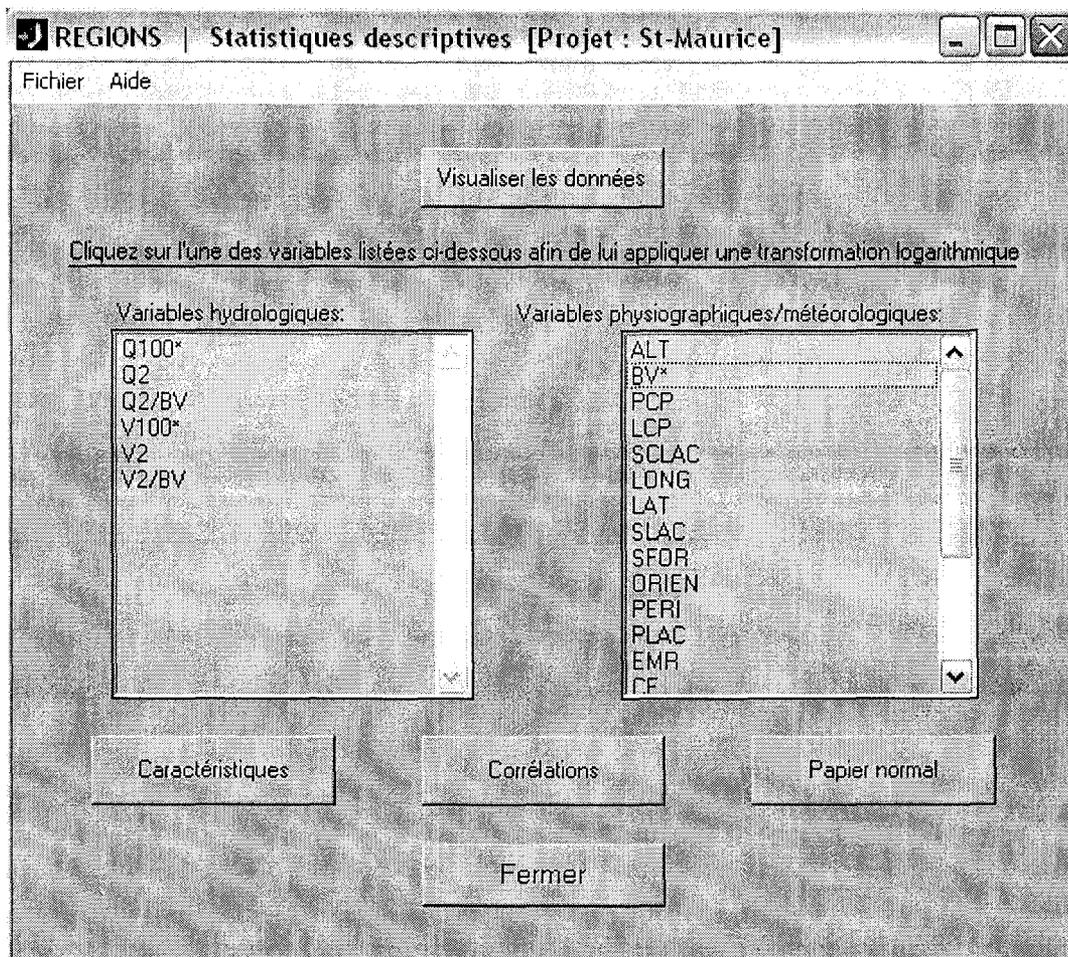


Figure 2.5 Fenêtre de statistiques descriptives

Voici maintenant la liste des boutons contenus dans la fenêtre avec une courte description de l'action à laquelle ils correspondent:

- "Visualiser les données" : ce bouton ouvre une fenêtre (voir Figure 2.6) affichant le contenu de la base de données associée au projet en cours (i.e. noms/numéros des stations ainsi que les valeurs des variables hydro et physio/météo considérées)
- "Caractéristiques" : ce bouton ouvre une fenêtre (voir Figure 2.7) donnant les caractéristiques de base (i.e. la moyenne, l'écart-type, le coefficient de variation, la médiane, le minimum, le maximum, le coefficient d'asymétrie et le coefficient

d'aplatissement ou "kurtosis") pour un certain nombre de variables préalablement sélectionnées par l'utilisateur

- "Corrélations" : ce bouton ouvre une fenêtre (voir Figure 2.8) affichant la matrice de corrélation pour les variables préalablement sélectionnées par l'utilisateur. Cette fenêtre est munie de "checkbox" contrôlant l'affichage des corrélations, des corrélations au carré, des probabilités au dépassement de même que des concomitances (i.e. le nombre d'observations concordantes pour chaque paire de variables)
- "Papier normal" : ce bouton permet de valider graphiquement l'hypothèse de normalité pour l'une ou l'autre des variables à l'étude (voir exemple sur la Figure 2.9). On pourra appliquer au besoin une transformation logarithmique sur les variables plus problématiques afin de rectifier la situation (on sait par exemple que les débits s'ajustent généralement mieux à la normalité en coordonnées logarithmiques). L'utilité de cette fonction découle de l'hypothèse de multinormalité des variables canoniques sur laquelle est basée la technique d'Analyse des Corrélations Canoniques (cf. section 2.6).
- "Fermer" : ce bouton permet tout simplement de quitter le module de statistiques descriptives tout en rendant le contrôle au menu principal

REGIONS | Visualisation des données [Projet : St-Maurice]

Fichier Aide

#	Station	Q100 m ³ /s	Q2 m ³ /s	Q2/BV m ³ /s/km ²	V100 m ³	V2 m ³	V2/BV m ³ /km ²	ALT m	BV km ²
1	040401	580.9	272.2	0.1056	1357	719	0.279	229	2577
2	040202	804.4	518.1	0.0947	2433.9	1418.2	0.2593	122	5470
3	040203	151	65.5	0.0504	459.5	253	0.1946	183	1300
4	040405	248.4	123.9	0.0623	817.9	431.7	0.2169	137	1990
5	040406	198.9	95.5	0.0729	514.9	337.1	0.2573	162	1310
6	040814	185.6	96	0.0733	497	259.2	0.1979	158	1310
7	041903	162.8	121.4	0.0575	630.1	447.2	0.2119	320	2110
8	043008	86.5	38.8	0.0388	349.4	168.5	0.1685	290	1000
9	043012	500.3	256.2	0.0989	968.7	577.7	0.2231	274	2590
10	050117	295.2	149.1	0.0898	610	352.9	0.2126	289	1660
11	050119	345.1	216.6	0.1165	922.9	491.2	0.2644	168	1860
12	050125	216.4	108.3	0.1031	488	245.5	0.2338	122	1050
13	050126	257.8	135.3	0.0973	615.3	347.3	0.2499	366	1390
14	050135	516	223.8	0.1435	807.9	440.3	0.2822	183	1560
15	050301	453.2	207.9	0.079	1236	570.4	0.2169	243	2630
16	050304	891.2	551.8	0.1201	2201.1	1344.7	0.2926	15	4596
17	050409	311	168.4	0.2656	415.4	274.1	0.4323	162	634
18	050701	95.6	69.4	0.1891	206.9	117.4	0.3199	27	387
19	052201	265.5	86.9	0.1581	347	171.5	0.3124	244	549
20	052203	307.4	243.1	0.1678	715.2	394.6	0.3107	152	1270
21	052202	355.1	178	0.1328	690.8	368.4	0.2749	53	1340
22	052601	199.9	118.3	0.1149	499.6	300	0.2913	76	1030
23	052801	256	151.4	0.1059	696.1	401.4	0.2807	15	1430
24	061901	2650.6	1398.6	0.0914	6633.7	4088.5	0.2672	111	15300
25	061905	1707.1	1056.1	0.0935	5256.5	2971.8	0.263	305	11300
26	062101	1559.1	1065.5	0.1226	4252.3	2766.4	0.3183	116	8690
27	062102	2020.9	1241.7	0.1291	4554.3	2832.6	0.295	117	9620
28	080101	309.8	180.9	0.0492	1265.3	765.8	0.2081	305	3680
29	080705	1099.1	527.5	0.0524	3579.8	1942.3	0.2296	305	8460
30	080707	2522.7	1333	0.0604	8535.9	5061.8	0.2295	250	22060

Fermer

Figure 2.6 Fenêtre de visualisation des données

REGIONS | Caractéristiques de base des variables [Projet : St-Maurice]

Fichier Aide

Variables	Moyenne	Ecart-type	Coeff. var.	Médiane	Minimum	Maximum	Asymétrie
Ln(Q100)	6.01	0.94	0.16	5.74	4.46	7.88	0.56
Q2	365.64	412.72	1.13	179.45	38.80	1398.60	1.62
Q2/BV	0.10	0.05	0.45	0.10	0.04	0.27	1.46
Ln(V100)	6.95	0.98	0.14	6.63	5.33	9.05	0.73
V2	1028.87	1266.73	1.23	436.00	117.40	5061.80	1.94
V2/BV	0.26	0.05	0.20	0.26	0.17	0.43	1.07
ALT	183.63	97.70	0.53	168.00	15.00	366.00	-0.03
Ln(BV)	7.75	1.01	0.13	7.47	5.91	10.00	0.57
PCP	2.19	1.55	0.71	2.05	0.22	6.57	0.86
LCP	144.70	98.75	0.68	119.00	36.00	461.00	1.67
SCLAC	1351.73	2495.60	1.85	545.00	1.00	12000.00	3.44
LONG	74.39	2.30	0.03	73.44	72.00	79.00	0.74
LAT	47.18	1.22	0.03	47.00	45.49	50.00	0.65
SLAC	267.33	332.63	1.24	156.00	12.00	1270.00	2.19
SFOR	3573.67	4544.88	1.27	1580.00	256.00	19900.00	2.29
ORIEN	238.73	83.89	0.35	270.00	0.00	293.00	-1.89
PERT	446.57	313.22	0.70	303.50	108.00	1360.00	1.49
PLAC	7.14	3.20	0.45	6.25	1.80	15.40	1.03
EMR	438.93	112.75	0.26	427.00	200.00	671.00	0.31
CF	2.18	0.37	0.17	2.16	1.52	3.21	0.82
NTMAX5	12.99	2.39	0.18	12.95	8.30	17.80	0.19
NTMNS	0.72	1.74	2.41	0.80	-3.80	4.40	-0.12
NPT5	13.07	4.87	0.37	12.24	6.21	24.81	0.85
NPT12-05	411.11	45.76	0.11	408.09	343.22	491.50	0.19
NPT06-11	514.98	101.48	0.20	543.05	280.00	633.39	-1.40
MSN15	18.92	2.52	0.13	18.31	14.16	23.47	0.41

Sauvegarder Fermer Imprimer

Figure 2.7 Fenêtre d'affichage des caractéristiques de base

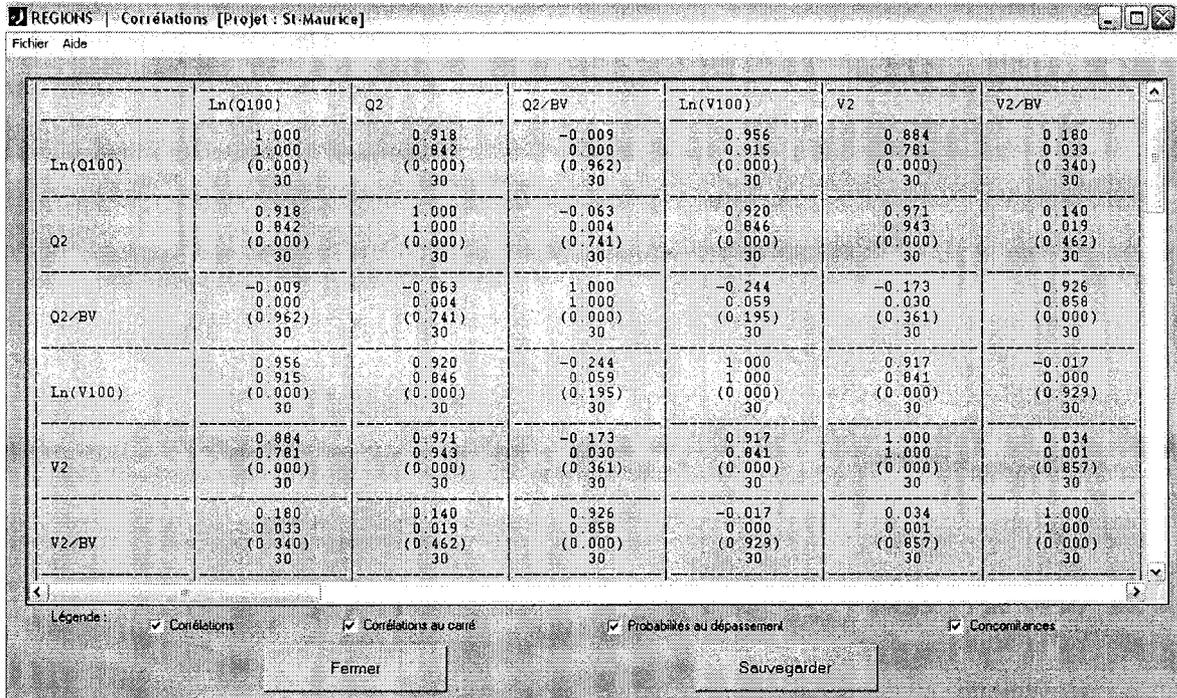


Figure 2.8 Fenêtre d'affichage de la matrice de corrélation

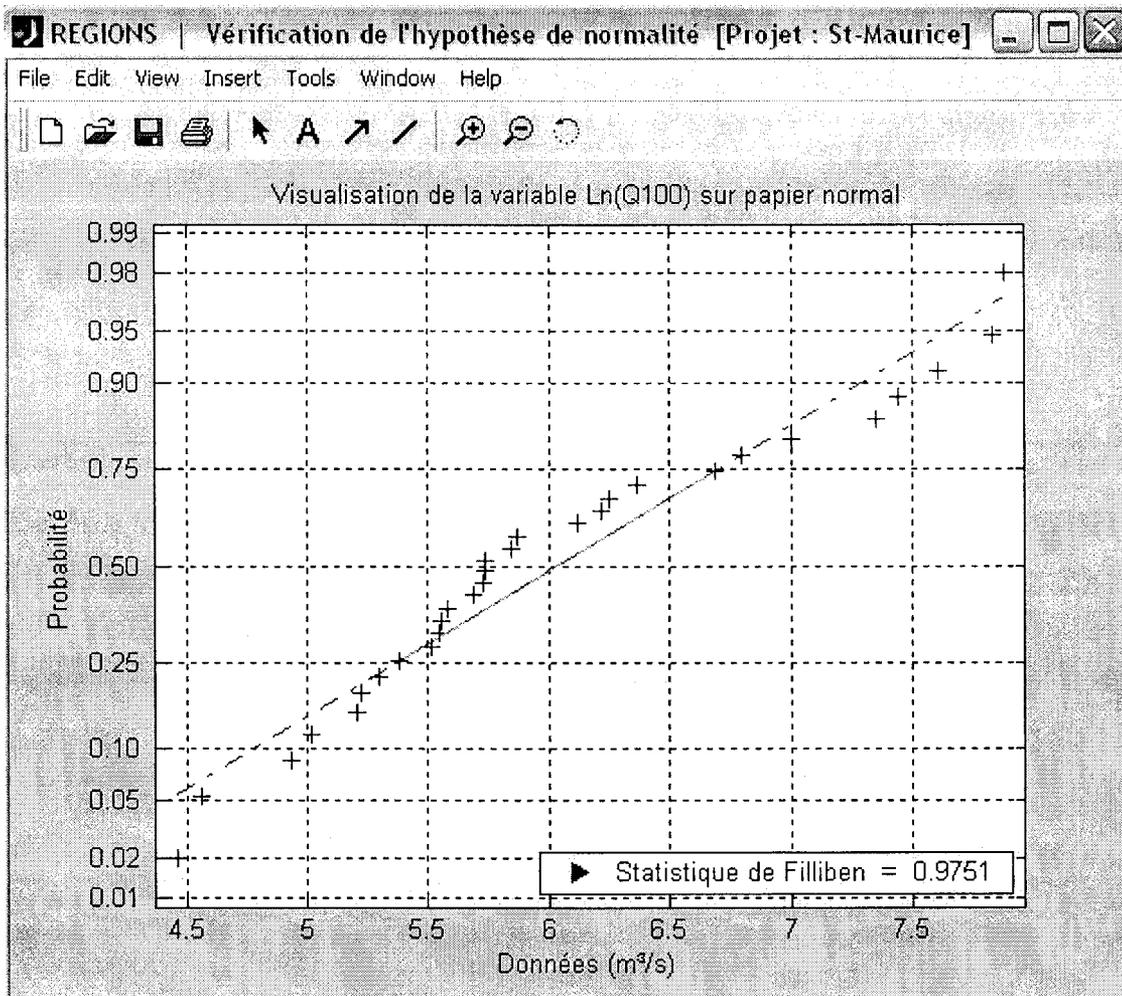


Figure 2.9 Fenêtre de validation graphique de l'hypothèse de normalité

2.4 Module de classification hiérarchique

2.4.1 Description générale du module

Ce module, présenté à la Figure 2.10, a été conçu pour appliquer la méthode de classification hiérarchique, dans le but de former des regroupements homogènes de stations de type géographique (par opposition aux regroupements de type voisinage), mais pas nécessairement contigus. On y accède à partir du menu principal en cliquant sur le bouton "Classification hiérarchique". L'idée générale la méthode consiste à établir le degré de similitude entre les différentes stations par le calcul de distances se basant sur les valeurs

observées pour certaines variables. Ceci permet de construire un arbre de classification (voir exemple de la Figure 2.12), où l'on associe les stations selon un algorithme de regroupement bien précis. Par exemple, en choisissant d'associer systématiquement les voisins les plus proches, on forme un premier regroupement avec les deux stations les plus proches (i.e. celles qui sont séparées par la plus petite distance), que l'on considère ensuite comme un nouvel individu qui sera comparé au reste des stations, i.e. qu'on répète ce processus d'association deux par deux de façon à former la hiérarchie de regroupement telle qu'illustrée sur l'arbre de classification. De là, en précisant un seuil de coupe, on se trouve à définir une partition de l'ensemble des stations, i.e. les classes. Le nombre de classes obtenu correspond au nombre de points d'intersection entre la ligne de coupe (illustrée en pointillés sur la Figure 2.12) et l'arbre. On détermine les éléments appartenant à chaque classe en suivant les différentes branches de l'arbre vers le bas, à partir de chaque point.

2.4.2 Choix des paramètres de la classification

Afin de faire exécuter le module, il faut tout d'abord choisir les variables d'intérêt pour la classification (i.e. les variables sur lesquelles on basera le calcul des distances) parmi le bassin de variables physiographiques/météorologiques disponibles. C'est ce à quoi sert le bouton "Choisir les variables", où les variables sélectionnées sont reportées dans la boîte d'édition située juste au-dessous de ce dernier. L'étude des corrélations pouvant être utile pour faire ce choix, c'est pourquoi on retrouve l'option "Corrélations" dans le menu défilant dénommé "Outils", celle-ci donnant accès à la même matrice de corrélations que celle obtenue à partir du module de statistiques descriptives (voir section 2.3).

De son côté, le "checkbox" intitulé "Pondération", lorsqu'il est activé, permet de tenir compte de l'importance relative des variables de classement par rapport à la variable que l'on veut régionaliser, et ce à l'aide de coefficients de pondérations basés sur les corrélations observées entre les variables sélectionnées et la variable d'intérêt. Quant à l'encadré intitulé "Type de standardisation", il permet comme son nom l'indique de choisir la transformation à appliquer aux données avant d'enclencher la procédure de classification (on utilise habituellement la centrée réduite, où l'on soustrait d'abord la moyenne pour ensuite diviser

par l'écart-type), généralement afin d'éliminer les effets d'échelle pouvant exister entre les différentes variables.

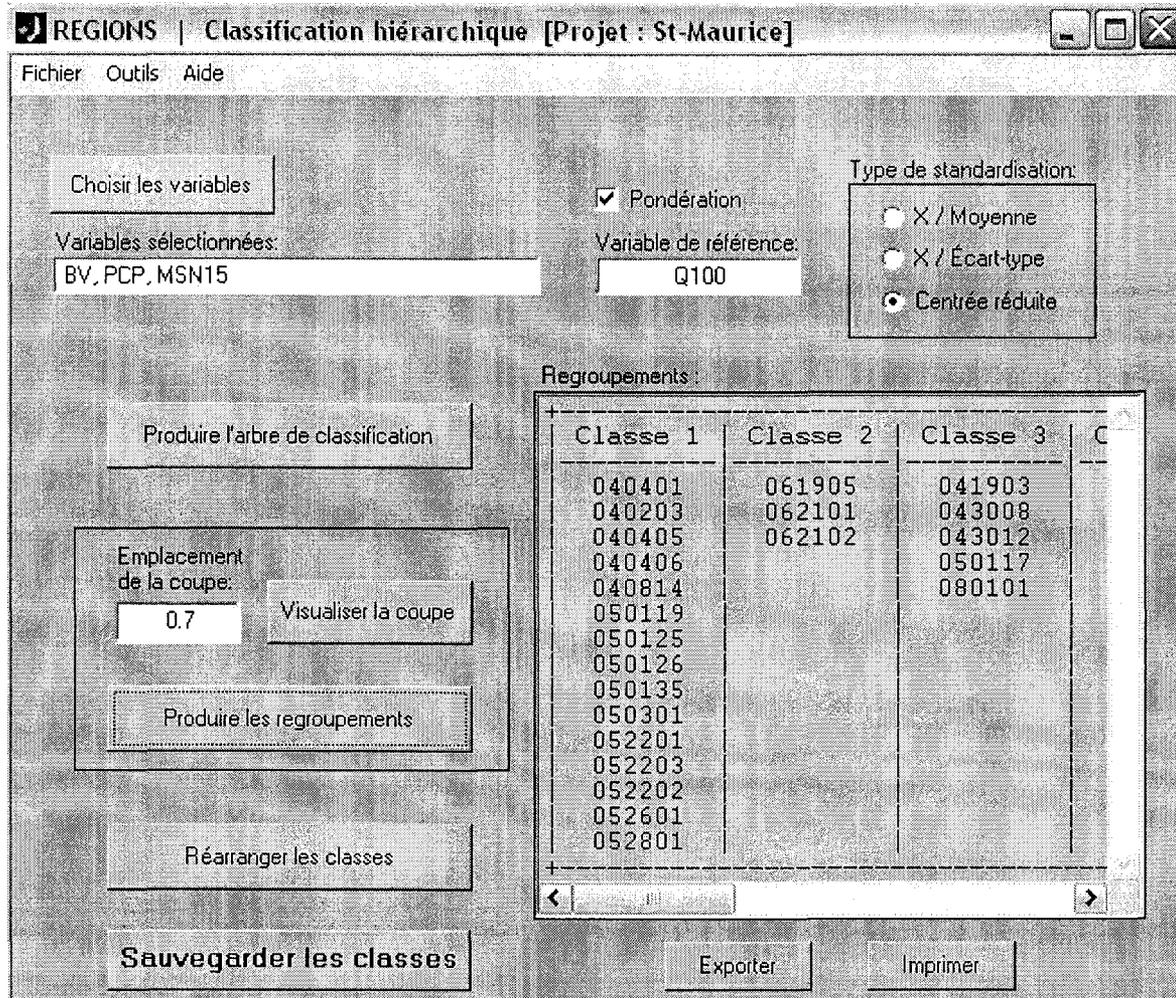


Figure 2.10 Fenêtre de classification hiérarchique

Une fois que le choix des variables de classement, de la variable de pondération (s'il y a lieu) ainsi que du type de standardisation ont été complétés, on doit ensuite appuyer sur le bouton "Produire l'arbre de classification", ce qui a pour effet d'ouvrir la fenêtre de sélection des paramètres de la classification, telle qu'illustrée à la Figure 2.11. Cette fenêtre présente deux menus déroulant ("popup menu") permettant de sélectionner le type de distance et l'algorithme de regroupement à considérer pour la production de l'arbre de classification. Les choix présentés dans ces menus correspondent aux options offertes présentement par

MATLAB. De brèves descriptions de ces distances de même que des algorithmes de regroupement sont données au Tableau 2.1 et au Tableau 2.2 respectivement. Dans le cas de la distance de Minkowski, il est possible de spécifier une valeur différente de l'exposant que celle inscrite par défaut (2) dans la boîte d'édition située au bas de la fenêtre de sélection. On obtient finalement l'arbre de classification en appuyant sur le bouton "OK".

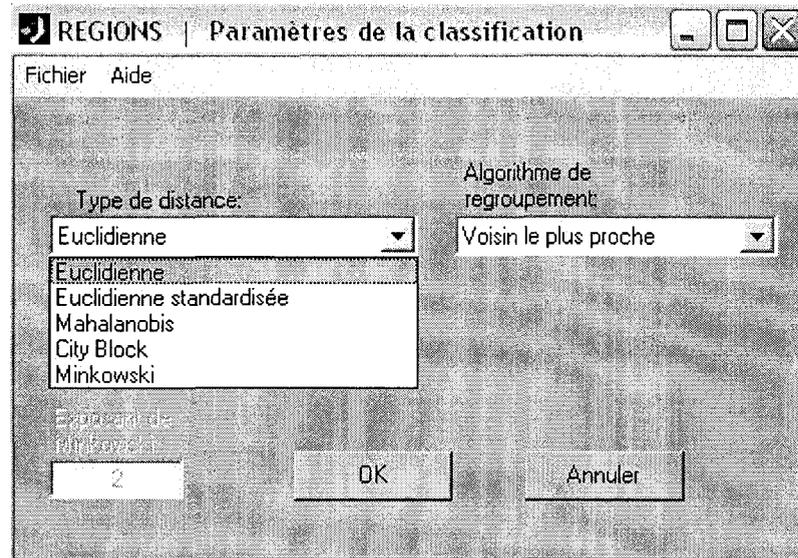


Figure 2.11 Fenêtre de sélection des paramètres de la classification

On remarquera que les paramètres sous-jacents à l'obtention de l'arbre de classification en question sont reportés au bas du graphique, tel qu'on peut le constater sur la Figure 2.12. Les nombres reportés en abscisse réfèrent aux différentes stations de l'étude selon leur ordre d'entrée dans le fichier de données, i.e. l'ordre dans lequel sont listées les stations dans la fenêtre de visualisation des données (voir Figure 2.6). Il est à noter que celle-ci peut être appelée pour consultation à partir du menu défilant "Outils". L'échelle de valeurs que l'on retrouve en ordonnée est quant à elle subordonnée au type de distance employée et définit par conséquent les valeurs admissibles pour le seuil de coupe.

Tableau 2.1 Description des différents types de distances

Type de distance	Expression de la distance entre les vecteurs x_r et x_s *	Commentaires
Euclidienne	$\sqrt{(x_r - x_s) \cdot (x_r - x_s)^T}$	la définition la plus usuelle
Euclidienne standardisée	$\sqrt{(x_r - x_s) D^{-1} (x_r - x_s)^T}$	D représente la matrice diagonale formée des éléments v_k^2 , où v_k^2 représente la variance de la variable k ($k = 1, \dots, n$)
Mahalanobis	$\sqrt{(x_r - x_s)^T V^{-1} (x_r - x_s)}$	V représente la matrice de covariance
City Block	$\sum_{k=1}^n x_{rk} - x_{sk} $	n représente le nombre de variables
Minkowski	$\sqrt[p]{\sum_{k=1}^n x_{rk} - x_{sk} ^p}$	cas $p = 1 \Leftrightarrow$ City Block cas $p = 2 \Leftrightarrow$ Euclidienne

* Soit X notre matrice de départ, que l'on traite comme étant formée de m vecteurs-lignes x_1, x_2, \dots, x_m (i.e. $m \equiv$ nombre de stations) comprenant chacun n éléments. On s'intéresse alors à la distance entre deux stations quelconques, soit r et s (compris entre 1 et m , $r \neq s$)

Tableau 2.2 Description des différents algorithmes de regroupement

Algorithme de regroupement	Définition mathématique*	Commentaires
Voisin le plus proche	$\min(\text{dist}(x_{ri}, x_{sj}))$	$i \in (1, \dots, n_r), j \in (1, \dots, n_s)$
Voisin le plus éloigné	$\max(\text{dist}(x_{ri}, x_{sj}))$	$i \in (1, \dots, n_r), j \in (1, \dots, n_s)$
Distance moyenne	$\frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} \text{dist}(x_{ri}, x_{sj})$	—
Distance centroïde	$\text{dist}(\bar{x}_r, \bar{x}_s)$, où $\bar{x}_r = \frac{1}{n_r} \sum_{i=1}^{n_r} x_{ri}$	il s'agit de la distance entre les centroïdes des deux groupes (\bar{x}_s est défini de façon similaire à \bar{x}_r)
Méthode de Ward	$\frac{n_r n_s \cdot \text{dist}(\bar{x}_r, \bar{x}_s)^2}{n_r + n_s}$	\bar{x}_r et \bar{x}_s représentent toujours les centroïdes des deux groupes

* Ces formules visent à définir la distance entre deux groupes de stations, disons r et s . Soit n_r le nombre de stations dans le groupe r et n_s le nombre de stations dans le groupe s , et soit x_{ri} le $i^{\text{ème}}$ élément du groupe r (représenté par un vecteur-ligne de n variables). Ces formules font également référence au type de distance (dénotée dist) employée par l'utilisateur (voir les définitions au Tableau 2.1)

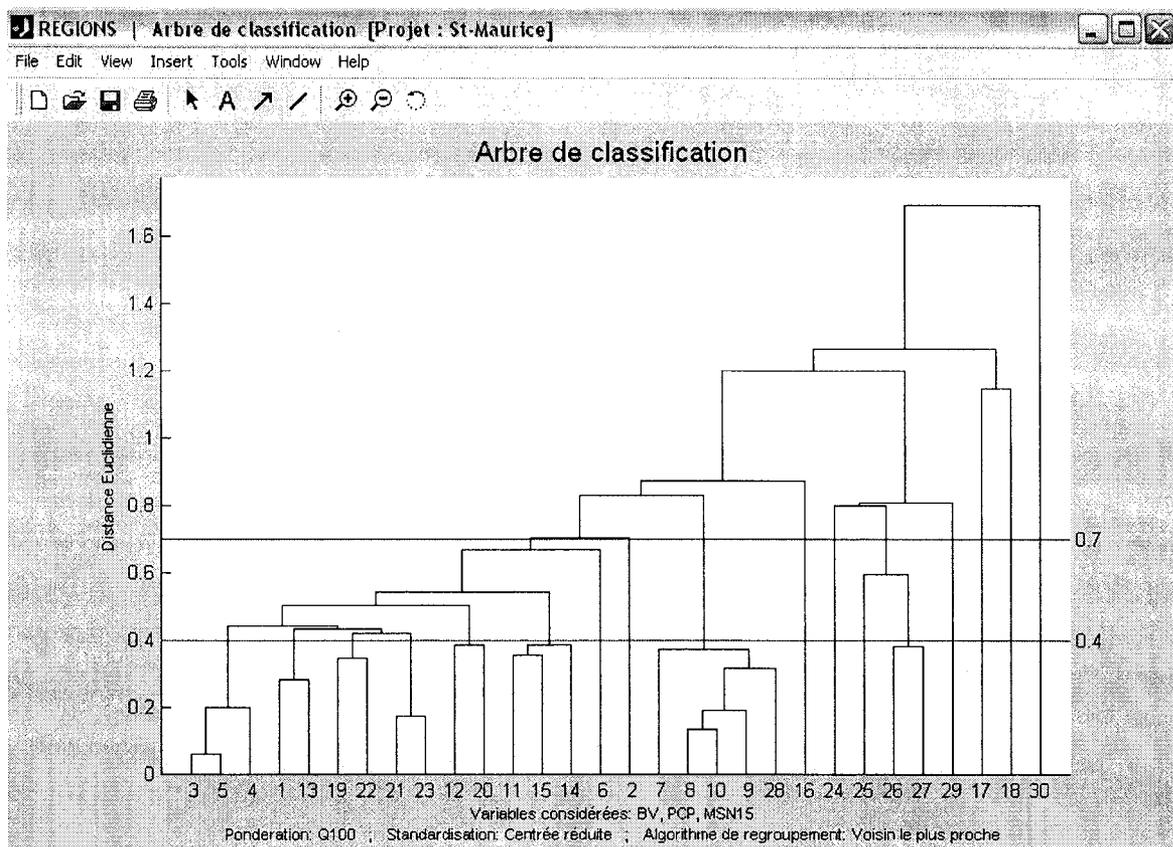


Figure 2.12 Fenêtre de représentation de l'arbre de classification

2.4.3 Obtention et réarrangement des classes

Rendu à ce stade, il ne reste donc plus qu'à revenir au module de classification de la Figure 2.10 afin de définir le seuil de coupe selon lequel on produira les classes homogènes de stations. On doit tout d'abord entrer sa valeur dans la boîte d'édition dénommée "Emplacement de la coupe". Puis, si on le désire, on peut se donner une idée des classes obtenues en fonction de ce seuil à l'aide du bouton "Visualiser la coupe", ce qui produit une ligne rouge (illustrée en pointillés dans ce document) correspondant à l'emplacement de la coupe sur l'arbre de classification. On peut aussi comparer plusieurs seuils de coupe (voir Figure 2.12) en entrant simplement une nouvelle valeur dans la boîte d'édition et en réappuyant sur le bouton "Visualiser la coupe". Le bouton "Produire les regroupements" permet quant à lui d'afficher dans la boîte d'affichage intitulée "Regroupements" les

regroupements obtenus selon le seuil de coupe spécifié. Cette opération peut elle aussi être répétée indéfiniment en modifiant l'emplacement de la coupe.

Le bouton "Exporter" permet d'exporter le contenu de la boîte d'affichage des regroupements dans un fichier texte où seront également reportés les paramètres ayant mené à ce classement (i.e. les variables considérées, le type de distance, l'algorithme de regroupement et le seuil de coupe). Le module ayant été conçu pour pouvoir être réexécuté en entier (on peut décider à tout moment de modifier les paramètres et de produire un nouvel arbre de classification) autant de fois qu'on le désire, le bouton "Exporter" devient alors utile pour garder une trace des ces différents essais. Le bouton "Imprimer" a essentiellement la même fonction que le bouton "Exporter", excepté que l'information sauvegardée dans le fichier texte est cette fois simplement envoyée à l'imprimante. Mentionnons enfin que ces deux boutons ont leur équivalent dans le menu défilant "Fichier" (options "Exporter les regroupements" et "Imprimer les regroupements").

Le bouton "Réarranger les classes" permet de modifier manuellement les classes obtenues, ce qui peut être utile pour tenir compte de certaines contraintes pratiques (e.g. éliminer les classes ne contenant pas assez d'éléments) ou encore de considérations motivées par l'expérience de l'hydrologue. Ce bouton appelle en fait le module de définition manuelle des classes, i.e. qu'il s'agit du même module qui est appelé lorsqu'on appuie sur le bouton "Régions imposées" du menu principal (afin de créer de toute pièce les regroupements homogènes). Or, la description de ce module fait l'objet de la section suivante. Les modifications apportées aux classes sont reportées automatiquement dans la boîte d'affichage du module de classification après avoir quitté le module de définition manuelle.

Finalement, le bouton "Sauvegarder les classes" permet de quitter le module de classification tout en sauvegardant dans le fichier de projet les regroupements de stations tels qu'ils apparaissent dans la boîte d'affichage à ce moment. Ce sont donc ces régions homogènes qui seront considérées ultérieurement lorsque que l'utilisateur passera à l'étape de l'estimation régionale. Bien entendu, rien n'empêche ce dernier de redéfinir ces régions en exécutant à nouveau le module.

2.5 Module de définition manuelle des classes

Ce module, présenté à la Figure 2.13, permet à l'utilisateur de définir lui-même les regroupements homogènes de stations qu'il souhaite considérer à l'intérieur de la méthodologie d'estimation régionale. On y accède soit en appuyant sur le bouton "Régions imposées" du menu principal, soit lorsque qu'on souhaite réarranger les classes produites à la suite d'une procédure de classification hiérarchique (voir section précédente). Dans ce dernier cas, les regroupements obtenus au module de classification sont reportés automatique dans la boîte d'affichage des classes (intitulée "Regroupements effectués") du module de définition manuelle. Si aucun regroupement n'a été préalablement défini, la boîte d'affichage apparaît vide tout simplement.

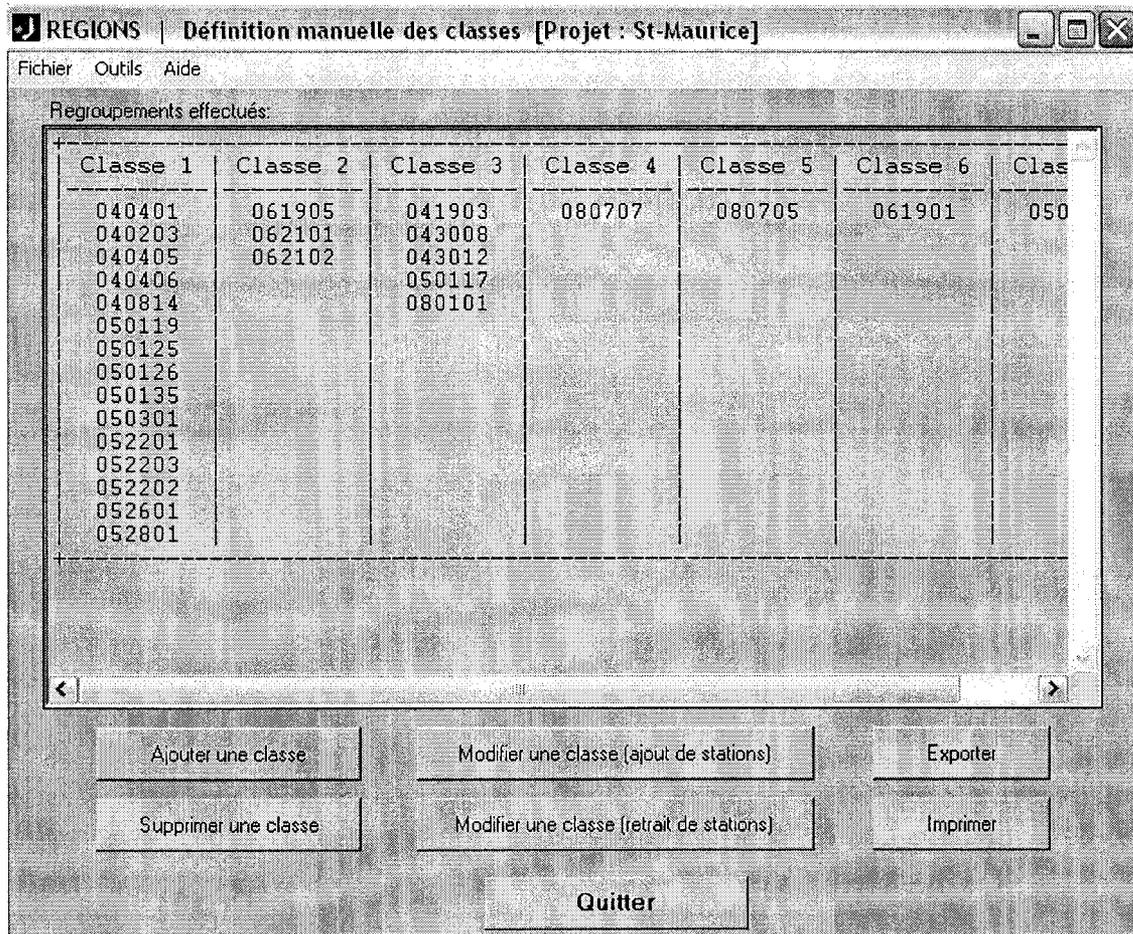


Figure 2.13 Fenêtre de définition manuelle des classes

Voici maintenant la liste des boutons contenus dans la fenêtre avec une courte description de l'action à laquelle ils correspondent:

- "Ajouter une classe" : ce bouton permet de créer une nouvelle classe dont on doit choisir les éléments parmi les stations qui n'ont pas encore été regroupées. Si toutes les stations sont déjà regroupées, il faut tout d'abord en libérer quelques unes au moyen du bouton "Supprimer une classe" ou encore "Modifier une classe (retrait de stations)"
- "Supprimer une classe" : ce bouton permet de supprimer une ou plusieurs des classes existantes
- "Modifier une classe (ajout de stations)" : ce bouton permet d'ajouter quelques stations (sélectionnées parmi celles encore disponibles) à l'une des classes existantes
- "Modifier une classe (retrait de stations)" : ce bouton permet de retirer quelques stations de l'une des classes existantes
- "Exporter" : ce bouton permet d'exporter le contenu de la boîte d'affichage vers un fichier texte (il a son équivalent dans le menu défilant "Fichier" sous l'appellation "Exporter les regroupements")
- "Imprimer" : ce bouton permet d'envoyer vers l'imprimante le contenu de la boîte d'affichage (il a son équivalent dans le menu défilant "Fichier" sous l'appellation "Imprimer les regroupements")
- "Quitter" : ce bouton permet de quitter le module tout en sauvegardant dans le fichier de projet les regroupements ainsi définis (ou modifiés) par l'utilisateur

Notons enfin que la fenêtre de visualisation des données (Figure 2.6) peut être appelée à partir menu défilant "Outils".

2.6 Module d'analyse canonique des corrélations (approche par optimisation du paramètre alpha)

2.6.1 Description générale du module

Ce module, présenté à la Figure 2.14, permet d'effectuer la détermination de voisinages hydrologiques homogènes pour des stations non jaugées à l'aide de la méthode d'analyse des corrélations canoniques (tel que décrit dans le rapport du projet C5 de la Chaire ou encore dans Ouarda *et al.* (2001)). On y accède à partir du menu principal en cliquant sur le bouton "ACC (approche par optimisation du paramètre alpha)". En bref, l'analyse des corrélations canoniques est un outil d'analyse statistique multivariée qui a pour but de faire ressortir les relations de dépendance pouvant exister entre deux ensembles de variables aléatoires (dans le cas qui nous intéresse, il s'agit de l'ensemble des variables hydrologiques *vs* l'ensemble des variables physiographiques/météorologiques). Pour cela, on crée de nouvelles paires de variables pour chaque ensemble (il s'agit en fait de combinaisons linéaires de chaque ensemble de variables) que l'on désigne comme variables canoniques, de sorte que la corrélation entre les variables canoniques d'une même paire soit maximisée, tandis que la corrélation entre les variables de paires différentes est nulle. On peut ainsi inférer sur la valeur des variables canoniques associées à l'ensemble hydrologique en connaissant la valeur des variables canoniques associées à l'ensemble physiographique/météorologique. On peut également calculer la distance séparant deux variables canoniques, ce qui nous permet de d'établir les voisinages hydrologiques. Dans cette approche, le paramètre "alpha" contrôle la taille du voisinage et doit être choisi de façon à maximiser la probabilité d'inclure des sites similaires au site cible dans le voisinage tout en minimisant à la fois la probabilité d'inclure des sites dissimilaires au site cible dans le voisinage.

2.6.2 Choix des paramètres de l'analyse des corrélations canoniques

La première étape nécessaire au fonctionnement du module consiste donc à spécifier les ensembles de variables hydrologiques et physiographiques/météorologiques à considérer, ce qui peut être fait à l'aide des boutons "Caractéristiques hydrologiques" et "Caractéristiques physiographiques / météorologiques" respectivement. Précisons qu'un minimum de deux variables hydrologiques est requis (pour la représentation en deux dimensions de l'ellipsoïde de voisinage) tandis que le nombre de variables physiographiques/météorologiques ne peut être inférieur au nombre de caractéristiques hydrologiques. Il est important toutefois de limiter le nombre de caractéristiques physio/météo car si on en utilise trop, on peut se retrouver éventuellement avec des ellipsoïdes de voisinage très petits et par conséquent pratiquement inutilisables (puisque plus le nombre de variables de comparaison est élevé, plus il est difficile de trouver des stations qui se ressemblent). Donc, si l'on dispose d'un grand bassin de variables physio/météo, il est recommandé d'en éliminer quelques unes au départ, ce qui peut être fait par exemple en examinant la matrice des corrélations (qui est disponible via l'option "Corrélations" du menu défilant "Outils") pour déterminer les variables les plus corrélées aux caractéristiques hydrologiques d'intérêt, tout en permettant d'identifier les variables redondantes. Bien entendu, l'expérience de l'hydrologue peut également entrer en ligne de compte à cette étape.

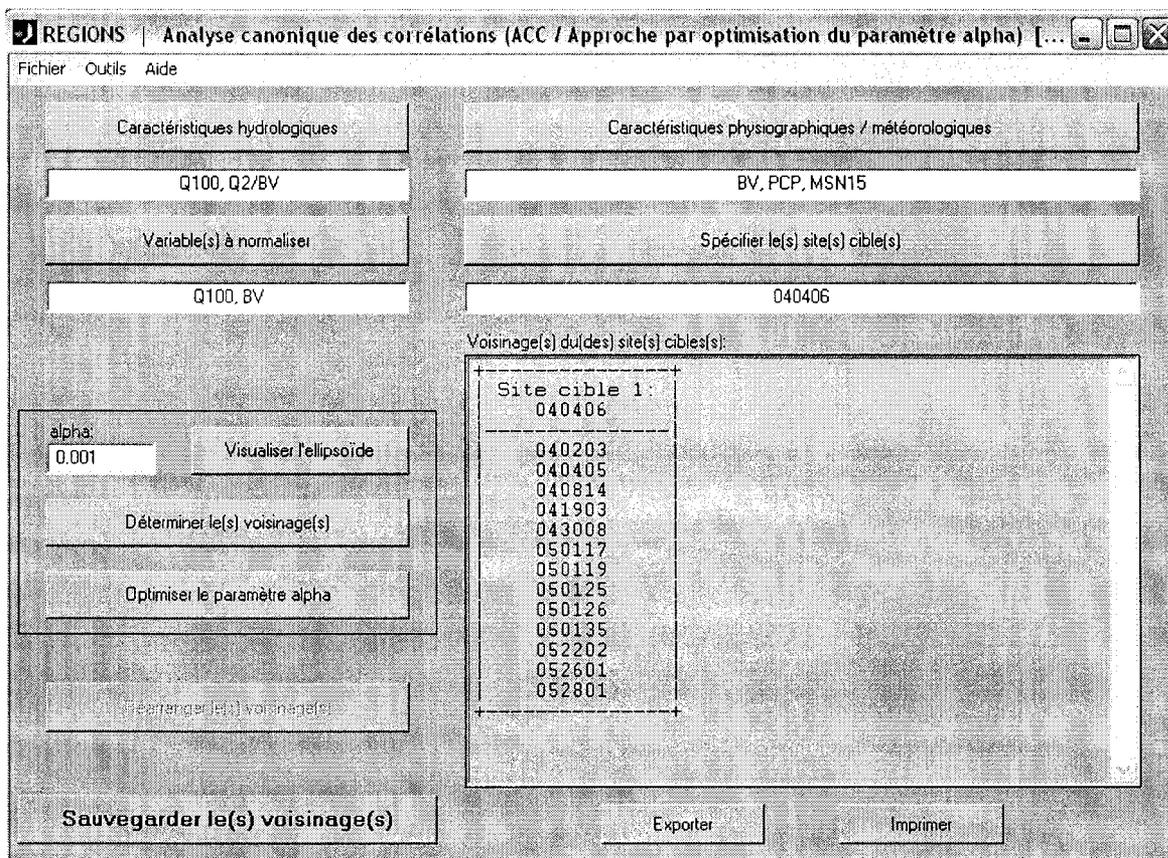


Figure 2.14 Fenêtre d'analyse canonique des corrélations

Le bouton "Variables à normaliser" sert quant à lui à indiquer, s'il y a lieu, les variables auxquelles on souhaite appliquer une transformation logarithmique afin de les rendre plus conformes à l'hypothèse de normalité. La normalité de chaque variable n'est pas une condition absolument nécessaire à l'application de la technique d'analyse canonique, quoiqu'elle représente une bonne façon de s'assurer que la distribution des variables canoniques sera multinormale (sachant que le calcul de la taille du voisinage repose pour sa part sur cette hypothèse). L'option "Vérifier la normalité des variables retenues" du menu défilant "Outils" permet à l'utilisateur de vérifier graphiquement la normalité d'une variable quelconque qui se trouve alors présentée sur papier normal (tout comme au module de statistiques descriptives, voir la Figure 2.9 à la section 2.3), et ce autant en coordonnées réelles que logarithmiques.

Il faut ensuite utiliser le bouton "Spécifier le(s) site(s) cible(s)" pour préciser quelles sont les stations non jaugées (i.e. les sites cibles) pour lesquelles on cherche à déterminer les voisinages.

Enfin, il reste à spécifier la valeur du paramètre "alpha" (dans la boîte d'édition du même nom) qui contrôle la grandeur de l'ellipsoïde englobant les stations appartenant au voisinage de la station cible. Plus il est petit, plus le voisinage sera grand (lorsque sa valeur est nulle, on se trouve à inclure toutes les stations). Il représente un niveau de confiance ; par exemple, pour une valeur de 0.05, on s'attend à ce que 95% des sites ayant un comportement similaire au site non jaugé soient inclus dans l'ellipse, ce qui en laisse 5% à l'extérieur. Plus on agrandit l'ellipse, plus on inclut de sites similaires mais plus on risque également d'inclure des sites dont le comportement est dissemblable au site cible, i.e. que l'on contamine inutilement le voisinage, ce qui se traduit progressivement par une dégradation de la qualité de l'estimation pour les variables hydrologiques. On cherche donc une valeur du paramètre "alpha" résultant en un bon compromis entre le nombre de stations représentatives incluses et le degré de contamination du voisinage.

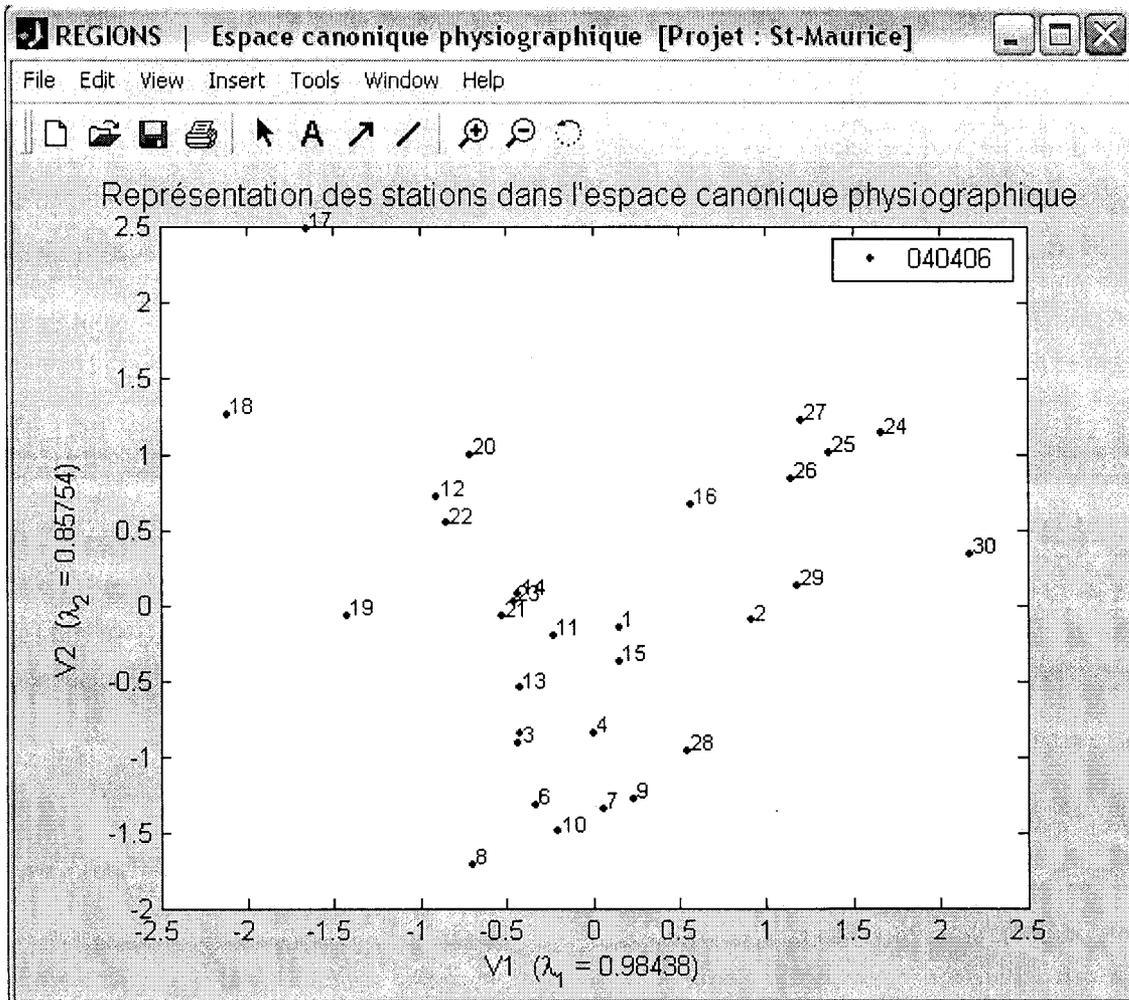


Figure 2.15 Fenêtre de position des stations dans l'espace canonique physiographique

2.6.3 Obtention des voisinages et optimisation du paramètre "alpha"

Une fois que toutes les informations requises ont été entrées, on doit appuyer sur le bouton "Déterminer le(s) voisinage(s)" afin d'enclencher la procédure d'analyse canonique. Au terme de celle-ci sont produits les graphiques de représentation des différentes stations dans les espaces canoniques physiographique (variables V1 et V2) et hydrologique (variables W1 et W2), tels qu'illustrés par la Figure 2.15 et la Figure 2.16, tandis que les voisinages obtenus pour chaque site cible sont reportés dans la boîte d'affichage intitulée "Voisinage(s) du(des) site(s) cible(s)". Les stations représentées dans l'espace canonique sont identifiées par des numéros référant à leur ordre d'entrée dans le fichier de données, i.e. l'ordre dans

lequel sont listées les stations dans la fenêtre de visualisation des données (voir Figure 2.6). Il est à noter que celle-ci peut être appelée pour consultation à partir du menu défilant "Outils" (en sélectionnant l'option "Visualiser les données"). Les valeurs λ_1 et λ_2 que l'on peut apercevoir sur la Figure 2.15 de même qu'à la Figure 2.16 représentent pour leur part les corrélations existant entre les variables V1 et W1 puis V2 et W2 respectivement.

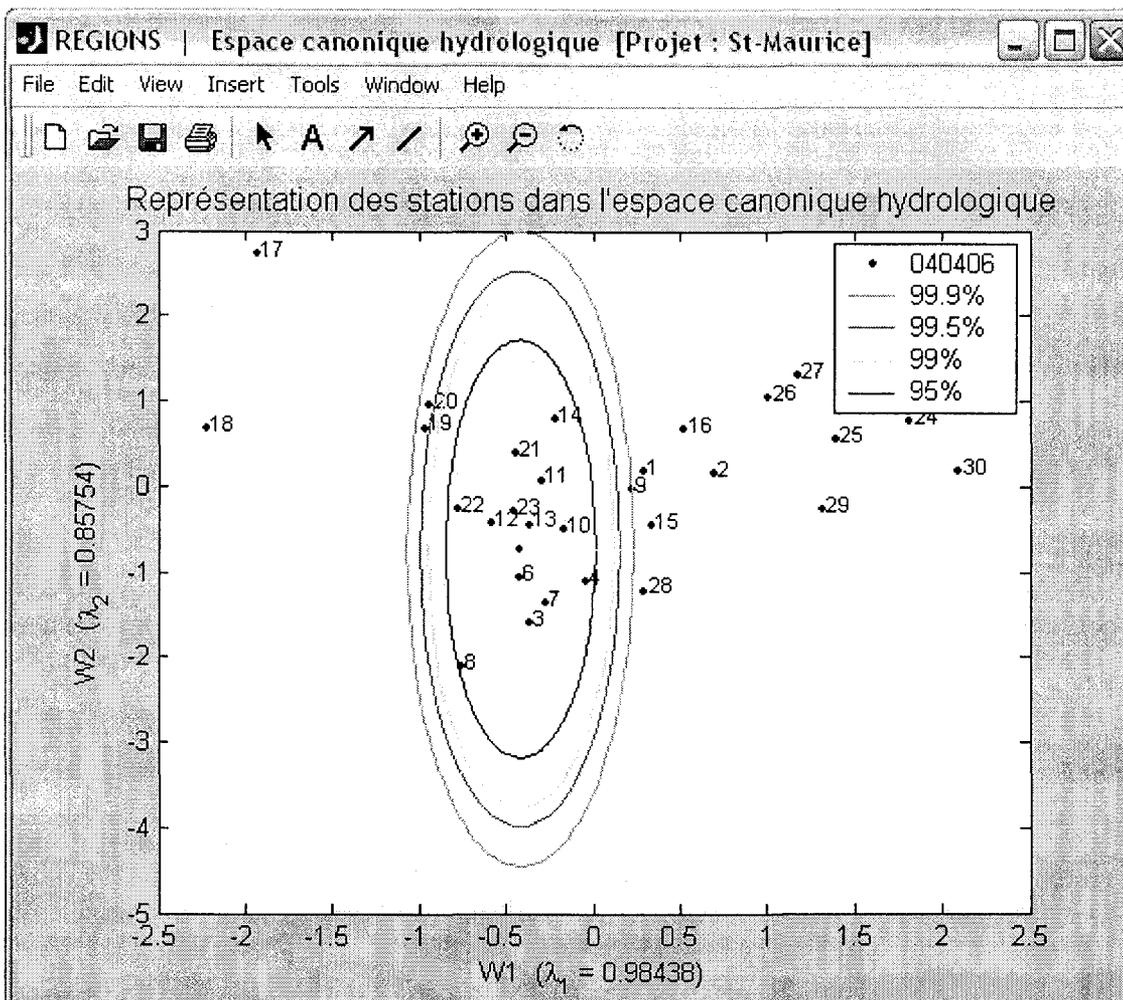


Figure 2.16 Fenêtre de position des stations dans l'espace canonique hydrologique

Si l'on modifie la valeur du paramètre "alpha", il faut réappuyer sur le bouton "Déterminer le(s) voisinage(s)" afin de mettre à jour les voisinages dans la boîte d'affichage. On peut aussi visualiser les ellipsoïdes obtenues selon différentes valeurs du paramètre "alpha" à l'aide du bouton "Visualiser l'ellipsoïde". Les ellipses apparaissent alors sur le graphique de

représentation des stations dans l'espace hydrologique (Figure 2.16), avec un point rouge marquant l'emplacement de la station cible et les niveaux de confiance correspondants rapportés dans la légende. Ceci permet donc à l'utilisateur de tester l'influence du paramètre "alpha" par rapport à la détermination des voisinages hydrologiques dans le but d'optimiser sa valeur.

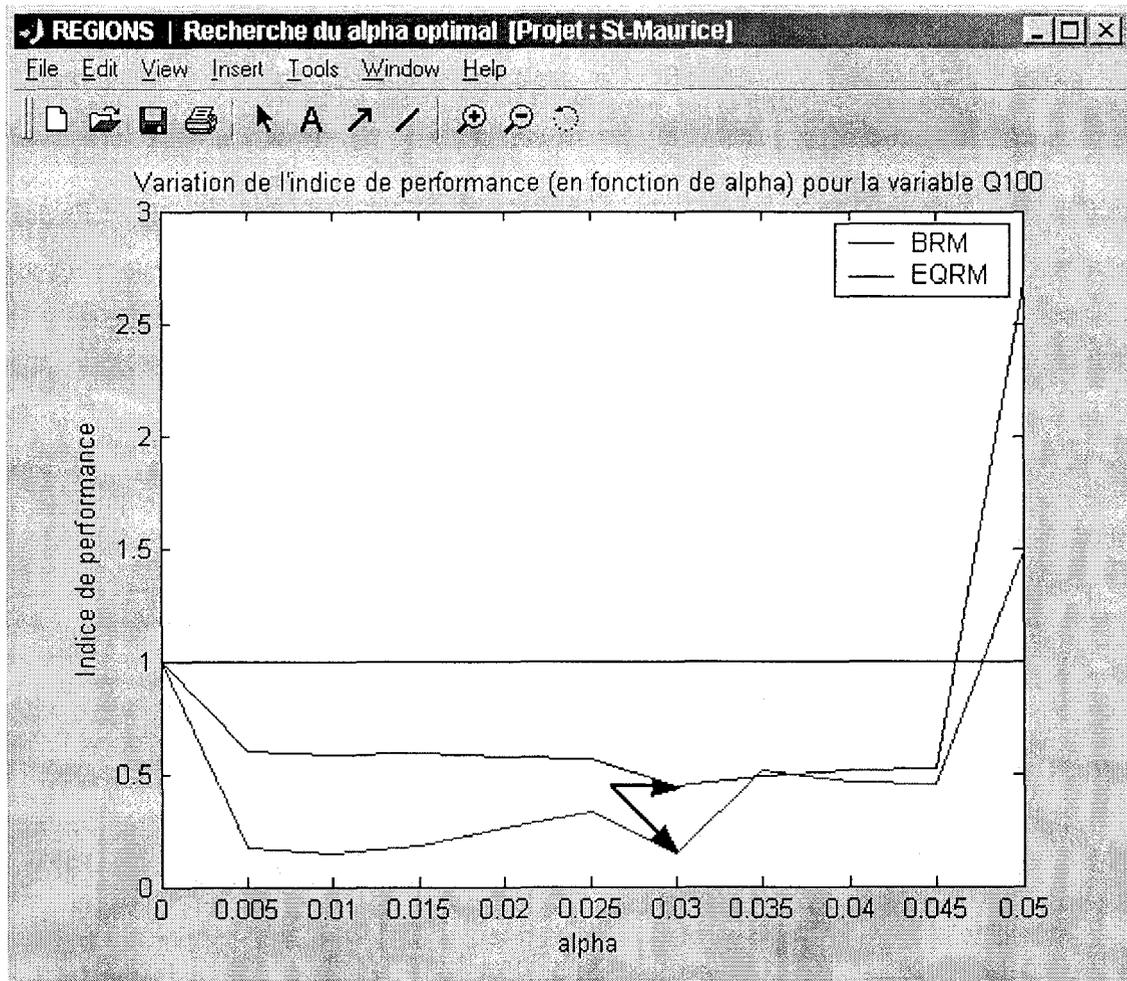


Figure 2.17 Fenêtre de représentation graphique des indices de performance en fonction du paramètre "alpha"

Cependant, il peut aussi se servir du bouton "Optimiser le paramètre alpha" pour l'aider dans cette recherche d'une valeur optimale pour le paramètre (i.e. une valeur qui maximise la probabilité d'inclure des stations similaires au site cible tout en minimisant la probabilité

d'accepter des stations non désirables dans le voisinage). Ce bouton lance l'exécution d'une procédure de type "jackknife", i.e. qu'elle considère tour à tour chacun des sites jaugés comme non jaugés, les autres sites servant à établir les voisinages et à estimer les variables hydrologiques au site considéré délibérément comme non jaugé (afin de pouvoir comparer les estimations obtenues avec les mesures déjà enregistrées au site). En répétant cette procédure pour différentes valeurs de "alpha", on calcule alors pour chaque valeur deux indices de performance basés sur le biais relatif moyen (BRM) de même que sur l'erreur quadratique relative moyenne (EQRM). Ces indices sont ensuite reportées sur un graphique en fonction du paramètre "alpha" (voir exemple de la Figure 2.17). On obtiendra en fait un graphique pour chaque variable hydrologique considérée dans l'analyse. On cherchera donc sur ces graphiques une valeur de "alpha" qui semble minimiser à la fois les indices BRM et EQRM (telle qu'indiquée par les flèches à la Figure 2.17), et ce pour chacune des variables à régionaliser. On constate ainsi que la procédure d'optimisation demeure quelque peu subjective, ce pourquoi seuls les graphiques d'indices de performance sont présentés au terme de cette dernière au lieu de suggérer à l'utilisateur une valeur optimale définitive. Pour obtenir plus de détails sur le choix du "alpha" optimal sur la base de l'interprétation des graphiques de performance, le lecteur est référé à Ouarda *et al* (1998).

Revenons maintenant à la fenêtre du module d'analyse canonique de la Figure 2.14. Le bouton "Exporter" permet d'exporter le contenu de la boîte d'affichage des voisinages dans un fichier texte où seront également reportés les paramètres sur la base desquels ces voisinages ont été déterminés (i.e. les variables hydro et physio/météo considérées, les variables normalisées s'il y a lieu, de même que la valeur employée pour le paramètre "alpha"). Le module ayant été conçu de façon à pouvoir être réexécuté en entier (on peut décider à tout moment de modifier les paramètres d'entrée pour produire de nouveaux voisinages) autant de fois qu'on le désire, le bouton "Exporter" devient alors utile pour garder une trace de ces différents essais. Le bouton "Imprimer" a essentiellement la même fonction que le bouton "Exporter", excepté que l'information sauvegardée dans le fichier texte est cette fois simplement envoyée à l'imprimante. Mentionnons enfin que ces deux

boutons ont leur équivalent dans le menu défilant "Fichier" (options "Exporter le(s) voisinage(s)" et "Imprimer le(s) voisinage(s)").

Le bouton "Réarranger le(s) voisinage(s)" n'est pas fonctionnel pour l'instant. Il a été placé là dans l'optique d'offrir à l'utilisateur la possibilité de modifier manuellement les voisinages obtenus au terme de l'analyse canonique (comme cela est présentement possible pour les regroupements qui sont produits dans le module de classification hiérarchique), ce qui devrait être complété éventuellement. Finalement, le bouton "Sauvegarder le(s) voisinage(s)" permet de quitter le module d'analyse canonique tout en sauvegardant dans le fichier de projet les voisinages tels qu'ils apparaissent dans la boîte d'affichage à ce moment. Ce sont donc ces régions homogènes qui seront considérées ultérieurement lorsque que l'utilisateur passera à l'étape de l'estimation régionale. Bien entendu, rien n'empêche ce dernier de redéfinir ces régions en exécutant à nouveau le module.

2.7 Module d'analyse canonique des corrélations (approche par classification)

2.7.1 Description générale du module

Ce module, présenté à la Figure 2.18, correspond à une définition révisée (par rapport à la méthode décrite à la section 2.6) d'un voisinage hydrologique homogène. En effet, cette méthode (décrite par Girard et al. (2000)) est basée sur le même principe général (analyse canonique des corrélations) que la méthode décrite à la section 2.6 et permet donc aussi de faire ressortir les relations de dépendance pouvant exister entre deux ensembles de variables aléatoires. Cependant, l'approche par classification se base sur une règle dérivée de la théorie de la classification, c'est-à-dire que la détermination du voisinage ne dépend plus du paramètre "alpha". La détermination du voisinage se fait plutôt en assignant une observation W (score canonique) à la population multinormale appropriée. On y accède à partir du menu principal en cliquant sur le bouton "ACC (approche par classification)".

2.7.2 Choix des paramètres de l'analyse des corrélations canoniques

La première étape nécessaire au fonctionnement du module consiste donc à spécifier les ensembles de variables hydrologiques et physiographiques/météorologiques à considérer, ce qui peut être fait à l'aide des boutons "Caractéristiques hydrologiques" et "Caractéristiques physiographiques / météorologiques" respectivement. Précisons qu'un minimum de deux variables hydrologiques est requis (pour la représentation en deux dimensions de l'ellipsoïde de voisinage) tandis que le nombre de variables physiographiques/météorologiques ne peut être inférieur au nombre de caractéristiques hydrologiques. Il est important toutefois de limiter le nombre de caractéristiques physio/météo car si on en utilise trop, on peut se retrouver éventuellement avec des ellipsoïdes de voisinage très petits et par conséquent pratiquement inutilisables (puisque plus le nombre de variables de comparaison est élevé, plus il est difficile de trouver des stations qui se ressemblent). Donc, si l'on dispose d'un grand bassin de variables physio/météo, il est recommandé d'en éliminer quelques unes au départ, ce qui peut être fait par exemple en examinant la matrice des corrélations (qui est disponible via l'option "Corrélations" du menu défilant "Outils") pour déterminer les variables les plus corrélées aux caractéristiques hydrologiques d'intérêt, tout en permettant d'identifier les variables redondantes. Bien entendu, l'expérience de l'hydrologue peut également entrer en ligne de compte à cette étape.

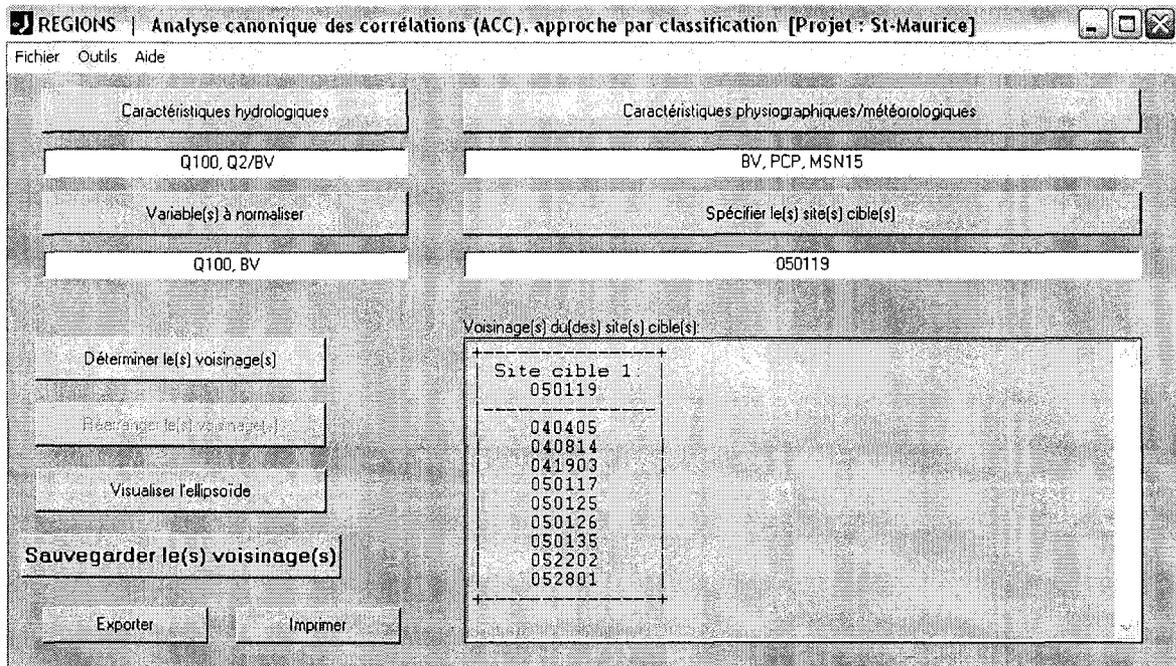


Figure 2.18 Fenêtre d'analyse canonique des corrélations (approche par classification)

Le bouton "Variables à normaliser" sert quant à lui à indiquer, s'il y a lieu, les variables auxquelles on souhaite appliquer une transformation logarithmique afin de les rendre plus conformes à l'hypothèse de normalité. La normalité de chaque variable n'est pas une condition absolument nécessaire à l'application de la technique d'analyse canonique, quoiqu'elle représente une bonne façon de s'assurer que la distribution des variables canoniques sera multinormale (sachant que le calcul de la taille du voisinage repose pour sa part sur cette hypothèse). L'option "Vérifier la normalité des variables retenues" du menu défilant "Outils" permet à l'utilisateur de vérifier graphiquement la normalité d'une variable quelconque qui se trouve alors présentée sur papier normal (tout comme au module de statistiques descriptives, voir la Figure 2.9 à la section 2.3), et ce autant en coordonnées réelles que logarithmiques.

Il faut ensuite utiliser le bouton "Spécifier le(s) site(s) cible(s)" pour préciser quelles sont les stations non jaugées (i.e. les sites cibles) pour lesquelles on cherche à déterminer les voisinages.

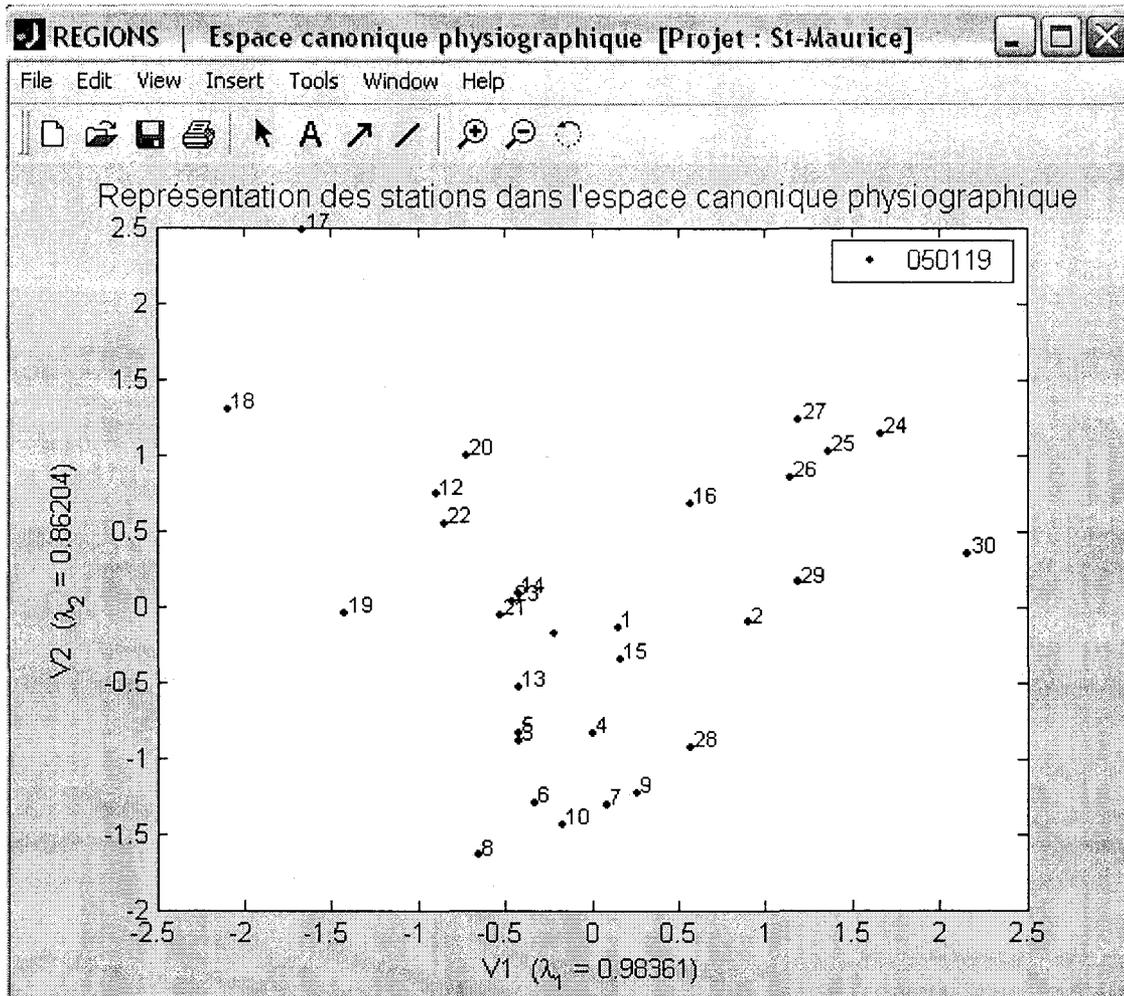


Figure 2.19 Fenêtre de position des stations dans l'espace canonique physiographique

2.7.3 Obtention des voisinages

Une fois que toutes les informations requises ont été entrées, on doit appuyer sur le bouton "Déterminer le(s) voisinage(s)" afin d'enclencher la procédure d'analyse canonique. Au terme de celle-ci sont produits les graphiques de représentation des différentes stations dans les espaces canoniques physiographique (variables V1 et V2) et hydrologique (variables W1 et W2), tels qu'illustrés par la Figure 2.19 et la Figure 2.20, tandis que les voisinages obtenus pour chaque site cible sont reportés dans la boîte d'affichage intitulée "Voisinage(s)

du(des) site(s) cible(s)". Les stations représentées dans l'espace canonique sont identifiées par des numéros référant à leur ordre d'entrée dans le fichier de données, i.e. l'ordre dans lequel sont listées les stations dans la fenêtre de visualisation des données (voir Figure 2.6). Il est à noter que celle-ci peut être appelée pour consultation à partir du menu défilant "Outils" (en sélectionnant l'option "Visualiser les données"). Les valeurs λ_1 et λ_2 que l'on peut apercevoir sur la Figure 2.19 de même qu'à la Figure 2.20 représentent pour leur part les corrélations existant entre les variables V1 et W1 puis V2 et W2 respectivement.

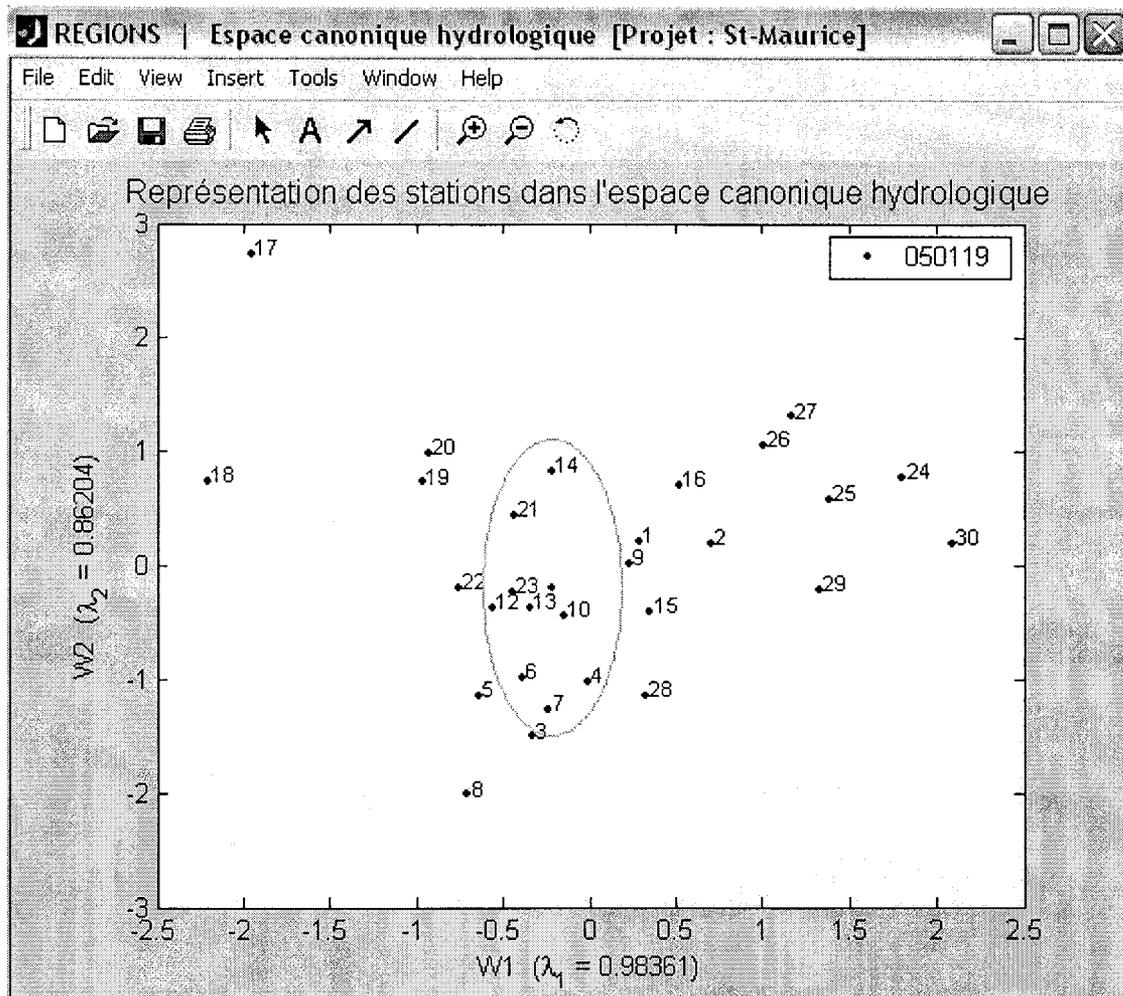


Figure 2.20 Fenêtre de position des stations dans l'espace canonique hydrologique

Revenons maintenant à la fenêtre du module d'analyse canonique de la Figure 2.18. Le bouton "Exporter" permet d'exporter le contenu de la boîte d'affichage des voisinages dans

un fichier texte où seront également reportés les paramètres sur la base desquels ces voisinages ont été déterminés (i.e. les variables hydro et physio/météo considérées, les variables normalisées s'il y a lieu, de même que la valeur employée pour le paramètre "alpha"). Le module ayant été conçu de façon à pouvoir être réexécuté en entier (on peut décider à tout moment de modifier les paramètres d'entrée pour produire de nouveaux voisinages) autant de fois qu'on le désire, le bouton "Exporter" devient alors utile pour garder une trace de ces différents essais. Le bouton "Imprimer" a essentiellement la même fonction que le bouton "Exporter", excepté que l'information sauvegardée dans le fichier texte est cette fois simplement envoyée à l'imprimante. Mentionnons enfin que ces deux boutons ont leur équivalent dans le menu défilant "Fichier" (options "Exporter le(s) voisinage(s)" et "Imprimer le(s) voisinage(s)").

Le bouton "Réarranger le(s) voisinage(s)" n'est pas fonctionnel pour l'instant. Il a été placé là dans l'optique d'offrir à l'utilisateur la possibilité de modifier manuellement les voisinages obtenus au terme de l'analyse canonique (comme cela est présentement possible pour les regroupements qui sont produits dans le module de classification hiérarchique), ce qui devrait être complété éventuellement. Finalement, le bouton "Sauvegarder le(s) voisinage(s)" permet de quitter le module d'analyse canonique tout en sauvegardant dans le fichier de projet les voisinages tels qu'ils apparaissent dans la boîte d'affichage à ce moment. Ce sont donc ces régions homogènes qui seront considérées ultérieurement lorsque que l'utilisateur passera à l'étape de l'estimation régionale. Bien entendu, rien n'empêche ce dernier de redéfinir ces régions en exécutant à nouveau le module.

2.8 Module de régression multiple

2.8.1 Choix du type de régions homogènes

On accède à ce module à partir du menu principal en appuyant sur le bouton "Régression multiple". Cependant, avant que ne s'ouvre la fenêtre du module proprement dit, l'utilisateur

voit d'abord apparaître la fenêtre illustrée à la Figure 2.21, intitulée "Choix du type de DRH pour la régression" où l'abréviation DRH signifie Détermination des Régions Homogènes. On doit y préciser de quelle façon les régions que l'on souhaite considérer dans l'analyse ont été déterminées. S'il s'agit de regroupements géographiques (non contigus) qui ont été obtenus soit à l'aide du module de classification hiérarchique (section 2.4) ou encore par le biais du module de définition manuelle des classes (section 2.5), on doit appuyer sur le bouton radio "Classification hiérarchique (ou Régions imposées)" ; s'il s'agit de voisinages hydrologiques obtenus grâce au module d'analyse canonique des corrélations, on doit plutôt appuyer sur le bouton radio "Analyse des Corrélations Canoniques (approche par classification)" (section 2.7) ou "Analyse des Corrélations Canoniques (approche par optimisation du paramètre alpha)" (section 2.6). Il est important de préciser ici que rien n'empêche que des régions homogènes de ces deux types puissent exister en même temps pour un projet donné, mais on devra toutefois les traiter séparément au moment de l'estimation régionale. Une fois que le bon bouton radio a été sélectionné, il n'y a plus qu'à appuyer sur le bouton "OK" pour passer au module de régression multiple.

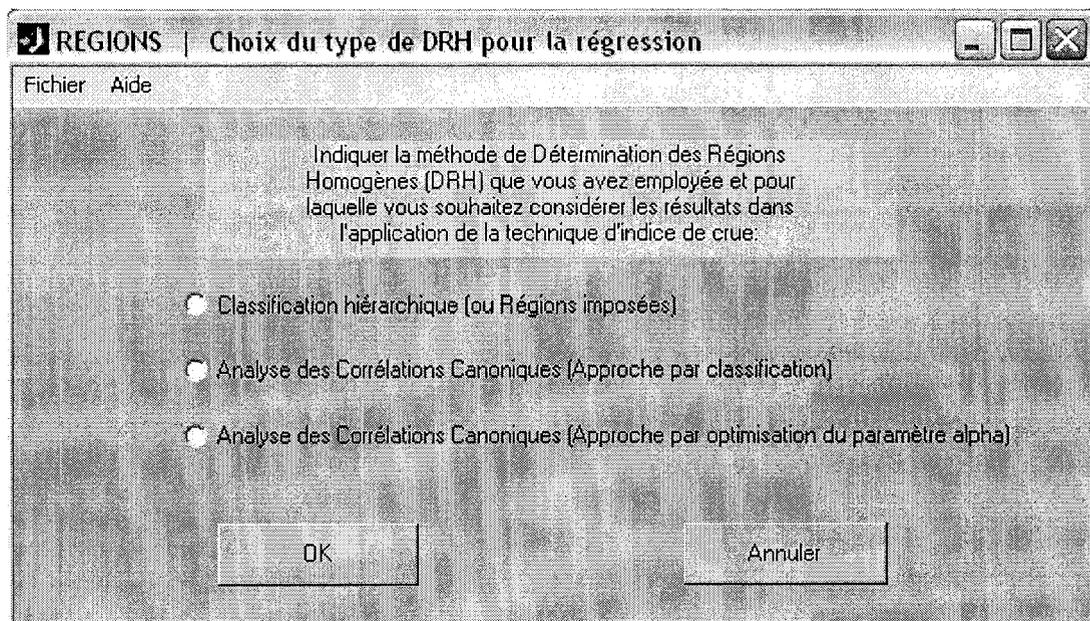


Figure 2.21 Fenêtre de spécification du type de régions homogènes à considérer au moment de l'exécution du module de régression multiple

2.8.2 Sélection du modèle régressif

Le module est présenté à la Figure 2.22 dans le cas où l'utilisateur a opté pour l'option « Analyse des corrélations canoniques (Approche par classification) ». Lorsque l'utilisateur opte pour des regroupements de type géographique, l'interface du module demeure pratiquement la même, excepté que le bouton "Choisir les variables physiographiques / météorologiques à inclure dans le modèle de régression" est scindé en deux boutons dénommés "Choisir les variables à régionaliser" et "Choisir les variables explicatives". Cette différence est due au fait que, contrairement au module d'analyse canonique des corrélations, les variables que l'on cherche à régionaliser ne sont pas explicitées directement dans les modules de classification hiérarchique et de définition manuelle des classes. Les variables à régionaliser sont donc reportées directement dans la boîte d'affichage "Variables à régionaliser" (conformément à ce qui a été entré au module d'analyse canonique) lorsqu'on utilise des voisinages, mais si ce n'est pas le cas, on doit les spécifier à l'aide du bouton "Choisir les variables à régionaliser".

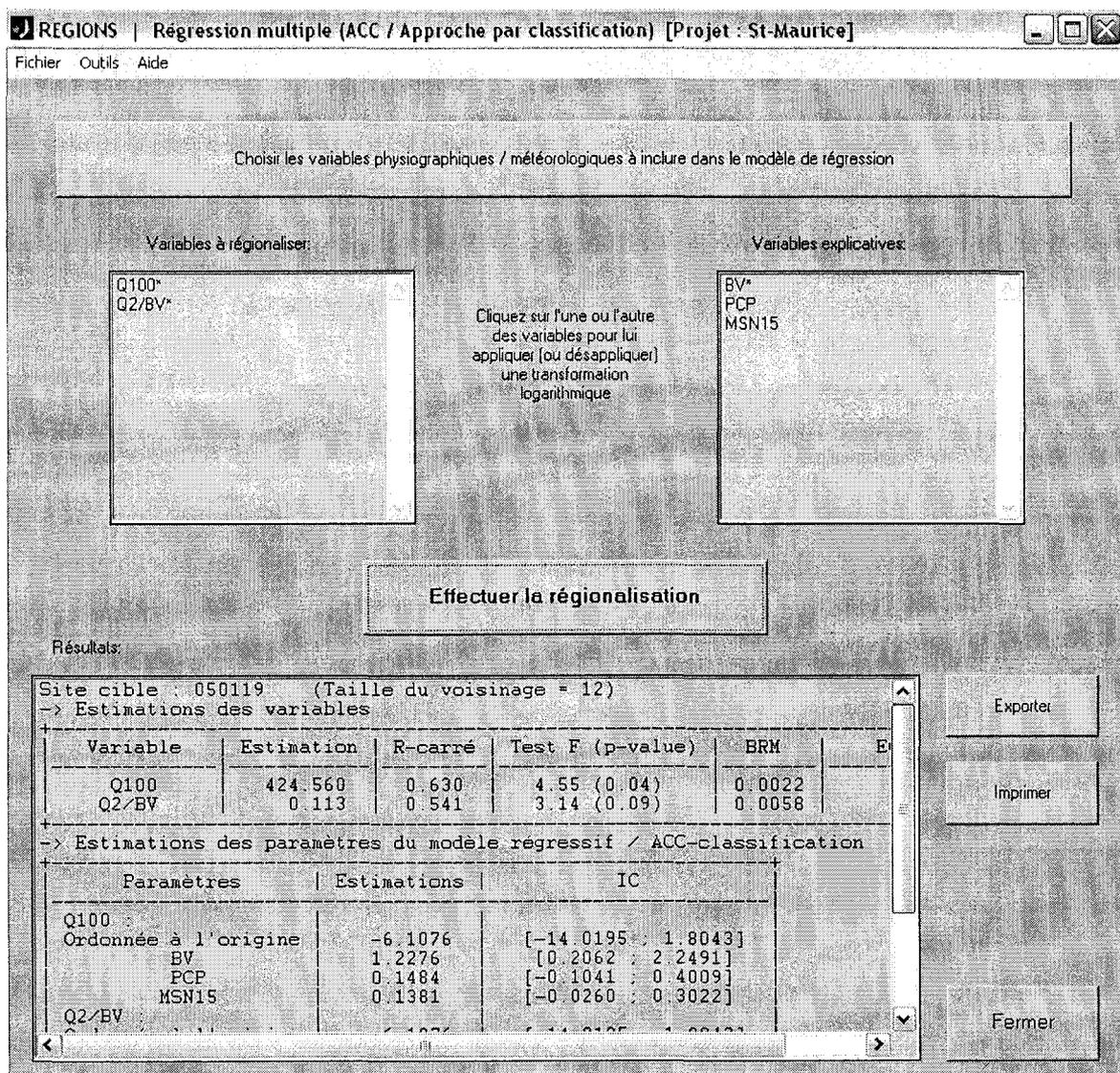


Figure 2.22 Fenêtre de régression multiple (option voisinages – ACC/approche par classification)

Les variables physiographiques/météorologiques utilisées pour l'analyse canonique sont elles aussi reportées automatiquement dans la boîte d'affichage "Variables explicatives", tout comme le seront les variables de classement employées au module de classification hiérarchique, le cas échéant. Bien qu'il soit naturel de considérer les variables ayant servi à la définition des régions homogènes comme prédicteurs dans le modèle régressif, il est toujours possible de modifier cette liste à l'aide des bouton "Choisir les variables

physiographiques / météorologiques à inclure dans le modèle de régression" et "Choisir les variables explicatives", selon le cas.

Tout comme au module de statistiques descriptives, il est possible de cliquer sur l'une ou l'autre des variables présentes dans les deux boîtes d'affichage afin de lui appliquer une transformation logarithmique (ce qui est de nouveau indiqué par un astérisque vis-à-vis les variables concernées). À ce propos, le menu défilant "Outils" offre l'option "Vérifier la normalité des variables retenues" permettant de valider graphiquement (tout comme au module de statistiques descriptives, voir la Figure 2.9 à la section 2.3) l'hypothèse de normalité pour l'une ou l'autre des variables retenues en la présentant sur papier normal (en coordonnées réelles et logarithmiques).

Une fois que les variables réponses (hydrologiques) et les variables explicatives (physiographiques/météorologiques) ont été sélectionnées, il suffit d'appuyer sur le bouton "Effectuer la régionalisation" afin d'obtenir les estimations recherchées. Une étape supplémentaire s'ajoute toutefois si l'on considère des regroupements de type géographique puisque l'on demandera alors à l'utilisateur pour quels sites (i.e. les stations non jaugées) l'estimation doit être effectuée (car cette discrimination n'est pas requise lors de l'exécution des modules de classification hiérarchique et de définition manuelle des classes).

2.8.3 Description des résultats de la régionalisation

Les résultats de la régression sont reportés dans la boîte d'affichage intitulée "Résultats". On y présente un tableau pour chaque site cible. Au-dessus de chaque tableau est indiqué la taille du voisinage (ou de la classe) associé au site cible. Or, en régression, il est recommandé que la taille de l'échantillon ne soit pas inférieure au nombre de paramètres du modèle (i.e. nombre de prédicteurs + 1) multiplié par 3, sans quoi on peut émettre de sérieuses réserves sur la qualité de l'estimation. Des avertissements préviennent l'utilisateur à cet effet chaque fois que cette condition n'est pas respectée. Dans le tableau même, on retrouve donc les estimations des variables d'intérêt à la première colonne, accompagnées de quelques statistiques relatives à la performance du modèle régressif, soit la valeur du

coefficient de détermination (le R^2), la valeur du test F global de signification, le biais relatif moyen ainsi que l'écart quadratique relatif moyen (ces deux dernières statistiques étant calculées sur la base des écarts entre valeurs prédites et mesurées). Il est aussi possible de valider le modèle régressif obtenu en cliquant dans le menu "outils", sous "Validation du modèle régressif". Ensuite, deux options sont possibles dans le but de vérifier les deux hypothèses à la base d'un modèle régressif : "Vérifier la normalité des résidus" et "Vérifier l'homogénéité des variances".

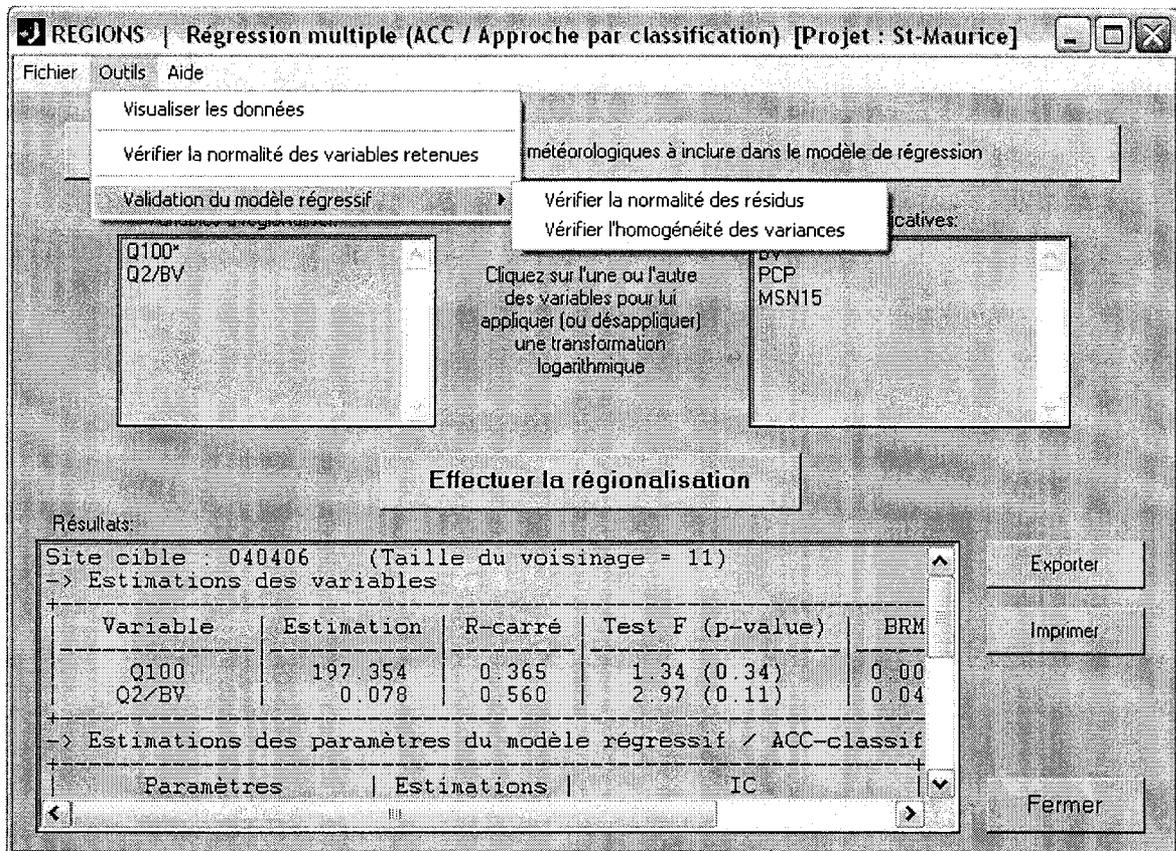


Figure 2.23 Fenêtre de régression multiple (options de validation du modèle régressif)

En ce qui concerne la normalité des résidus, on obtient, pour chaque site cible, un graphique représentant les quantiles des résidus en fonction de ceux de la distribution normale. L'hypothèse de normalité des résidus sera vérifiée si les points semblent distribués le long de la droite. Pour ce qui est de l'homogénéité des variances, un graphique représentant les

résidus en fonction des valeurs prédites est affiché pour chaque site cible. L'hypothèse d'homogénéité des variances est vérifiée si les points ne semblent pas distribués selon un certain patron. En effet, les points devraient être distribués de façon aléatoire autour de la valeur 0 (en ordonnée).

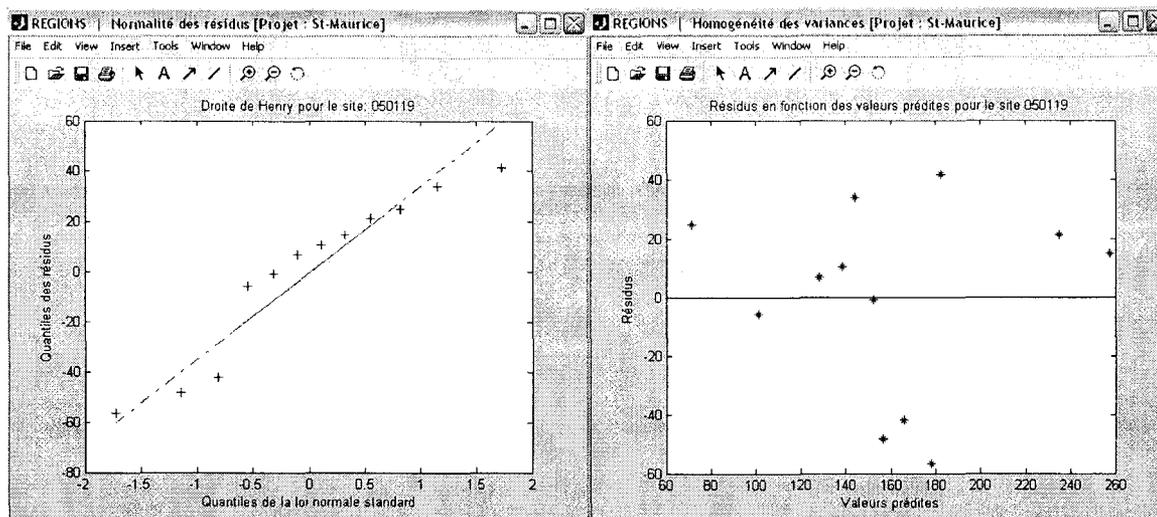


Figure 2.24 Graphiques de validation du modèle régressif

Il est important de noter que ces outils donnent des indications importantes concernant la validité du modèle régressif, mais ces dernières sont toutefois quelque peu subjectives.

De plus, un tableau, suivant celui présentant l'estimation des variables, présente l'estimation des paramètres du modèle régressif. En effet, on y retrouve, pour chaque variable hydrologique d'intérêt, les estimations (avec intervalles de confiance) de l'ordonnée à l'origine et des paramètres associés aux différentes variables explicatives choisies.

Enfin, on précise tout au bas de la boîte d'affichage les variables explicatives employées de même que la taille minimale recommandée pour chaque voisinage/classe.

Le bouton "Exporter" permet d'exporter le contenu de la boîte d'affichage "Résultats" dans un fichier texte, tandis que le bouton "Imprimer" l'envoie directement à l'imprimante. Ces deux boutons possèdent leur équivalent dans le menu défilant "Fichier" (options "Exporter

les résultats" et "Imprimer les résultats". Ces boutons peuvent s'avérer utile lorsqu'on veut faire une étude des résultats obtenus en fonction des variables incluses dans le modèle de régression, puisqu'il permet de garder une trace de ces différents essais. Finalement, lorsque l'utilisateur en a terminé avec ce module, il n'a qu'à appuyer sur le bouton "Fermer" pour le quitter et revenir au menu principal.

2.9 Module de l'indice de crue

La méthode de l'indice de crue (Dalrymple, 1960) pour l'estimation régionale des crues repose sur l'hypothèse de base que les données aux différents sites d'une région homogène sont indépendantes et suivent la même distribution statistique à un facteur d'échelle près. Conséquemment, on suppose alors que certaines caractéristiques statistiques (tel que le coefficient de variation ou le coefficient d'asymétrie) sont constantes à l'intérieur d'une même région. La méthode de l'indice de crue est composée des étapes suivantes :

1. Standardisation des données : À chaque site i , les données (Q_i^T) sont standardisées en les divisant par la variable choisie comme étant l'indice de crue ($INDC_i$). La moyenne est souvent utilisée, bien que la médiane puisse aussi être utilisée.

$$X_i^T = \frac{Q_i^T}{INDC_i}$$

2. Calage des paramètres de la distribution régionale : La distribution moyenne standardisée de la région, $F_r(x)$, est alors déterminée en regroupant les données standardisées des différents sites, puis le quantile adimensionnel $X_r(T)$ de la distribution de X_r est estimé par :

$$\hat{X}(T) = F_r^{-1}\left(1 - \frac{1}{T}; \hat{\theta}\right)$$

Où T est la période de retour, et $\hat{\theta}$ est le vecteur des paramètres estimés pour la distribution choisie.

3. Estimation des quantiles de la distribution régionale en un point non jaugé :

L'indice de crue au site cible S est estimé par régression sur les caractéristiques physiographiques/météorologiques des bassins, et le quantile désirée $Q_s(T)$ au site cible est obtenue en multipliant l'indice de crue estimé avec le quantile standardisé correspondant.

2.9.1 Choix du type de régions homogènes

On accède à ce module à partir du menu principal en appuyant sur le bouton "Indice de crue". Cependant, avant que ne s'ouvre la fenêtre du module proprement dit, l'utilisateur voit d'abord apparaître la fenêtre illustrée à la Figure 2.25, intitulée "Choix du type de DRH pour la régression" où l'abréviation DRH signifie Détermination des Régions Homogènes. On doit y préciser de quelle façon les régions que l'on souhaite considérer dans l'analyse ont été déterminées. S'il s'agit de regroupements géographiques (non contigus) qui ont été obtenus soit à l'aide du module de classification hiérarchique (section 2.4) ou encore par le biais du module de définition manuelle des classes (section 2.5), on doit appuyer sur le bouton radio "Classification hiérarchique (ou Régions imposées)" ; s'il s'agit de voisinages hydrologiques obtenus grâce au module d'analyse canonique des corrélations, on doit plutôt appuyer sur le bouton radio "Analyse des corrélations (approche par classification)" (section 2.7) ou "Analyse des corrélations (approche par optimisation du paramètre alpha)" (section 2.6). Il est important de préciser ici que rien n'empêche que des régions homogènes de ces deux types puissent exister en même temps pour un projet donné, mais on devra toutefois les traiter séparément au moment de l'estimation régionale. Une fois que le bon bouton radio a été sélectionné, il n'y a plus qu'à appuyer sur le bouton "OK" pour passer au module de l'indice de crue.

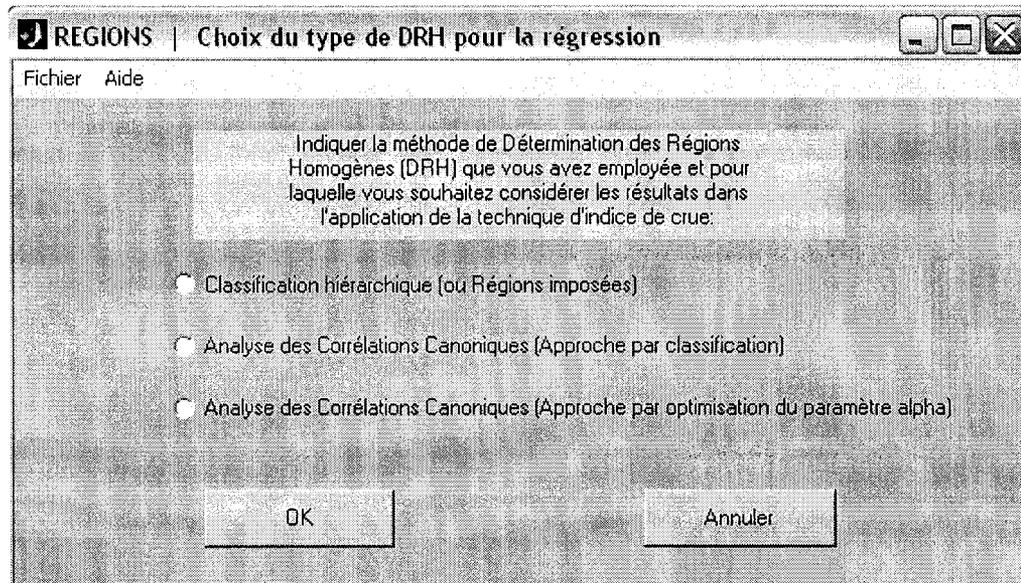


Figure 2.25 Fenêtre de spécification du type de régions homogènes à considérer au moment de l'exécution du module de l'indice de crue

2.9.2 Sélection du modèle régressif

Le module est présenté à la Figure 2.24 dans le cas où l'utilisateur a opté pour l'option « Analyse des corrélations canoniques (Approche par classification) ». Lorsque l'utilisateur opte pour des regroupements de type géographique, l'interface du module demeure pratiquement la même, excepté que le bouton "Choisir les variables physiographiques / météorologiques à inclure dans le modèle de régression" est scindé en deux boutons dénommés "Choisir les variables à régionaliser" et "Choisir les variables explicatives". Cette différence est due au fait que, contrairement aux modules d'analyse canonique des corrélations (par classification ou par optimisation du paramètre alpha), les variables que l'on cherche à régionaliser ne sont pas explicitées directement dans les modules de classification hiérarchique et de définition manuelle des classes. Les variables à régionaliser sont donc reportées directement dans la boîte d'affichage "Variables à régionaliser" (conformément à ce qui a été entré au module d'analyse canonique) lorsqu'on utilise des voisinages, mais si ce n'est pas le cas, on doit les spécifier à l'aide du bouton "Choisir les variables à régionaliser".

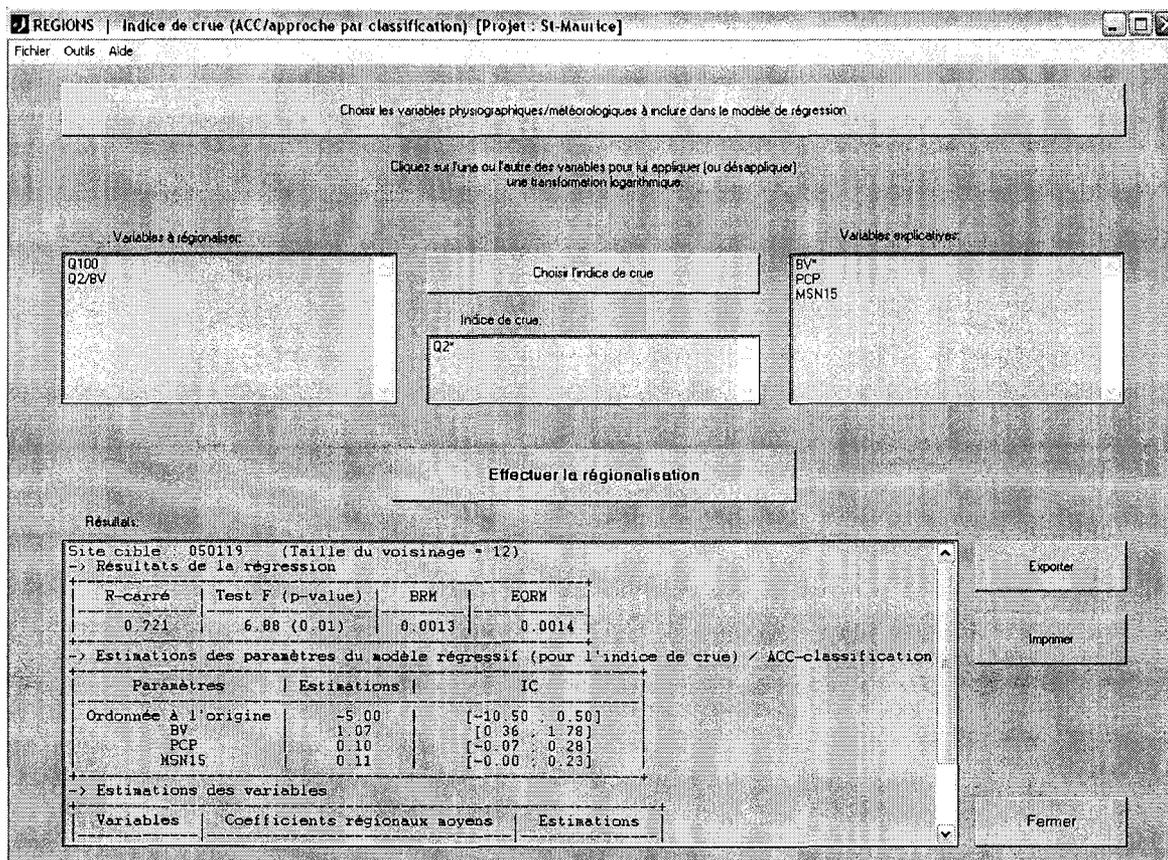


Figure 2.26 Fenêtre de l'indice de crue (option voisinages – ACC/approche par classification)

Les variables physiographiques/météorologiques utilisées pour l'analyse canonique sont elles aussi reportées automatiquement dans la boîte d'affichage "Variables explicatives", tout comme le seront les variables de classement employées au module de classification hiérarchique, le cas échéant. Bien qu'il soit naturel de considérer les variables ayant servi à la définition des régions homogènes comme prédicteurs dans le modèle régressif, il est toujours possible de modifier cette liste à l'aide des boutons "Choisir les variables physiographiques / météorologiques à inclure dans le modèle de régression" et "Choisir les variables explicatives", selon le cas.

Ensuite, à l'aide du bouton « Choisir l'indice de crue », il est possible de choisir parmi les variables hydrologiques celle qui sera utilisée comme étant l'indice de crue. Le quantile de période de retour 2 ans est souvent utilisé, puisqu'il correspond à la médiane.

Tout comme au module de statistiques descriptives, il est possible de cliquer sur l'une ou l'autre des variables présentes dans les deux boîtes d'affichage afin de lui appliquer une transformation logarithmique (ce qui est de nouveau indiqué par un astérisque vis-à-vis les variables concernées). À ce propos, le menu défilant "Outils" offre l'option "Vérifier la normalité des variables retenues" permettant de valider graphiquement (tout comme au module de statistiques descriptives, voir la Figure 2.9 à la section 2.3) l'hypothèse de normalité pour l'une ou l'autre des variables retenues en la présentant sur papier normal (en coordonnées réelles et logarithmiques).

Une fois que les variables réponses (hydrologiques), les variables explicatives (physiographiques/météorologiques) et la variable qui sera utilisée comme indice de crue ont été sélectionnées, il suffit d'appuyer sur le bouton "Effectuer la régionalisation" afin d'obtenir les estimations recherchées. Une étape supplémentaire s'ajoute toutefois si l'on considère des regroupements de type géographique puisque l'on demandera alors à l'utilisateur pour quels sites (i.e. les stations non jaugées) l'estimation doit être effectuée (car cette discrimination n'est pas requise lors de l'exécution des modules de classification hiérarchique et de définition manuelle des classes).

2.9.3 Description des résultats de la régionalisation

Les résultats de la régression sont reportés dans la boîte d'affichage intitulée "Résultats". On y présente tout d'abord un tableau donnant les résultats de la régression effectuée pour estimer la variable choisie comme indice de crue. Dans le tableau même, on retrouve donc quelques statistiques relatives à la performance du modèle régressif, soit la valeur du coefficient de détermination (le R^2), la valeur du test F global de signification, le biais relatif moyen ainsi que l'écart quadratique relatif moyen (ces deux dernières statistiques étant calculées sur la base des écarts entre valeurs prédites et mesurées). Tel que mentionné à la section 2.8.3, il est aussi possible de vérifier la validité du modèle régression en choisissant, dans le menu "outils", l'option "Validation du modèle régressif" (voir Figures 2.23 et 2.24 de la section 2.8.3).

Au-dessus du premier tableau de résultats indiqué la taille du voisinage (ou du regroupement) associé au site cible. Or, en régression, il est recommandé que la taille de l'échantillon ne soit pas inférieure au nombre de paramètres du modèle (i.e. nombre de prédicteurs + 1) multiplié par 3, sans quoi on peut émettre de sérieuses réserves sur la qualité de l'estimation. Des avertissements préviennent l'utilisateur à cet effet chaque fois que cette condition n'est pas respectée. Ensuite, un tableau présente l'estimation des paramètres du modèle régressif pour l'indice de crue. En effet, les estimations (avec intervalle de confiance) de l'ordonnée à l'origine et des paramètres associés aux différentes variables explicatives choisies sont présentées.

Finalement, un dernier tableau présente les coefficients régionaux moyens et les estimations des variables réponses choisies. Les coefficients régionaux correspondent à la moyenne des X_i^T (étape 2 de la méthode de l'indice de crue) pour chaque site i faisant partie de la région homogène associée au site cible. Pour obtenir l'estimation d'une variable, on multiplie le coefficient régional moyen associé à cette variable par l'estimation de l'indice de crue obtenue par régression pour le site cible. La répartition de ces coefficients régionaux peut être affichée pour chaque site cible. On y accède par le menu "Outils", sous "Voir la répartition des coefficients régionaux" (Figure 2.27).

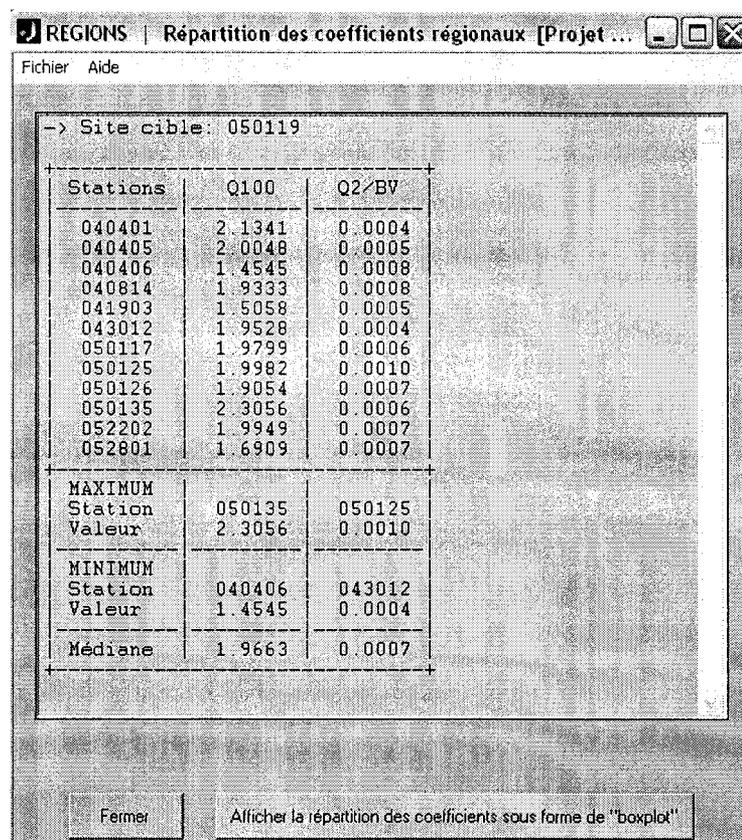


Figure 2.27 Fenêtre de répartition des coefficients régionaux (module d'indice de crue)

À l'aide de ces coefficients, on peut donc juger de la qualité du coefficient régional moyen pour un site cible (et ainsi, répondre à la question suivante, par exemple : est-il représentatif des différents coefficients régionaux?) et ainsi juger aussi de la qualité de l'estimation. À noter que ces coefficients peuvent aussi être affichés sous forme de "boxplot", ce qui permet de juger visuellement de la situation. Par ailleurs, lorsqu'on clique sur le bouton "Afficher la répartition des coefficients sous forme de boxplot", le graphique est affiché simultanément pour chacune des variables.

Enfin, on précise, tout au bas de la boîte d'affichage, les variables explicatives et l'indice de crue employés de même que la taille minimale recommandée pour chaque voisinage/classe.

Le bouton "Exporter" permet d'exporter le contenu de la boîte d'affichage "Résultats" dans un fichier texte, tandis que le bouton "Imprimer" l'envoie directement à l'imprimante. Ces

deux boutons possèdent leur équivalent dans le menu défilant "Fichier" (options "Exporter les résultats" et "Imprimer les résultats". Ces boutons peuvent s'avérer utile lorsqu'on veut faire une étude des résultats obtenus en fonction des variables incluses dans le modèle de régression, puisqu'il permet de garder une trace de ces différents essais. Finalement, lorsque l'utilisateur en a terminé avec ce module, il n'a qu'à appuyer sur le bouton "Fermer" pour le quitter et revenir au menu principal.

2.10 Module de combinaison bayésienne de l'information régionale et locale

Ce module a été conçu pour améliorer l'estimation des quantiles en utilisant à la fois l'information régionale (obtenue par régression linéaire) et les observations locales. Pour pouvoir utiliser ce module, trois quantiles doivent être inclus dans les variables hydrologiques du projet. On accède à ce module en cliquant sur le bouton "Régression multiple régionale+information locale" dans la fenêtre principale de *REGIONS*. Tel que pour la méthode de l'indice de crue ou la régression multiple, l'utilisateur doit tout d'abord spécifier la méthode utilisée pour déterminer les régions homogènes (voir Figure 2.21 à la section 2.8).

Aussitôt après, une fenêtre apparaît pour demander à l'utilisateur de choisir trois quantiles (par ordre croissant de période de retour ou par ordre décroissant de probabilité d'occurrence) parmi les variables hydrologiques et d'en préciser les probabilités associées (Figure 2.28).

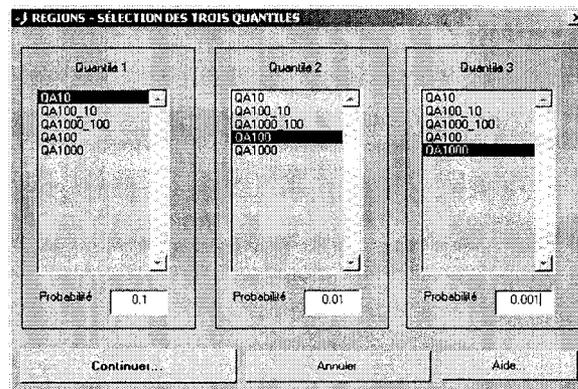


Figure 2.28 Spécification des trois quantiles et de leurs probabilités respectives

Les probabilités saisies dans les boîtes de texte doivent impérativement correspondre au quantile choisi dans la liste au dessus (Figure 2.28). Le programme effectue ensuite une analyse régionale (calcul des quantiles et des différences de quantiles par régression linéaire multiples et correction du biais dû à la transformation logarithmique; calcul de la matrice de variance-covariance des résidus de la régression) et en affiche les résultats (Figure 2.29). Pour plus de détails concernant la théorie derrière ce module, le lecteur pourra se rapporter à Seidou *et al.* (2005a, 2005b).

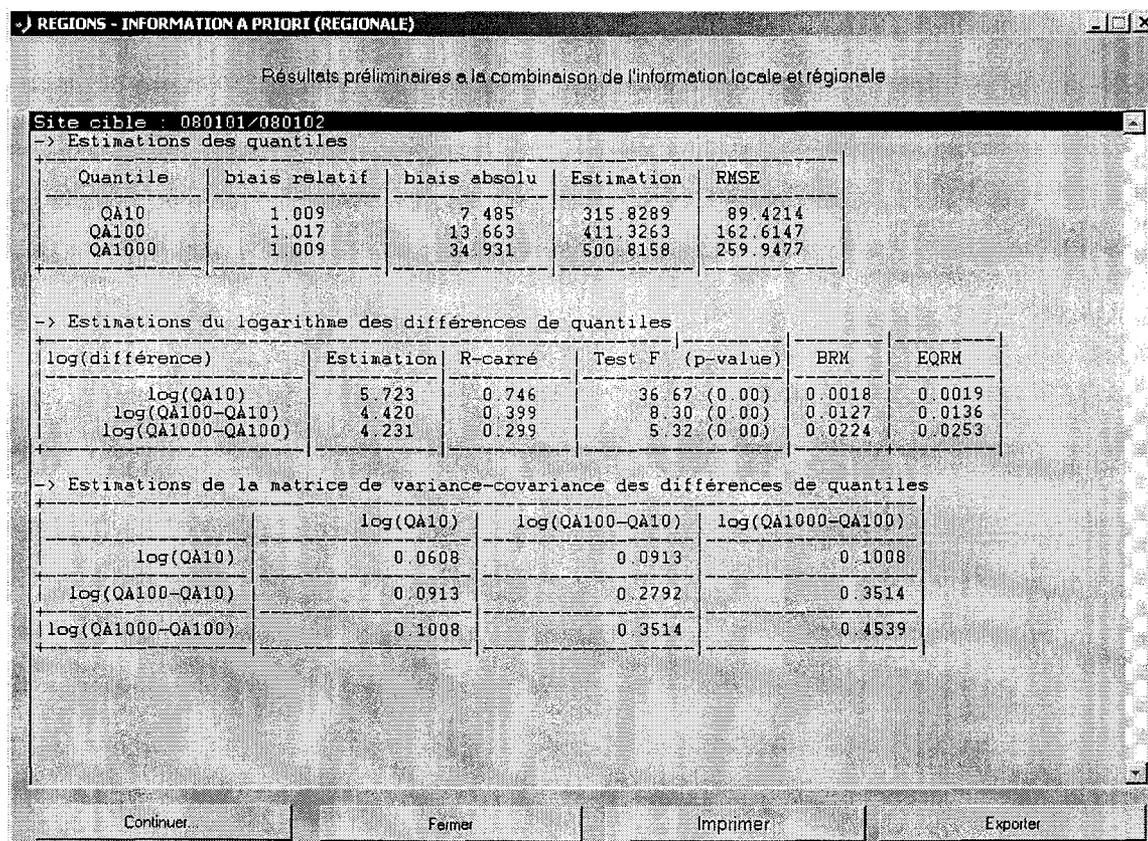


Figure 2.29 Affichage des résultats de l'analyse régionale

Lorsque l'utilisateur clique sur le bouton "continuer", il accède à la fenêtre permettant d'appliquer la méthode bayésienne empirique de combinaison de l'information locale et régionale (Figure 2.30). Il doit alors charger les données locales (Figure 2.31), qui doivent être dans un fichier texte (*.txt), en une seule colonne.

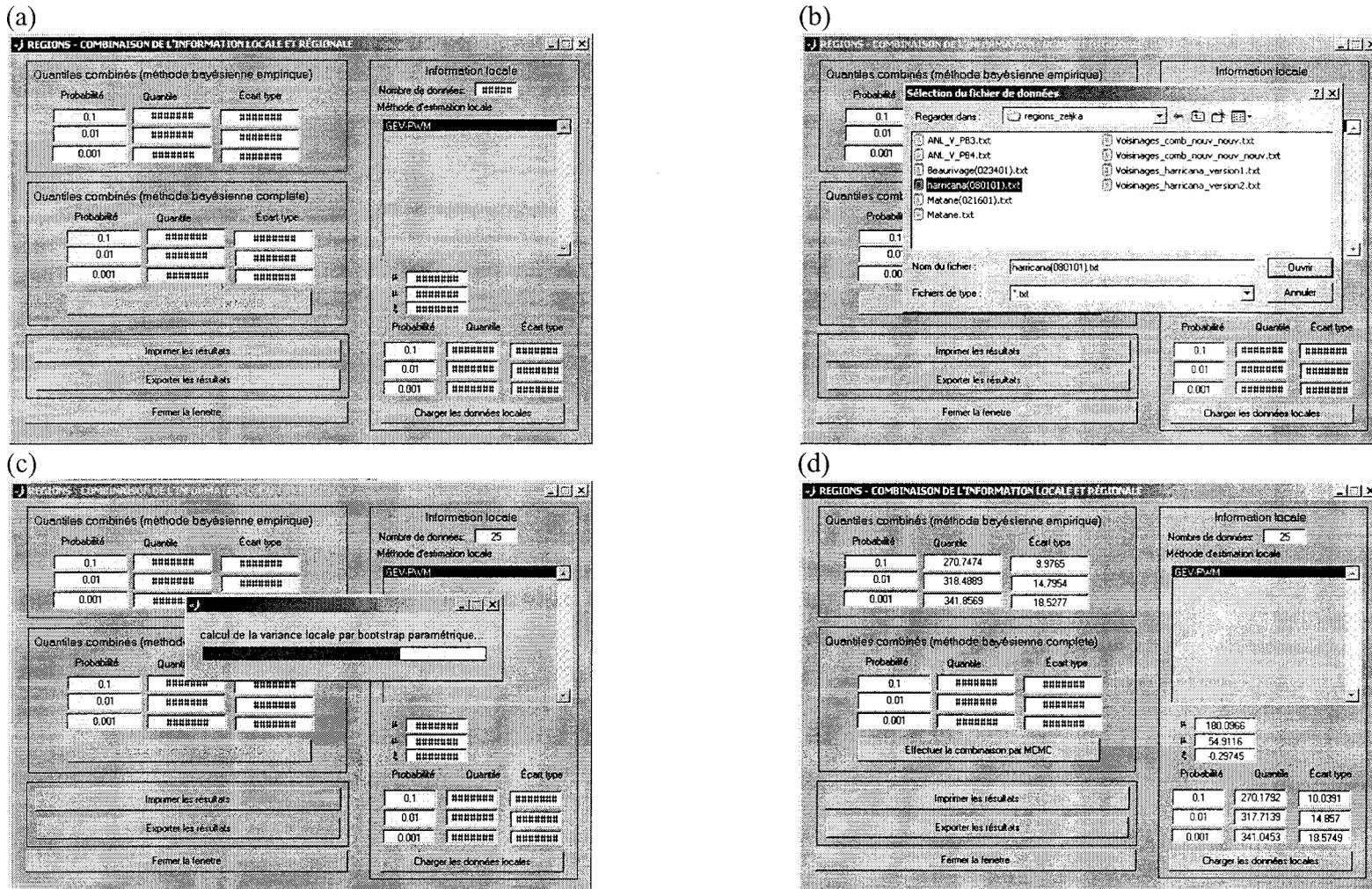


Figure 2.30 Méthode bayésienne empirique de combinaison de l'information locale et régionale: a) fenêtre d'accueil; b) chargement des données locales; c) calcul des paramètres et des quantiles locaux par bootstrap paramétrique; d) affichage des résultats

Une fois la méthode bayésienne empirique appliquée, l'utilisateur a le choix d'utiliser la méthode bayésienne complète en cliquant sur "Effectuer la combinaison par MCMC". La fenêtre affichée à la Figure 2.31 apparaît et lui permet de fixer le nombre total d'itérations désiré ainsi que la "période de chauffe" (nombre d'itérations qui ne seront pas prises en compte lors de l'inférence sur les paramètres).

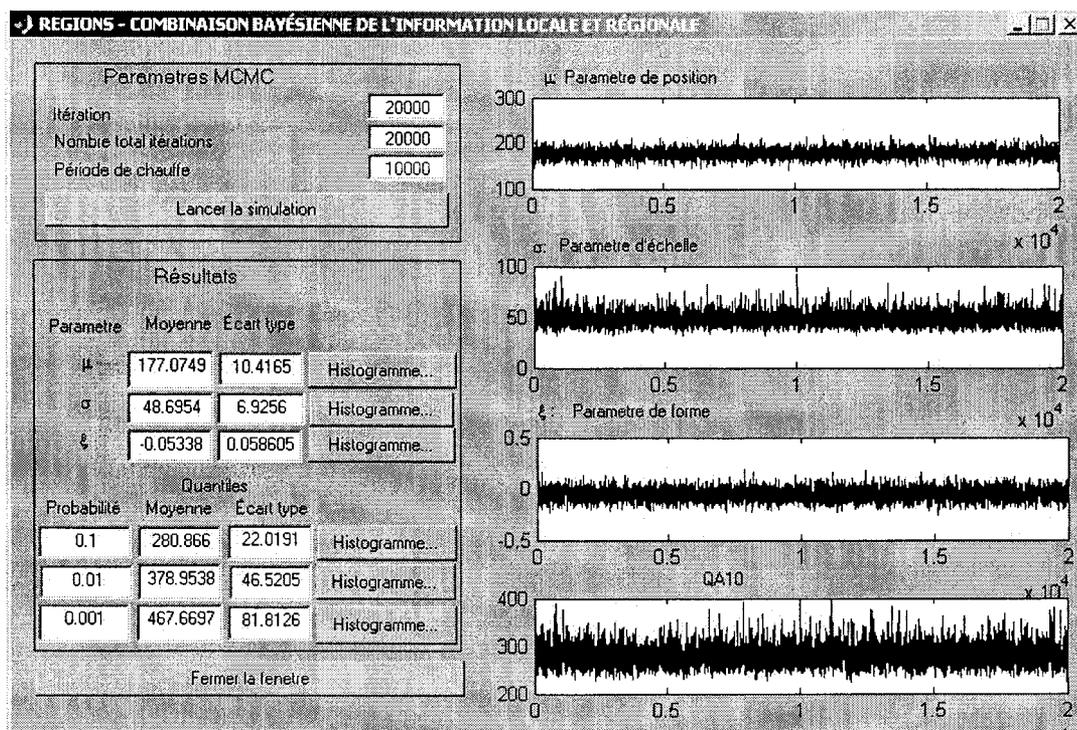


Figure 2.31 Combinaison de l'information locale et régionale par MCMC

Une fois les paramètres de l'algorithme MCMC fixés, l'utilisateur clique sur le bouton "Lancer la simulation" et attend entre quelques secondes et quelques minutes (selon le nombre d'itérations choisis) l'affichage des résultats. Sur la partie droite de la fenêtre, il peut voir l'évolution des paramètres de la distribution GEV et du premier quantile ce qui lui permet de juger visuellement de la convergence de l'algorithme. Il peut recommencer si nécessaire en augmentant le nombre d'itérations ou le "temps de chauffe". L'estimation des quantiles et des paramètres est alors affichée à gauche de l'écran. L'utilisateur peut voir les histogrammes d'un paramètre ou d'un quantile en cliquant sur le bouton "histogramme" située à la droite de la valeur obtenue (Figure 2.32).

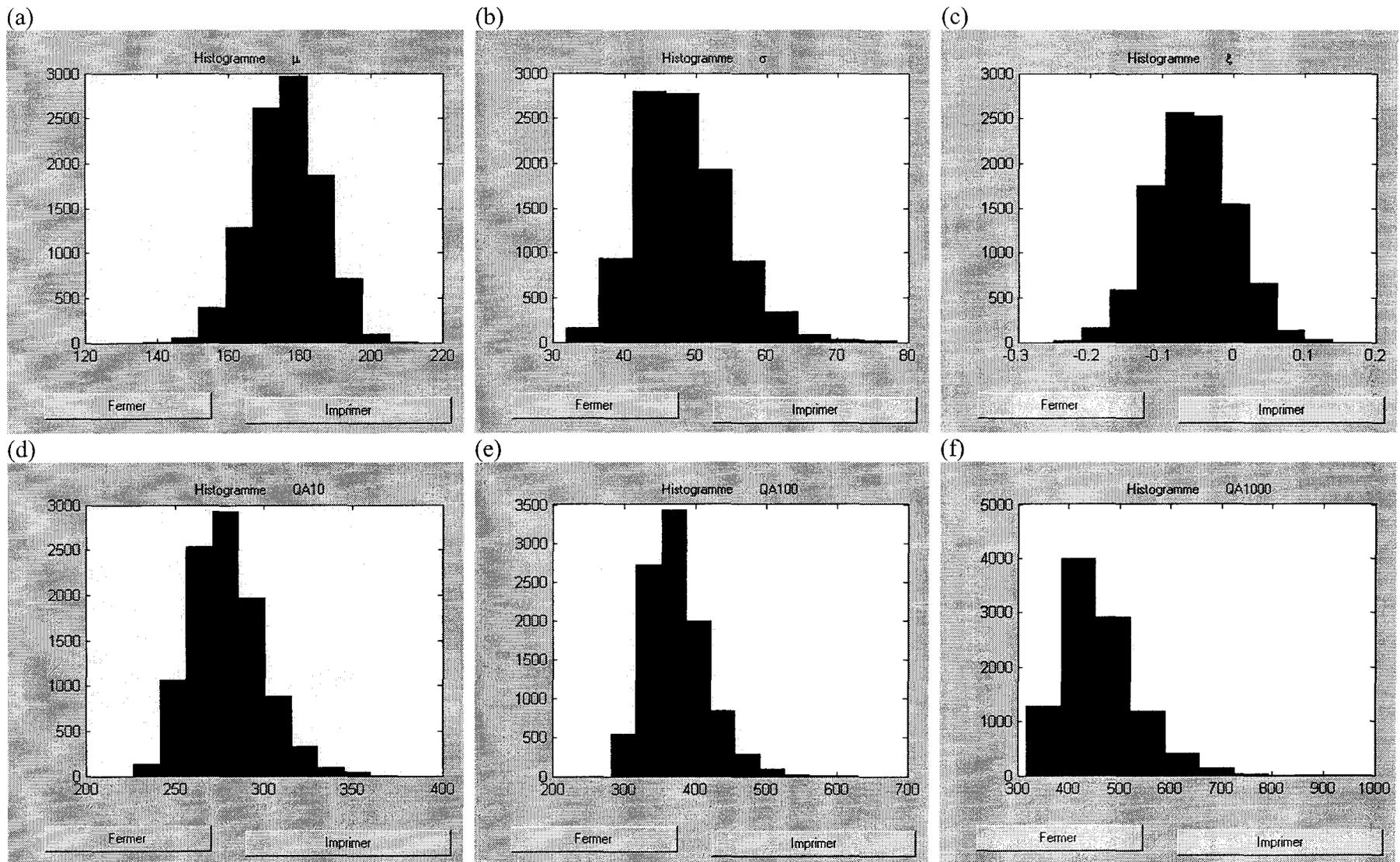
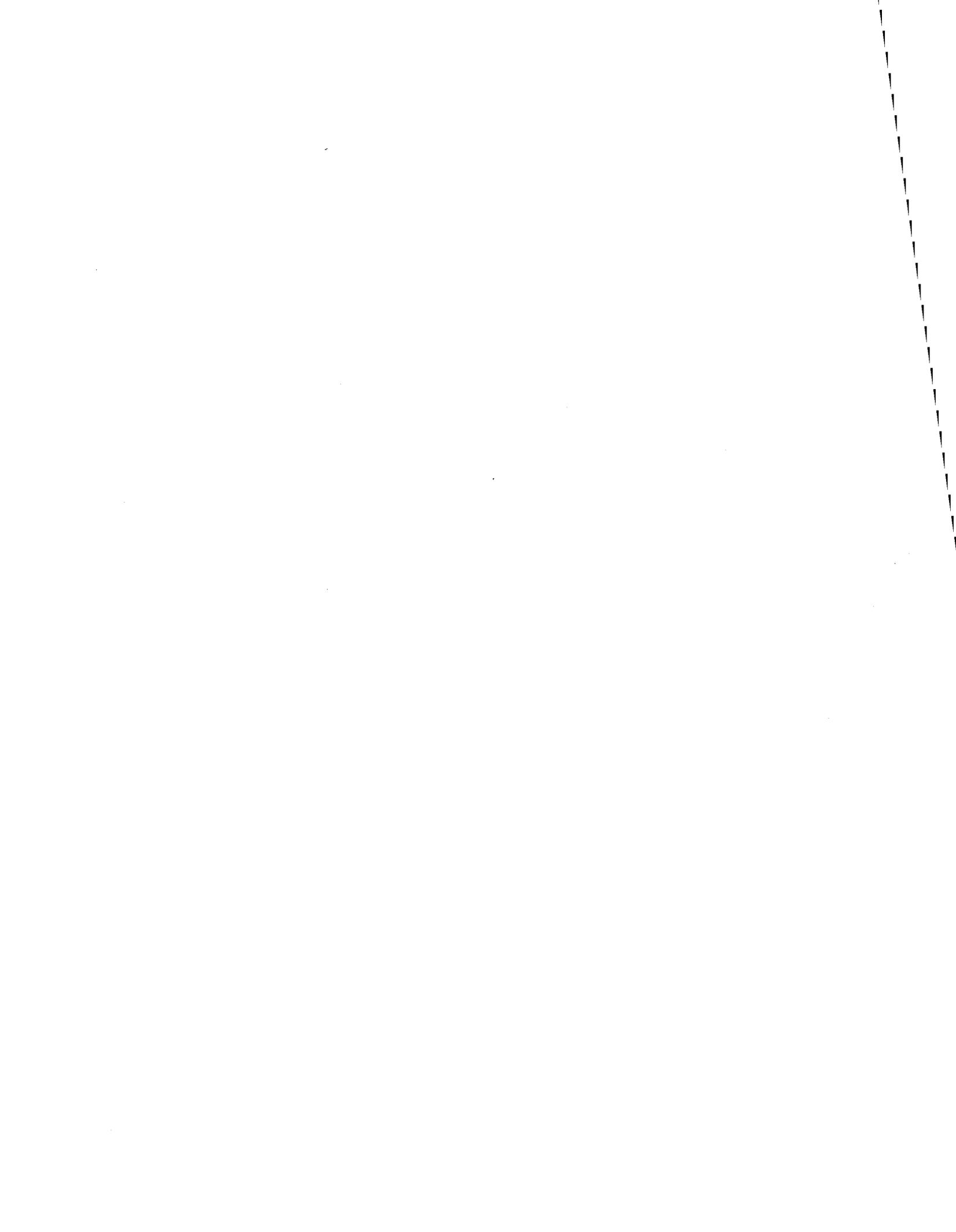


Figure 2.32 Histogrammes des paramètres et des quantiles: a) μ ; b) σ ; c) ξ ; d) q_{10} ; e) q_{100} ; f) q_{1000}



2.11 Fenêtre de commentaires

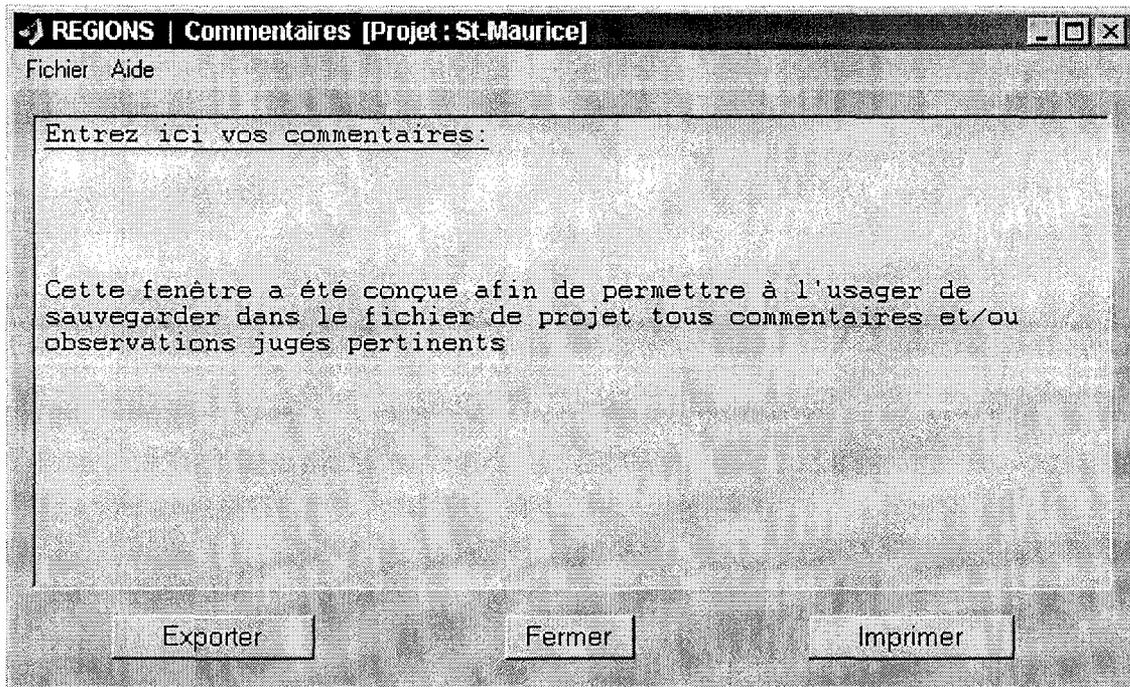


Figure 2.33 Fenêtre de commentaires

La fenêtre de commentaires, présentée à la Figure 2.33 est accessible à partir de toutes les fenêtres de *REGIONS* (excepté les fenêtres graphiques) via l'option "Commentaires" de leur menu défilant "Fichier", sauf pour le menu principal où l'on retrouve cette option dans le menu défilant "Projet". Elle sert, comme son nom l'indique, à consigner des annotations relatives au projet en cours, celles-ci étant sauvegardées automatiquement dans le fichier de projet. Le bouton "Exporter" permet d'exporter le contenu de cette fenêtre dans un fichier texte et le bouton "Imprimer" lui dirige ce contenu vers l'imprimante. Le bouton "Fermer" permet de quitter la fenêtre. Notons en terminant que les trois boutons de la fenêtre ont leur équivalent du même nom dans le menu défilant "Fichier".



RÉFÉRENCES

Dalrymple, T., 1960 : Flood frequency analysis. *US Geol. Surv. Water Supply Pap.*, 1543A.

Girard, C, T.B.M.J. Ouarda et B. Bobée, 2000 : Une approche par classification à la constitution de voisinages homogènes basés sur l'ACC. INRS-ETE, Rapport de recherche No R-576

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

Ouarda, T.B.M.J., Haché, M., Bobée, B. (1998). Rapport final du projet C5 : Régionalisation des événements hydrologiques extrêmes. INRS-Eau, Rapport de recherche No R-534.

Seidou, O., Ouarda, T.B.M.J., Barbet, M., Bruneau, P. and Bobée, B. (2005a). Une méthode bayésienne de combinaison de l'information locale et régionale en analyse fréquentielle des crues. Rapport de recherche R-789, ISBN 2-82146-498-2, INRS-ETE, Québec.

Seidou, O., Ouarda, T.B.M.J., Barbet, M., Bruneau, P. and Bobée, B. (2005b). A bayesian combination of local and regional information flood frequency analyses. to be submitted to *Water Resources Research*

The MathWorks, Inc. (2000). *Using Matlab, Version 6.*

