

UNIVERSITE DU QUEBEC

THESE

présentée

à

L'INSTITUT NATIONAL DE LA RECHERCHE SCIENTIFIQUE

comme exigence partielle

de la

maîtrise es Sciences (Eau)

par

Olivier Cazaillet

Ingénieur hydraulicien

CONTRIBUTION A LA RATIONALISATION D'UN RESEAU DE MESURE
DE QUALITE DE L'EAU: APPLICATION AU RESEAU DE MESURE
DU BASSIN DE LA RIVIERE YAMASKA

Septembre 1977

REMERCIEMENTS

Nous tenons à remercier particulièrement monsieur Bernard Bobée qui nous a proposé le choix du sujet de ce mémoire, nous a dirigé tout au long de sa réalisation et nous a conseillé dans toutes les corrections finales de ce travail.

Nous tenons à souligner notre reconnaissance à l'égard de monsieur Marius Lachance qui nous a aidé dans l'établissement des programmes de calcul sur ordinateur et nous a aidé à utiliser les programmes d'ajustement de lois statistiques et d'analyse factorielle des correspondances mis au point à l'INRS-Eau.

Nous remercions aussi madame Lise Potvin pour ses conseils lors de l'étude géologique du bassin versant de la rivière Yamaska et monsieur André Tessier pour tous les renseignements qu'il nous a fournis sur les problèmes de qualité et d'analyse chimique de l'eau.

Notre gratitude va également à messieurs Pierre Campbell et Daniel Cluis, professeurs à l'INRS-Eau qui ont bien voulu revoir et commenter la première version de ce travail ainsi qu'à M. Jean Rousselle, professeur à l'école Polytechnique qui a été correcteur externe de cette thèse.

Nous exprimons également notre gratitude à monsieur André Parent pour l'élaboration des figures et des graphiques, à madame Francine Bordeleau pour le travail de dactylographie de ce mémoire ainsi qu'à monsieur Magella Cantin pour les travaux d'édition.

Je tiens aussi à témoigner ma reconnaissance à tous ceux qui ont rendu mon séjour au Québec agréable et m'ont aidé de leur amitié.

TABLE DES MATIERES

REMERCIEMENTS	i
LISTE DES TABLEAUX	viii
LISTE DES FIGURES	xi
LISTE DES SYMBOLES	xiii
INTRODUCTION	1
1. PRINCIPES DE LA RATIONALISATION D'UN RESEAU DE MESURE	
DE QUALITE DE L'EAU	3
1.1 Schéma de réalisation d'un réseau de mesure	4
1.2 Les objectifs et la planification du réseau	7
1.2.1 Choix des paramètres de qualité	9
1.2.2 Choix des caractéristiques statistiques	9
1.2.3 Choix des critères de qualité.....	10
1.2.4 Choix des précisions et des niveaux de confiance	11
1.2.5 Choix des méthodes statistiques	12
1.2.6 Choix des méthodes de mesure.....	12
1.2.7 Choix de la densité et de la localisation des stations.....	12
1.3 Réalisation du plan d'échantillonnage	13
1.3.1 Caractérisation des paramètres de qualité.....	14
1.3.1.1 Paramètres de qualité.....	14
1.3.1.2 Caractéristiques statistiques	15
1.3.1.3 Précision, niveau de confiance	17
1.3.1.4 Critères de qualité	17
1.3.2 Densité et localisation des stations	19
1.3.2.1 Répartition des stations du réseau.....	19
1.3.2.2 Choix de l'emplacement d'une station.....	23
1.3.2.3 Exemples de méthodes utilisées	24

TABLE DES MATIERES (SUITE)

1.3.3	Dates et fréquences d'échantillonnage	26
1.3.3.1	Cas de séries de mesures aléatoires.....	27
1.3.3.2	Cas de séries de mesures non aléatoires.....	28
1.3.3.3	Estimation de la variabilité temporelle.....	31
1.4	Collecte, transport, analyse des échantillons.....	32
1.4.1	Limite de détection et précision d'analyse.....	33
1.4.2	Méthode d'analyse.....	33
1.5	Traitement des données.....	34
1.6	Cas d'un réseau existant.....	35
2.	METHODES STATISTIQUES	37
2.1	Caractéristiques statistiques d'une série d'observations.....	37
2.1.1	Moyenne, écart-type, coefficient de variation	38
2.1.1.1	Intervalle de confiance sur la moyenne.....	38
2.1.1.2	Précision sur la moyenne.....	39
2.1.2	Maximum, minimum	40
2.1.3	Quantile.....	40
2.1.3.1	Intervalle de confiance d'un quantile.....	41
2.1.3.2	Précision sur un quantile	43
2.1.4	Coefficient d'asymétrie.....	45
2.1.5	Coefficient d'auto-correlation.....	48
2.2	Structure des séries d'observations.....	49
2.2.1	Généralités.....	50
2.2.2	Test de normalité.....	50
2.2.3	Processus de Markov d'ordre 1.....	52
2.2.3.1	Vérifications.....	54
2.2.3.2	Prédiction.....	54
2.2.3.3	Contenu d'information de la moyenne d'une série chronologique.....	55

TABLE DES MATIERES (SUITE)

2.2.4	Ajustement de lois statistiques	58
2.2.4.1	Intérêt des lois	58
2.2.4.2	Conditions d'application et choix des lois	59
2.2.4.3	Ajustement des lois	60
2.2.4.4	Intervalles de confiance	62
2.2.4.5	Lois utilisées	63
2.3	Relations entre séries d'observations. Analyse globale des séries d'observation	64
2.3.1	Corrélation - Régression linéaire	65
2.3.1.1	Corrélation	65
2.3.1.2	Régression linéaire	66
2.3.1.3	Contenu d'information de la moyenne régionale	68
2.3.2	Analyse des correspondances	71
3.	DESCRIPTION DU BASSIN VERSANT DE LA RIVIERE YAMASKA ET DU RESEAU DE MESURE DE QUALITE	76
3.1	Description du bassin versant	76
3.1.1	Physiographie	76
3.1.2	Géologie	77
3.1.3	Pédologie	79
3.1.4	Hydrographie	79
3.1.5	Population, industrie, utilisation du sol	81
3.1.6	Problèmes de qualité de l'eau sur le bassin	81
3.2	Description du réseau de mesures	83
3.2.1	Localisation et densité du réseau	83
3.2.2	Dates et fréquence de mesures.....	83
3.2.3	Paramètres de qualité mesurés	87

TABLE DES MATIERES (SUITE)

3.3 Objectifs	89
3.3.1 Objectifs du réseau actuel	89
3.3.2 Objectifs proposés	91
4. ETUDE GENERALE DE LA QUALITE DE L'EAU ET DU RESEAU DE	
MESURE DU BASSIN VERSANT DE LA RIVIERE YAMASKA	93
4.1 Généralités	94
4.2 Etude descriptive générale	95
4.2.1 Moyennes, écarts-types, maxima, coefficients de variation	96
4.2.2 Précision sur la moyenne	105
4.2.3 Dépassement de critères de qualité	108
4.2.4 Représentation graphique	112
4.2.5 Résumé et conclusion	114
4.3 Représentation synthétique des paramètres de qualité	117
4.3.1 Indépendance des observations	118
4.3.2 Etude des paramètres de qualité persistents	120
4.3.2.1 Adaptation d'un processus de Markov d'ordre 1	121
4.3.2.2 Réduction de la fréquence de mesure	126
4.3.2.3 Conclusion	129
4.3.3 Etude des paramètres de qualité aléatoires	129
4.3.3.1 Coefficients d'asymétrie - Test de normalité	130
4.3.3.2 Ajustement de loi	133
4.3.3.3 Conclusion	140

TABLE DES MATIERES (SUITE)

4.4	Etude de la variabilité spatiale du réseau	140
4.4.1	Relations entre stations	141
4.4.1.1	Etude des coefficients de corrélation	141
4.4.1.2	Régressions linéaires	150
4.4.1.3	Contenu d'information de la moyenne régionale	153
4.4.1.4	Conclusion	154
4.4.2	Etude globale des ions majeurs	155
4.4.2.1	Réalisation de l'analyse des correspondances	155
4.4.2.2	Résultats	156
4.4.2.3	Etudes des "points-paramètres"	157
4.4.2.4	Etude des "points-stations"	162
4.4.2.5	Etude conjointe des points-paramètres et des points-stations	164
4.5	Conclusion	167
5.	ETUDE D'UN OBJECTIF: EAU POTABLE	171
5.1	Définition des objectifs	171
5.1.1	Choix des objectifs	171
5.1.2	Définition des objectifs	172
5.2	Réponses du réseau actuel	174
5.2.1	Connaissance de la moyenne	174
5.2.2	Ajustements de lois statistiques	174
5.2.3	Intervalles de confiance	192
5.2.4	Probabilité de dépassement de critères de qualité	196
5.2.5	Précision atteinte sur le quantile 90 au niveau de confiance 95%	198

TABLE DES MATIERES (SUITE)

5.3	Correction du réseau pour atteindre les objectifs visés	199
5.3.1	Correction de la fréquence de mesure: cas de la moyenne	199
5.3.2	Correction de la fréquence de mesure: cas du quantile 90	200
5.4	Conclusions	201
CONCLUSION		203
REFERENCES		205
ANNEXE A-1		208
ANNEXE A-2		225
ANNEXE A-3		239

LISTE DES TABLEAUX

3.1	Liste des stations de mesure	85
3.2	Paramètres de qualité étudiés (unité de mesure, symbole, limite de détection)	88
3.3	Critères de qualité utilisés	90
4.1	Moyennes des paramètres de qualité par station	97
4.2	Ecart-types des paramètres de qualité par station.....	98
4.3	Maxima observés des paramètres de qualité à chaque station.....	99
4.4	Classement relatif des cours d'eau d'après l'étude des moyennes, écarts-types et maxima	101
4.5	Coefficients de variation des paramètres de qualité à chaque station	104
4.6	Précisions sur la moyenne des paramètres de qualité à chaque station (en pourcentage de la moyenne) au niveau de confiance 95%..	106
4.7	Maxima observés, situation des stations par rapport aux critères de qualité envisagés	109
4.8	Coefficients d'auto-corrélation de chaque série d'observations	119
4.9	Coefficients d'auto-corrélation des résidus des séries d'observations ($x_{i+1} - t_1 x_i$) à chaque station et pour chaque paramètre	123
4.10	Tests de normalité sur les résidus des séries d'observations ($x_{i+1} - t_1 x_i$) à chaque station et pour chaque paramètre de qualité	124
4.11	Contenu d'information de la moyenne des séries d'observations	127
4.12	Nombre de valeurs indépendantes effectives (permettant d'obtenir la même information sur la moyenne du paramètre de qualité).....	128
4.13	Coefficients d'asymétrie des paramètres de qualité par station	131
4.14	Tests de normalité	132

LISTE DES TABLEAUX (SUITE)

4.15 Valeurs observées classées aux stations 16 et 17 du paramètre conductivité et probabilité empirique correspondante	134
4.16 Ajustement de la loi Gamma par la méthode du maximum de vraisemblance (station 16, Conductivité)	136
4.17 Ajustement de la loi Gamma par la méthode du maximum de vraisemblance (station 17, Conductivité)	137
4.18 Coefficients de corrélation significatifs entre stations pour les paramètres du groupe "ions majeurs"	142
4.19 Coefficients de corrélation significatifs entre stations pour les paramètres du groupe "substances nutritives"	143
4.20 Couples de stations à forts coefficients de corrélation	144
4.21 Analyse factorielle sur les ions majeurs: interprétation des résultats	165
5.1 Valeurs classées et probabilités empiriques (station 12, Manganèse)	176
5.2 Caractéristiques de l'échantillon. Test d'indépendance des observations (station 12, Manganèse)	177
5.3 Ajustement de la loi Gamma par la méthode du maximum de vraisemblance (station 12, Manganèse)	178
5.4 Ajustement de la loi Pearson type 3 par la méthode des moments avec la correction d'asymétrie	179
5.5 Ajustement de la loi Log-Gamma par la méthode des moments appliquée à la série des valeurs observées	180
5.6 Ajustement de la loi Log-Pearson type 3 par la méthode des moments sur la série des valeurs observées	181
5.7 Intervalles de confiance de la loi Gamma à 95% Manganèse, station 12 (méthodes 1 et 2)	193

LISTE DES TABLEAUX (SUITE)

5.8	Probabilités de dépassement de critères de qualité (manganèse, stations 12, 15, 16; fer, stations 12, 18)	196
5.9	Précision sur le quantile 90 au niveau de confiance (95%) (manganèse, stations 12, 15, 16; fer, stations 12, 18)	198
5.10	Nombre de mesures indépendantes nécessaires à l'obtention d'une précision de 20% au niveau de confiance 95% sur la moyenne des paramètres fer et manganèse aux stations 12, 15, 16 et 18	199
5.11	Nombre de mesures indépendantes nécessaires à l'obtention d'une précision de 25% sur le quantile 90, au niveau de confiance 95% (fer station 12, 18; manganèse stations 12, 15, 16)	200

LISTE DES FIGURES

1.1	Schéma de planification d'un réseau d'acquisition de données de qualité de l'eau	5
2.1	Schéma d'étude des séries d'observations	51
3.1	Localisation des stations d'échantillonnage du réseau	84
3.2	Schéma du réseau de stations	86
4.1	Distribution spatiale des concentrations moyennes d'ions majeurs pour la période d'étude (février 1974 - novembre 1975)	113
4.2	Ajustement des lois normale et Gamma (conductivité, station 16)	138
4.3	Ajustement des lois normale et Gamma (conductivité, station 17)	139
4.4	Schéma des groupes de stations fortement corrélées	149
4.5	Analyse factorielle des correspondances des ions majeurs dans le plan des axes 1 et 2: étude des points paramètres	158
4.6	Analyse factorielle des correspondances des ions majeurs dans le plan des axes 1 et 2: étude des points stations	159
4.7	Visualisation des résultats de l'analyse des correspondances appliquée aux ions majeurs	166
5.1	Ajustement des lois Gamma et Pearson type 3 (manganèse, station 12)	182
5.2	Ajustement des lois log-Gamma et log-Pearson type 3 (manganèse, station 12).....	183
5.3	Ajustement des lois Gamma et Pearson type 3 (manganèse, station 15)	184
5.4	Ajustement des lois log-Gamma et log-Pearson type 3 (manganèse, station 15)	185
5.5	Ajustement des lois Gamma et Pearson type 3 (manganèse, station 16)	186

LISTE DES FIGURES (SUITE)

5.6	Ajustement des lois log-Gamma et log-Pearson type 3 (manganèse, station 16)	187
5.7	Ajustement des lois Gamma et Pearson type 3 (fer, station 12)	188
5.8	Ajustement des lois log-Gamma et log-Pearson type 3 (fer, station 12)	189
5.9	Ajustement des lois Gamma et Pearson type 3 (fer, station 18)	190
5.10	Ajustement des lois log-Gamma et log-Pearson type 3 (fer, station 18)	191
5.11	Comparaison des intervalles de confiance à 95% de la loi Gamma, méthodes 1 et 2 (manganèse, station 12)	194
5.12	Intervalles de confiance à 95% de la loi log-Pearson type 3 (fer, station 12)	195

LISTE DES SYMBOLES

$F(x)$	Fonction de répartition d'une variable aléatoire
$f(x)$	Fonction densité de probabilité d'une variable aléatoire
$E(X)$	Espérance mathématique de la variable aléatoire X
$\text{var}(X)$	Variance de la variable aléatoire X
N	Nombre d'observations à une station
n	Nombre de stations
$(x_1, x_2, \dots, x_N)$	Echantillon de N observations à une station
μ_k	Moment centré d'ordre k de la population
m_k	Moment centré d'ordre k de l'échantillon
$\hat{m}_k$	Moment centré non biaisé d'ordre k
μ	Moyenne de la population
$\bar{x}$	Moyenne de l'échantillon
σ	Ecart-type de la population
s	Ecart-type estimé

LISTE DES SYMBOLES (SUITE)

C_v	Coefficient de variation de la population
$\bar{C}_v$	Coefficient de variation estimé
q_p	Quantile d'ordre p de la population (probabilité au dépassement 1 - p)
Q_p	Quantile estimé d'ordre p
C_s	Coefficient d'asymétrie de la population
$\bar{C}_s$	Coefficient d'asymétrie de l'échantillon
$\hat{C}_s, (X_s)_1, (C_s)_2, (C_s)_3$	Estimateurs de C_s
X_p	Événement de probabilité au dépassement donnée
$\hat{X}_p$	Estimateur de X_p
τ_1	Coefficient d'auto-corrélation d'ordre 1
$x_{i+1} = \tau_1 x_i + \varepsilon_{i+1}$	Expression d'un processus de Markov d'ordre 1
τ_ε	Coefficient d'auto-corrélation d'ordre 1 de la série des résidus ε_i
$I(m_k, N)$	Contenu d'information du moment d'ordre k d'une population estimé à partir d'un échantillon de taille N

LISTE DES SYMBOLES (SUITE)

I_r	Contenu d'information relatif de deux séries de même taille
I, I_1, I_2	Contenus d'information
ρ	Coefficient de corrélation
ρ_{ij}	Coefficient de corrélation de la station i avec la station j
r	Coefficient de corrélation estimé
Y_i	Valeur prédite par une régression linéaire
a, b	Coefficients de la droite de régression
$\bar{\rho}$	Coefficient de corrélation moyen de l'ensemble des stations du réseau
$\bar{y}_i$	Moyenne estimée à la station i
$\bar{Y}$	Moyenne régionale estimée
$\sigma_{\bar{Y}}^2$	Variance de la moyenne régionale
l, l_0	Erreurs d'échantillonnage
ϵ, ϵ_0	Précisions
α	Seuil de signification d'un test

LISTE DES SYMBOLES (SUITE)

$(1 - \alpha)$	Niveau de confiance d'un intervalle de confiance ou d'une précision atteinte
u	Variable normale centrée réduite
d.d.l.	Abréviation pour "degré de liberté"
t_v	Variable de Student à v d.d.l.

INTRODUCTION

Les problèmes de qualité de l'eau se posent aujourd'hui dans tous les pays, de façon cruciale.

Les utilisations de l'eau sont très variées et à des usages distincts correspondent autant de contraintes de qualité plus ou moins strictes. On évalue la qualité de l'eau pour telles de ses utilisations au moyen des données que fournissent les réseaux de mesure. Toute la connaissance qu'on obtient sur cette qualité dépend donc de la planification adéquate de l'acquisition des données. Les réseaux sont très coûteux en matériel et en général leur durée d'exploitation couvre souvent plusieurs années; le plus grand soin doit donc être apporté à leur conception.

L'objet de ce travail est de mettre en évidence l'importance d'une planification rigoureuse des réseaux de mesure de qualité. On s'attache à montrer comment une définition très précise des objectifs de qualité auxquels le réseau doit répondre permet de planifier ce réseau de façon rationnelle, c'est-à-dire de réaliser le meilleur compromis entre le coût et l'efficacité.

Le premier chapitre envisage de façon théorique la rationalisation d'un réseau: la démarche et les problèmes qui se posent. Dans la suite nous réalisons l'étude d'un réseau de mesure déjà implanté sur le bassin versant de la rivière Yamaska et dont la durée d'opération est de deux ans. Ce réseau fut construit sans objectifs précis tel qu'on l'entend dans ce travail. Nous effectuons cependant le traitement des données recueillies et leur interprétation pour en tirer le maximum d'information. Le souci d'une

rationalisation ultérieure éventuelle du réseau oriente cette étude et justifie les diverses méthodes statistiques utilisées.

Un dernier chapitre montre enfin comment la définition précise d'un objectif (étude des paramètres fer et manganèse par rapport à l'eau d'alimentation) permet la détermination rationnelle des caractéristiques du plan d'échantillonnage.

CHAPITRE 1

PRINCIPES DE LA RATIONALISATION D'UN RESEAU

DE MESURE DE QUALITE DE L'EAU

INTRODUCTION: nécessité de la rationalisation

Les campagnes de mesure sont coûteuses en personnel sur le terrain ou dans les laboratoires, en frais de conservation, de transport, d'analyse des échantillons et en coût de traitement des données.

On constate souvent que les données récoltées ne permettent pas de répondre aux questions qu'on se pose sur la qualité de l'eau pour tel ou tel usage. Il arrive que les paramètres pertinents n'aient pas été mesurés alors qu'on dispose par contre de paramètres dont la connaissance est peu utile; parfois, c'est la fréquence des mesures qui est insuffisante et ne permet pas de suivre correctement la variabilité temporelle du paramètre désiré. Il se peut que la précision sur les résultats finaux déçoive ou, au contraire, soit inutilement grande eu égard à l'utilisation qu'on veut faire des résultats.

Toutes ces lacunes ont pour cause commune le manque de souci de rationalisation dans l'élaboration du plan d'échantillonnage.

Le manque de planification qui accompagne certaines campagnes de mesure est dû à l'absence de définition claire des buts que le réseau doit atteindre ou à une définition trop vague de ses objectifs tel que "étude de la qualité générale de l'eau", comme le signale WILSON (1974).

Le but de tous les réseaux et campagnes de mesure est certes d'évaluer "la qualité de l'eau". Mais que signifie en soi cette notion? On peut consommer l'eau, pratiquer des sports nautiques, irriguer des champs, approvisionner une industrie alimentaire, etc... Les critères de qualité varieront selon l'usage et la notion de qualité de l'eau acquiert un sens clair en liaison avec un usage précis.

De même un réseau de mesure de qualité de l'eau doit avoir pour finalité de donner des renseignements sur la qualité de l'eau en vue de telle ou telle utilisation. Il doit répondre à un (ou des) objectif précis. Et ce n'est qu'une fois ces objectifs clairement définis qu'on peut bâtir un réseau rationnel d'acquisition de données.

En conclusion on peut dire que la rationalisation d'un réseau de mesure consiste en son implantation et en son exploitation adéquate de façon à atteindre au moindre coût et avec le maximum d'efficacité les objectifs qui ont présidé à sa mise en place.

1.1 Schéma de réalisation d'un réseau

On donne à la figure 1.1 un plan général d'acquisition de données de qualité qui indique les diverses étapes de la démarche à suivre.

Quand on a identifié les diverses utilisations de l'eau dont on se préoccupe, dans un certain cadre géographique (bassin versant, cours d'eau, etc...), comme, par exemple, l'alimentation en eau municipale ou industrielle, ainsi que les problèmes existants ou susceptibles de se poser, comme les rejets urbains ou l'eutrophisation des lacs, on est en mesure de définir précisément les objectifs que le réseau doit atteindre.

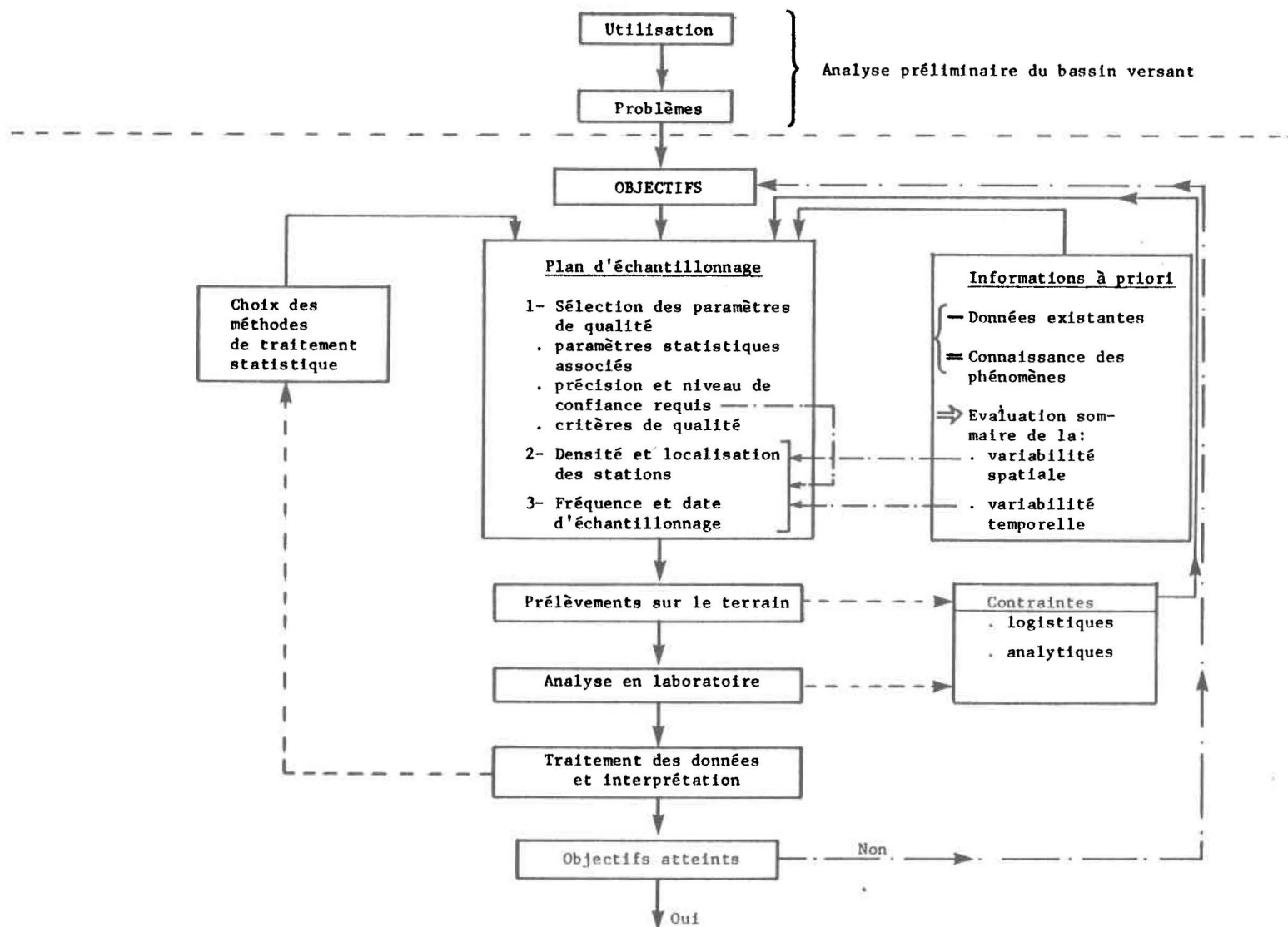


Figure 1-1. Schéma de planification d'un réseau d'acquisition de données de qualité de l'eau

Nous envisagerons plus loin, en 1.2, toute l'implication directe et indirecte de la définition des objectifs à tous les niveaux de la planification du programme d'acquisition de données.

On peut ensuite établir le plan d'échantillonnage qui comprend la connaissance de trois composantes essentielles:

- la sélection des paramètres de qualité à mesurer et des caractéristiques statistiques afférentes (moyennes, quantiles) pour lesquelles on effectuera le traitement de données et qui permettront de quantifier les résultats (quoi mesurer);
- la densité et la localisation des stations d'échantillonnage (où mesurer);
- la fréquence des mesures de chaque paramètre à chaque station (quand mesurer).

Par ailleurs la réalisation du plan d'échantillonnage dépend de divers facteurs comme:

- l'existence et la pertinence de l'information disponible à priori sur le bassin qui permet une première évaluation de la variabilité temporelle et spatiale des paramètres nécessaire à l'élaboration du plan d'échantillonnage;
- des contraintes logistiques et analytiques;
- le choix des méthodes de traitement de données; ce choix peut, par exemple, nécessiter l'acquisition de données concomitantes à des stations différentes.

WILSON (1974) signale qu'il est important de consacrer tous les efforts et le temps nécessaire à la planification rationnelle et rigoureuse

des réseaux. Cette réflexion préliminaire permettra par la suite l'acquisition de données pertinentes et des économies importantes en ressources humaines et financières.

Quand le plan d'échantillonnage est complètement établi, on réalise les opérations successives de prélèvements sur le terrain, d'analyse en laboratoire, et, finalement, de traitement des données. On est alors en mesure de constater si les objectifs sont atteints et de modifier éventuellement le plan d'échantillonnage. Cette modification tient compte des données acquises et, dans ce sens, la rationalisation du réseau a un caractère dynamique.

Nous allons détailler les différents stades de ce schéma de rationalisation.

1.2 Les objectifs et la planification du réseau

C'est la définition précise des objectifs qui conditionne la rationalisation adéquate du réseau, c'est-à-dire sa réalisation au moindre coût pour une efficacité maximale dans ses réponses.

La définition des objectifs s'effectue souvent à partir des utilisations de l'eau sur le bassin et des problèmes susceptibles de se poser. Les objectifs auxquels un réseau peut répondre sont très divers. Chacun est spécifique quant aux paramètres à mesurer, aux précisions à atteindre, etc...

Plusieurs usages spécifiques de l'eau pour lesquels un réseau peut apporter l'information nécessaire peuvent être cités:

- alimentation en eau potable;
- vie aquatique (à différents niveaux selon que l'on veut conserver, par exemple, la présence de salmonidés ou simplement celle de la barbotte);

- récréation, sports nautiques;
- eau pour l'irrigation;
- eau pour un type d'industrie donné.

On peut aussi se préoccuper d'évaluer diverses pollutions:

- rejets ponctuels humains ou industriels;
- rejets diffus d'origine agricole;

Des objectifs assez généraux comme l'alimentation en eau potable devront être traduits en pratique de façon précise par des énoncés du type suivant: "connaissance, dans les zones habitées du bassin susceptibles de recevoir des prises d'eau, de la concentration moyenne de fer avec une précision de 20% sur cette concentration moyenne à un niveau de confiance de 95%".

Le choix de l'objectif intervient dans l'organisation du réseau à tous les stades de la planification:

- choix des paramètres de qualité;
- choix des caractéristiques statistiques (moyenne, quantile, etc...) dont la connaissance permet de répondre au problème posé;
- choix des précisions à obtenir sur les caractéristiques statistiques pour un certain niveau de confiance (d'où l'on peut déduire par des méthodes statistiques les fréquences de mesure);
- choix des critères d'évaluation de la qualité (normes);
- choix des méthodes statistiques de traitement des données;
- choix des méthodes analytiques de mesure;
- choix de la densité et de la localisation des stations du réseau.

1.2.1 Choix des paramètres de qualité

Connaissant l'objectif du réseau, on peut déterminer les paramètres physico-chimiques ou biologiques dont la connaissance est strictement nécessaire et suffisante pour atteindre cet objectif.

Ainsi il sera probablement difficile de répondre convenablement à des objectifs concernant l'alimentation en eau potable ou la récréation si on a omis de mesurer la présence de coliformes fécaux. Par contre, la mesure des bicarbonates, de la silice ou de la demande chimique en oxygène (D.C.O.) pourrait être une perte de temps et d'argent dans ce cas. Il est important de définir à priori les objectifs à atteindre et d'en déduire les paramètres à mesurer sans tenir compte des facilités qu'on a pour mesurer tel ou tel paramètre, car on serait alors tenté de mesurer des paramètres inutiles, sous le faux prétexte qu'un jour leur connaissance pourrait être utile. Ainsi la mesure systématique des ions majeurs, qu'on pratique souvent, n'apporte que peu d'information vis-à-vis de nombreux usages de l'eau.

Des études se poursuivent pour associer à chaque utilisation de l'eau un ensemble de paramètres pertinents. Ce type d'études est très intéressant vis-à-vis du problème de la rationalisation des réseaux de qualité et le développement de cette tendance est souhaitable.

1.2.2 Choix des caractéristiques statistiques

Selon l'objectif et le paramètre de qualité considéré on s'intéressera à la valeur moyenne, à la variance, aux valeurs extrêmes, etc... de ce paramètre de qualité; il existe en effet un grand nombre de caractéristiques statistiques qu'on peut étudier.

Ainsi, pour un objectif d'alimentation en eau potable, on peut s'intéresser à la concentration moyenne de manganèse au cours de la période de basses eaux ou au quantile 90 de la distribution des concentrations de ce paramètre au cours de l'année, c'est-à-dire à la concentration de manganèse qui n'est dépassée que 10% des fois, de façon à comparer cette valeur à la norme à ne pas dépasser pour que l'eau soit potable.

Remarquons que le choix des caractéristiques statistiques aura une incidence sur le choix des méthodes qui devront être utilisées dans le traitement des données. Ainsi l'étude d'un quantile de la distribution des concentrations d'un paramètre chimique nécessite la connaissance de la distribution statistique associée à ce paramètre. On devra alors estimer la loi de distribution à partir de l'échantillon. Cette étude n'est, par contre, pas nécessaire si on s'intéresse uniquement à la moyenne ou à la médiane d'un paramètre.

1.2.3 Choix des critères de qualité

Les critères de qualité relatifs à un même paramètre de qualité varient suivant l'utilisation de l'eau. Par exemple, la norme relative au fer est plus stricte pour l'eau de boisson que pour l'eau d'alimentation des usines de pâte à papier.

Le type de critère de qualité à respecter peut induire le choix des caractéristiques statistiques à étudier ainsi que la précision à atteindre sur ces caractéristiques. Il influe donc indirectement sur les méthodes de traitement statistique appropriées à l'étude des caractéristiques statistiques choisies et sur la fréquence des mesures qui dépend de la précision qu'on veut atteindre. Un critère de qualité peut, par exemple, fixer une concentration limite qui ne doit pas être dépassée en moyenne durant une cer-

taine période ou qui ne doit pas être dépassée dans plus de 5% des cas avec une certaine précision sur cette évaluation ($5\% \pm 0.5\%$).

1.2.4 Choix des précisions et des niveaux de confiance

Un objectif comme l'alimentation en eau potable peut nécessiter par exemple la connaissance de la concentration de mercure dépassée au plus dans 5% des cas avec une très bonne précision car la toxicité de cet élément peut être forte.

La théorie statistique permet de calculer avec quelle probabilité cette précision est atteinte; c'est ce qu'on appelle le niveau de confiance. En général, on le fixe à 80, 90 ou 95%. On cherche, par exemple, à obtenir une caractéristique statistique avec une précision ϵ qu'on se fixe pour un niveau de confiance de 95%: c'est-à-dire qu'il n'y a que 5 chances sur 100 que l'erreur d'échantillonnage soit plus grande que celle qu'on tolère. Pour une variabilité donnée cette contrainte permet de déterminer le nombre (N) d'observations (supposées indépendantes) à recueillir afin de satisfaire cet objectif de précision. Ainsi la connaissance de la précision désirée pour un certain niveau de confiance est nécessaire pour déterminer la fréquence de mesures du paramètre de qualité. Plus la précision désirée est grande, plus grand devra être le nombre de mesures pour l'atteindre. Il est donc fondamental, quand on définit les objectifs du réseau, d'associer à chaque paramètre de qualité pertinent la caractéristique statistique qui servira à le quantifier, accompagnée de la précision désirée à un certain niveau de confiance. La définition précise des objectifs permettra de justifier telle ou telle précision désirée.

1.2.5 Choix des méthodes statistiques

Le choix des caractéristiques statistiques et de leur précision, des critères de qualité sur les paramètres de qualité correspondants, implique un choix des méthodes statistiques pertinentes.

Des méthodes plus ou moins fines, longues et complexes seront à mettre en oeuvre selon la précision à atteindre, la quantité de données à traiter, selon que l'objectif du réseau demande une connaissance globale de la variabilité spatiale des paramètres sur le bassin ou seulement la connaissance de quelques caractéristiques statistiques en des points précis.

1.2.6 Choix des méthodes de mesure analytique

Il existe souvent plusieurs méthodes de mesure d'un même paramètre de qualité. Ces méthodes diffèrent dans leurs exigences (équipement, personnel), leur facilité de mesure, leur précision, leur justesse, leur limite de détection et donc dans leur coût. Il est clair que la définition des objectifs permettra de faire un choix des méthodes appropriées, ainsi qu'une sélection des modes de conservation, de transport des échantillons que nécessite l'emploi de ces méthodes. Des coûts importants sont impliqués à ce stade de la planification et il importe donc de bien savoir dans quel but précis on réalise telle ou telle analyse.

1.2.7 Choix de la densité et de la localisation des stations

L'objectif visé par l'implantation du réseau permet en général de déterminer le nombre et l'emplacement des stations.

Par exemple, si on étudie l'impact des rejets d'eaux résiduaires, on installera des stations en aval des sources de rejet et, éventuellement, en amont, comme témoins. Si on recherche des zones homogènes pour le comporte-

ment de certains paramètres de qualité de façon à découper un bassin versant en régions susceptibles d'être consacrées à la récréation, à l'agriculture, au développement industriel, un réseau uniformément réparti sur le bassin pourra dans ce cas être adéquat. Le nombre de stations dépendra aussi, pour une grande part, de la variabilité spatiale des paramètres de qualité.

On comprend, après cette discussion, toute l'importance d'une définition aussi précise que possible des objectifs d'un réseau. Il est nécessaire que ces objectifs soient quantifiés.

L'objectif "qualité de l'eau d'alimentation" se traduira par un ensemble d'énoncés, relatifs aux divers paramètres de qualité pertinents, du type suivant: "connaissance du quantile q_p (concentration dépassée dans $(1 - p)\%$ des cas) de la distribution du fer sur le bassin versant avec une précision de 20% au niveau de confiance 95%."

Remarquons de plus que la définition précise des objectifs et la planification qui pourra s'en suivre permettront de chiffrer correctement les coûts du réseau. Réciproquement les moyens financiers dont on dispose à priori joueront un grand rôle. Il faudra peut-être faire un compromis entre l'objectif à atteindre et les coûts que sa réalisation implique. Ces coûts interviendront de façon discriminante pour effectuer des sélections au niveau des paramètres de qualité, des méthodes et caractéristiques statistiques, des méthodes d'analyse et surtout des précisions requises qui détermineront le nombre d'échantillons à prélever.

1.3 Réalisation du plan d'échantillonnage

Pour établir le plan d'échantillonnage on doit choisir les paramètres à mesurer, fixer la densité et la localisation des stations et, enfin

déterminer la fréquence de mesure et les dates de prélèvement. Nous discutons en détail ces éléments dans ce qui suit.

1.3.1 Les paramètres de qualité

La détermination des paramètres de qualité nécessite l'identification:

- des paramètres physico-chimiques, bactériologiques, etc...;
- des caractéristiques statistiques associées à chaque paramètre;
- des précisions et niveaux de confiance pour les caractéristiques choisies;
- des critères de qualité pertinents à chaque paramètre en fonction des utilisations considérées.

1.3.1.1 Choix des paramètres

On ne doit pas choisir des paramètres en raison de la facilité de leur mesure, sous le prétexte que ces données pourraient devenir utiles, mais l'on doit effectuer ce choix en fonction des objectifs du réseau. WILSON (1974) souligne d'autre part que le choix des paramètres doit être fait sans ambiguïté. On sait, en effet, que de nombreuses substances se présentent dans l'eau sous diverses formes physiques (particulaires ou solubles) ou chimiques (cristallines ou précipitées pour la forme particulière, libres ou complexées pour la forme soluble, par exemple). C'est le problème de la spéciation d'un élément. Ainsi l'azote se présente sous forme inorganique soluble dans l'ammoniaque, les nitrates et les nitrites. On le rencontre aussi sous forme organique dans les organismes vivants ou morts, adsorbé sur les sédiments, en solution.

On s'intéresse fréquemment à un paramètre appelé "phosphore hydrolysable" qui représente la quantité de phosphore dosé après hydrolyse des

divers composés organiques ou inorganiques contenant du phosphore. Certains auteurs inclueront dans la définition de ce paramètre les orthophosphates qui sont pourtant déjà sous forme hydrolysée. Selon d'autres, les orthophosphates forment une classe à part.

Il est essentiel de préciser à quelle forme du paramètre on s'intéresse car le choix de la méthode analytique et la précision des résultats peuvent en dépendre. De nombreux paramètres de qualité sont sujets au problème de la spéciation. Citons les métaux, l'azote, le phosphore, le carbone et la silice.

Dans le même ordre d'idées, WILSON (1974) fait remarquer que certains paramètres de qualité représentent en réalité une classe de composés divers qui peuvent être présents ou non dans l'échantillon. C'est le cas des composés phénolés qui représentent un ensemble de diverses substances phénolées distinctes, des pesticides, etc... Il serait préférable d'indiquer à quel composé particulier on s'intéresse, de façon à pouvoir choisir la méthode de mesure appropriée, ou de spécifier si on s'intéresse particulièrement à un paramètre intégrateur comme, par exemple, l'azote total.

1.3.1.2 Les caractéristiques statistiques

On synthétise l'information désirée sur les paramètres de qualité au moyen de caractéristiques statistiques dont les plus classiques sont la moyenne, l'écart-type, le coefficient de variation, le maximum, un quantile donné, etc... Si on s'intéresse à un paramètre qui fluctue peu au cours de l'année, comme cela pourrait être le cas du calcium ou de la conductivité, l'étude de la moyenne pourra s'avérer suffisante, surtout si les valeurs de ces paramètres se trouvent bien en deçà des normes prescrites.

Si on considère un paramètre qui fluctue beaucoup, de façon aléatoire, ou dont les valeurs sont systématiquement élevées, et risquent de dépasser les normes prescrites, on peut considérer le maximum observé et le comparer à la norme. Mais il se peut qu'on n'ait pas échantillonné au moment où se sont produites les fortes valeurs ou, au contraire, que le maximum observé soit exceptionnel et ne constitue pas, bien qu'il dépasse la norme, un danger réel. Il sera bon alors d'étudier la loi statistique qui représente le mieux la distribution des valeurs du paramètre de qualité et de considérer comme caractéristique statistique d'étude un quantile q_p de cette distribution (valeur dépassée $(1 - p)\%$ des fois). C'est le quantile qu'on pourra alors comparer à la norme.

Ainsi à chaque paramètre de qualité peut être associée une ou plusieurs caractéristiques statistiques pertinentes qui décriront au mieux le paramètre de qualité par rapport à l'objectif étudié.

Le choix des caractéristiques statistiques est important pour éviter des calculs inutiles et surtout pour répondre correctement aux objectifs. Ainsi la moyenne renseignera très mal et pourra même induire en erreur sur un dépassement de normes, dans le cas d'un paramètre qui fluctue beaucoup. Si le réseau a été établi pour l'étude d'une moyenne, les données recueillies avec une certaine précision sur cette moyenne ne permettront pas, par la suite, de déterminer un quantile avec la précision suffisante car le nombre de mesures, suffisant pour atteindre une certaine précision sur la moyenne, sera trop faible dans le cas du quantile.

La détermination des caractéristiques statistiques a également une incidence sur le traitement ultérieur des données. Le problème du choix des caractéristiques statistiques doit donc se poser en même temps que le choix des paramètres de qualité.

1.3.1.3 Précision, niveau de confiance

De manière générale il est possible de calculer la précision associée à une caractéristique statistique donnée d'un paramètre de qualité lorsque sa distribution est connue.

Pour un même paramètre de qualité, la précision désirée sur une caractéristique statistique peut varier selon l'objectif poursuivi; on exigera ainsi une plus grande précision si le paramètre est relatif à l'eau de boisson que s'il est relatif à l'agriculture, les normes à respecter étant généralement plus strictes dans le premier cas que dans le second.

Il faudrait toujours indiquer la précision désirée car c'est grâce à cette connaissance qu'on pourra calculer la fréquence de mesure du paramètre de qualité.

Un facteur externe qui intervient de façon contraignante à ce niveau est la variabilité du paramètre de qualité. Plus la variabilité est grande, plus la fréquence d'échantillonnage devra être élevée pour atteindre une précision donnée.

Une précision doit être choisie de façon réaliste pour respecter les objectifs tout en tenant compte des contraintes extérieures comme la variabilité du paramètre ou le budget dont on dispose.

Le coût sera croissant, en effet, avec la précision requise et c'est généralement un compromis coût-précision qui permet de déterminer la fréquence d'échantillonnage adéquate.

1.3.1.4 Critères de qualité

Les critères de qualité dépendent de l'objectif étudié. La limite

tolérée pour la concentration de fer sera par exemple de 0.3 mg/l pour l'eau de boisson et de 2.6 mg/l pour l'eau avant traitement dans l'industrie des pâte à papier.

Les critères de qualité sont difficiles à définir. Ils varient d'un pays à l'autre et changent au fur et à mesure que la connaissance des phénomènes physico-chimiques et biologiques et les méthodes d'analyse progressent. Ils sont parfois difficiles à utiliser. En effet, faut-il penser que tel critère ne doit jamais être dépassé - et en ce cas on ne pourra pas toujours répondre à la question car il faudrait disposer d'un très grand nombre de valeurs du paramètre mesuré et pas seulement d'un échantillon de taille réduite - ou bien doit-on comprendre que la probabilité que ce critère soit dépassé ne doit pas excéder une valeur très faible comme par exemple 0.05 ou 0.01, ce qui serait peut-être plus réaliste?

Il est certain que la méconnaissance et la complexité des phénomènes rendent difficile la fixation de critères. Ainsi de très faibles concentrations dans l'eau de mercure sous forme organique pourront sembler inoffensives mais conduiront, si elles sont continues, à de fortes accumulations dans certains organes des poissons; elles seront alors toxiques, voire léthales. On sait que le mercure et d'autres éléments sont susceptibles d'accumulation le long de la chaîne trophique.

En conclusion, il semble qu'un effort important de recherche reste à accomplir dans ce domaine pour homogénéiser les critères entre pays, ou les adapter aux conditions particulières des pays, préciser les conditions de leur application, et même les créer dans certains cas.

Quoi qu'il en soit, le choix des critères de qualité doit être effectué lors de la planification du réseau car ceux-ci interviendront dans la détermination des caractéristiques statistiques à employer comme outils de

comparaison aux valeurs des critères, dans le choix des précisions à atteindre sur les caractéristiques et, par suite, dans le choix de la fréquence de mesures. Dans le cas du fer, cité ci-dessus, on étudiera par exemple le quantile 90 par rapport à la norme 0.3 mg/l pour l'eau de boisson et la moyenne par rapport à la norme 2.6 mg/l des industries de pâte à papier. La précision requise sur la moyenne sera plus facile à atteindre que la précision requise sur le quantile; la fréquence de mesure s'en ressentira.

1.3.2 Densité et localisation des stations

Il est clair que cette étape est déterminante pour le coût d'implantation et d'opération du réseau. Elle doit être réalisée soigneusement.

Nous allons dégager les éléments qui permettent la sélection:

- au niveau global de la répartition des stations du réseau dans leur ensemble;
- au niveau élémentaire de l'emplacement précis d'une station.

Nous décrirons ensuite deux types d'approche valables dans certains cas.

1.3.2.1 Répartition des stations du réseau

Divers facteurs vont influencer la dimension spatiale du réseau:

- a- l'objectif du réseau;
- b- la variabilité spatiale des paramètres étudiés;
- c- les caractéristiques diverses du bassin versant:
 - . caractéristiques physiques (taille, forme, relief)
 - . caractéristiques géologiques;

- . caractéristiques hydrographiques (forme du réseau hydrographique, types d'écoulements, nombre de sous-bassins, etc...);
- . autres caractéristiques comme la localisation des centres urbains, des zones de culture (fonction de l'engraisement et du drainage) et d'élevage, des zones industrielles.

a- Objectif du réseau

On a évoqué (en 1.2.7) la nécessité de connaître l'objectif visé pour déterminer l'implantation spatiale du réseau. Pour un objectif comme la qualité de l'eau de boisson on plantera des stations à proximité des prises d'eau. Pour l'étude de la vie aquatique ou pour la recherche des gradients de variation de certains paramètres à travers le bassin, il faudra répartir plus uniformément les stations sur l'ensemble du bassin.

b- Variabilité spatiale

Théoriquement une connaissance de base de la variabilité spatiale est nécessaire pour établir le réseau. Si un paramètre varie peu sur le bassin comme, par exemple, un paramètre chimique d'origine géologique sur un bassin à sous-sol homogène, un petit nombre de stations suffira à apprécier les faibles variations.

Si le paramètre varie beaucoup, un réseau plus dense sera nécessaire. Cela pourra être le cas d'un paramètre influencé par l'agriculture comme l'azote si l'utilisation du sol est très variée sur le bassin.

A priori, cette variabilité spatiale sera inconnue à moins de disposer de mesures réalisées dans le passé. On se référera alors, pour établir le réseau, à la connaissance des caractéristiques du bassin versant qui sont susceptibles d'avoir une influence comme l'utilisation du sol, les villes et industries, etc...

En général, il est conseillé de commencer l'exploitation du réseau avec une densité assez élevée de stations et de diminuer cette densité éventuellement au regard des premiers résultats. Lorsque l'on dispose de données, une analyse sommaire de la variabilité spatiale peut en général contribuer à établir le plan d'échantillonnage. La précision sur une caractéristique statistique spatiale comme la moyenne régionale sera fonction de la variabilité spatiale: le réseau sera plus ou moins dense selon la précision à atteindre et l'importance de la variabilité spatiale du paramètre.

En général, l'étude spatiale du réseau a pour but de détecter les points "problèmes" du bassin, plus que d'évaluer des caractéristiques statistiques spatiales globales. Au niveau spatial on approfondira alors la connaissance de ces points: par exemple, une portion de cours d'eau. Ainsi on peut chercher à évaluer sur quelle distance un cours d'eau parvient à réduire la demande biologique en oxygène (D.B.O.) que lui occasionne un égout urbain. Pour ce type d'étude SHERWANI et MOREAU (1975) décrivent une approche possible (cf. 1.3.2.3, a).

c- Caractéristiques du bassin versant

. caractéristiques physiques, géologiques, hydrographiques

Selon la taille, la forme allongée ou circulaire, les différents types de relief rencontrés sur le bassin, le réseau sera plus ou moins dense pour cerner les variations en cause.

Le bassin peut couvrir des régions géologiquement distinctes (zones de dépôts meubles, de roches sédimentaires, métamorphiques, ignées, etc...). Les concentrations en ions majeurs, en cations métalliques en seront influencées.

Dans le cas d'une étude globale sur le bassin, le réseau de mesure suivra le réseau hydrographique. Le nombre de stations augmentera avec le nombre de sous-bassins, la longueur des cours d'eau, les hétérogénéités diverses. On implante ainsi des stations à l'exutoire des sous-bassins, des lacs, des barrages-réservoirs, à un changement de régimes de cours d'eau, au niveau de l'interface eau douce-eau salée d'un estuaire, à un confluent, etc..., selon l'objectif du réseau.

On aura souvent intérêt à placer les stations de qualité à proximité des stations de mesure de débits. La connaissance du débit de la rivière est indispensable dans le calcul des débits massiques et utile pour l'explication de nombreux phénomènes comme l'augmentation ou la diminution des valeurs de certains paramètres en période d'étiage ou de crue.

Dans le cas des substances nutritives, l'augmentation relative des valeurs du paramètre étudié au cours d'une période d'augmentation de débit peut signifier une origine diffuse (engraissement des cultures, lessivage et drainage de surface), alors qu'une augmentation relative en période de décroissance du débit peut signifier une origine ponctuelle (égout, rejet d'usine, etc...). Par contre, pour certains ions majeurs, une augmentation des concentrations en période d'étiage signifierait une origine géologique (apport des nappes phréatiques).

utilisation du sol, urbanisation, industrialisation

Selon qu'on voudra étudier l'influence des rejets diffus de l'agriculture (dus à l'engraissement des cultures, au drainage et à l'érosion), des rejets ponctuels urbains ou industriels, le réseau se distribuera suivant l'utilisation du sol, les zones urbanisées, l'emplacement des industries.

En général, il pourra être utile d'installer une ou plusieurs stations témoins qui serviront de référence en aval de régions relativement homogènes en ce qui concerne la géologie, l'utilisation du territoire, etc..., et peu influencée par les activités humaines (INRS-EAU, 1976).

1.3.2.2 Choix de l'emplacement d'une station de mesure

Une fois que la répartition globale des stations du réseau est établie, il faut trouver l'emplacement précis de chaque station sur le cours d'eau. Ce choix n'est pas fait au hasard. Le problème essentiel qui se pose est celui de l'homogénéité de la section de mesure (WILSON, 1974).

L'hétérogénéité d'une section peut avoir pour cause des facteurs physiques, chimiques ou biologiques. En général, il s'agira d'un problème de mélange d'ordre physique. Si la section est hétérogène, on pourra prendre plusieurs relevés en divers points de la section. Mais il est de loin préférable de changer de section. L'homogénéité doit être vérifiée tant en période de crue qu'en période d'étiage, car la nature de la section peut varier suivant les épisodes hydrologiques.

Un autre problème peut être celui de l'accessibilité. S'il faut opérer depuis la rive, des difficultés importantes se poseront. De plus,

à proximité de la rive, les courants et la végétation seront source d'hétérogénéité.

Les mesures à partir de bateaux sont délicates (difficultés pour maintenir l'embarcation, etc...) et onéreuses (transport de l'embarcation, etc...).

Une bonne solution semble être, quand celà est possible, d'opérer la mesure à partir d'un pont, dans le centre du cours-d'eau, en intégrant éventuellement la prise de l'échantillon sur la profondeur du cours d'eau, sans toucher le fond.

De plus, les sections au droit d'un pont paraissent être assez homogènes. Des échantillonnages effectués à partir de ponts sur diverses rivières du Québec et pour plusieurs paramètres n'ont pratiquement jamais révélé de différences significatives entre les mesures prises au centre, à la gauche et à la droite de la section. Le brassage est, en effet, plus important sous un pont et les berges sont souvent dégagées à ce niveau.

1.3.2.3 Exemples de méthodes utilisées

a) Méthode de localisation des stations par l'étude des gradients de concentration (SHERWANI et MOREAU, 1975)

La distance entre stations dépendra des gradients de concentrations des paramètres. Pour obtenir une résolution spatiale uniforme, il faudrait avoir un réseau plus dense dans les zones de forts gradients et plus lâche dans les zones de faibles gradients. Pour un réseau dont l'un des objectifs est l'identification de la variabilité spatiale de plusieurs paramètres, c'est le paramètre dont les valeurs varient le plus rapidement qui doit servir de base à l'établissement du réseau. D'après SHERWANI et MOREAU (1975), ce paramètre sera souvent la DBO_5 si elle est mesurée par le réseau.

La distance entre la source de pollution et la première station d'échantillonnage devrait être telle que la concentration excessive soit réduite de moitié. La concentration excessive est la différence entre la concentration dans le courant après le point de rejet, à la fin de la zone de mélange et la concentration naturelle de base de la rivière. La distance correspond donc à la période de demie-vie de la concentration excessive. Pour les stations suivantes la distance serait doublée chaque fois.

Cette méthode nécessite la connaissance de la vitesse du courant et des équations cinétiques de transformation des paramètres étudiés. Elle conviendra pour l'étude de l'autoépuration d'un cours d'eau en fonction de certains types de paramètres de qualité.

b) Approche utilisée par l'INRS-Eau et le M.R.N. pour l'établissement des réseaux de mesure des bassins versants des rivières Yamaska et St-François

Chacun des deux bassins étudiés a été divisé en dix tronçons représentant un apport de 10% du débit d'étiage à l'embouchure. On a aussi divisé chaque bassin en dix tronçons de population humaine approximativement égale. Une station fut placée provisoirement à l'interface de chacun de ces tronçons et une autre à la décharge de chaque lac important. Puis ce premier réseau fut réduit en tenant compte de considérations pratiques comme la facilité d'échantillonnage et la proximité des stations les unes des autres.

On a ainsi obtenu un réseau de 19 stations sur le bassin de la Yamaska (4843 km²), ce qui représente une densité globale d'environ une station par 250 km².

Ce type d'approche convient à des réseaux d'analyse assez globale sur tout bassin versant. C'est une approche de type physique puisqu'elle ne tient pas compte de la variabilité spatiale des paramètres.

1.3.3 Dates et fréquences d'échantillonnage

La détermination de la fréquence de mesure d'un paramètre de qualité à une station dépend essentiellement de deux facteurs:

- la variabilité temporelle du paramètre de qualité;
- la précision à atteindre sur la caractéristique statistique associée pour un niveau de confiance donné.

Dans le cas où on peut considérer les mesures du paramètre de qualité comme aléatoires, il est possible d'estimer le nombre et donc, la fréquence des mesures à réaliser pour atteindre la précision voulue, compte-tenu de la variabilité des mesures. Mais généralement, la série temporelle des valeurs d'un paramètre de qualité sera composé de quatre types d'éléments:

- une tendance générale qui peut s'étendre sur plusieurs années (trend);
- des variations cycliques avec une période régulière ou variable (journalière, hebdomadaire, saisonnière, annuelle);
- des variations aléatoires dont les caractéristiques statistiques seront plus ou moins stables et, elles-aussi, éventuellement variables au cours de l'année;
- une persistance qui traduit la mémoire interne du phénomène en relation avec la fréquence des mesures (processus aléatoire).

Ainsi, selon l'importance relative de ces facteurs, on aura affaire à un phénomène aléatoire, à un processus stochastique, à un phénomène déterministe. En général une série chronologique sera une superposition de ces divers éléments avec prépondérance de l'un ou de l'autre (1.3.2.2).

Quoi qu'il en soit, l'estimation à priori de la variabilité temporelle sera nécessaire pour déterminer la fréquence des mesures, ce qui soulève certains problèmes évoqués en 1.3.3.3.

1.3.3.1 Cas des séries de mesures aléatoires

Quand les observations sont aléatoires, il est possible de calculer la fréquence de mesure qui permettra d'obtenir la caractéristique statistique étudiée (moyenne ou quantile) avec une précision donnée pour un certain niveau de confiance.

On montre en effet (cf, 2.1.1.2) que le nombre N d'observations qui permet d'estimer la moyenne avec une erreur maximale tolérée ℓ_0 au niveau de confiance $(1 - \alpha)$ est:

$$N \geq \left(\frac{t_{N-1}(\alpha/2) \cdot S}{\ell_0} \right)^2 \quad (1.1)$$

$t_{N-1}(\alpha/2)$: valeur de la variable de Student à $N-1$ degrés de liberté pour une probabilité au dépassement $\alpha/2$.

S : écart-type estimé de la population.

Par exemple si l'on désire connaître la concentration moyenne annuelle de fer à une station de mesure avec une erreur maximale de 0.5 mg Fe/l au niveau de confiance 95%, en supposant que l'écart-type estimé est $S = 0.1$, on trouve

$N \geq 18$. Cela correspond à une fréquence d'échantillonnage minimum d'une observation chaque trois semaines durant une année (observations indépendantes).

Un raisonnement analogue pourrait être fait à propos d'un quantile (cf. 2.1.3.2). Ainsi la fréquence de mesure d'un paramètre de qualité aléatoire dépend essentiellement:

- de la précision désirée à un certain niveau de confiance sur la caractéristique étudiée;
- dans le cas de la moyenne, de la variabilité temporelle du paramètre (écart-type S) qui est estimé à partir de l'information dont on dispose a priori.

La fréquence de mesure croît avec la précision désirée pour un niveau de confiance donné. Ceci met en évidence que la planification d'un réseau résulte toujours d'un compromis entre le coût (nombre de mesures) et la précision.

1.3.3.2 Cas des séries non totalement aléatoire

Trois cas se présentent:

- a- présence d'une tendance générale;
- b- présence de cycles;
- c- présence d'une persistance.

Dans chaque cas la composante aléatoire existe et a plus ou moins d'importance.

a- Tendance générale

L'étude des réseaux de détection des tendances générales sur plusieurs années est envisagée par SHIRWANI et MOREAU (1975, chap. 7).

Ce type de réseau ne requiert pas de fréquence de mesure élevées, mais des mesures durant une longue période (plusieurs années).

b- Cycles

Pour de nombreux paramètres de qualité, des variations aléatoires sont superposées aux variations cycliques. Dans ce cas, la date aussi bien que la fréquence d'échantillonnage doivent être choisies attentivement.

Supposons en effet qu'un paramètre soit soumis à des variations diurnes régulières (exemple: influence de la photosynthèse sur l'oxygène dissous); si on échantillonne ce paramètre toujours aux mêmes heures, on obtiendra une estimation biaisée de la moyenne.

On peut observer des cycles journaliers, hebdomadaires ou annuels. Deux cas se présenteront selon que l'amplitude des cycles sera relativement faible ou grande par rapport à celle des variations aléatoires.

WILSON (1974) explicite l'approche suivante pour le calcul de la moyenne annuelle. Le raisonnement peut s'appliquer à d'autres caractéristiques statistiques.

- * Si les variations cycliques sont sensiblement peu importantes vis-à-vis des variations aléatoires, on pourra les négliger en première approximation et on recueillera les échantillons de façon uniforme selon la fréquence calculée comme en 1.3.3.1.
- * Si les variations cycliques sont importantes, on corrigera leurs effets en procédant de la manière suivante:

- Si les variations journalières sont importantes:

- . On fait varier l'heure du jour auquel est pris l'échantillon de façon que chaque portion du jour soit représentée de façon à peu près égale dans l'ensemble final des résultats à partir desquels sera calculée la moyenne annuelle.
- . ou bien on collecte, pour chaque jour où une analyse est requise, au moins six échantillons à des intervalles de temps réguliers dans la journée et qui sont composés en proportions égales pour donner l'échantillon à analyser. Ce cas convient si l'échantillonnage peut être automatisé.

- Si les variations hebdomadaires sont importantes:

Comme précédemment, deux approches sont possibles. La plus simple consiste à échantillonner de façon que, dans l'ensemble des échantillons, chaque jour de la semaine soit également représenté.

c- Effet de persistance

Il peut se produire que les observations d'une même série ne soient pas indépendantes. La $i^{\text{ème}}$ observation dépend de la $(i - 1)^{\text{ème}}$, voire de plusieurs autres observations précédentes. Dans certains cas classiques cet effet de persistance peut être traduit par un processus de Markov d'ordre 1 (cf 2.2.3):

$$x_{i+1} = r x_i + \epsilon_{i+1}$$

- r : coefficient d'autocorrélation d'ordre 1 de la série des x_i
- ϵ_{i+1} : composante aléatoire indépendante de x_{i+1} .

Soit une série de N observations décrites par un processus de Markov d'ordre 1. On montre (cf. 2.2.3.3) qu'on peut trouver un nombre N' d'observations indépendantes tel que :

- l'effet de persistance ait disparu;
- la série des N' observations permette d'obtenir la même précision sur la moyenne que la série des N observations pour un niveau de confiance donné.

En première approximation, on a :

$$N' = N \cdot \frac{1 - r}{1 + r} \quad (1.2)$$

1.3.3.3 Estimation de la variabilité temporelle

La détermination de la fréquence de mesure, qui permet le calcul de la moyenne avec une certaine précision, nécessite la connaissance de la variabilité temporelle (cf. 1.3.3.1).

D'après WILSON (1974), l'estimation de l'écart-type S devrait être obtenue à partir de mesures étendues sur approximativement la même période que celle utilisée pour le calcul de la moyenne avec une fréquence d'échantillonnage hebdomadaire. Mais ces mesures ne seront vraisemblablement pas disponibles; on utilisera alors les données existantes. Comme l'imprécision sur S sera assez importante, Wilson conseille de commencer les mesures avec une fréquence approximativement double de celle calculée. Cette fréquence pourra être réduite en cours d'opération si les résultats sont satisfaisants. En fait, tout cela dépendra des objectifs et des moyens dont on dispose. Il peut advenir qu'aucune donnée ne soit disponible pour estimer S . On pourra alors utiliser des résultats obtenus sur d'autres cours d'eau semblables à celui étudié et faire des transpositions.

La meilleure approche consiste à réaliser un pré-échantillonnage lorsque les objectifs fixés, le temps et les moyens financiers dont on dispose le permettent. La présence et surtout l'amplitude des cycles journaliers ou hebdomadaires ne pourra être estimée qu'en recourant à des tests expérimentaux (six mesures par jour durant plusieurs jours, par exemple, pour les cycles journaliers). Remarquons enfin que l'écart-type qui représente la variabilité temporelle peut changer au cours de l'année de façon importante en raison d'effets hydrologiques (selon la saison, par exemple). Il peut alors être intéressant de faire varier en conséquence la fréquence de mesures au cours de l'année de façon à obtenir une précision identique sur le paramètre étudié, pour chaque tranche de temps considérée. Ici intervient le choix de la date d'échantillonnage. On cherchera à échantillonner plus spécialement au cours d'époques de crue ou d'étiage où l'on risque d'observer des valeurs extrêmes des paramètres étudiés.

En conclusion à l'étude des dimensions spatiales et temporelles du réseau, soulignons que le plan d'échantillonnage ne peut être optimal simultanément pour tous les paramètres. On devra réaliser un compromis ou effectuer une classification hiérarchique des paramètres suivant leur importance pour l'objectif visé.

1.4 Collecte, transport, analyse des échantillons

Les objectifs de précision désirés influencent le choix des méthodes analytiques. Ce choix a de plus une grande importance, car la validité des résultats obtenus en laboratoire est essentielle pour effectuer une analyse adéquate.

Deux problèmes généraux, soulevés par WILSON (1974), conditionnent le choix de l'analyse en laboratoire:

- la limite de détection;
- la précision de l'analyse.

1.4.1 Limite de détection et précision de l'analyse

On suggère en général que la limite de détection d'une substance soit approximativement égale à 10% du plus petit critère de qualité relatif à ce paramètre. Cette limite de détection sera gouvernée par la précision des analyses aux basses concentrations.

La précision à obtenir sur une mesure analytique conditionnera le choix de la méthode à employer. Cette précision devra être suffisante tout en restant réaliste. En effet, comme le signale WILSON (1974): à la limite de détection les erreurs dues seulement au hasard équivalent à 60% du résultat. Aussi est-il suggéré que la précision relative d'un résultat décroisse avec la valeur du résultat en rapport avec le critère de qualité.

Sur cette base, cet auteur propose la définition suivante pour exprimer la précision désirée:

"l'erreur sur le résultat analytique ne devrait pas excéder la plus grande des deux valeurs: limite de détection ou un pourcentage donné p du résultat".

La limite de détection et la valeur p varieront selon le paramètre de qualité et le critère afférent.

1.4.2 Méthode d'analyse

Connaissant la précision et la limite de détection, on choisit la méthode d'analyse. Il est clair que réciproquement les méthodes d'analyse sont une contrainte pour la précision et la limite de détection.

WILSON évoque le problème important que peut poser l'analyse d'échantillons effectués par des laboratoires différents. On constate en effet que, bien qu'utilisant des méthodes identiques, des laboratoires distincts obtiennent souvent des résultats différents avec des précisions différentes. Cela provient du fait que, d'un laboratoire à l'autre, les techniciens de laboratoire, les analyses, les réactifs employés, l'instrumentation, les temps de transport et de stockage, le mode de conservation, etc... varient.

Pour remédier à ces problèmes, il est suggéré que des contrôles efficaces aient lieu au sein d'un même laboratoire et entre laboratoires.

1.5 Le traitement des données.

Les diverses méthodes statistiques utilisables seront, en partie, développées au deuxième chapitre de ce travail. Rappelons que le choix des méthodes est, pour une part, conditionné par l'objectif (cf. 1.2.5).

Le traitement sur ordinateur sera souvent très efficace, en gain de temps et en précision de calcul, vu le grand nombre de données à traiter. De plus, certaines méthodes ne sont utilisables qu'avec traitement informatique en raison de leur complexité de calcul.

Le choix des méthodes statistiques pourra avoir une répercussion sur la conception du réseau. Certaines méthodes globales comme l'analyse des correspondances nécessitent l'emploi de tableaux de données complets; les études de corrélation entre stations ou entre paramètres nécessitent que les mesures soient prises à des dates identiques; des analyses temporelles demandent une fréquence de mesure la plus constante possible pour mettre en évidence des tendances ou des cycles.

Le choix des méthodes statistiques devra donc être pensé dès l'élabo-

ration du plan d'échantillonnage et l'on doit se demander avant d'acquérir les données l'usage qui en sera fait.

1.6 Cas d'un réseau existant

Il arrive souvent qu'on soit en présence d'un réseau qui est en service depuis quelques temps. Le choix des paramètres, des stations, des fréquences de mesure a été effectué et, en général, ce choix n'aura pas tenu compte de l'ensemble des considérations développées précédemment. Un but d'étude consiste alors à rationaliser le réseau par la définition d'objectifs pertinents pour les différents usages de l'eau des régions étudiées. On peut considérer que ce travail s'intègre dans le cadre de la figure 1-1 avec une information à priori plus ou moins importante. L'analyse de cette information permet d'évaluer jusqu'à quel point le réseau atteint les objectifs et de définir les éventuelles modifications du plan d'échantillonnage. Le résultat sera une réduction ou une augmentation du nombre de paramètres de qualité mesurés, de leur fréquence, du nombre de stations où la mesure doit être effectuée.

Il se peut aussi qu'un réseau change d'objectifs. Un réseau d'appréciation et de mesure de la qualité de l'eau peut être utilisé par la suite de façon plus sporadique comme réseau de contrôle de qualité. Diverses méthodes statistiques permettent d'apprécier la situation d'un réseau et d'étudier ses réductions possibles (contenu de l'information, régressions linéaires entre stations, entre paramètres, analyse de variance, analyse factorielle, etc...). Ces méthodes seront envisagées au deuxième chapitre.

Nous considérons, dans ce travail, le réseau de mesures du bassin versant de la rivière Yamaska, après deux années d'opération. L'étude de l'infor-

mation fournie par le réseau et la mise en oeuvre de diverses méthodes statistiques visant à sa réduction ou à son amélioration montreront la nécessité d'une démarche rationnelle fondée sur la définition d'objectifs précis si l'on veut obtenir des résultats pertinents et si l'on veut évaluer correctement les performances du réseau. Nous tenterons de mettre en application cette démarche dans le cadre de l'objectif eau potable.

CHAPITRE 2

METHODES STATISTIQUES

Ce chapitre décrit de manière synthétique les différentes notions et méthodes statistiques utilisées au cours de ce travail.

On envisage successivement

- les caractéristiques statistiques relatives à une série de mesures;
- la caractérisation de la structure d'une série de mesures;
- les inter-relations entre séries de mesure et les analyses globales d'ensemble des séries ou des tableaux de données.

2.1 Caractéristiques statistiques d'une série d'observations.

On appelle population l'ensemble de toutes les valeurs possibles d'une caractéristique donnée, par exemple d'un paramètre de qualité. On ne dispose pour l'étude de cette population que d'un échantillon qui est une série comprenant un nombre fini d'observations. Les caractéristiques statistiques de la population sont estimées à partir de celles de l'échantillon. C'est ce qui conduit à des erreurs d'échantillonnage et à la nécessité de déterminer la précision atteinte dans l'estimation d'une caractéristique statistique de la population. Les caractéristiques statistiques considérées dans cette étude sont la moyenne, l'écart-type, le coefficient de variation, les valeurs extrêmes, les quantiles, le coefficient d'asymétrie et le coefficient d'auto-corrélation. On suppose dans la suite (sauf spécification précise dans le texte) que l'échantillon des valeurs observées est aléatoire, c'est-à-dire que ces observations sont tirées au hasard, de façon indépendante,

dans la population. Tous les symboles utilisés sont définis au début de ce travail.

2.1.1 Moyenne, écart-type, coefficient de variation

Soit $(x_1, x_2, \dots, x_i, \dots, x_N)$ un échantillon aléatoire de N observations. Le moyenne de la population μ est estimée par la moyenne de l'échantillon $\bar{x}$ telle que: $\bar{x} = \sum_{i=1}^N x_i / N$ (2.1)

L'écart-type de la population σ est estimé par la quantité:

$$S = \left[\sum_i (x_i - \bar{x})^2 / (N - 1) \right]^{\frac{1}{2}} \quad (2.2)$$

Le coefficient de variation de la population $C_v = \sigma / \mu$ est estimé par

$$\bar{C}_v = S / \bar{x} \quad (2.3)$$

L'écart-type de la moyenne de l'échantillon est $\sqrt{\frac{\sigma^2}{N}}$. Il est estimé par la quantité $\sqrt{\frac{S^2}{N}}$.

2.1.1.1 Intervalle de confiance sur la moyenne

Dans le cas de grands échantillons ou dans le cas de petits échantillons, lorsque les observations proviennent d'une population normale, on montre que la quantité:

$$t_{N-1} = \frac{\bar{x} - \mu}{\sqrt{S^2/N}}$$

suit une loi de Student à $(N - 1)$ degrés de liberté.

L'intervalle de confiance de la moyenne au niveau de confiance $(1 - \alpha)$ est alors:

$$\bar{x} - t_{N-1}(\alpha/2) \sqrt{S^2/N} \leq \mu \leq \bar{x} + t_{N-1}(\alpha/2) \sqrt{S^2/N} \quad (2.5)$$

où $t_{N-1}(\alpha/2)$ est la variable de Student de probabilité au dépassement $\alpha/2$.

On remarque que l'intervalle de confiance est fonction de la taille de l'échantillon, de la variabilité de la variable aléatoire (paramètre de qualité), prise en compte par S , et du niveau de confiance choisi.

2.1.1.2 Précision sur la moyenne

L'erreur d'échantillonnage commise dans l'estimation de la moyenne de la population à partir de l'échantillon est:

$$\ell = |\bar{x} - \mu| \quad (2.6)$$

D'après (2.3), il vient que l'erreur d'échantillonnage pour un niveau de signification $(1 - \alpha)$ fixé est:

$$\ell = t_{N-1}(\alpha/2) \sqrt{S^2/N} \quad (2.7)$$

Le nombre minimal de mesures N_0 à effectuer pour commettre une erreur au plus égale à une valeur fixée ℓ_0 avec un niveau de confiance $(1 - \alpha)$ est, d'après (2.7), donnée par:

$$N_0 = \left(\frac{t_{N-1}(\alpha/2) \cdot S}{\ell_0} \right)^2$$

La précision relative ϵ , obtenue sur la moyenne est donnée par:

$$\epsilon = \frac{|\bar{x} - \mu|}{\bar{x}} \quad (2.9)$$

Le nombre de mesures N_0 permettant d'obtenir la précision $\varepsilon_0 \frac{\ell_0}{\bar{x}}$

$$N_0 = \left(\frac{t_{N-1} (\alpha/2) \cdot \overline{C}_v}{\varepsilon_0} \right)^2 \quad (2.10)$$

Remarquons que l'on peut exprimer l'intervalle de confiance de la moyenne au niveau de confiance $(1 - \alpha)$ en fonction de la précision ε atteinte sur la moyenne au niveau $(1 - \alpha)$; d'après 2.5, 2.7 et 2.9

$$\bar{x} (1 - \varepsilon) \leq \mu \leq \bar{x} (1 + \varepsilon) \quad (2.11)$$

2.1.2 Maximum, minimum

Le maximum d'un échantillon est la plus forte valeur observée, mais il ne donne aucune indication sur la valeur maximale observable de la population. En pratique il est préférable de s'intéresser surtout à la prédiction des valeurs qui ont une faible probabilité au dépassement (valeur dépassée dans 1% des cas, par exemple). Ce problème est résolu par l'étude de la distribution statistique que l'on peut ajuster aux valeurs observées. Le maximum de l'échantillon ne donne qu'une indication grossière sur les valeurs élevées et peut être trompeur: les erreurs d'échantillonnage sont en effet grandes au niveau des valeurs extrêmes, et cela d'autant plus que la variabilité du paramètre est importante. Le minimum est égal à la limite de détection du point de vue de l'observation.

2.1.3 Quantile

Soit F la fonction de répartition d'une variable aléatoire X . Le quantile d'ordre p de la distribution, q_p , est tel que:

$$F(q_p) = p$$

Le quantile empirique Q_p de l'échantillon est la valeur dépassée $(1 - p)\%$ des fois.

On démontre (FOURGEAUD et FUCHS, 1972) que le quantile empirique Q_p est distribué suivant une loi normale de moyenne $\overline{Q_p}$ et de variance $\text{var}(Q_p)$ qui sont données par:

$$\left\{ \begin{array}{l} \overline{Q_p} = q_p \\ \text{var}(Q_p) = \frac{p(1-p)}{N \cdot f(q_p)^2} \end{array} \right. \quad (2.12)$$

où f est la densité de probabilité de la variable aléatoire. Ce résultat est valable de façon asymptotique, c'est-à-dire quand N est suffisamment grand.

2.1.3.1 Intervalle de confiance d'un quantile

D'après le résultat précédent, quand N est suffisamment grand, la quantité:

$$u = \frac{Q_p - q_p}{\sqrt{\text{var}(Q_p)}} \quad (2.13)$$

suit une loi normale centrée réduite. Si N est faible, la tendance vers la normalité est moins forte et le résultat précédent est approximatif.

On voit d'après (2.12) que le calcul de $\text{var}(Q_p)$ fait intervenir $f(q_p)$ qui est inconnu, puisque q_p est inconnu. On estimera donc $\text{var}(Q_p)$ par la quantité $\widehat{\text{var}}(Q_p)$ définie par:

$$\widehat{\text{var}}(Q_p) = \frac{1}{[f(Q_p)]^2} \frac{p(1-p)}{N} \quad (2.14)$$

Dans ces conditions, l'intervalle de confiance du quantile q_p au niveau de confiance $(1 - \alpha)$ s'écrit d'après (2.13) et (2.14):

$$Q_p - \frac{u(\alpha/2)}{f(Q_p)} \sqrt{\frac{p(1-p)}{N}} \leq q_p \leq Q_p + \frac{u(\alpha/2)}{f(Q_p)} \sqrt{\frac{p(1-p)}{N}} \quad (2.15)$$

En pratique, on dispose au départ d'un échantillon de N valeurs; on peut alors estimer les paramètres de la loi de distribution de la population par différentes méthodes (méthode des moments, méthode du maximum de vraisemblance), ce qui permet de tracer la courbe théorique donnant la valeur d'un événement pour une probabilité au dépassement donné. De cette courbe on déduira Q_p pour une probabilité p donnée. Connaissant les paramètres de la loi, on calculera la valeur correspondante de la fonction densité de probabilité: $f(Q_p)$. On disposera alors de tous les éléments pour le calcul de l'intervalle de confiance de quantile, au niveau de confiance désiré.

Remarque: La formule (2.15) permet de déterminer l'intervalle de confiance d'une loi de distribution en faisant varier p sur l'intervalle $(0, 1)$. Ce calcul est approché: il n'est strictement valable que lorsque $N \rightarrow \infty$. Mais sa réalisation est simple puisque seule la connaissance de f (densité de probabilité) suffit. Il serait intéressant d'étudier - suivant les lois de distribution - les écarts qui existent entre cette méthode et d'autres méthodes de calcul d'intervalle de confiance, soit par les statistiques d'ordre (BOBEE et MORIN, 1973), soit encore par le calcul de la variance asymptotique d'un événement de probabilité au non dépassement donné à par-

tir des moments de la loi (BOBEE, 1973) (cf. 5.2.3).

2.1.3.2 Précision sur un quantile

L'erreur d'échantillonnage sur le quantile Q_p est

$$\ell = |Q_p - q_p| \quad (2.16)$$

D'après (2.13) et (2.14), l'erreur d'échantillonnage pour un niveau de signification $(1 - \alpha)$ est:

$$\ell = u(\alpha/2) \cdot \frac{1}{f(Q_p)} \sqrt{\frac{p(1-p)}{N}} \quad (2.17)$$

Le nombre minimal de mesures N_0 à effectuer pour commettre une erreur au plus égale à une valeur fixée ℓ_0 , avec un niveau de confiance $(1 - \alpha)$, sera (d'après 2.17):

$$N_0 = p(1-p) \left(\frac{u(\alpha/2)}{\ell_0 f(Q_p)} \right)^2 \quad (2.18)$$

Soit ϵ la précision relative obtenue sur le quantile Q_p , définie par:

$$\epsilon = \frac{|Q_p - q_p|}{Q_p} \quad (2.19)$$

Le nombre de mesures N_0 permettant d'obtenir la précision ϵ_0 fixée au niveau de confiance $(1 - \alpha)$ sera:

$$N_0 = p(1-p) \left(\frac{u(\alpha/2)}{\epsilon_0 \cdot f(Q_p) \cdot Q_p} \right)^2 \quad (2.20)$$

Pour une population connue (loi statistique), un quantile Q_p et un niveau de confiance $(1 - \alpha)$ donnés, les valeurs p , Q_p , $f(Q_p)$ et $u(\alpha/2)$ sont déterminées dans l'équation 2.20. On peut écrire:

$$N = \frac{k}{\epsilon^2} \quad \text{avec} \quad k = p(1 - p) \left(\frac{u(\alpha/2)}{Q_p f(Q_p)} \right)^2$$

Donc, si on connaît la précision ϵ_1 atteinte avec un nombre d'observations N_1 , on pourra calculer le nombre d'observations N_2 nécessaires à l'obtention d'une précision désirée ϵ_2 sur un même quantile et pour un même niveau de confiance par l'équation:

$$N_2 = N_1 \left(\frac{\epsilon_1}{\epsilon_2} \right)^2 \quad (2.21)$$

Ainsi, dès qu'on dispose d'un nombre de données suffisant pour estimer les paramètres de la distribution statistique représentative de la population (paramètre de qualité, par exemple) et, donc, de la densité de probabilité f , on peut déterminer si la précision requise dans la connaissance du quantile q_p est atteinte (équation 2.17) ou s'il faut augmenter la fréquence de mesure et de quelle façon (équations 2.20 et 2.21). Rappelons cependant que l'ajustement des lois statistiques requiert les conditions d'homogénéité et d'indépendance des observations (cf. 2.2.4.2) et que les formules de calcul (2.17) et (2.20) ne sont strictement valables que pour des valeurs de N suffisamment élevées, bien qu'en pratique elles soient utilisées pour des tailles d'échantillons réduites.

2.1.4 Coefficient d'asymétrie

Le calcul du coefficient d'asymétrie d'une population est important pour le choix d'une distribution adéquate et, dans certains cas (loi à 3 paramètres, méthode des moments), pour le calcul de l'ajustement de cette distribution.

Le coefficient d'asymétrie (C_s) de la population est tel que:

$$C_s = \frac{\mu_3}{\mu_2^{3/2}}$$

Le coefficient d'asymétrie de l'échantillon est:

$$\overline{C}_s = \frac{m_3}{m_2^{3/2}} \quad (2.22)$$

$\overline{C}_s$ est un estimateur biaisé de C_s (en effet $E(\overline{C}_s) \neq C_s$). Il existe diverses corrections de $\overline{C}_s$ qui donnent une meilleure estimation de C_s .

La première correction consiste à utiliser les estimateurs non biaisés des moments d'ordre 3 et d'ordre 2.

$$(C_s)_1 = \frac{m_3}{m_2^{3/2}} \quad (2.23)$$

avec
$$\hat{m}_3 = \frac{N}{(N-1)(N-2)} \sum_{i=1}^N (x_i - \bar{x})^3$$

$$\hat{m}_2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

$$(C_s)_1 = \frac{\sqrt{N(N-1)}}{N-2} \overline{C_s} \quad (2.24)$$

$(C_s)_1$ sous-évalue systématiquement C_s , aussi peut-on utiliser une deuxième correction expérimentale $(C_s)_2$:

$$(C_s)_2 = \frac{\sqrt{N(N-1)}}{(N-2)} \left(1 + \frac{8.5}{N} \right) \overline{C_s} \quad (2.25)$$

Dans le cas de la loi Pearson type 3, BOBEE (1975) propose une estimation plus précise de C_s par la formule suivante :

$$(C_s)_3 = \left[\left(1 + \frac{6.51}{N} + \frac{20.2}{N^2} \right) + \left(\frac{1.48}{N} + \frac{6.77}{N^2} \right) \overline{C_s^2} \right] \overline{C_s} \quad (2.26)$$

Test sur le coefficient d'asymétrie

FISHER (1931) montre que le coefficient d'asymétrie C_s d'un échantillon de taille N tiré d'une population normale, défini par $\hat{C}_s = \hat{m}_3 / \hat{m}_2^{3/2}$, est distribué suivant une loi normale de moyenne nulle et de variance $\text{var}(\hat{C}_s)$ telle que :

$$\text{var}(\hat{C}_s) = \frac{6N(N-1)}{(N-2)(N+1)(N+3)} \quad (2.27)$$

Ce résultat permet de tester si le coefficient d'asymétrie estimé d'une population à partir d'une série de N observations est différent ou non de zéro, pour un niveau de signification α donné. Ce test n'est théoriquement valable que pour des échantillons tirés d'une population normale mais on l'applique aussi pour d'autres populations.

$$u = \frac{0 - C_s}{\sqrt{\text{var}(C_s)}} \quad (2.28)$$

où u est une variable normale centrée réduite.

Si, pour un niveau de signification α on a $|u| \leq u(\alpha/2)$, on accepte que l'hypothèse $C_s = 0$ est vraie.

Lorsqu'on estime C_s par $(C_s)_1$ et non par $\hat{C}_s$, on utilise la relation:

$$(C_s)_1 = \frac{N^2}{(N-1)(N-2)} \hat{C}_s \quad (2.29)$$

Lorsque $N = 20$, par exemple, la zone d'acceptation de l'hypothèse $C_s = 0$ est:

au seuil	$\alpha = 0.05$	$-1.17 < (C_s)_1 < 1.17$
	$\alpha = 0.01$	$-1.54 < (C_s)_1 < 1.54$



_____ C_s n'est pas significativement différent de 0

+++++ C_s est significativement différent de 0 (niveau 5%)

+++++ C_s est très significativement différent de 0 (niveau 1%)

Dans ce travail on a utilisé l'estimateur $(C_s)_1$ pour le calcul des asymétries (tableau 4.13). La correction $(C_s)_3$ est employée lors de l'ajustement des lois à 3 paramètres (Pearson type 3 et log-Pearson type 3) par la méthode des moments.

2.1.5 Coefficient d'auto-corrélation

Quand les observations utilisées ne sont pas indépendantes, les résultats établis dans les paragraphes précédents ne sont plus exacts et l'ajustement d'une distribution statistique n'est plus possible. L'indépendance peut être vérifiée par des tests basés sur les corrélations sérielles.

Le test d'ANDERSON (1941) utilise le coefficient d'auto-corrélation d'ordre 1

$$t_1 = \left[\sum_{i=1}^{N-1} x_i x_{i+1} + x_1 x_N - \left(\sum_{i=1}^N x_i \right)^2 / N \right] / \left[\sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2 / N \right] \quad (2.30)$$

Pour un échantillon de N valeurs, tiré d'une population normale, t_1 est distribué suivant une loi normale de moyenne $\overline{t_1}$ et de variance $\text{var}(t_1)$ telle que:

$$\overline{t_1} = - \frac{1}{N-1} \quad (2.31)$$

$$\text{var}(t_1) = \frac{N-2}{(N-1)^2} \quad (2.32)$$

Ce résultat nous permet de tester l'indépendance d'une série d'observations. Il n'est théoriquement valide que pour des échantillons tirés d'une loi normale mais on l'utilise aussi pour d'autres populations.

$$\text{Posons } u = \frac{t_1 - \bar{t}_1}{\sqrt{\text{var}(t_1)}} \quad (2.33)$$

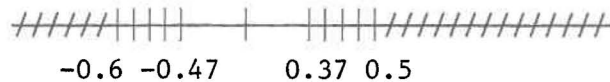
où u est une variable normale centrée réduite

Si pour un niveau de signification α fixé on a $|u| \leq u(\alpha/2)$, on accepte l'hypothèse de nullité du coefficient d'auto-corrélation, c'est-à-dire que les observations sont indépendantes.

Pour $N = 22$, par exemple, on obtient les zones d'acceptation suivantes:

$$\alpha = 0.05 \quad -0.47 < t_1 < 0.37$$

$$\alpha = 0.01 \quad -0.60 < t_1 < 0.50$$



———— t_1 non significativement différent de 0

||||| t_1 significativement différent de 0 (niveau 5%)

t_1 très significativement différent de 0 (niveau 1%)

2.2 Structure des séries d'observation

Au lieu de s'intéresser à des caractéristiques statistiques déterminées, on peut vouloir étudier la série d'observations dans son ensemble.

Si les observations sont indépendantes, on peut ajuster une distri-

bution statistique. Dans le cas contraire on tente d'identifier la partie déterministe de la structure. Nous n'avons pas étudié ici le problème des tendances à long-terme ou des cycles car la durée des observations (moins de deux ans) ne s'y prêtait pas.

2.2.1 Généralités

La méthode d'étude de la structure des séries s'est posée dans ce travail suivant le schéma représenté à la figure 2.1. Quand il y a dépendance, on envisage l'étude par un processus de Markov d'ordre 1; ce qui permet de conserver l'effet chronologique. Cependant la reconstitution des valeurs à partir d'une telle méthode n'est réaliste que si l'auto-corrélation d'ordre 1 est assez élevée (ce qui est rarement le cas).

Quand il y a indépendance, le test de "W" (HAHN G.J. and S.S. SHAPIRO, 1967) permet, dans le cas de petits échantillons, de tester l'hypothèse de normalité. Si cette hypothèse est rejetée, il faut ajuster d'autres lois que la distribution normale. L'ajustement d'une loi perd tout effet chronologique mais permet d'évaluer la probabilité d'événements extrêmes avec, éventuellement, les intervalles de confiance associés à ces estimations.

On étudiera successivement dans ce qui suit:

- le test de normalité "W";
- l'adaptation d'un processus de Markov;
- l'ajustement de lois.

2.2.2 Test de normalité

Ce test est efficace pour tester si un échantillon provient d'une distribution normale, spécialement quand la taille de l'échantillon est réduite

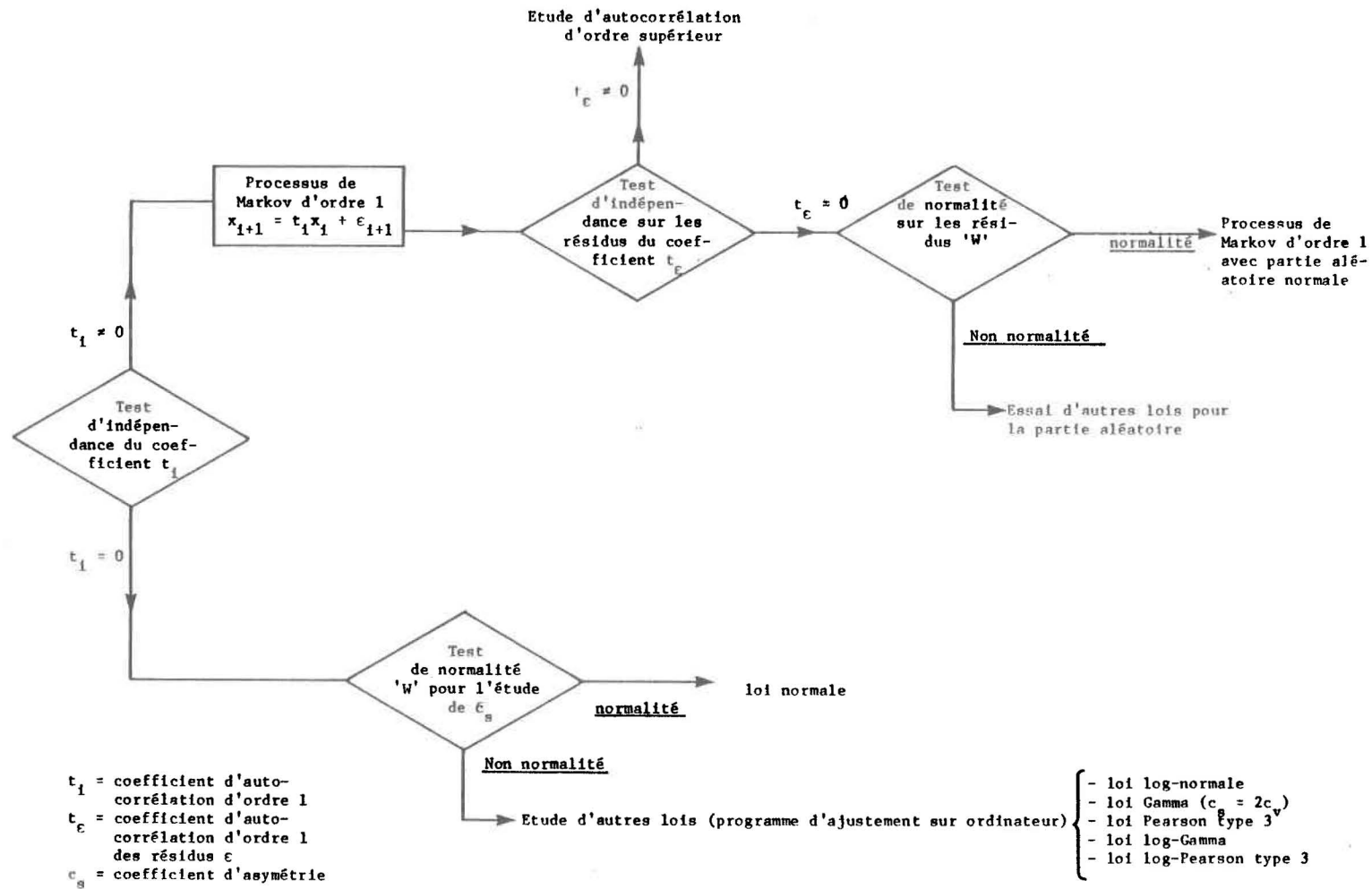


Figure 2-1. Schéma d'étude des séries d'observations

(HAHN G.J. and S.S. SHAPIRO, 1967). Ce test indique si, pour un niveau de confiance donné, on peut rejeter ou non l'hypothèse de normalité. La procédure est la suivante:

- on classe les N valeurs de l'échantillon par ordre croissant;
- on calcule la quantité $SS = \sum_{i=1}^N (x_i - \bar{x})^2$;
- on détermine

$$B = \sum_{i=1}^k a_{N-i+1} (x_{N-i+1} - x_i)$$

où k est tel que $k = \frac{N}{2}$ si N est pair

$$k = \frac{N-1}{2} \text{ si } N \text{ est impair}$$

a_{N-i+1} est un coefficient donné par des tables.

- on calcule W_c tel que:

$$W_c = \frac{B^2}{SS}$$

Une autre table donne, selon la taille N d'un échantillon tiré d'une loi normale, la valeur de W ayant une probabilité au non dépassement donnée.

Soit $W(\alpha)$ la valeur correspondant à la probabilité α pour une taille N .

Si $W_c < W(\alpha)$, la probabilité que l'échantillon considéré provienne d'une loi normale est inférieure à α .

2.2.3 Processus de Markov d'ordre 1

On teste (cf. 2.1.5) la dépendance des observations en considérant le coefficient d'auto-corrélation d'ordre 1. S'il est significativement différent de zéro, on considère qu'il existe une dépendance entre les observations

successives prises deux à deux. Théoriquement, il faudrait tester si les coefficients d'auto-corrélation d'ordre supérieur (ordre 2, ordre 3, ...) sont nuls ou, au contraire, significativement différents de zéro, de façon à déterminer de combien d'observations précédentes dépend une observation quelconque i . Dans le cadre de cette étude, l'intervalle de temps séparant les mesures étant d'un mois, donc assez large, nous avons seulement considéré l'ordre 1. Plus précisément, nous avons fait l'hypothèse d'un processus de Markov d'ordre 1 décrit par l'équation:

$$x_{i+1} = t_1 x_i + \varepsilon_{i+1} \quad (2.34)$$

où ε_{i+1} est une variable aléatoire indépendante de x_{i+1} .

Soient μ et σ^2 la moyenne et la variance de la population. La relation (2.34) peut s'écrire:

$$\varepsilon_{i+1} = x_{i+1} - t_1 x_i \quad (2.35)$$

$$\text{avec} \quad E(\varepsilon_{i+1}) = \bar{\varepsilon} = (1 - t_1) \mu \quad (2.36)$$

$$\text{var}(\varepsilon_{i+1}) = \sigma_{\varepsilon}^2 = (1 - t_1^2) \sigma^2 \quad (2.37)$$

$$\text{Posons} \quad \eta_{i+1} = \frac{\varepsilon_{i+1} - \bar{\varepsilon}}{\sigma_{\varepsilon}} \quad (2.38)$$

η_{i+1} est une variable aléatoire centrée réduite

L'équation (2.34) du processus de Markov s'écrit en tenant compte de 2.36, 2.37 et 2.38:

$$x_{i+1} = \mu + t_1 (x_i - \mu) + \sigma \sqrt{1 - t_1^2} \eta_{i+1} \quad (2.39)$$

η_{i+1} peut être, par exemple, une variable normale centrée réduite.

2.2.3.1 Vérifications

Pour vérifier qu'il est justifié de ne pas étudier le processus de Markov à un ordre supérieur à un (1), on considère la quantité $\varepsilon_{i+1} = x_{i+1} - t_1 x_i$ qui ne doit plus être auto-correlée. On étudie le coefficient d'auto-corrélation d'ordre 1, t_ε , de la série des ε_i . Si t_ε n'est pas significativement différent de zéro, on conclura à l'indépendance des ε_i .

Dans ce cas on pourra chercher la loi de distribution des résidus ε_i . Entre autre on peut tester l'hypothèse de normalité avec le test de "W" (cf. 2.2.2). Si cette hypothèse est rejetée, on peut envisager l'ajustement de lois plus complexes (log-normale, gamma, Pearson type 3, ...).

2.2.3.2 Prédiction

L'intérêt de l'ajustement d'un processus de Markov d'ordre 1 à une série d'observations est de permettre d'envisager la prédiction des états futurs du phénomène.

On voit, en effet, d'après l'équation (3.34) que, si on connaît une valeur initiale x_0 , on peut reconstituer la série des états successifs du processus $x_1, x_2, \dots$. En réalité cette reconstitution aura un sens selon l'importance de la partie déterministe relativement à la partie aléatoire du processus.

La quantité t_1^2 représente le pourcentage de variance expliquée par la partie déterministe du processus. Ainsi, pour un processus tel que $t_1 = 0.7$, 49% de variance totale seulement sont expliqués par la partie déterministe (donc 51% sont aléatoires). Cependant $t_1 = 0.7$ est une valeur très significativement différente de zéro pour des échantillons de 22 observations comme

ceux qui sont utilisés dans ce travail. On conçoit que la prédiction ne sera envisageable que si l'auto-corrélation est forte ($t_1 > 0.85$). La partie aléatoire du processus sera alors réduite à de petites fluctuations. Dans les autres cas où l'effet déterministe est indéniable (t_1 significativement différent de zéro), sans être pour autant prépondérant de façon nette sur l'effet aléatoire, tout essai de reconstitution sera affecté d'une grande imprécision dû à la forte partie de la variance résiduelle, non prise en compte par la partie déterministe.

En conclusion, on peut dire que l'effet de persistance est, à moins d'être très fort, un handicap pour l'étude d'une série d'observations puisqu'il ne permet en définitive ni une étude statistique (ajustement de loi), ni une véritable étude chronologique.

2.2.3.3 Application: contenu d'information de la moyenne d'une série chronologique

On appelle contenu d'information d'un moment d'ordre k , la quantité

$$I(m_k, N) = \frac{1}{\text{var}(m_k)} \quad (2.40)$$

où m_k est l'estimateur du moment d'ordre k de la population, calculé à partir de l'échantillon de taille N .

Quand une série est persistante, il existe un lien déterministe entre deux observations successives. Du point de vue étudié ici, on dit qu'il y a répétition de l'information. Pour l'estimation de la moyenne de la population par exemple, un nombre inférieur d'observations serait suffisant, ce nombre étant tel que le lien déterministe entre les observations ait disparu.

Le contenu d'information de la moyenne, estimée à partir d'une série de N observations indépendantes, est:

$$I_1 = \frac{1}{\text{var}(\bar{x})} = \frac{1}{\sigma^2/N} \quad (2.41)$$

où σ^2 est la variance de la population.

Pour une série de N observations dépendantes, le contenu d'information de la moyenne peut s'écrire:

$$I_2 = \frac{1}{\sigma_m^2} \quad (2.42)$$

où σ_m^2 est la variance de la moyenne estimée.

On peut définir le contenu d'information relatif (I_r) de la série dépendante par rapport à la série aléatoire prise comme référence par:

$$I_r = \frac{I_2}{I_1}$$

$$I_r = \frac{\sigma^2/N}{\sigma_m^2} \quad (2.43)$$

a. Cas d'un processus de Markov d'ordre 1

Dans le cas où la série d'observations répond à un modèle de Markov d'ordre 1, on démontre (MATALAS et LANGBEIN, 1962) que l'estimateur de la moyenne a pour variance:

$$\sigma_m^2 = \frac{\sigma^2}{N} \left[\frac{1+t_1}{1-t_1} - \frac{2}{N} t_1 \frac{(1-t_1^N)}{(1-t_1)^2} \right] \quad (2.44)$$

D'après (2.43) et (2.44), il vient:

$$I_r = \left[\frac{1+t_1}{1-t_1} - \frac{2}{N} \frac{t_1(1-t_1^N)}{(1-t_1)^2} \right]^{-1} \quad (2.45)$$

Le nombre N' d'observations indépendantes suffisant pour obtenir la même information sur la moyenne que celle fournie par la série dépendante (markovienne d'ordre 1) de N observations est tel que le contenu d'information de la moyenne estimé à partir des N' observations soit identique à celui estimé à partir des N observations dépendantes.

Le contenu d'information I'_1 d'une série de N' observations indépendantes est, d'après (2.41)

$$I'_1 = \frac{1}{\sigma^2/N'}$$

Pour conserver l'information, il faut avoir (d'après 2.42):

$$I'_1 = I_2$$

$$\frac{1}{\sigma^2/N'} = \frac{1}{\sigma^2/N} \left[\frac{1+t_1}{1-t_1} - \frac{2}{N} \frac{t_1(1-t_1^N)}{(1-t_1)^2} \right]^{-1}$$

Soit, d'après (2.45):

$$\underline{\underline{N' = N \cdot I_r}} \quad (2.46)$$

Comme $I_r \leq 1$, on a toujours $N' \leq N$. Les N' observations doivent être choisies avec une fréquence telle qu'elles soient indépendantes.

b. Précision sur la moyenne

D'après (2.7), l'erreur commise sur la moyenne avec un échantillon de N' observations indépendantes est, au niveau de confiance $(1 - \alpha)$:

$$\ell = \left| \frac{t(\alpha/2) \cdot S}{\sqrt{N'}} \right| \quad (2.47)$$

On en déduit que la précision sur la moyenne obtenue à partir d'une série auto-correlée de N observations (telles que $N' = N \cdot I_r$) est, au niveau de confiance $(1 - \alpha)$:

$$\ell = \left| \frac{t(\alpha/2) \cdot S}{\sqrt{N \cdot I_r}} \right| \quad (2.48)$$

où I_r est donné par (2.45).

Le second terme du crochet de l'équation (2.45) est, en première approximation, négligeable devant le premier. La précision sur la moyenne s'écrit alors d'après (2.48), au niveau de confiance $(1 - \alpha)$:

$$\ell \approx \left| \frac{t(\alpha/2) \cdot S}{\sqrt{N \cdot \frac{1 - t_1}{1 + t_1}}} \right|$$

La précision sur la moyenne obtenue à partir d'une série d'observations dépendantes est moins bonne que la précision obtenue à partir d'une série d'observations indépendantes de même taille.

2.2.4 Ajustement de lois statistiques

2.2.4.1 Intérêt des lois

L'intérêt de l'ajustement d'une loi statistique est de réaliser une synthèse de la connaissance qu'on peut avoir sur une population aléatoire d'un

échantillon. On condense l'information donnée par les N observations dans la connaissance des deux ou trois paramètres estimés de la loi. La connaissance de la loi permet la prédiction d'événements extrêmes; par exemple, des événements de probabilité au dépassement très faible. On sera aussi en mesure de déterminer les intervalles de confiance de la loi à un certain niveau de confiance; ce qui permet d'évaluer l'intervalle de confiance, à ce niveau, associé à l'estimation d'un événement quelconque. Par contre on perd dans ce type d'étude tout effet chronologique.

2.2.4.2 Conditions d'application et choix des lois

Deux conditions de base sont requises pour l'ajustement d'une loi à partir d'un échantillon de N observations:

- les observations doivent être indépendantes. La vérification de l'indépendance est réalisée grâce à des tests dont la mise en oeuvre repose sur l'étude des coefficients d'auto-corrélation (test de WALD-WOLFOWITZ, 1943 ; ANDERSON, 1941);
- l'échantillon doit être représentatif du phénomène étudié, c'est-à-dire que les observations doivent être des réalisations de la même variable aléatoire et doivent provenir de la même population (test de MANN-WHITNEY, 1947).

En général plusieurs lois sont susceptibles d'être ajustées à un échantillon pour déterminer la population statistique dont peuvent provenir les valeurs prises par un paramètre de qualité. Si aucune considération purement théorique ne permet d'effectuer de choix entre ces lois, certaines considérations physiques peuvent cependant servir de guide; on sait, par exemple,

qu'un paramètre de qualité est borné inférieurement par la limite de détection et l'on peut également utiliser le fait que le coefficient d'asymétrie des paramètres chimiques est, en général, positif ou nul.

2.2.4.3 Ajustement des lois

On dispose de deux méthodes principales d'ajustement:

- la méthode des moments qui consiste à estimer les moments de la population, qui sont fonctions des paramètres de la loi, par les valeurs correspondantes de l'échantillon;
- la méthode du maximum de vraisemblance qui consiste à maximiser, par rapport aux paramètres de la loi, la fonction vraisemblance des observations.

Ces deux méthodes ne conduisent pas en général aux mêmes résultats. Selon le type de loi envisagée, la précision recherchée et la taille de l'échantillon, on emploiera de préférence l'une ou l'autre méthode.

La méthode du maximum de vraisemblance est souvent optimale mais, dans le cas de lois pour lesquelles l'intervalle de définition de la variable aléatoire dépend d'un paramètre à estimer (loi Pearson type 3, par exemple), l'optimalité de la méthode n'est vraie que de manière asymptotique. Pour des échantillons de petite taille, la méthode des moments est alors préférable.

On utilise souvent le coefficient d'asymétrie dans la mise en oeuvre de la méthode des moments pour les lois à trois paramètres. Ce coefficient est difficile à estimer correctement et des biais importants peuvent s'introduire. Diverses corrections de biais ont été proposées, dont celle de BOBEE et ROBITAILLE (1975) qui est utilisée dans cette étude pour l'ajustement des lois (cf. 2.1.1).

Notons enfin que lorsqu'on considère une distribution déduite par transformation logarithmique, log-Pearson type 3 par exemple, il est préférable d'égaliser les moments de l'échantillon original à ceux de la distribution log-Pearson type 3 (BOBEE, 1975) plutôt que d'égaliser les moments de la série des logarithmes des valeurs observées à ceux de la distribution Pearson type 3. BOBEE et ROBITAILLE (1977) ont étudié pour les lois Pearson type 3, gamma, log-Pearson type 3 et log-gamma, les influences diverses sur l'ajustement des lois du choix des méthodes d'ajustement et de l'estimation du coefficient d'asymétrie pour plusieurs séries de données. C'est ainsi que, compte-tenu des résultats de cette étude et compte-tenu des caractéristiques propres aux paramètres de qualité, on a réalisé au cours de ce travail, pour un certain nombre de paramètres, les ajustements suivants:

- loi gamma, méthode du maximum de vraisemblance;
- loi Pearson type 3, méthode des moments, correction d'asymétrie $(C_s)_3$ (cf. 2.1.4);
- loi log-gamma, méthode des moments appliquée à la série des valeurs observées;
- loi log-Pearson type 3, méthode des moments appliquée à la série des valeurs observées.

L'ajustement dans ces différents cas est réalisé par un programme de calcul mis au point à l'INRS-Eau.

Pour effectuer un choix final de la loi qui décrira le paramètre de qualité, le meilleur moyen semble être de tracer les diverses lois calculées et de regarder celle qui s'ajuste le mieux aux observations. En effet les divers tests (Chi-2, KOLMOGOROV-SMIRNOV) sont très indulgents et ne sont pas discriminants pour le choix des lois.

2.2.4.4 Intervalles de confiance

L'intervalle de confiance d'un événement X_p , de probabilité au dépassement donnée, à un certain niveau de confiance, est déterminé par le programme de calcul utilisé pour l'ajustement des lois, selon le principe suivant: quand la taille N de l'échantillon des observations est suffisamment grande, l'estimateur $\hat{X}_p$ (donnée par la loi ajustée) de la vraie valeur X_p est distribuée suivant une loi normale de moyenne X_p avec une variance $\text{var}(\hat{X}_p)$. On calcule $\text{var}(\hat{X}_p)$ selon la loi envisagée (BOBEE, 1973). L'intervalle de confiance X_p au niveau de confiance $(1 - \alpha)$ est alors:

$$\hat{X}_p - u_{\alpha/2} \sqrt{\text{var}(\hat{X}_p)} \leq X_p \leq \hat{X}_p + u_{\alpha/2} \sqrt{\text{var}(\hat{X}_p)} \quad (2.49)$$

On peut aussi envisager de calculer l'intervalle de confiance de la loi pour un niveau de confiance donné d'après les remarques faites précédemment sur les quantiles des distributions (cf. 2.1.3.1). Dans ce cas, l'intervalle de confiance d'un quantile q_p , au niveau de confiance $(1 - \alpha)$, est donné par l'équation 2.15.

Les deux méthodes font appel à la même hypothèse "asymptotique": on suppose que l'estimateur $\hat{X}_p$ est distribué normalement avec pour moyenne X_p et pour variance une quantité $\text{var}(\hat{X}_p)$. Ce résultat n'est rigoureux que si N est suffisamment grand. Par contre, ces deux méthodes diffèrent dans le calcul de $\text{var}(\hat{X}_p)$. L'intérêt de la deuxième méthode est que le calcul de cette variance est relativement simple. Cependant il reste à comparer théoriquement les deux procédés de façon à déterminer l'importance relative des approximations réalisées par l'un et par l'autre au cours des diverses étapes de calcul (approximation asymptotique, calcul de la variance). On donne un exemple de comparaison des deux méthodes en 5.2.3.

2.2.4.5 Lois utilisées. loi gamma (cas de l'asymétrie positive)

$$f(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)} e^{-\alpha x} x^{\lambda-1} \quad \begin{cases} 0 \leq x < \infty \\ \alpha > 0, \lambda > 0 \end{cases} \quad (2.50)$$

On a :

$$\mu = \frac{\lambda}{\alpha} \quad \sigma^2 = \frac{\lambda}{\alpha^2} \quad C_s = \frac{2}{\sqrt{\lambda}} \quad C_s = 2 C_v$$

. loi Pearson type 3 (cas de l'asymétrie positive)

$$f(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)} e^{-\alpha(x-m)} (x-m)^{\lambda-1} \quad \begin{cases} m \leq x < \infty \\ \alpha > 0, \lambda > 0 \end{cases} \quad (2.51)$$

On a :

$$\mu = \frac{\lambda}{\alpha} + m \quad \sigma^2 = \frac{\lambda}{\alpha^2} \quad C_s = \frac{2}{\sqrt{\lambda}}$$

. loi log-Pearson type 3

Cette loi est telle que le logarithme de la variable aléatoire suive une loi Pearson type 3.

Si $y = \log x$ suit une loi Pearson type 3, x suit une loi log-Pearson type 3 dont la fonction densité est :

$$g(x) = \frac{|\alpha|}{\Gamma(\lambda)} e^{-\alpha(\log x - m)} [\alpha(\log x - m)]^{\lambda-1} \cdot \frac{1}{x \ln 10} \quad (2.52)$$

$\log$ = logarithme décimal

$\ln 10$ = logarithme népérien de 10

avec:

$$\begin{cases} \lambda > 0 \\ \alpha > 0 & 10^m < x < \infty \\ \alpha < 0 & 0 < x < 10^m \end{cases}$$

loi log-gamma

Si $y = \log x$ suit une loi gamma, x suit une loi log-gamma de densité:

$$g(x) = \frac{|\alpha|}{\Gamma(\lambda)} e^{-\alpha \log x} (\alpha \log x)^{\lambda-1} \frac{1}{x \ln 10} \quad (2.53)$$

avec:

$$\begin{cases} \lambda > 0 \\ \alpha > 0 & x > 1 \\ \alpha < 0 & 0 < x < 1 \end{cases}$$

2.3 Relations entre séries d'observations. Analyse globale des séries d'observations.

L'étude de la dimension spatiale du réseau (variabilité spatiale, répartition géographique et évolution de certaines caractéristiques statistiques) s'effectue par la comparaison des séries d'observations, deux à deux ou dans leur ensemble.

Ce type d'étude permet de justifier la nécessité d'augmenter le réseau ou, au contraire, la possibilité de le réduire. Les deux types de méthodes envisagées sont:

- étude des corrélations entre stations, régressions linéaires, contenu d'information de la moyenne régionale;
- analyse factorielle des correspondances.

2.3.1 Corrélation - Régressions linéaires

2.3.1.1 Corrélation

Le coefficient de corrélation entre deux séries d'observations

(x_i) et (y_i) , $(i = 1, \dots, N)$ est:

$$r = \frac{\sum_{i=1}^N x_i y_i - N \bar{x} \bar{y}}{\sqrt{\left(\sum_{i=1}^N x_i^2 - N \bar{x}^2\right) \left(\sum_{i=1}^N y_i^2 - N \bar{y}^2\right)}} \quad (2.54)$$

$$-1 \leq r \leq +1$$

Deux hypothèses sont à respecter pour l'interprétation de cette quantité:

- les séries d'observations sont approximativement distribuées normalement;
- la dépendance entre les séries doit être linéaire pour être caractérisée par le coefficient de corrélation.

Le coefficient de corrélation mesure une dépendance linéaire mais ne permet pas de déceler une dépendance non linéaire si elle existe. Par ailleurs un fort coefficient de corrélation peut être le résultat d'une "corrélation fictive" que seul un examen graphique permet de déceler (exemple: répartition des observations en deux nuages de points distincts).

. test sur le coefficient de corrélation

Le coefficient r est un estimateur de la vraie valeur ρ qu'on obtiendrait à partir des populations. Une valeur relativement élevée de r , surtout quand la taille des échantillons est faible, ne permet pas de conclure nécessairement

rement à une liaison linéaire significative. Il faut tester si ρ est significativement différent de zéro. On montre que la quantité:

$$t = \frac{r}{\sqrt{1 - r^2} / \sqrt{N - 2}} \quad (2.55)$$

suit une loi de Student à $(N - 2)$ degrés de liberté.

Si, pour un niveau de signification α on a $|t| \leq t_{N-2} (\alpha/2)$, on accepte l'hypothèse que $\rho = 0$.

Connaissant N , on peut calculer la valeur critique r_0 que r doit dépasser pour que l'hypothèse $\rho \neq 0$ soit acceptée (d'après 2.55), on a:

$$|r_0| = \frac{t_{N-2} (\alpha/2)}{\sqrt{[t_{N-2} (\alpha/2)]^2 + (N-2)}} \quad (2.56)$$

Ainsi, pour $N = 20$, au niveau de signification $\alpha = 5\%$,

$$t_{N-2} (\alpha/2) = 2.10$$

$$|r_0| = .445$$

Pour toute valeur calculée r inférieure à 0.445, on conclura que ρ n'est pas significativement différent de zéro au niveau 5%.

2.3.1.2 Régression linéaire

Si on dispose d'un fort coefficient de corrélation, significativement différent de zéro, entre deux variables X et Y et que la corrélation ne semble pas fictive d'après un examen graphique, on calcule la régression linéaire d'une variable en fonction de l'autre, de la façon suivante:

On considère Y en fonction de X .

Soient (x_i) et (y_i) $i = 1, \dots, N$ les valeurs observées

Y_i la valeur prédite par la régression est

$$Y_i = a + b (x_i - \bar{x})$$

L'équation de la droite est:

$$Y = a + b (X - \bar{x}) \quad (2.57)$$

a et b sont les coefficients de la droite de régression, déterminés par la minimisation de la somme des carrés des écarts $(Y_i - y_i)$ entre la valeur prédite et la valeur observée.

On montre que:

$$a = \bar{y} = \sum_{i=1}^N y_i / N$$

$$b = \frac{\sum_{i=1}^N x_i y_i - N \bar{x} \bar{y}}{\sum_{i=1}^N x_i^2 - N \bar{x}^2}$$

Intervalle de confiance et précision

Quand on reconstitue une valeur Y_i à partir d'une observation x_i , avec l'équation de régression, il est important de chiffrer la précision de cette reconstitution.

La quantité:

$$t = \frac{y_i - Y_i}{S \sqrt{1 + \frac{1}{N} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^N (x_i - \bar{x})^2}}} \quad (2.58)$$

suit une loi de Student à $(N - 2)$ degrés de liberté

avec:

$$S^2 = \frac{1}{N - 2} \sum_{i=1}^N (y_i - Y_i)^2$$

On en déduit l'intervalle de confiance au niveau $(1 - \alpha)$ d'une valeur y_i pour x_i fixée:

$$Y_i - t_{N-2}(\alpha/2) \sqrt{1 + \frac{1}{N} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \leq y_i \leq Y_i + t_{N-2}(\alpha/2) \sqrt{1 + \frac{1}{N} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \quad (2.59)$$

Quand x varie, on obtient les intervalles de confiance au niveau $(1 - \alpha)$ des valeurs reconstituées qui sont des hyperboles d'équation:

$$y = a + b (x - \bar{x}) \pm t_{N-2} (\alpha/2) S \sqrt{1 + \frac{1}{N} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \quad (2.60)$$

D'après (2.59), l'erreur maximale ℓ_0 commise dans la reconstitution d'une valeur y à partir d'une observation x_0 au niveau de confiance $(1 - \alpha)$ sera:

$$\ell_0 = t_{N-2} (\alpha/2) \cdot S \sqrt{1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \quad (2.61)$$

La valeur reconstituée est (d'après 2.57):

$$Y_0 = a + b (x_0 - \bar{x})$$

La précision (en pourcentage) de la reconstitution est:

$$\varepsilon_0 = 100 \cdot \frac{\ell_0}{Y_0}$$

2.3.1.3 Contenu d'information de la moyenne régionale

Soit un réseau de n stations. L'information apportée par les n stations sur la moyenne régionale d'un paramètre de qualité est-elle redondante et, dans ce cas, quel sera le nombre suffisant n' de stations indépendantes?

Soient $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_n$ les moyennes à chaque station.

La moyenne régionale est:

$$\bar{Y} = \sum_{i=1}^N \bar{y}_i / n$$

Si ρ_{ij} est le coefficient de corrélation de la station i avec la station j , on définit le coefficient de corrélation moyen $\bar{\rho}$ par:

$$\bar{\rho} = \sum_{i>j} \sum \rho_{ij} / \frac{n(n-1)}{2}$$

On suppose que chaque série est aléatoire, comporte N observations, et que leur variance commune est σ^2 .

La variance de la moyenne régionale est donnée par:

$$\frac{\sigma^2}{\bar{Y}} = \frac{\sigma^2}{Nn} [1 + \bar{\rho} (n-1)]$$

Quand toutes les stations sont complètement indépendantes les unes des autres, $\bar{\rho} = 0$ et le contenu d'information est:

$$I_1 = \frac{1}{\text{var}(\bar{Y})} = \frac{1}{\sigma^2/Nn} \quad (2.62)$$

Quand les séries ne sont pas complètement indépendantes, $\bar{\rho}$ est différent de zéro et le contenu d'information est (MATALAS et LANGBEIN, 1962):

$$I_2 = \left[\frac{\sigma^2}{Nn} [1 + \bar{\rho} (n-1)] \right]^{-1} \quad (2.63)$$

Le contenu d'information relatif est:

$$I_r = \frac{I_2}{I_1} = [1 + \bar{\rho} (n-1)]^{-1} \quad (2.64)$$

Le réseau de stations complètement indépendantes est pris comme référence. On suppose $\rho_{ij} \geq 0$ quels que soient i et j .

Quand $\bar{\rho} = 0$ et $\rho_{ij} = 0$ quels que soient i et j , le contenu d'information I est n fois l'information à une station; chaque station contribue pleinement à l'information.

Si par contre $\bar{\rho} = 1$, $\rho_{ij} = 1$ quels que soient i et j , $I_r = \left(\frac{\sigma^2}{N}\right)^{-1}$, le contenu d'information du réseau sur la moyenne régionale se réduit à celui d'une seule station.

Quand $0 < \bar{\rho} < 1$, on obtiendra la même information sur la moyenne régionale avec un nombre n' de stations tel que:

- les n' stations sont indépendantes ($\rho_{ij} = 0$ quels que soient i et j) de sorte qu'il n'y a plus duplication d'information;
- la variance de la moyenne régionale est conservée. On a la même information sur cette moyenne.

Le contenu d'information de ces n' stations sera:

$$I_1' = \frac{1}{\sigma^2/Nn'} \quad (\text{indépendance}) \quad (2.65)$$

$$I_1' = I_2 \quad (\text{conservation du contenu d'information})$$

Il vient d'après (2.63) et (2.65) et (2.64):

$$\underline{\underline{n' = n \cdot I_r}} \quad (2.66)$$

$$I_r = [1 + \bar{\rho} (n - 1)]^{-1}$$

$$\bar{\rho} > 0 \quad I_r < 1 \quad n' < n$$

Les n' stations devront être choisies indépendantes.

Remarquons que ce calcul fait appel à des hypothèses difficiles à respecter dans la réalité (les n séries sont supposées aléatoires et de variance commune) et qu'il n'est valable que vis-à-vis de la moyenne régionale. C'est un calcul mathématique qui est seulement indicatif.

En effet (2.66) s'écrit:

$$n' = \frac{n}{1 + \bar{\rho} (n - 1)}$$

Quand $n \rightarrow \infty$ $n' \rightarrow 1/\bar{\rho}$

ainsi, pour $\bar{\rho} = 0.5$, $n' = 2$ stations au maximum;

pour $\bar{\rho} = 0.3$, $n' = 3$ alors qu'en réalité $n = 19$ stations

(réseau Yamaska)

La réduction paraît très grande mais ces trois stations doivent être indépendantes. Par ailleurs on conseille dans le calcul de $\bar{\rho}$ d'utiliser les valeurs (ρ_{ij}) qui sont les limites inférieures des intervalles de confiance à 95% de ρ_{ij} qu'il est possible de calculer. On obtiendra alors une estimation sécuritaire de $\bar{\rho}$ et donc de n' .

En pratique, essentiellement en raison des proximités géographiques, il est souvent difficile, voire impossible, de déterminer des stations réelles qui ne soient pas corrélées entre elles. Cependant les calculs précédents donnent une bonne idée du niveau de redondance de l'information relative à l'estimation d'une moyenne régionale.

2.3.2 Analyse des correspondances

L'analyse des correspondances (BENZECRI, 1973) permet d'obtenir des

représentations synthétiques de vastes ensembles de valeurs numériques et d'en dégager les principales caractéristiques. Un réseau de mesures fournit un grand nombre de données que les techniques classiques de traitement ne peuvent considérer que de façon très partielle (exemple: étude d'un paramètre à une station, étude de deux stations pour un paramètre, étude d'une caractéristique statistique d'un paramètre pour l'ensemble des stations). Les méthodes multivariées comme l'analyse des correspondances permettent, par contre, une approche globale et synthétique.

L'analyse des correspondances nous permet de faire ressortir simultanément les relations inter-stations, inter-paramètres, stations-paramètres: c'est donc une méthode descriptive très puissante.

Description de la méthode

Considérons un tableau de dimension $n \times p$ (n lignes, p colonnes). Soit k_{ij} un élément de ce tableau; k_{ij} pourrait, par exemple, être la moyenne des observations du paramètre j à la station i . On effectue, sur ce tableau, les transformations suivantes:

$$\begin{aligned} k_{ij} &\rightarrow p_{ij} = \frac{k_{ij}}{k} & k &= \sum_i \sum_j k_{ij} \\ p_{ij} &\rightarrow r_{ij} = \frac{p_{ij} - p_{i.} p_{.j}}{\sqrt{p_{i.} p_{.j}}} & \text{avec} & \begin{cases} p_{i.} = \sum_j p_{ij} \\ p_{.j} = \sum_i p_{ij} \end{cases} \end{aligned}$$

Ces transformations sont symétriques par rapport à i et j . On associe au tableau des p_{ij} , deux nuages de points:

- dans R^p , on construit un nuage de n points M_i ayant pour coordonnées

$$x_{ij} = \frac{p_{ij}}{p_{i.} \sqrt{p_{.j}}} \quad j = 1, \dots, p$$

- dans R^n , on construit un nuage de p points P_J ayant pour coordonnées

$$y_{ij} = \frac{p_{ij}}{p_{.j} \sqrt{p_{i.}}} \quad i = 1, \dots, n$$

Considérons le nuage des p points P_J dans R^n . On peut ajuster ce nuage au sens des moindres carrés par un hyperplan de dimension r ($r \leq p-1$), déterminé par r axes orthogonaux dits axes factoriels. Les axes factoriels ont la propriété d'extraire progressivement le plus d'information possible concernant les proximités entre les points p_j , c'est-à-dire concernant la variance du nuage. Ainsi le premier axe factoriel prend en compte la plus grande partie de la variabilité du nuage (axe d'allongement du nuage). Le deuxième axe factoriel prend en compte la plus grande variance restante, dans une direction orthogonale à celle du premier axe, etc... En pratique on utilise surtout les trois premiers axes.

Le terme r_{ij} , transformé de p_{ij} , est le terme général d'une matrice R ($n \times p$).

Soit $V_1 = R \times R'$ (R' : transposée de R)

On montre que chaque valeur propre λ_i de V_1 représente la variance du nuage de points calculée le long de l'axe factoriel qui a pour direction le vecteur propre associé à λ_i . A la plus grande valeur propre λ_1 est associée le premier axe factoriel. La permutation des indices i et j transpose la matrice R : l'analyse du nuage des n points M_i de R^p conduit à l'étude de la matrice $V_2 = R' \times R$. On montre que cette matrice a les mêmes valeurs propres que V_1 et que les vecteurs propres sont liés par une relation simple à ceux de V_1 .

Donc, dans un premier temps, la méthode consiste dans le calcul des valeurs propres et des vecteurs propres de la matrice $R \times R'$ (ou $R' \times R$).

Dans un deuxième temps on calcule les projections des points, P_j par exemple, sur les divers axes factoriels. La projection d'un point sur le $i^{\text{ème}}$ axe factoriel est le facteur i relatif à ce point. En pratique on se limite aux axes factoriels qui expliquent la plus grande partie de la variance du nuage (entre trois et cinq). Puis on projette les points P_j sur les plans formés par ces axes factoriels pris deux à deux. On se limite en général au plan formé par les deux premiers axes factoriels et, éventuellement, pour aider à l'interprétation, on considère les projections du nuage sur les plans formés par les axes 1 et 3, ainsi que 2 et 3.

Pour l'interprétation des représentations graphiques obtenues sur ces plans factoriels on utilise les propriétés suivantes de l'analyse des correspondances:

- en raison des transformations initiales du tableau de départ, des points (lignes ou colonnes) qui ont des composantes proportionnelles se trouvent confondues dans le nuage de points et donc aussi en projection. On pourra ainsi mettre en évidence les similitudes de comportement de points M_i ou de points P_j , selon que leurs projections seront confondues ou voisines dans la représentation graphique;
- en raison des propriétés (effet barycentrique) des matrices V_1 et V_2 , il est possible de superposer sur un même plan de projection les points M_i et les points P_j . On démontre que, dans cette représentation simultanée, la proximité entre un point M_i et un groupe de points P_j signifie que le point M_i caractérise le comportement du groupe de points P_j .

En résumé, la mise en oeuvre de la méthode d'analyse des correspondances est la suivante:

- calcul de la matrice R ;
- calcul de la matrice $V_1 = R \times R'$;
- détermination des valeurs propres et vecteurs propres de V ;
- calcul des projections des points M_i et P_j sur les axes factoriels choisis;
- représentation graphique des projections des points M_i et P_j sur les plans factoriels choisis.

On appelle parfois "observations" les points M_i et "variables" les points P_j .

En conclusion, la méthode permet:

- de regrouper les observations qui ont un comportement semblable;
- de regrouper les variables qui ont un comportement semblable;
- d'identifier les variables qui expliquent le comportement d'un groupe d'observations.

Cette brève synthèse de la méthode d'Analyse Factorielle des Correspondances montre son intérêt pour l'analyse des grands tableaux de données. Une description plus complète de cette méthode et de ses applications possibles est donnée par BENZECRI (1973).

CHAPITRE 3

DESCRIPTION DU BASSIN VERSANT DE LA RIVIERE YAMASKA ET DU RESEAU DE MESURE DE QUALITE - OBJECTIFS PROPOSES

INTRODUCTION

Le bassin de la rivière Yamaska est un des bassins hydrographiques québécois les plus détériorés du point de vue de la qualité des eaux. Les collectivités humaines, les industries et surtout l'utilisation agricole sont cause de cet état (MALO-LALANDE, 1976). Des études diverses ont été entreprises sur ce bassin et un réseau de mesure a été mis en place dans le cadre d'une étude conjointement menée par le Ministère des Richesses naturelles (M.R.N. - Qualité des eaux) et l'Institut National de la Recherche Scientifique (INRS-Eau).

3.1 Description du bassin versant

Le bassin versant de la rivière Yamaska est borné à l'est par le bassin de la rivière Saint-François, à l'ouest et au sud par celui de la rivière Richelieu et au nord par le fleuve Saint-Laurent. Le bassin s'étend sur une superficie de 4910 km².

3.1.1 Physiographie

On distingue trois zones principales qui sont, du nord au sud (MALO-LALANDE, 1976):

- la région dite des "Basses terres", plaine formée de sédiments marins (argile et sable) où le terrain s'incline graduellement vers le fleuve Saint-Laurent. Son altitude varie entre 7.5 et 90 mètres

environ. Il s'agit des meilleures terres agricoles de la province. Les sédiments argileux de la mer Champlain qui recouvrent en majorité cette région sont particulièrement sujets à l'érosion;

- la région dite du "Piedmont" qui constitue la zone de transition entre les Basses-terres et la région des Appalaches. Cette région ondulée à fortement valonnée, est formée de "till" appalachien et de minces dépôts marins reposant sur du "till". L'altitude varie de 90 à 180 mètres;
- la zone montagneuse formée par une partie de la chaîne des Appalaches (partie sud-est du bassin). Le relief y est très accidenté. On note la présence de nombreux blocs erratiques, des dépôts morainiques et fluvio-glaciaires.

La zone plane (Basses-terres et Piedmont) voit sa régularité perturbée par la présence des structures montérégiennes isolées (Rougemont, Yamaska, Brôme, Shefford).

Les trois zones ont subi les effets de la glaciation; seules les deux premières furent sujettes à l'invasion marine de la mer Champlain.

3.1.2 Géologie

. Roches consolidées

Le bassin chevauche deux provinces géologiques distinctes, séparées par la faille de Champlain ou faille de Logan, qui s'oriente nord-est (MALO-LALANDE, 1976).

- * Les Basses-terres du nord-ouest du bassin sont formées de roches sédimentaires faiblement plissées (âge ordovicien). Ces roches comprennent surtout des "shales" (schistes argileux), ainsi que des grès et des roches carbonatées (calcaires et dolomies). Les affleurements sont peu nombreux et souvent localisés près des cours d'eau.

Les collines montérégiennes sont formées de roches ignées intrusives. Elles occupent une infime partie du territoire.

- * Les Appalaches sont constituées de roches fortement plissées (relief accentué). On rencontre des roches sédimentaires (âge paleozoïque, Ordovicien, silurien et Dévonien) presque toutes métamorphisées (schistes argileux et ardoises, grès, "siltstone"), ainsi que des roches volcaniques du même âge en étroite association avec les roches sédimentaires. Cet assemblage de roches sédimentaires et volcaniques est envahi d'intrusion de roches ignées ultra-basiques et acides.

Dépôts meubles

- * Les dépôts meubles de la région des Basses-terres sont constitués en majorité d'argiles marines. Celles-ci recouvrent environ 40% du bassin de la Yamaska. Leur épaisseur moyenne peut atteindre 75 mètres. Elles sont composées, pour 70 à 75%, de fines particules des roches laurentiennes charriées par les glaciers (minéraux: quartz, feldspath, amphibole). Pour 25 à 30% elles sont formées de minéraux argileux et de carbonates qui proviennent de la désagrégation des roches sous-jacentes des Basses-terres.

Ces argiles marines reposent sur des tills glaciaires ou des dépôts fluvio-glaciaires.

- * Les dépôts meubles dominants de la région des Appalaches sont des "tills" glaciaires (sédiments désassortis et non stratifiés transportés et déposés sous l'action des glaciers). La composition minéralogique des tills appalachiens varie rapidement localement en raison de la grande diversité des roches sous-jacentes.

3.1.3 Pédologie

On trouve une grande variété de sols: sols dérivés d'alluvions fluviales, de dépôts lacustres, de sédiments de la mer Champlain, de "till" appalachien et dépôts fluvio-glaciaires; on trouve aussi des sols organiques et des alluvions non différenciées qui sont associées aux plaines d'inondation. Ces sols sont riches en matière organique et possèdent une fertilité remarquable. Les sols dérivés de la mer Champlain sont plus aptes à la grande culture que ceux dérivés des "tills" et des dépôts fluvio-glaciaires.

Les Basses-terres ont un grand potentiel agricole et conviennent aux grandes cultures. La qualité des sols du Piedmont est plus pauvre. Le potentiel agricole du plateau appalachien est très faible en raison d'une topographie accidentée et de la nature pierreuse des sols (MALO-LALANDE, 1976).

3.1.4 Hydrographie

Le bassin de la rivière Yamaska a une forme compacte qui se termine par un exutoire effilé (cf. Figure 3.1). La tête du bassin est située dans les Appalaches. Etant donné l'orientation sud-nord du bassin, les conditions sont réunies pour que se développent des crues subites et intenses lors de la fonte des neiges. Au centre du bassin, à 5.3 milles en amont de la ville de Sainte-Hyacinthe, se rencontrent deux rivières d'égale importance qui forment la trame du réseau:

- la rivière Yamaska proprement dite avec un bassin versant de 653 milles carrés;
- la rivière Noire avec un bassin versant de 593 milles carrés.

La rivière Yamaska, dans son sous-bassin amont, se divise en trois branches d'importance équivalente:

- . la rivière Yamaska-Nord issue du lac Waterloo (bassin versant: 125 milles carrés) qui traverse Granby;
- . la rivière Yamaska-Sud-Est qui traverse Cowansville (bassin versant: 170 milles carrés);
- . le tronçon principal de la Yamaska, désigné du nom de Yamaska-Centre (bassin versant: 190 milles carrés), issu du lac Brôme qui est le plus important lac du bassin.

Ces trois branches sont assez semblables en ce qui concerne la physiographie de leurs bassins. La vie économique et industrielle se concentre sur les rives de la rivière Yamaska.

La rivière Noire ne comporte pas de tributaires aussi importants que ceux de la Yamaska. La partie montagneuse ne concerne qu'une petite partie de son cours. Il y a peu de forêts mais un important réseau de drainage collecte les eaux d'irrigation. Les lacs sont inexistants (sauf lac Roxton, 0.63 milles carrés). Les étiages sont sévères et les crues plus accusées que sur la Yamaska, mais le bassin ne comporte pas d'agglomération majeure et les besoins en eau sont moindres.

Dans son cours inférieur (en aval de Saint-Hyacinthe), la rivière Yamaska ne reçoit que des tributaires peu importants ayant toutes les caractéristiques des rivières de plaines. Elle se jette dans le lac Saint-Pierre.

3.1.5 Population, industrie, utilisation du sol

L'agriculture, ainsi que la production animale, sont très développées sur le bassin. Signalons l'industrie laitière (Coopérative de Granby qui manipule 75% du lait industriel du Québec) et la production porcine qui occasionnent des pollutions organiques importantes dans l'eau.

La région la plus industrialisée est celle de Granby (Yamaska-Nord) (textile, caoutchouc, industries alimentaires, tabac, produits métallurgiques et électriques). Les régions de Saint-Hyacinthe (textiles) et de Sorel (industrie lourde) sont aussi industrialisées. Saint-Hyacinthe et Granby sont les deux villes les plus importantes. Notons aussi l'importance de Cowansville, Waterloo, Farnham et Acton Vale.

Si la zone plane possède un fort potentiel agricole, la zone montagneuse est plutôt vouée aux loisirs et à l'exploitation forestière.

3.1.6 Problèmes de qualité de l'eau sur le bassin

• pollution agricole

La pollution agricole provient des fumiers, des engrais à base de phosphore et d'azote et des insecticides divers. Cette forme de pollution affecte la zone plane du bassin. Une étude de l'INRS-Eau (1976) sur l'impact de l'agriculture conclue que les concentrations en substances nutritives et en sédiments en suspension trouvées dans plusieurs sous-bassins agricoles dépassent les normes généralement acceptées.

• pollution domestique et industrielle

Les conclusions de l'étude de l'O.P.D.Q. (1973) sont les suivantes:

- cent cinquante-six réseaux d'égouts municipaux se déversent directement, sans traitement, dans les cours d'eau. Sur sept postes d'épuration existant, deux seulement fonctionnent d'une façon efficace;
- cent cinquante-cinq industries, sur les cent quatre-vingts inventoriées, envoient leurs déchets dans les réseaux d'égouts (usines d'abattage et de préparation de la viande, de transformation de produits laitiers, de conserverie, d'embouteillage d'eaux gazeuses, de textiles et produits chimiques).

D'après l'O.P.D.Q., dans l'ensemble du bassin de la Yamaska, la pollution d'origine industrielle produit des effets beaucoup plus néfastes que la pollution d'origine domestique.

déchets solides

Actuellement quelques 110,000 tonnes de déchets solides sont déversées quotidiennement sur le territoire, selon des techniques "archaïques" et peu efficaces (35 dépotoirs municipaux sur 37 sont à ciel ouvert, les dépotoirs particuliers abondent dans les autres municipalités où on ne procède pas à l'enlèvement des déchets). (O.P.D.Q., 1973).

Un rapport de l'INRS-Eau (1976) sur les indices de danger environnemental sur les bassins de la Yamaska et de la Saint-François conclut que:

- "les poissons capturés dans la rivière Yamaska sont dans une condition physiologique misérable par rapport à ceux capturés dans la rivière Saint-François. Cet état physiologique té-

moigne des pressions considérables exercées sur la santé des organismes aquatiques de ce bassin".

- Il y a présence de mercure et de plomb dans l'eau et les sédiments de la Yamaska. Les concentrations de mercure et de plomb sont élevées dans les organismes capturés.

3.2 Description du réseau de mesure

3.2.1 Localisation et densité du réseau

La carte de la figure 3.1 donne la position des 19 stations du réseau. Le mode d'implantation du réseau a été décrit en 1.3.2.3 b. Le tableau 3.1 donne la position exacte des stations de mesure (généralement situées sur des ponts). La figure 3.2 donne une représentation schématique du réseau qui situe les stations les unes par rapport aux autres.

3.2.2 Dates et fréquences des mesures

Les mesures ont commencé en novembre 1973 de façon épisodique jusqu'en février 1974. Elles se sont alors poursuivies régulièrement jusqu'en novembre 1975. Pour traiter des tableaux de données complets, nous n'avons considéré que les observations effectuées mensuellement de février 1974 à novembre 1975. On a vérifié que l'intervalle de temps séparant deux observations consécutives était régulier et à peu près égal à 30 jours. On note cependant l'absence d'observations, pour l'ensemble du réseau, en janvier et en septembre 1975. On dispose ainsi d'un ensemble maximum de 20 observations à chaque station pour chaque paramètre de qualité considéré.

Pour les prélèvements sur le terrain, le réseau a été divisé en deux parties:

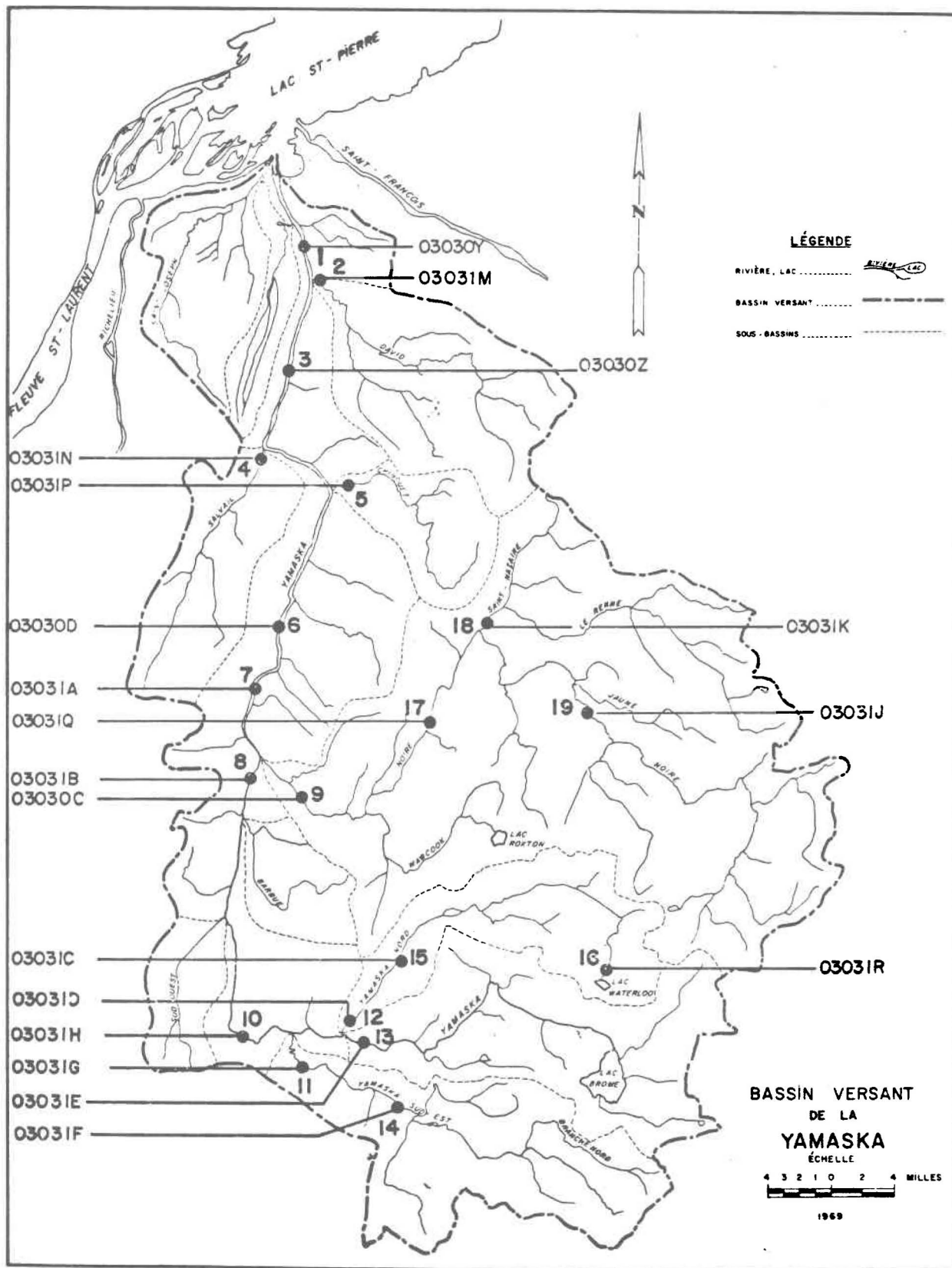


FIGURE 3.1: Localisation des stations d'échantillonnage du réseau.

TABLEAU 3.1: Liste des stations de mesure

Station	Rivière	Localisation
1- 03030Y	Yamaska	au pont-route à Yamaska
2- 03031M	David	au pont-route à l'embouchure
3- 03030Z	Yamaska	au pont-route de Saint-Marcel
4- 03031N	Salvail	au pont-route à son embouchure
5- 03031P	Chibouet	au pont-route à Saint-Hughes
6- 03030D	Yamaska	au pont-route 20
7- 03031A	Yamaska	au pont-route à Douville
8- 03031B	Yamaska	au pont-route à 4.0 km en amont de la Rivière Noire
9- 03030C	Noire	au pont-route à Saint-Pie
10- 03031H	Yamaska	au pont-route à 1.6 km en aval de Farnham
11- 03031G	Yamaska Sud-Est	à 10.9 km en aval de Brigham
12- 03031D	Yamaska-Nord	au pont-route à 1.9 km en aval de Saint-Alphonse de Granby
13- 03031E	Yamaska	au pont-route à 5.1 km en aval d'Adamsville
14- 03031F	Yamaska Sud-Est	au pont-route à 7.7 km en aval de Cowansville
15- 03031C	Yamaska-Nord	au pont-route à 7.7 km en aval de Granby
16- 03031R	Yamaska-Nord	au pont-route 39 à 2.9 km en aval du lac Waterloo
17- 03031Q	Noire	au pont-route de Saint-Hector de Bagot
18- 03031K	Saint-Nazaire	au pont-route à 2.4 km de son embouchure
19- 03031J	Noire	au pont-route à 2.8 km en aval de Roxton-Falls

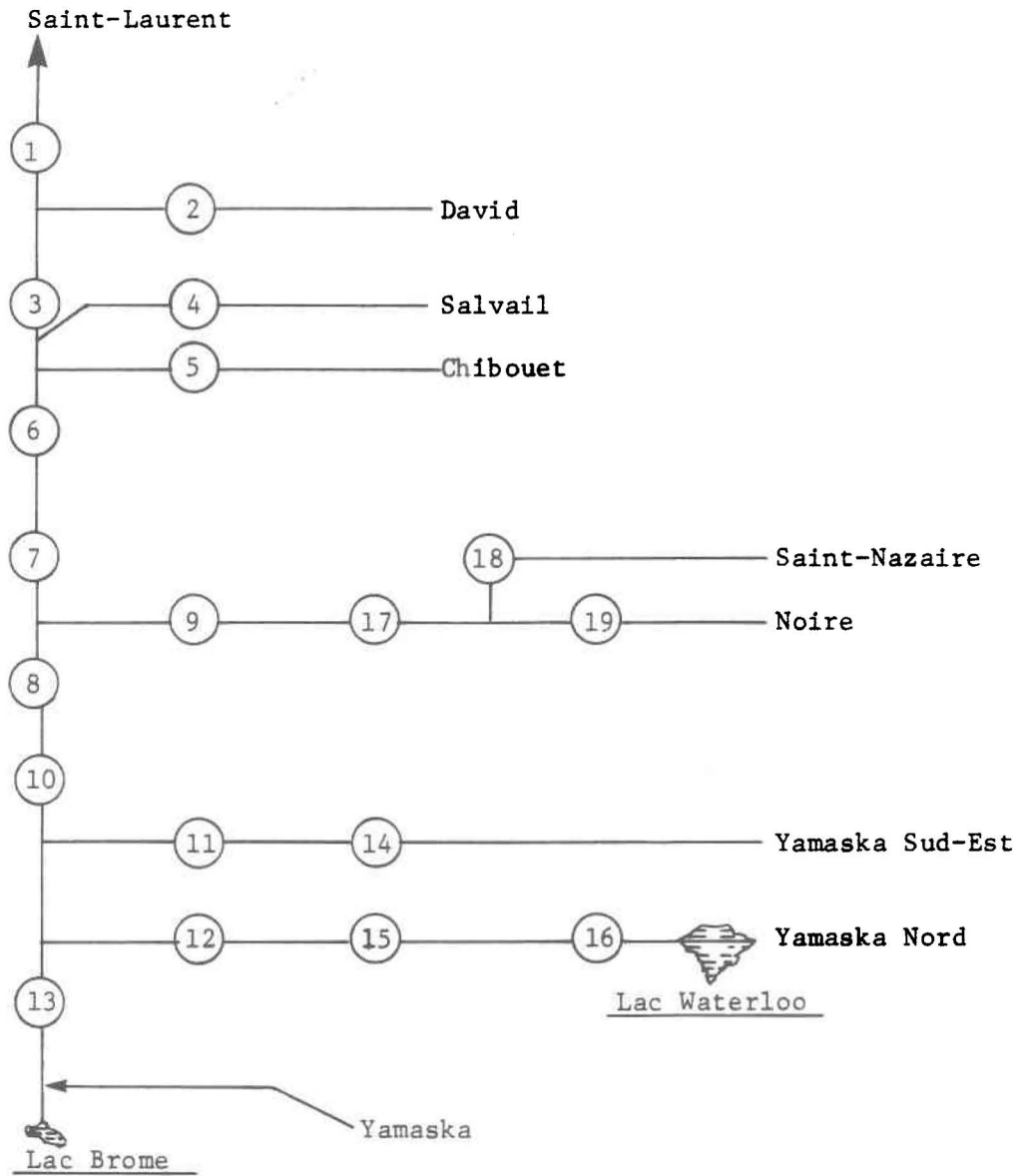


Figure 3-2. Schéma du réseau de stations

- la partie en amont (stations 10 à 19) était visitée en premier. Les relevés étaient effectués dans un délai allant de un à trois jours;
- la partie en aval du bassin (stations 1 à 9) était visitée une semaine plus tard. Les relevés étaient effectués dans un délai allant de un à trois jours.

Il existe donc un intervalle de temps important entre les relevés de la partie en amont du bassin et ceux de la partie en aval. Pour réaliser des études chronologiques des observations (auto-corrélation), on a reconstitué les observations manquantes de janvier et septembre 1975 par interpolation linéaire; c'est-à-dire que si i est l'indice de la mesure manquante, $x_i = (x_{i-1} + x_{i+1}) / 2$.

3.2.3 Paramètres de qualité mesurés

Initialement le réseau mesurait 29 paramètres de qualité sur des échantillons filtrés et non filtrés. Nous avons retenu une liste de 15 paramètres (tableau 3.2) comportant:

- des ions majeurs (échantillons filtrés), plus la conductivité et la silice (échantillons non filtrés);
- diverses formes d'azote et de phosphore (échantillons filtrés);
- fer et manganèse (échantillons filtrés).

Le tableau 3.2 donne les unités de mesure des paramètres de qualité, les numéros qui leur correspondent dans le traitement informatique, ainsi que leur limite de détection et les symboles qu'on utilisera parfois dans la suite de ce travail pour les désigner. Toute valeur inférieure à la limite de détection a été supposée, pour les fins de l'analyse, égale à la moitié de cette limite de détection.

TABLEAU 3.2: Paramètres de qualité étudiés
(unité de mesure, symbole, limite de détection).

Paramètre	Symbole	Unité de mesure	Limite de détection	Numéro de code
Silice	SiO ₂	mg SiO ₂ /l	0.1	1
Conductivité	Cond.	µmhos/cm	0.1	2
Azote total	N _T	mg N/l	0.01	3
Azote ammoniacal	NH ₃	mg N/l	0.02	4
Nitrates + Nitrites	NO ₃	mg N/l	0.1	5
Phosphore total	P _T	mg P/l	0.01	6
Phosphore hydrolysable	P _h	mg P/l	0.01	7
Calcium	Ca	mg Ca/l	0.1	8
Fer	Fe	mg Fe/l	0.01	9
Magnésium	Mg	mg Mg/l	0.05	10
Manganèse	Mn	mg Mn/l	0.05	11
Sodium	Na	mg Na/l	0.05	12
Potassium	K	mg K/l	0.005	13
Chlorures	Cl	mg Cl/l	0.1	14
Sulfates	SO ₄	mg SO ₄ /l	0.5	15

Le choix des 15 paramètres sur les 29 effectivement mesurés est fondé sur l'intérêt de ces paramètres (exemple: fer et manganèse qui causent des problèmes sur le bassin pour la consommation d'eau) et sur des critères pratiques. On a, en effet, retenu les paramètres pour lesquels les séries d'observations étaient les plus complètes.

Le tableau 3.3 donne les critères relatifs aux divers paramètres qui ont été retenus pour cette étude en fonction des divers usages de l'eau.

En conclusion, on dispose d'un réseau de:

- 19 stations;
- 15 paramètres de qualité mesurés à chaque station;
- 20 mesures mensuelles relatives à chaque paramètre (de février 1974 à novembre 1975).

soit un total de $N = 5700$ observations.

3.3 Objectifs

3.3.1 Objectifs du réseau actuel

D'après l'étude de l'INRS-Eau (1976) les buts qui ont présidé à la construction du réseau sont les suivants:

- d'une part obtenir une connaissance générale de la qualité de l'eau en ce qui concerne les substances nutritives et certains métaux traces;
- d'autre part mettre au point des méthodes d'analyse relatives aux substances nutritives, à des paramètres biologiques (potentiel de fertilité, potentiel autotrophe, potentiel hétérotrophe, ...), à des paramètres chimiques (capacité de complexation cationique, etc...).

TABLEAU 3.3: Critères de qualité utilisés.

Paramètre	Critère de qualité	Usage	Origine du critère
Fer (mg Fe/l)	0.3	eau potable	Norme canadienne
Manganèse (mg Mn/l)	0.05	eau potable	Norme canadienne
Azote inorganique soluble* (mg N/l)	0.1 acceptable 0.3 critique	vie aquatique	Critère INRS-Eau
Azote ammoniacal * (mg N/l)	0.1 critique	vie aquatique	Critère Nisbet et Vernaux, critère INRS-Eau
Phosphore hydrolysable * (mg P/l)	0.009 acceptable 0.027 critique	vie aquatique	Critère INRS-Eau
Chlorures (mg Cl/l)	250	eau potable	Norme canadienne
Sulfates (mg SO ₄ /l)	320	eau potable	Critère M.R.N. - qualité des eaux

* On se reportera à l'étude intégrée de la qualité des eaux des bassins versants des rivières Saint-François et Yamaska (INRS-EAU, 1976) pour l'explication du choix de ces critères.

C'est, pour une grande part, un objectif de recherche qui a déterminé l'établissement du réseau et aucun objectif précis, au sens où nous l'avons défini dans le premier chapitre de ce travail, n'a été proposé comme justification et base de la construction du réseau. C'est ainsi qu'un certain nombre de paramètres (silice, conductivité, ions majeurs) a été mesuré sans être utilisé par la suite. Par contre, bien que l'on dispose d'observations nombreuses, on ne peut répondre que très partiellement à des objectifs concrets comme la qualité de l'eau pour l'alimentation ou pour la récréation. Les paramètres biologiques fondamentaux pour ces objectifs par exemple coliformes totaux ou fécaux n'ont, en effet, pas été mesurés.

3.3.2 Objectifs proposés

Nous nous proposons de réaliser notre étude dans le cadre de deux utilisations précises de l'eau:

- eau potable;
- support de la vie aquatique.

Nous ne disposons pas de tous les paramètres pertinents pour répondre convenablement à des objectifs reliés à ces usages de l'eau, mais nous pouvons donner des réponses partielles en considérant les paramètres effectivement mesurés. Pour l'alimentation en eau potable nous disposons des paramètres fer et manganèse dont on sait que la présence est excessive sur le bassin. Ces deux éléments confèrent à l'eau un goût désagréable et peuvent tacher le linge, la porcelaine ou se déposer dans les tuyaux et canalisations et, finalement, les boucher.

Pour l'étude de la vie aquatique nous disposons de l'azote et du phosphore sous diverses formes. Des critères ont été suggérés par l'INRS-Eau en rapport avec les problèmes d'eutrophisation qu'occasionne l'excès de ces substances.

Signalons enfin que l'étude générale du réseau de mesures après ses deux années de fonctionnement (chapitre 4) pourrait entrer dans le cadre de l'objectif:

"Connaissance de base de la qualité générale de l'eau sur l'ensemble du bassin".

Un objectif aussi imprécis serait à rejeter pour l'établissement d'un réseau rationnel (cf. chapitre 1). Cependant, dans notre cas, où le réseau est déjà constitué, il est nécessaire de faire le point sur l'information reçue et sur les "performances" du réseau. Ce n'est qu'en connaissant l'état actuel complet de l'information que l'on sera en mesure de préciser l'état futur du réseau en vue de tel ou tel objectif. Cette partie peut être considérée comme une analyse sommaire des données existantes.

CHAPITRE 4

ETUDE GENERALE DE LA QUALITE DE L'EAU ET DU RESEAU DE MESURE DU BASSIN VERSANT DE LA RIVIERE YAMASKA.

Le but de ce chapitre est de faire le point sur l'information obtenue par le réseau après deux années de fonctionnement.

Cette étude pourrait correspondre à un objectif de connaissance de la base de la qualité générale de l'eau sur le bassin versant de la rivière Yamaska. Cette connaissance d'ensemble des paramètres de qualité mesurés, ainsi que celle des performances du réseau, sont nécessaires si on veut être en mesure de faire des recommandations sur l'état futur du réseau en vue de tel ou tel objectif plus précis. Nous envisageons par ailleurs, dans certaines parties du chapitre, les paramètres de qualité relatifs aux substances nutritives dans le cadre de l'usage vie aquatique et les paramètres fer et manganèse pour l'usage "alimentation en eau".

L'étude du réseau et des paramètres de qualité mesurés peut s'effectuer au moyen de méthodes statistiques qui synthétisent l'information pour la rendre plus aisément exploitable et pour en faciliter l'interprétation. Dans un premier temps on se borne au calcul des caractéristiques statistiques descriptives générales (moyenne, variance, précision sur la moyenne, extrême, etc...) qui permettent cependant une première caractérisation des paramètres de qualité et donnent une indication sur le comportement du réseau. Cette connaissance est également nécessaire pour les traitements plus approfondis; mais il est clair que l'information fournie par ces caractéristiques statistiques reste très sommaire.

Dans un deuxième temps on essaie donc de réaliser une synthèse et une représentation plus complète de l'information. On ne se contente plus d'une caractéristique statistique, mais on envisage une synthèse de l'information disponible sur un paramètre de qualité au moyen d'un opérateur mathématique (fonction du temps, processus aléatoire, loi statistique selon le cas).

Enfin l'étude du réseau, pour être complète, doit ajouter à la connaissance de l'information locale, à chacune des stations, celle d'une information au niveau de l'ensemble du bassin versant qui rend compte de la variabilité spatiale des paramètres.

4.1 Généralités

Le réseau fournit un ensemble de 5700 observations (cf. 3.2.3). L'analyse statistique traite uniformément toutes ces données, mais l'interprétation des nombreux résultats de cette analyse devra être effectuée en tenant compte de divers éléments tels que la localisation relative des stations du réseau, le type de paramètre, etc... Les stations du réseau peuvent être regroupées par cours d'eau ou par sous-bassin. Rappelons la disposition du réseau.

<u>Rivière</u>	<u>Numéro des stations</u>
Yamaska	1, 3, 6, 7, 8, 10, 13
Noire	9, 17, 19
St-Nazaire	18
Yamaska Sud-Est	11, 14
Yamaska Nord	12, 15, 16
David	2
Salvail	4
Chibouet	5

La localisation exacte des stations est donnée au chapitre 3 (tableau 3.1).

La rivière St-Nazaire (station 18) est un affluent de la rivière Noire et draine un sous-bassin agricole. Les rivières David, Salvail, Chibouet drainent des sous-bassins agricoles des Basses-Terres et se jettent dans le cours aval de la rivière Yamaska.

On peut aussi former à priori des groupes de paramètres de qualité susceptibles d'avoir des comportements voisins. On distingue ainsi trois groupes.

Groupes	Paramètres
Groupe I (essentiellement ions majeurs)	Conductivité, Silice, Calcium, Magnésium, Sodium, Potassium, Chlorures, Sulfates
Groupe II (substances nutritives)	Azote total, Azote ammoniacal, Nitrates, Phosphore total, Phosphore hydrolysable
Groupe III	Fer, Manganèse

Cette classification à priori n'exclut pas la possibilité de variations à l'intérieur d'un même groupe ou de comportements semblables d'un groupe à l'autre.

4.2 Etude descriptive générale

Cette première étude caractérise les observations recueillies sur les paramètres de qualité et les performances du réseau au moyen des caractéristiques statistiques suivantes:

- la moyenne statistique qui permet une première quantification du problème et permet un classement relatif des diverses stations et cours d'eau les uns par rapport aux autres (cf. 4.2.1). Une représentation graphique permet une visualisation de ces résultats (cf. 4.2.4);

- l'écart-type qui traduit la gamme de variabilité temporelle des paramètres de qualité (cf. 4.2.1);
- le coefficient de variation qui met en évidence les paramètres de qualité à forte variabilité temporelle (cf. 4.2.1);
- la précision atteinte sur la moyenne statistique qui rend compte de la performance du réseau dans la connaissance de cette caractéristique (cf. 4.2.2);
- le maximum observé qui permet aussi un classement relatif des stations du bassin et qui, comparé aux critères de qualité envisagés, révèle l'existence de problèmes de qualité actuels ou potentiels (cf. 4.2.3).

4.2.1 Moyennes, Ecart-types, Extrêmes, Coefficients de variation

Pour chaque paramètre de qualité et à chaque station, on a calculé la moyenne et l'écart-type (cf. chapitre 2, équations 2.1 et 2.2) et on a extrait la valeur maximale observée. Ces résultats sont donnés respectivement dans les tableaux 4.1, 4.2 et 4.3. Ces tableaux, et les suivants, sont constitués de façon qu'une colonne corresponde à un paramètre de qualité (la symbolisation est donnée dans le tableau 3.2) et qu'une ligne corresponde à une station (numérotage de 1 à 19, cf. tableau 3.1). Pour étudier ces tableaux le plus globalement possible, nous avons adopté la méthode suivante.

On classe les stations par ordre décroissant des valeurs de la caractéristique étudiée (moyenne, écart-type ou maximum). On affecte alors chaque station d'un poids relatif à son classement. Le classement est réalisé pour chaque paramètre de qualité et pour chacune des trois caractéristiques statistiques étudiées.

TABLEAU 4.1: Moyennes des paramètres de qualité par station.

	SiO ₂	Cond.	N _T	NH ₃	NO ₃	P _T	P _h	Ca	Fe	Mg	Mn	Na	K	Cl	SO ₄
1	4.19	270.70	1.34	.36	.61	.11	.09	27.39	.08	6.07	.04	20.27	3.38	25.28	19.48
2	5.28	426.60	1.05	.33	.33	.10	.07	46.91	.10	12.84	.06	28.76	5.63	26.83	28.18
3	4.00	261.45	1.32	.34	.63	.12	.10	25.73	.08	5.72	.03	20.12	3.22	25.92	19.07
4	7.58	694.30	1.64	.50	.87	.10	.08	43.08	.10	21.19	.06	86.25	7.40	90.95	29.10
5	5.47	465.20	2.15	.84	.86	.24	.19	58.27	.13	11.17	.08	28.80	5.86	32.77	38.46
6	3.87	223.12	1.27	.41	.53	.12	.09	22.96	.10	4.31	.06	18.20	3.64	22.52	16.05
7	4.00	215.70	1.12	.36	.48	.09	.07	22.54	.11	4.18	.10	15.61	2.67	19.14	16.14
8	3.60	216.10	1.10	.33	.48	.12	.09	20.91	.10	3.82	.08	18.63	2.48	23.03	14.78
9	4.38	169.00	1.01	.30	.43	.08	.06	21.89	.13	3.28	.05	7.61	2.42	9.29	14.18
10	3.48	187.95	.91	.21	.38	.11	.09	17.11	.09	2.67	.06	17.34	1.89	19.60	13.88
11	3.48	165.30	.80	.14	.38	.07	.05	17.91	.11	2.54	.05	11.41	1.51	14.74	12.28
12	4.82	311.85	3.00	1.33	.35	.52	.39	22.43	.25	3.56	.21	40.11	4.46	42.80	24.23
13	3.46	139.10	.44	.06	.18	.01	.01	14.14	.06	2.53	.04	10.92	1.26	16.65	10.46
14	4.61	167.60	1.00	.36	.34	.13	.11	16.98	.14	2.50	.16	13.30	1.52	17.80	12.35
15	4.34	276.10	2.11	.82	.27	.35	.28	21.13	.22	3.36	.17	31.78	4.17	35.54	24.01
16	5.20	290.60	2.27	.94	.49	.59	.51	24.37	.15	3.66	.22	26.99	3.89	40.40	20.30
17	4.72	157.45	.82	.23	.30	.07	.05	21.17	.17	3.11	.05	6.65	2.20	7.46	15.61
18	5.90	221.75	1.17	.34	.49	.15	.13	29.07	.21	3.74	.08	10.74	2.71	12.02	22.26
19	4.35	131.35	.58	.09	.22	.05	.03	17.12	.20	2.73	.06	5.70	1.48	6.70	10.73

Pour les unités, se référer au tableau 3.2.

TABLEAU 4.2: Ecart-types des paramètres de qualité par station.

	SiO ₂	Cond.	N _T	NH ₃	NO ₃	P _T	P _h	Ca	Fe	Mg	Mn	Na	K	Cl	SO ₄
1	1.60	61.88	.57	.39	.32	.05	.05	5.47	.03	1.84	.02	8.69	.91	11.54	9.81
2	2.85	105.87	.67	.42	.26	.04	.04	13.07	.06	3.86	.04	11.13	3.49	12.89	13.35
3	1.50	58.93	.58	.42	.30	.06	.06	5.57	.03	1.73	.02	9.29	.82	12.27	9.10
4	3.71	357.78	.67	.49	.60	.07	.08	12.33	.07	8.76	.04	67.41	5.37	77.61	12.43
5	2.21	139.18	.73	.72	.41	.39	.34	19.00	.08	4.73	.08	17.31	2.63	13.95	23.58
6	1.59	60.32	.51	.35	.24	.08	.07	4.70	.03	1.10	.03	11.01	3.36	11.58	7.62
7	1.45	45.38	.44	.27	.24	.05	.05	4.57	.06	1.08	.13	7.45	.63	8.10	9.46
8	1.72	38.24	.39	.27	.24	.06	.06	4.42	.05	.74	.07	9.27	.60	8.67	6.16
9	1.60	45.10	.38	.25	.22	.03	.04	6.78	.06	.96	.03	3.00	.76	3.22	7.90
10	1.63	64.62	.34	.25	.13	.08	.09	3.25	.03	.44	.03	11.55	.53	10.23	6.52
11	1.90	41.01	.29	.17	.21	.05	.04	3.57	.04	.50	.02	7.21	.41	8.38	4.78
12	1.25	105.19	2.38	1.68	.30	.38	.36	6.14	.13	1.07	.13	22.86	2.44	17.93	13.02
13	1.67	31.99	.19	.08	.15	.01	.00	2.13	.03	.59	.03	4.03	.29	5.76	5.06
14	1.12	55.65	.52	.52	.19	.14	.17	4.51	.07	.59	.19	7.46	.51	10.36	4.37
15	1.31	110.60	2.02	1.18	.24	.29	.26	5.47	.15	.86	.11	22.54	2.62	18.74	14.22
16	1.89	86.21	1.09	.69	.39	.52	.47	7.27	.07	1.04	.17	11.38	1.80	18.39	11.23
17	2.51	40.16	.28	.26	.17	.05	.04	5.77	.08	.83	.03	3.25	1.00	2.47	9.92
18	2.37	62.51	.40	.40	.20	.17	.22	7.73	.11	1.17	.04	6.05	.98	7.45	13.65
19	2.23	39.07	.24	.15	.16	.03	.02	5.23	.12	.82	.07	3.44	.46	3.01	6.93

Pour les unités, se référer au tableau 3.2.

TABLEAU 4.3: Maxima observés des paramètres de qualité à chaque station.

	SiO ₂	Cond.	N _T	NH ₃	NO ₃	P _T	P _h	Ca	Fe	Mg	Mn	Na	K	Cl	SO ₄
1	7.40	410.00	2.35	1.30	1.33	.20	.19	39.40	.14	11.00	.10	38.00	4.90	53.00	51.00
2	10.70	580.00	2.65	1.60	.90	.23	.17	68.00	.23	20.00	.17	47.00	15.00	66.00	67.00
3	7.30	368.00	2.44	1.40	1.15	.28	.26	37.00	.13	10.00	.09	45.00	4.60	50.00	49.00
4	18.00	1680.00	2.92	1.38	2.48	.37	.39	65.00	.25	38.00	.18	310.00	25.00	360.00	65.00
5	11.00	703.00	3.37	3.00	2.26	1.63	1.34	98.00	.33	17.00	.30	72.00	11.00	55.00	114.00
6	7.30	350.00	2.58	1.50	.95	.31	.28	30.50	.15	6.60	.14	40.00	17.50	48.00	40.00
7	7.10	295.00	2.01	1.10	.86	.21	.21	31.00	.30	6.80	.58	34.40	3.70	40.00	41.00
8	6.70	276.00	1.70	.90	.90	.26	.26	27.00	.20	5.00	.31	49.00	3.70	47.00	33.00
9	8.60	256.00	1.74	.87	.90	.14	.15	35.00	.24	4.90	.11	12.00	3.90	14.60	41.00
10	6.10	371.00	1.74	1.00	.68	.32	.36	21.00	.16	3.40	.16	47.50	2.90	55.00	33.00
11	6.30	276.00	1.52	.53	.70	.17	.17	27.00	.19	3.50	.10	28.50	2.40	44.00	27.00
12	7.10	560.00	9.81	5.20	1.17	1.50	1.41	40.00	.50	6.50	.52	90.00	12.00	75.00	53.00
13	6.40	229.00	.88	.32	.52	.06	.02	18.50	.12	4.30	.10	24.00	1.90	36.00	27.00
14	6.50	322.00	2.42	2.00	.68	.54	.75	28.00	.30	3.60	.68	30.00	2.50	49.50	25.00
15	6.80	494.00	9.02	4.00	.70	1.11	1.01	30.50	.55	4.70	.50	84.00	10.00	78.00	49.00
16	9.50	460.00	4.18	2.40	2.03	1.70	1.57	39.10	.32	5.80	.67	52.50	10.00	84.00	52.00
17	9.00	228.00	1.62	.93	.61	.20	.17	31.90	.36	4.40	.12	12.50	4.90	13.90	42.00
18	10.10	345.00	2.20	1.80	.86	.64	.98	40.00	.49	6.00	.17	27.40	5.00	35.00	57.00
19	8.70	221.00	1.34	.66	.63	.09	.08	27.60	.45	4.40	.33	14.40	2.30	17.00	32.00

Pour les unités, se référer au tableau 3.2.

Au regard des tableaux de pondération, on constate que:

- certains groupes de stations ont des comportements semblables vis-à-vis de certains groupes de paramètres de qualité;
- les groupes de paramètres de qualité qui se dégagent sont:
 - les ions majeurs (groupe I) et les nitrates;
 - les substances nutritives (groupe II), sauf les nitrates, avec parfois des distinctions à faire entre les formes d'azote et celles de phosphore;
 - le fer et le manganèse (groupe III).

La classification à priori faite en 4.1 est ainsi justifiée:

- les groupes de stations mis en évidence sont souvent formés de stations qui appartiennent au même cours d'eau;
- les groupements obtenus sont les mêmes quelle que soit la caractéristique statistique envisagée: moyenne, écart-type ou maximum.

Ces résultats nous permettent de construire le tableau 4.4 de classement relatif des stations et cours d'eau les uns par rapport aux autres en fonction des trois groupes de paramètres de qualité dégagés. Plusieurs conclusions sont à tirer de ce tableau.

- la rivière Yamaska-Nord est caractérisée par les fortes valeurs moyennes et maximales et la plus grande variabilité des paramètres relatifs aux substances nutritives, ainsi que du fer et du manganèse. Les concentrations en ions majeurs y sont élevées. Ce classement est relatif puisqu'il est effectué par rapport aux autres cours

TABLEAU 4.4: Classement relatif des cours d'eau d'après l'étude des moyennes, écart-types et maxima.

Valeurs relatives des caractéristiques statistiques, moyenne, écart-type, maximum	Ions majeurs et Nitrates paramètres: 1, 2, 5, 8, 10, 12, 13, 14, 15	Substances nutritives (sauf nitrates) paramètres: 3, 4, 6, 7	Fer et Manganèse paramètres: 9, 11
Fortes	David (2), Salvail (4), Chibouet (5).	<u>Yamaska-Nord (12,15,16)</u>	<u>Yamaska-Nord (12,15,16)</u> St-Nazaire (18), (Fe)
Intermédiaires à Fortes	<u>Yamaska-Nord (12,15,16)</u>	Salvail (4), (N)* Chibouet (5), (P) Yamaska Sud-Est (14) (P)	Noire (9,17,19) (Fe) Yamaska Sud-Est (11,14) (Mn)
Intermédiaires	Yamaska (1,3,6,7,8,10)	Yamaska (1,3,6,7,8,10)	David (2), Salvail (4) Chibouet (5)
Faibles à Intermédiaires	Yamaska Sud-Est (11,14) Noire (9,17,19)	Yamaska Sud-Est (11) Noire (9,17,19)	Yamaska (1,3,6,7,8,10)
Faibles	Yamaska (13)	Yamaska (13)	Yamaska (13)

Légende: - les numéros entre parenthèses dans le tableau correspondent aux stations concernées
 - Fe: fer Mn: manganèse P: formes de phosphore N: formes d'azote
 * exemple: Salvail (4), (N) signifie que la station 4 de la rivière Salvail a des caractéristiques intermédiaires à fortes en ce qui concerne les formes d'azote (sauf nitrates).

d'eau du bassin. Si on étudie les valeurs observées par rapport aux critères de qualité utilisés (cf. 4.2.4), on constate que les concentrations sont fortes en valeur absolue, ce qui démontre que la rivière Yamaska-Nord possède un niveau d'enrichissement élevé en substances nutritives et de fortes concentrations en fer et en manganèse.

- Les rivières David, Salvail, Chibouet qui drainent des bassins agricoles se distinguent par leurs fortes concentrations en ions majeurs et des concentrations élevées en azote pour la rivière Salvail et en phosphore pour la rivière Chibouet. La station 4 de la rivière Salvail se détache nettement des autres stations par des valeurs élevées des paramètres suivants: silice, conductivité, nitrates, magnésium, sodium, potassium, chlorures ainsi que, à un degré moindre, calcium et sulfates. A la station 5, située sur la rivière Chibouet, on observe plus spécialement des valeurs élevées de calcium, sulfates et nitrates. Remarquons cependant que les plus fortes concentrations d'ions majeurs observées à ces stations restent loin d'atteindre les valeurs maximales permises par les critères relatifs à l'eau potable.
- Le tronçon principal de la rivière Yamaska occupe une position intermédiaire sur le tableau 4.4, à l'exception de la station 13. Cela s'explique par le fait que ce cours d'eau principal intègre les différents effets des autres cours d'eau du bassin.
- La rivière Noire semble moins chargée en substances nutritives ainsi qu'en ions majeurs que les autres cours d'eau du bassin.

- Enfin, la station 13 (en aval d'Adamsville), en amont de la rivière Yamaska, est caractérisée par des faibles valeurs moyennes et maximales et par une faible variabilité pour l'ensemble des paramètres de qualité étudiés. Cette station est, par ailleurs, peu influencée par des activités humaines ou agricoles. Cette situation particulière pourrait donc lui faire jouer un rôle d'indicateur ou de témoin pour le bassin.

L'étude des concentrations moyennes classées permet d'établir des rapprochements entre des paramètres de qualité qui se comportent de façon assez voisine d'une station à l'autre:

- . phosphore total et phosphore hydrolysable;
- . potassium et sulfates;
- . sodium et chlorures;
- . azote total et ammoniacque pour les fortes et faibles valeurs essentiellement.

Le tableau 4.5 indique les coefficients de variation de chaque série d'observations (cf. chapitre 2, équation 2.3). On peut tenter de classer les stations par rapport à ces coefficients de variation comme on l'a fait pour les moyennes et écarts-types. Mais on obtient des résultats légèrement différents dont l'interprétation n'est pas aisée. Le coefficient de variation permet, par contre, de mettre en évidence certains paramètres de qualité qui se distinguent des autres par des valeurs élevées de cette caractéristique, ce qui traduit une forte variabilité du paramètre de qualité eu égard à sa valeur moyenne. C'est ainsi le cas de l'azote ammoniacal, du phosphore total, du phosphore hydrolysable et du manganèse. A un fort coefficient de variation correspond une

TABLEAU 4.5: Coefficients de variation des paramètres de qualité à chaque station.

M: moyenne - S: écart-type

	SiO ₂	Cond.	N _T	NH ₃	NO ₃	P _T	P _h	Ca	Fe	Mg	Mn	Na	K	Cl	SO ₄
1	.38	.23	.42	1.09	.52	.47	.57	.20	.36	.30	.58	.43	.27	.46	.50
2	.54	.25	.63	1.27	.78	.45	.54	.28	.62	.30	.75	.39	.62	.48	.47
3	.37	.23	.44	1.23	.47	.48	.57	.22	.43	.30	.55	.46	.25	.47	.48
4	.49	.52	.41	.97	.68	.75	1.09	.29	.73	.41	.73	.78	.73	.85	.43
5	.40	.30	.34	.85	.47	1.66	1.81	.33	.64	.42	1.01	.60	.45	.43	.61
6	.41	.27	.40	.86	.46	.66	.81	.20	.32	.26	.59	.60	.92	.51	.47
7	.36	.21	.39	.77	.49	.53	.71	.20	.53	.26	1.32	.48	.24	.42	.59
8	.48	.18	.36	.81	.50	.46	.65	.21	.49	.19	.83	.50	.24	.38	.42
9	.37	.27	.37	.83	.51	.38	.60	.31	.43	.29	.58	.39	.31	.35	.56
10	.47	.34	.37	1.17	.36	.71	.99	.19	.33	.16	.59	.67	.28	.52	.47
11	.55	.25	.36	1.23	.55	.63	.82	.20	.37	.20	.43	.63	.27	.57	.39
12	.26	.34	.79	1.26	.86	.72	.93	.27	.51	.30	.63	.57	.55	.42	.54
13	.48	.23	.43	1.31	.84	.96	.69	.15	.43	.23	.63	.37	.23	.35	.48
14	.24	.33	.53	1.43	.56	1.07	1.53	.27	.50	.24	1.14	.56	.34	.58	.35
15	.30	.40	.96	1.43	.88	.83	.93	.26	.70	.26	.69	.71	.63	.53	.59
16	.36	.30	.48	.74	.80	.89	.91	.30	.49	.28	.79	.42	.46	.46	.55
17	.53	.26	.35	1.13	.55	.68	.88	.27	.49	.27	.60	.49	.45	.33	.64
18	.40	.28	.34	1.17	.40	1.13	1.73	.27	.53	.31	.47	.56	.36	.62	.61
19	.51	.30	.41	1.59	.74	.56	.73	.31	.61	.30	1.13	.60	.31	.45	.65
M	.42	.29	.46	1.11	.60	.74	.92	.25	.50	.28	.74	.54	.42	.48	.52
S	.09	.08	.16	.25	.17	.31	.38	.05	.12	.06	.25	.12	.19	.12	.09

Pour les unités, se référer au tableau 3.2.

mauvaise précision relative sur la moyenne (cf. chapitre 2, équation 2.9) du fait de la forte variabilité relative. Nous préférons établir un classement des paramètres de qualité selon la précision atteinte sur la moyenne (cf. 4.2.2). Ce classement, qui traduit en même temps la performance du réseau vis-à-vis du paramètre de qualité considéré, semble plus significatif qu'un classement basé sur le coefficient de variation.

En conclusion, il s'avère qu'une étude sommaire des caractéristiques statistiques telles que moyenne, écart-type, coefficient de variation et maximum des séries de valeurs observées permet :

- d'obtenir des indications quantitatives sur les paramètres de qualité et leur variabilité;
- d'établir des groupes de paramètres et de stations;
- d'effectuer un classement global des différentes stations du bassin selon les paramètres de qualité envisagés.

Cependant ce type d'étude est fastidieux (les classements et regroupements sont longs à réaliser), subjectif (choix des regroupements) et incomplet. Il est en effet possible que des séries aient les mêmes moyenne et écart-type mais soient très différentes au niveau des valeurs individuelles. Par ailleurs les comportements spécifiques sont laissés de côté pour faire place aux tendances plus générales.

4.2.2 Précision sur la moyenne

Le tableau 4.6 indique les précisions relatives (exprimées en pourcentage) atteintes sur la moyenne de chaque paramètre de qualité à chaque station, (cf. chapitre 2, équations 2.6, 2.7, 2.9) au niveau de confiance 95%. Le tableau 4.6 nous renseigne sur les performances du réseau selon les paramètres de qua-

TABLEAU 4.6: Précisions sur la moyenne des paramètres de qualité à chaque station
(en pourcentage de la moyenne) au niveau de confiance 95%.
M: moyenne - S: écart-type

	SiO ₂	Cond.	N _T	NH ₃	NO ₃	P _T	P _h	Ca	Fe	Mg	Mn	Na	K	Cl	SO ₄
1	17.91	10.70	19.72	51.05	24.57	21.79	26.63	9.34	17.07	14.18	26.97	20.06	12.63	21.37	23.56
2	25.27	11.61	29.69	59.57	36.29	21.24	25.06	13.04	29.05	14.07	35.05	18.12	28.97	22.48	22.16
3	17.52	10.55	20.68	57.55	22.16	22.55	26.85	10.13	20.22	14.14	25.66	21.60	11.89	22.15	22.33
4	22.94	24.12	19.05	45.47	32.04	34.91	50.92	13.39	34.33	19.34	34.10	36.58	33.96	39.94	19.99
5	18.92	14.00	15.93	40.01	22.16	77.60	84.57	15.26	30.08	19.82	47.19	28.13	21.01	19.92	28.69
6	19.27	12.65	18.67	40.35	21.45	30.98	38.10	9.58	15.09	11.97	27.56	28.30	43.18	24.06	22.23
7	16.97	9.85	18.30	36.13	23.01	24.62	33.42	9.48	24.89	12.13	61.94	22.34	11.03	19.80	27.44
8	22.31	8.28	16.63	38.06	23.62	21.73	30.32	9.90	22.75	9.10	38.64	23.29	11.28	17.62	19.49
9	17.11	12.49	17.45	39.06	23.76	17.60	27.91	14.49	20.27	13.64	27.13	18.43	14.63	16.22	26.05
10	21.93	16.09	17.30	54.95	16.75	33.10	46.54	8.89	15.62	7.70	27.61	31.18	12.98	24.42	21.98
11	25.52	11.61	16.98	57.48	25.55	29.60	38.21	9.32	17.47	9.16	20.24	29.56	12.76	26.60	18.20
12	12.14	15.79	37.09	59.09	40.41	33.87	43.44	12.81	23.70	13.99	29.72	26.67	25.59	19.60	25.14
13	22.55	10.76	20.03	61.50	39.49	45.00	32.38	7.04	20.29	10.92	29.28	17.28	10.82	16.18	22.62
14	11.32	15.54	24.63	67.05	26.16	49.98	71.81	12.42	23.52	11.07	53.52	26.26	15.77	27.24	16.57
15	14.09	18.75	44.75	67.08	41.42	38.65	43.29	12.12	32.89	12.03	32.30	33.19	29.44	24.68	27.70
16	16.96	13.88	22.51	34.48	37.66	41.70	42.62	13.96	23.11	13.23	37.01	19.74	21.60	21.31	25.89
17	24.85	11.94	16.25	52.72	25.59	32.03	41.23	12.76	22.88	12.43	27.90	22.84	21.17	15.47	29.74
18	18.81	13.19	16.04	54.77	18.82	52.89	80.75	12.44	24.75	14.69	21.77	26.37	16.96	29.00	28.69
19	23.93	13.92	19.33	74.53	34.51	26.02	34.20	14.28	28.50	13.97	53.08	28.18	14.57	21.03	30.23
M	19.5	13.50	21.60	52.20	28.20	34.50	43.10	11.60	23.50	13.00	34.60	25.20	19.50	22.60	24.10
S	4.30	3.60	7.70	11.70	7.80	14.50	17.70	2.30	5.50	3.00	11.50	5.40	9.10	5.70	4.00

lité et les stations considérées. Les résultats obtenus montrent, en effet, que la précision atteinte varie pour un paramètre d'une station à l'autre et, à une station donnée, selon les paramètres. On peut cependant tenter un classement des paramètres de qualité selon les précisions atteintes en moyenne sur l'ensemble des stations.

Paramètres	Précision atteinte dans l'estimation de la moyenne
Conductivité, Calcium, Magnésium	12 à 15%
Silice, Azote total, Fer, Sodium, Potassium, Chlorures, Sulfates	20 à 25%
Nitrates	30%
Phosphore total, Manganèse	35%
Phosphore hydrolysable	43%
Azote ammoniacal	52%

On remarque (cf. 4.2.1) qu'on obtient la plus mauvaise précision sur les paramètres qui ont les plus forts coefficients de variation (phosphore total et hydrolysable, manganèse, azote ammoniacal). Si on désire atteindre une précision de 20% sur la moyenne de l'ensemble des paramètres, il faut augmenter la fréquence de mesure des substances nutritives et du manganèse alors qu'on peut réduire celle du calcium, du magnésium et de la conductivité. Signalons qu'une étude plus approfondie des précisions atteintes ou des fréquences de mesure nécessaires implique que l'on considère les effets de persistance des séries d'observation (cf. 4.3.1); en effet les calculs précédents supposent que

les séries étudiées sont constituées de valeurs indépendantes, ce qui peut être admissible seulement en première approximation. A partir des tableaux (4.1) et (4.6), on peut déterminer les intervalles de confiance à 95% de la moyenne pour chaque paramètre de qualité à chaque station d'après la formule (2.11).

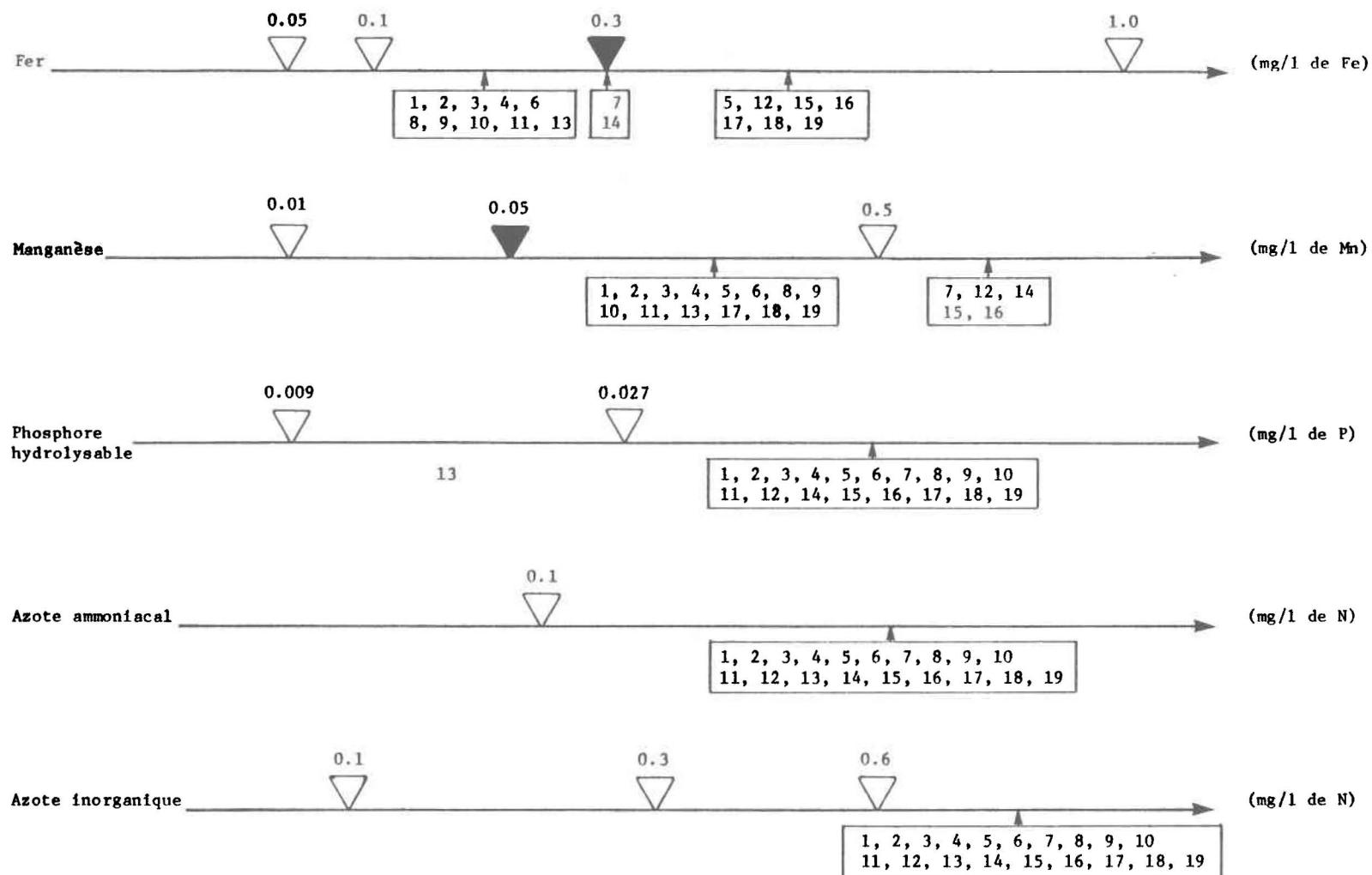
4.2.3 Dépassements de critères de qualité

Pour évaluer la qualité de l'eau sur le bassin, en fonction de certains usages, il faut comparer les valeurs observées des paramètres de qualité à des critères de qualité établis pour ces usages.

Nous comparons ici les valeurs maximales observées (tableau 4.3) à des critères de qualité relatifs aux objectifs:

- eau potable (normes du Canada pour le fer et le manganèse);
- vie aquatique (critères sur les substances nutritives adoptés par l'INRS-Eau).

Le tableau 4.7 indique le classement des stations du bassin en fonction de la valeur maximale observée du paramètre de qualité considérée par rapport aux critères de qualité utilisés. Les critères repérés par un triangle noir sont ceux qui sont retenus dans cette étude (cf. chapitre 3, tableau 3.3). On a repéré par un triangle blanc d'autres critères adoptés, de façon à mieux situer le problème. Le tableau suivant indique ces critères.



NOTE: Pour le code des stations, voir la figure 3.2.

Tableau 4-7. Maxima observés, situation des stations par rapport aux critères de qualité envisagés.

Paramètre	Critère de qualité (limites supérieures)			Usage
Fer (mg Fe/l)	Objectif	Canada	0.05	Eau potable
		O.M.S.	0.1	
	Acceptable	Canada	0.3 (norme retenue ici)	
		O.M.S.	1.0	
Manganèse (mg Mn/l)	Objectif	Canada	0.01	Eau potable
		O.M.S.	0.05	
	Acceptable	Canada	0.05 (norme retenue ici)	
		O.M.S.	0.5	
Azote inorganique soluble (mg N/l)	Acceptable (INRS-Eau)	0.1	Vie aquatique	Qualité générale de l'eau
	Critique (INRS-Eau)	0.3		
	Limite supérieure M.R.N.	0.6		

Le critère de 0.1 mg/l d'azote sous forme ammoniacale représente, d'après NISBET et VERNAUX (1970), un seuil indicatif de pollution grossière du milieu par les rejets humains et animaux. Les concentrations limites pour l'azote inorganique et le phosphore hydrolysable (cf. chapitre 3, tableau 3.3) ont été choisies par l'INRS-Eau, en fonction des dangers d'eutrophisation. Signalons que les normes pour l'eau potable relatives aux ions majeurs (chlorures, sulfates) ne sont jamais dépassées.

Le tableau 4.7 appelle les commentaires suivants:

- des dépassements des normes du Canada pour l'eau potable sont observés à toutes les stations du bassin dans le cas du manganèse et à certaines stations (Yamaska-Nord, tronçon en amont de la rivière

Noire, rivières St-Nazaire et Chibouet) dans le cas du fer.

- des dépassements des critères adoptés pour le phosphore hydrolysable, l'azote ammoniacal et l'azote inorganique sont observés à toutes les stations du bassin, exception faite de la station 13 (Adamsville), dans le cas du phosphore hydrolysable.

Pour compléter cette étude, on peut calculer les pourcentages de dépassement des critères de qualité qui ont été observés aux diverses stations. On donne au tableau suivant les résultats globaux obtenus par l'INRS-Eau (1976) en ce qui concerne les dépassements de critères pour les substances nutritives. Ce tableau donne l'intervalle des dépassements observés (en pourcentage) pour l'ensemble des stations du bassin (sauf à la station 13).

Paramètre	Critère	Dépassements observés (en pourcentage)
Azote inorganique (mg N/l)	0.1	82 à 100
	0.3	71 à 100
Phosphore hydrolysable (mg P/l)	0.09	87 à 100
	0.027	56 à 100
Azote ammoniacal (mg N/l)	0.10	53 à 100

Notons enfin qu'aux stations de la rivière Yamaska-Nord (12, 15, 16) les dépassements de critère relatifs aux substances nutritives et au manganèse sont pratiquement systématiques.

En conclusion, il apparaît que des problèmes se posent sur le bassin de la rivière Yamaska pour l'eau potable, dans le cas du fer et du manganèse, et pour la vie aquatique en raison d'un trop fort enrichissement en substances

nutritives sous formes azotées et phosphorées. Le sous-bassin de la rivière Yamaska-Nord semble plus particulièrement touché par ces deux problèmes.

L'étude précédente reste cependant sommaire: on a prouvé l'existence de dépassements de critère, mais on n'a pas associé une probabilité à ces dépassements. On a donné quelques fréquences empiriques de dépassements de critère calculées sur les échantillons de valeurs observées mais cela ne permet pas de situer la probabilité réelle de ces dépassements au sein de la population des valeurs possibles du paramètre de qualité envisagé. Nous indiquons au chapitre 5 (5.2.4) une méthode rigoureuse permettant d'estimer la probabilité de dépassement d'un critère donné.

4.2.4 Représentation graphique

Il est important de pouvoir visualiser rapidement les valeurs moyennes et la variabilité spatiale moyenne des paramètres de qualité sur un bassin versant. Pour effectuer une représentation synthétique on peut adopter la méthode suivante appliquée dans le cas des ions majeurs: sur une carte du bassin on associe à chaque station d'échantillonnage un cercle dont le diamètre est proportionnel à la racine carrée de la conductivité; la surface du cercle est alors proportionnelle à la conductivité. Le cercle est divisé en sept parties (arcs). L'angle qui intercepte chaque arc est proportionnel à la concentration d'un des sept ions majeurs qui contribuent à la valeur de la conductivité à la station de mesure. Les ions majeurs sont divisés en:

- . quatre cations: sodium, potassium, calcium, magnésium
- . trois anions: chlorures, sulfate, carbonates

Le cercle se compose d'un demi-cercle pour les cations et d'un demi-cercle pour les anions: la condition de neutralité ionique implique en ef-

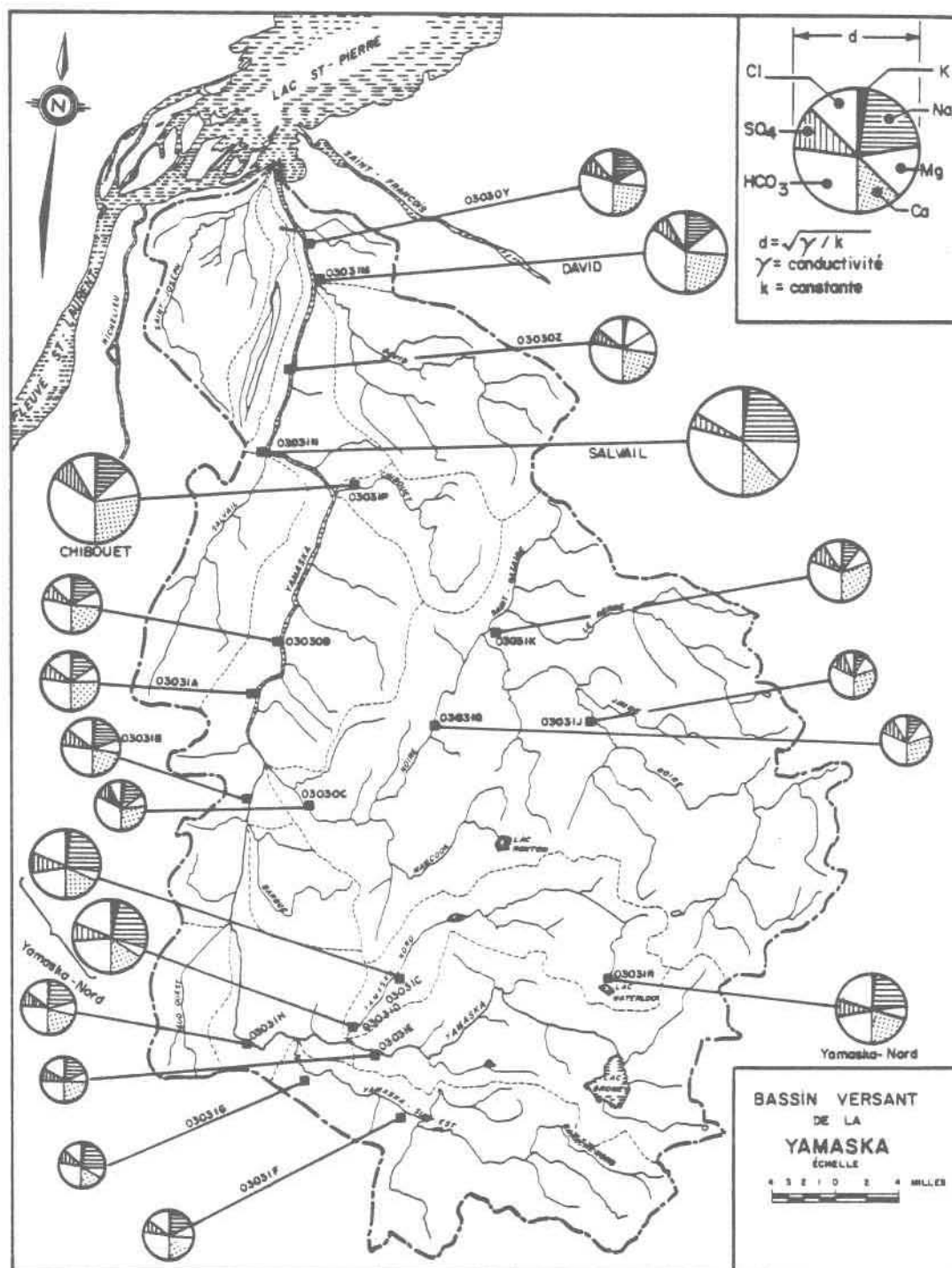


Figure 4.1. Distribution spatiale des concentrations moyennes d'ions majeurs pour la période d'étude : février 74 - novembre 75.

fet que la somme des charges positives soit égale à celle des charges négatives. Les concentrations sont exprimées en milliéquivalents par litre avec les facteurs de conversion suivants (HARNED et OWEN, 1958):

Ion	Force ionique meq/l par mg/l
Calcium	0.04990
Magnésium	0.08226
Sodium	0.04350
Potassium	0.02557
Bicarbonates	0.01639
Sulfate	0.02082
Chlorures	0.02821

A partir des résultats du tableau 4.1 on obtient la figure 4.1 où l'on observe de fortes valeurs de conductivité pour les stations des rivières Salvail, Chibouet et David et, à un degré moindre, pour les stations de la rivière Yamaska-Nord.

4.2.5 Résumé et conclusion

On a constaté l'existence de trois groupes de paramètres de qualité:

- groupe I : ions majeurs, silice, conductivité;
- groupe II : substances nutritives;
- groupe III: fer, manganèse

Un classement relatif des stations en fonction des valeurs obtenues pour la moyenne, le maximum et l'écart-type, pour les groupes de paramètres précédents permet de faire ressortir les points suivants: (tableau 4.4).

- La rivière Yamaska-Nord (stations 12, 15, 16) possède les plus fortes moyennes et maxima et la plus grande variabilité pour :

- . les substances nutritives;

- . le fer et le manganèse

Les ions majeurs sont aussi observés en quantités assez élevées relativement au reste du bassin.

- Les rivières David, Salvail, Chibouet (stations 2, 4, 5 respectivement) se distinguent par des valeurs moyennes et maximales de la conductivité et des divers ions majeurs les plus élevées du bassin. Ces concentrations restent cependant en deçà des normes prescrites pour l'eau de consommation.
- On constate une présence assez élevée de fer sur le sous-bassin de la rivière Noire (stations 9, 17, 19 de la rivière Noire et station 18 de la rivière St-Nazaire) et de manganèse aux stations (11, 14) de la rivière Yamaska Sud-Est.
- Les stations 1, 3, 6, 7, 8, 10 de la rivière Yamaska ainsi que 9, 17, 19 de la rivière Yamaska Sud-Est (sauf pour le manganèse) occupent, en ce qui concerne les concentrations moyennes, maximales et la variabilité des divers paramètres de qualité, une position intermédiaire dans l'ensemble des stations du bassin.
- La station 13 de la rivière Yamaska se distingue nettement des autres stations du bassin: on y observe les concentrations maximales et moyennes les plus faibles du bassin, ainsi que la plus faible variabilité, pour l'ensemble des paramètres étudiés.

- Les précisions obtenues sur la moyenne varient d'une station à l'autre pour un même paramètre et, surtout, suivant les paramètres de qualité à une même station. La précision que l'information recueillie permet d'atteindre pour la moyenne des paramètres conductivité, calcium et magnésium est de 12 à 15%; elle est de 20 à 25% pour la silice, l'azote total, le fer, le sodium, le potassium, les chlorures et les sulfates. La précision est mauvaise (supérieure à 30%) dans le cas du manganèse et des substances nutritives autres que l'azote total.
- Une étude des concentrations maximales observées par rapport aux critères de qualité retenus indique l'existence de dépassements de norme:
 - . pour l'eau de consommation domestique: des dépassements de la norme canadienne sont constatés pour le manganèse à toutes les stations et pour le fer aux stations 12, 15, 16 de la Yamaska-Nord, 17, 19 de la rivière Noire, 18 de la rivière St-Nazaire et 5 de la rivière Chibouet;
 - . pour la vie aquatique: des dépassements de critère sont à signaler pour les substances nutritives à toutes les stations du bassin (sauf station 13 pour le phosphore hydrolysable). On a constaté, de plus, que la fréquence de ces dépassements est élevée.

En conclusion, il semble qu'il existe des problèmes d'enrichissement excessif des eaux en substances nutritives sur l'ensemble du bassin (sauf station 13) et, plus spécialement, sur le sous-bassin de la rivière Yamaska-Nord.

Des problèmes dus au fer et surtout au manganèse se posent pour l'eau de consommation domestique. Dans ce cas encore, le sous-bassin de la rivière Yamas-ka-Nord semble le plus touché.

4.3 Représentation synthétique des paramètres de qualité

L'étude précédente (4.2) nous a permis de caractériser globalement les paramètres de qualité étudiés à l'aide de leur moyenne, leur écart-type et la valeur maxima observée. Une connaissance plus précise de la variabilité de ces paramètres, de leur évolution temporelle et une évaluation plus rigoureuse des dépassements de critère nécessitent une étude plus approfondie. Les caractéristiques étudiées en 4.2 (moyenne, écart-type, maximum) constituent une simplification importante de l'information qu'on peut tirer de l'échantillon des valeurs observées. Quand cela est possible, on obtient une information plus complète et synthétique en représentant, par une distribution statistique, la population des valeurs que peut prendre le paramètre de qualité considéré. On ne dispose pas de données suffisantes pour réaliser une étude temporelle; la durée de l'échantillonnage est trop courte pour déceler des tendances à long terme ou des cycles annuels. Par ailleurs, la fréquence des données ne permet pas de détecter d'éventuels cycles journaliers ou hebdomadaires. Il est, par contre, important pour effectuer une représentation par une loi statistique de la série d'observations d'un paramètre à une station d'étudier la dépendance des observations successives.

- Si les observations sont dépendantes, l'étude du paramètre de qualité peut être effectuée en considérant un processus markovien (cf. 4.3.2).
- Si les observations sont indépendantes, on peut synthétiser l'in-

formation de la série observée en ajustant une distribution statistique. Dans cette représentation on perd l'effet chronologique (qui n'a pas d'importance en raison du caractère aléatoire de la série) et l'on peut déterminer les quantiles de la distribution, ce qui permet l'étude rigoureuse des dépassements de critère de qualité (cf. 4.3.3).

Après avoir classé les paramètres de qualité selon leur niveau de dépendance, nous examinons leur comportement selon la méthode d'étude appropriée.

4.3.1 Indépendance des observations

Le tableau 4.8 donne les coefficients d'auto-corrélation des séries d'observation (cf. équation 2.30) en indiquant ceux qui sont significativement ou très significativement différents de zéro d'après le test d'Anderson (cf. 2.1.5).

On remarque sur le tableau 4.8 que, pour un paramètre donné, la dépendance des observations varie d'une station à l'autre. On peut cependant dégager des tendances générales sur le comportement moyen des paramètres de qualité.

Auto-corrélation	Paramètre
forte	Sulfates, Silice
intermédiaire à forte	Nitrates, Potassium
intermédiaire	Azote ammoniacal
faible à intermédiaire	Azote total, Phosphore total, Phosphore hydrolysable, Calcium, Manganèse, Sodium
faible	Conductivité, Fer, Magnésium, Chlo- rures

TABLEAU 4.8: Coefficients d'auto-corrélation de chaque série d'observations.

_____ valeur très significativement différente de 0 (test d'Anderson)

----- valeur significativement différente de 0

Les seuils de signification sont: $\alpha = 0.05$ $-.47 < r < .37$
 $\alpha = 0.01$ $-.60 < r < .50$

	SiO ₂	Cond.	N _T	NH ₃	NO ₃	P _T	P _h	Ca	Fe	Mg	Mn	Na	K	Cl	SO ₄
1	.32	.57	.52	.61	.26	.53	.57	.49	.24	.48	.21	.51	.54	.40	.64
2	.52	.20	.43	.41	.32	.40	.39	.18	.49	.06	.44	.34	.37	.02	.57
3	.51	.53	.42	.53	.17	.45	.46	.48	.26	.42	.38	.38	.57	.38	.67
4	.01	.49	.24	.43	.10	.06	.04	.27	.41	.28	.32	.53	.53	.52	.27
5	.12	.36	.09	.00	.06	.57	.57	.13	.39	.39	.55	.18	.62	.13	.42
6	.60	.36	.32	.24	.35	.45	.38	.32	.09	.30	.51	.32	.08	.38	.76
7	.55	.38	.35	.17	.51	.25	.05	.42	.29	.39	.19	.15	.34	.24	.71
8	.44	.17	.27	.12	.29	.17	.04	.17	.02	.00	.24	.01	.00	.03	.60
9	.43	.38	.37	.51	.23	.12	.14	.31	.21	.31	.05	.40	.64	.42	.60
10	.70	.35	.29	.27	.40	.09	.08	.30	.07	.28	.08	.40	.23	.25	.72
11	.56	.24	.33	.09	.54	.12	.09	.34	.33	.31	.12	.33	.26	.31	.69
12	.52	.43	.55	.37	.53	.44	.42	.64	.11	.58	.09	.39	.63	.13	.76
13	.73	.09	.45	.37	.74	.16	.27	.22	.07	.08	.40	.29	.11	.33	.46
14	.60	.25	.55	.58	.64	.05	.04	.52	.51	.46	.54	.16	.41	.20	.75
15	.45	.33	.38	.33	.73	.50	.51	.59	.60	.50	.05	.21	.58	.19	.62
16	.03	.35	.43	.10	.11	.25	.23	.36	.34	.25	.31	.06	.32	.07	.49
17	.46	.37	.34	.53	.64	.11	.27	.41	.17	.29	.41	.41	.31	.18	.71
18	.32	.19	.34	.39	.48	.22	.09	.35	.28	.26	.01	.29	.19	.03	.63
19	.54	.36	.23	.19	.65	.04	.12	.48	.37	.25	.26	.50	.08	.09	.70

Ce tableau nous permet de classer les paramètres de qualité en:

- paramètres persistants (forte auto-corrélation pour la majorité des séries du réseau);
- paramètres aléatoires (auto-corrélation non significativement différente de zéro, faible et faible à intermédiaire pour la majorité des séries d'observation).

Le classement conditionne le choix de la méthode d'étude:

- pour un paramètre persistant, on envisage une représentation par un processus markovien;
- pour un paramètre aléatoire, on envisage une représentation par l'ajustement de la loi statistique.

Ces deux types d'étude ont pour même but de synthétiser de façon simple le maximum d'information qu'on peut obtenir sur le paramètre considéré et de modéliser le phénomène. En pratique, seule l'étude des paramètres aléatoires permet une exploitation intéressante: reconstitution des caractéristiques statistiques diverses (moyenne, écart-type, quantile ...) et prévision des événements rare (cf. 2.2.4).

4.3.2 Etude des paramètres de qualité persistants

On envisage ici le cas de la silice et des sulfates qui présentent, sur l'ensemble des stations du réseau, des séries d'observation auto-correlées (cf. 4.3.1). Nous faisons l'hypothèse que les séries d'observations peuvent être représentées par des processus de Markov d'ordre 1 (cf. 2.2.3). Dans ce cas l'observation x_{i+1} dépend de l'observation précédente x_i et le processus est décrit par l'équation 2.34:

$$x_{i+1} = t_1 x_i + \varepsilon_{i+1}$$

où: t_1 est le coefficient d'auto-corrélation d'ordre 1 donné au tableau 4.8;

ε_{i+1} est la partie aléatoire du processus indépendante de x_{i+1} .

Les étapes de l'étude sont les suivantes:

- On essaie de réaliser une bonne adaptation du processus markovien au paramètre de qualité: on pourra alors envisager la prévision des valeurs futures du paramètre de qualité. On aura ainsi synthétisé par une équation du type 2.34 toute l'information disponible sur le paramètre de qualité.
- Si la reconstitution du paramètre de qualité par le processus se révèle insatisfaisante (écarts trop grands par rapport aux valeurs observées), ce qui rend impossible toute prévision, on cherchera alors la fréquence de mesures suffisante pour obtenir des observations indépendantes permettant d'étudier le paramètre de qualité comme paramètre aléatoire (cf. 4.3.3). On sait qu'il est possible de calculer le nombre d'observations indépendantes nécessaire et suffisant pour obtenir une information sur la moyenne équivalente à celle fournie par la série des valeurs dépendantes (cf. 2.2.3.3).

On a réalisé cette étude pour la silice à la station 10 et pour les sulfates à la station 6. Ces deux séries d'observation présentent les coefficients d'auto-corrélation les plus élevés pour ces paramètres.

4.3.2.1 Adaptation d'un processus de Markov d'ordre 1

On distingue deux étapes:

- a) Vérification de l'existence du processus et caractérisation;
- b) Simulation des valeurs prises par le paramètre de qualité suivant le processus.

a) Existence et détermination du processus

Si on a affaire à un processus de Markov d'ordre 1, le résidu ε_{i+1} de l'équation 2.34 doit être aléatoire. Pour cela on étudie l'indépendance de la série des valeurs ε_{i+1} telle que:

$$\varepsilon_{i+1} = x_{i+1} - t_1 x_i \quad (4.1)$$

où t_1 est le coefficient d'auto-corrélation de la série d'observations du paramètre de qualité étudié.

Le coefficient d'auto-corrélation d'ordre 1, t_ε , de la série des ε_{i+1} ne doit pas être significativement différent de zéro. On donne au tableau 4.9 les valeurs des coefficients d'auto-corrélation t_ε pour toutes les séries d'observation. Le test d'ANDERSON (cf. 2.1.5) montre qu'aucun de ces coefficients n'est significativement différent de zéro (sauf sulfates, station 12). Ainsi, dans le cas de la silice et des sulfates, l'effet de persistance n'existe pas pour les résidus et le processus de Markov d'ordre 1 semble pouvoir s'appliquer.

Pour déterminer complètement le processus, il faut trouver une loi statistique représentative de la population des résidus. On a testé (tableau 4.10) l'hypothèse de normalité des résidus par le test de W (cf. 2.2.2). L'hypothèse de normalité est acceptable pour l'ensemble des séries de résidus donc, en particulier, pour la silice et les sulfates, au niveau de confiance 5%.

TABLEAU 4.9: Coefficients d'auto-corrélation des résidus des séries d'observations ($x_{i+1} - t_1 x_i$)
à chaque station et pour chaque paramètre.

	SiO ₂	Cond.	N _T	NH ₃	NO ₃	P _T	P _h	Ca	Fe	Mg	Mn	Na	K	Cl	SO ₄
1	-.10	.19	.04	.19	.06	.10	.07	.22	.05	.13	.07	.18	.13	.08	.28
2	.12	.05	.04	.07	.02	.01	.02	.03	-.02	.02	.08	.08	-.06	-.00	.19
3	-.11	.18	-.01	.11	.04	.16	.11	.10	.03	.11	.14	.12	.20	.11	.28
4	.00	-.03	.02	.10	.03	-.01	-.01	.04	-.10	-.00	.09	.12	-.04	.11	.06
5	.01	.10	-.00	.00	-.01	.11	.13	.03	.01	.08	.14	.01	.28	.01	.18
6	.04	.09	-.04	.01	.04	.09	.01	.08	.01	.04	.16	.10	.00	.12	.41
7	.01	.11	-.03	.01	.05	.04	-.00	.16	-.01	.08	.03	.05	.05	-.01	.22
8	-.05	.05	-.05	.00	-.04	.06	.00	-.00	.01	-.00	.05	.00	.00	-.00	.13
9	-.06	.10	-.05	.05	-.02	-.02	-.03	.07	.01	.06	.01	.11	.23	.09	.15
10	.18	.16	.01	.00	-.00	-.04	-.02	.11	.01	.02	.01	.20	.05	.09	.29
11	.04	.06	.04	.02	.08	.05	.03	.15	-.09	.12	-.01	.14	.01	.10	.25
12	.24	.14	.21	.12	.16	.12	.18	.15	.04	.07	.01	.17	.38	.03	.46
13	.09	-.02	-.07	.06	.20	.03	.06	.08	.01	.02	-.05	.04	.03	.06	.07
14	.07	.10	.29	.25	.18	-.01	-.01	.22	.15	.19	.22	.07	.07	.06	.29
15	.16	.10	.17	.13	.34	.24	.25	.21	.32	.13	.01	.07	.28	.07	.23
16	.01	.19	.04	.02	.00	.07	.02	.18	-.07	.04	.10	.01	.06	.02	.03
17	.02	.07	.05	.06	.20	-.00	.04	.13	-.00	-.01	.09	.16	.07	.03	.34
18	-.00	.04	.04	.00	.07	-.06	.01	.10	.04	.02	.00	.06	.01	.00	.10
19	.14	.15	.05	.03	.23	-.01	.02	.23	-.02	.05	.00	.20	.02	.02	.32

TABLEAU 4.10: Tests de normalité sur les résidus des séries d'observations ($x_{i+1} - t_{1,x_1}$) à chaque station et pour chaque paramètre de qualité.

Seuil critique: $W_{5\%} = 0.91$ - .88: valeur inférieure au seuil

	SiO ₂	Cond.	N _T	NH ₃	NO ₃	P _T	P _h	Ca	Fe	Mg	Mn	Na	K	Cl	SO ₄
1	.99	.95	.98	<u>.81</u>	.93	.97	.93	.97	.98	.98	<u>.89</u>	.95	.99	.97	.95
2	.94	.94	<u>.90</u>	<u>.63</u>	.95	<u>.82</u>	<u>.89</u>	.94	.96	.97	<u>.82</u>	.94	<u>.73</u>	<u>.90</u>	.96
3	.96	.94	.99	<u>.81</u>	.94	<u>.89</u>	<u>.88</u>	.97	.97	.97	<u>.67</u>	.94	.97	.95	.96
4	.93	.95	.98	<u>.88</u>	.93	<u>.73</u>	<u>.68</u>	.96	.95	.96	<u>.82</u>	<u>.80</u>	<u>.64</u>	<u>.68</u>	.97
5	.98	.93	.94	<u>.87</u>	<u>.82</u>	<u>.45</u>	<u>.46</u>	.97	<u>.83</u>	.93	<u>.76</u>	.95	<u>.90</u>	.95	.95
6	.98	.96	.94	<u>.76</u>	.97	.91	<u>.89</u>	.94	.96	.96	<u>.88</u>	<u>.89</u>	<u>.48</u>	<u>.90</u>	.96
7	.98	.97	.99	<u>.89</u>	.99	.96	<u>.85</u>	.96	<u>.89</u>	.98	<u>.68</u>	.90	.96	.91	<u>.84</u>
8	.97	.95	.98	<u>.90</u>	.98	.96	<u>.88</u>	.97	.97	.97	<u>.84</u>	<u>.83</u>	.99	.93	.94
9	.94	.96	.97	<u>.88</u>	.97	.98	.96	.97	.95	.96	<u>.86</u>	.98	.96	.93	.92
10	.93	.97	.96	<u>.78</u>	.95	.92	<u>.81</u>	.97	.95	.96	<u>.82</u>	.94	.97	<u>.87</u>	.98
11	.98	.95	<u>.88</u>	<u>.80</u>	<u>.89</u>	<u>.81</u>	<u>.87</u>	.92	.98	.96	.94	<u>.87</u>	.93	<u>.73</u>	.96
12	.93	.98	.94	<u>.86</u>	<u>.90</u>	.95	.93	.92	.95	.97	.94	.98	<u>.88</u>	.97	.95
13	.93	.95	.96	<u>.74</u>	.95	<u>.79</u>	<u>.71</u>	.96	.99	<u>.89</u>	<u>.84</u>	<u>.78</u>	.92	<u>.75</u>	<u>.85</u>
14	.98	.94	<u>.88</u>	<u>.72</u>	.94	<u>.67</u>	<u>.52</u>	.93	.98	.93	<u>.73</u>	.92	.94	<u>.85</u>	.96
15	.91	.92	<u>.72</u>	<u>.82</u>	.94	<u>.89</u>	<u>.90</u>	.98	<u>.87</u>	.96	<u>.84</u>	<u>.78</u>	<u>.80</u>	<u>.90</u>	.98
16	<u>.93</u>	.95	.93	.92	<u>.59</u>	<u>.88</u>	<u>.86</u>	.92	.98	.97	<u>.81</u>	.91	<u>.81</u>	<u>.90</u>	<u>.82</u>
17	<u>.97</u>	.98	.94	.92	.93	<u>.88</u>	.92	.93	.96	.94	.91	.96	<u>.89</u>	.91	.91
18	.95	.94	.96	<u>.76</u>	.96	<u>.71</u>	<u>.46</u>	<u>.82</u>	.93	.94	.93	.93	<u>.89</u>	<u>.71</u>	.92
19	.96	.98	.88	<u>.57</u>	.97	.94	<u>.84</u>	.96	.95	.98	<u>.62</u>	.94	.93	<u>.72</u>	.91

En conclusion, on peut donc considérer que la représentation des paramètres silice et sulfate par un processus de Markov d'ordre 1 à résidu normal est acceptable.

b) Simulation du paramètre de qualité

En utilisant l'équation (2.39), on peut déterminer les coefficients des équations représentant le processus de Markov pour les deux cas étudiés:

$$\text{Silice, station 10: } x_{i+1} = 3.56 + 0.7 (x_i - 3.56) + 1.16 \eta_{i+1}$$

$$\text{Sulfates, station 6: } x_{i+1} = 16.86 + 0.76(x_i - 16.86) + 5.11 \eta_{i+1}$$

où η_{i+1} est une variable normale centrée réduite.

Dans les deux cas précédents on a reconstitué la série des valeurs avec le processus aléatoire en simulant une variable aléatoire normale centrée réduite.

Dans cette reconstitution la quantité t_1^2 représente la variance expliquée par la partie déterministe du processus. Par exemple, dans le cas des sulfates à la station 6, on a: $t_1 = 0.76$, $t_1^2 = 0.58$. Ainsi, 58% de la variance totale est expliquée par la partie déterministe et donc, 42% provient de la partie aléatoire du processus (ϵ_{i+1}). Ainsi, quand t_1^2 n'est pas très élevé, la partie aléatoire de l'équation joue un rôle prépondérant et nuit à la qualité de la reconstitution. On l'a vérifié en simulant une centaine de processus, correspondant à la même série, avec une centaine de vecteurs normaux différents. On a constaté qu'il était hasardeux d'envisager toute forme de reconstitution de ce genre, sauf lorsque l'on est en présence d'un très fort coefficient d'auto-corrélation t_1 , ce qui n'est pas le cas pour la plupart des séries de mesures dont nous disposons.

Ainsi, pour la silice et les sulfates, la reconstitution envisagée est imprécise. Il est cependant possible, à partir des données disponibles, d'examiner la réduction de la fréquence de mesure conduisant à des séries de valeurs indépendantes.

4.3.2.2 Réduction de la fréquence de mesure

Le problème vient, en fait, de ce que le réseau est mal adapté à la mesure des paramètres silice et sulfates. Il faut:

- soit augmenter la fréquence de mesure jusqu'à ce que les données obtenues permettent la modélisation de la partie déterministe, de façon suffisamment précise;
- soit réduire cette fréquence jusqu'à l'obtention d'une série de valeurs indépendantes qu'on pourra traiter comme une série aléatoire (cf. 4.3.3).

Il est clair que la deuxième méthode est la plus simple et la plus économique. Cette méthode conduit à réduire la redondance de l'information due à la partie déterministe du phénomène (cf. 2.2.3.3). On a montré (cf. 2.2.3.3 a)) qu'on pouvait obtenir une information équivalente (c'est-à-dire une même variance) sur la moyenne d'un paramètre de qualité, obtenue avec une série persistente de N observations, au moyen d'une série de N' observations indépendantes telle que:

$$N' = N I \qquad N' < N$$

où I est le contenu relatif d'information (cf. 2.2.3.3 a)).

On a calculé le contenu d'information de la moyenne de toutes les séries (cf. tableau 4.11), d'après l'équation (2.45). Le contenu d'information

TABLEAU 4.11: Contenu d'information de la moyenne des séries d'observations.

	SiO ₂	Cond.	N _T	NH ₃	NO ₃	P _T	P _h	Ca	Fe	Mg	Mn	Na	K	Cl	SO ₄
1	1.00	.30	.34	.27	1.00	.33	.30	.36	1.00	.37	1.00	.35	.32	.44	.24
2	.34	1.00	.42	.44	1.00	.45	.46	1.00	.36	1.00	.41	1.00	1.00	1.00	.30
3	.35	.33	.43	.33	1.00	.40	.39	.37	1.00	.43	.47	.47	.30	.47	.22
4	1.00	.36	1.00	.42	1.00	1.00	1.00	1.00	.44	1.00	1.00	.33	.33	.34	1.00
5	1.00	1.00	1.00	1.00	1.00	.30	.30	1.00	.46	.46	.32	1.00	.26	1.00	.42
6	.27	1.00	1.00	1.00	1.00	.40	.47	1.00	1.00	1.00	.35	1.00	1.00	.47	.17
7	.31	.46	1.00	1.00	.35	1.00	1.00	.43	1.00	.46	1.00	1.00	1.00	1.00	.20
8	.41	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.27
9	.42	.46	1.00	.35	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.45	.24	.43	.28
10	.20	1.00	1.00	1.00	.45	1.00	1.00	1.00	1.00	1.00	1.00	.44	1.00	1.00	.19
11	.31	1.00	1.00	1.00	.32	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.21
12	.33	.42	.31	.48	.33	.41	.43	.24	1.00	.29	1.00	.46	.25	1.00	.16
13	.18	1.00	.40	1.00	.18	1.00	1.00	1.00	1.00	1.00	.45	1.00	1.00	1.00	.39
14	.27	1.00	.32	.29	.24	1.00	1.00	.34	.35	.39	.32	1.00	.44	1.00	.17
15	.40	1.00	.47	1.00	.18	.36	.34	.28	.27	.35	1.00	1.00	.29	1.00	.26
16	1.00	1.00	.42	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.37
17	.39	1.00	1.00	.33	.24	1.00	1.00	.44	1.00	1.00	.44	.44	1.00	1.00	.20
18	1.00	1.00	1.00	.46	.37	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.25
19	.32	1.00	1.00	1.00	.24	1.00	1.00	.37	.48	1.00	1.00	.36	1.00	1.00	.20

TABLEAU 4.12: Nombre de valeurs indépendantes effectives (permettant d'obtenir la même information sur la moyenne du paramètre de qualité).

	SiO ₂	Cond.	N _T	NH ₃	NO ₃	P _T	P _h	Ca	Fe	Mg	Mn	Na	K	Cl	SO ₄
1	—	7	8	6	—	7	7	8	—	8	—	8	7	10	5
2	7	—	9	10	—	10	10	—	8	—	9	—	—	—	7
3	8	7	9	7	—	9	9	8	—	9	10	10	7	10	5
4	—	8	—	9	—	—	—	—	10	—	—	7	7	7	—
5	—	—	—	—	—	7	6	—	10	10	7	—	6	—	9
6	6	—	—	—	—	9	10	—	—	—	8	—	—	10	4
7	7	10	—	—	8	—	—	9	—	10	—	—	—	—	4
8	9	—	—	—	—	—	—	—	—	—	—	—	—	—	6
9	9	10	—	8	—	—	—	—	—	—	—	10	5	10	6
10	4	—	—	—	10	—	—	—	—	—	—	10	—	—	4
11	7	—	—	—	7	—	—	—	—	—	—	—	—	—	5
12	7	9	7	10	7	9	9	5	—	6	—	10	6	—	4
13	4	—	9	—	4	—	—	—	—	—	10	—	—	—	9
14	6	—	7	6	5	—	—	8	8	9	7	—	10	—	4
15	9	—	10	—	4	8	8	6	6	8	—	—	6	—	6
16	—	—	9	—	—	—	—	—	—	—	—	—	—	—	8
17	9	—	—	7	5	—	—	10	—	—	10	10	—	—	4
18	—	—	—	10	8	—	—	—	—	—	—	—	—	—	6
19	7	—	—	—	5	—	—	8	11	—	—	8	—	—	4

—: la série est déjà composée d'observations indépendantes.

est égal à l'unité quand le coefficient d'auto-corrélation de la série n'est pas significativement différent de zéro (série aléatoire). On donne, au tableau 4.12, les nombres de mesure effectifs N' correspondants.

Dans le cas de la silice, station 10, et des sulfates, station 6, on trouve $N' = 4$, ce qui correspond à une fréquence de mesure bi-annuelle durant deux années. On a vérifié sur les données existantes qu'il y a effectivement indépendance des valeurs pour cette fréquence.

4.3.2.3 Conclusion

Quand les observations relatives à un paramètre de qualité sont dépendantes, on a intérêt à réduire la fréquence de mesures, de façon à obtenir des observations indépendantes. Le tableau 4.12 indique, dans le cas des stations auto-correlées du réseau, le nombre de mesures indépendantes permettant d'obtenir la même information sur la moyenne du paramètre de qualité correspondant.

4.3.3 Etude des paramètres de qualité dont les observations sont aléatoires

Une série d'observations indépendantes peut être représentée par une loi statistique (cf. 2.2.4). La loi résume toute la connaissance qu'on peut avoir du paramètre de qualité. Elle sera d'une grande utilité pour déterminer:

- la valeur du paramètre de qualité correspondant à une certaine probabilité au dépassement (par exemple, valeur dépassée 5% des fois) ainsi que l'intervalle de confiance sur cette valeur à un niveau donné;
- la probabilité au dépassement d'un critère de qualité quelconque afférent à ce paramètre.

Ces deux problèmes réciproques sont importants dans l'étude de la qualité de l'eau, pour un objectif d'utilisation fixé. On distingue ici deux groupes de paramètres de qualité:

- ceux pour lesquels on peut ajuster une loi normale;
- ceux pour lesquels il faut faire choix d'un autre type de loi, en raison d'une valeur significativement différente de zéro du coefficient d'asymétrie.

L'hypothèse de normalité est testée par le test de W (cf. 2.2.2). On a, de plus, calculé le coefficient d'asymétrie (cf. 2.1.6) qui est nul dans le cas de la loi normale et dont la connaissance est indicative pour le choix d'autres types de lois. On se contentera d'exposer les principes et les résultats de l'ajustement dans le cas de la conductivité aux stations 16 et 17. On consacrera, en effet, le chapitre 5 à l'étude des applications de l'ajustement de loi pour le fer et le manganèse pour l'usage "eau potable".

4.3.3.1 Coefficient d'asymétrie - Test de normalité

Le tableau 4.13 donne les valeurs des coefficients d'asymétrie des séries d'observations indépendantes (d'après le tableau 4.8), calculés d'après la formule (2.23). Le tableau 4.14 indique, pour ces mêmes séries, le résultat du test de normalité W (cf. 2.2.2). Les cases vides de ces tableaux correspondent aux séries auto-correlées pour lesquelles les calculs seraient erronés en raison de la dépendance des valeurs de la série. L'analyse de ces tableaux permet de mettre en évidence les conclusions suivantes:

- asymétrie: Les séries relatives aux paramètres fer, calcium, magnésium, conductivité et azote total semblent avoir de

TABLEAU 4.13: Coefficients d'asymétrie des paramètres de qualité par station.

	SiO ₂	Cond.	N _T	NH ₃	NO ₃	P _T	P _h	Ca	Fe	Mg	Mn	Na	K	Cl	SO ₄
1	.24				-.19				-.34		<u>1.21</u>				
2		-.41			.54			-.01		.37		-.20		<u>1.42</u>	
3					-.69				-.25						
4	.66		-.12		1.16	<u>2.57</u>	<u>2.93</u>	.28		.13	<u>1.36</u>				1.1
5	.49	-.30	-.20	<u>1.57</u>	<u>1.94</u>			.16				.87		.27	
6		.65	1.15	<u>2.01</u>	-.36			.23	-.09	.55		.76	<u>4.06</u>		
7			.59	<u>1.27</u>		.60	<u>1.54</u>				<u>3.13</u>	1.01	-.19	.87	
8		-.31	.22	.91	-.32	.71	<u>1.37</u>	-.50	.29	-.26	<u>2.33</u>	<u>1.87</u>	.03	.89	
9					.46	-.10	.75	.19	.02	.30	.90				
10		<u>1.52</u>	.67	<u>2.06</u>		1.03	<u>1.93</u>	-.21	-.29	.00	<u>1.76</u>		.43	<u>2.45</u>	
11		.99	1.08	<u>1.30</u>		1.05	<u>1.29</u>	.38	.14	-.41	.06	<u>1.35</u>	.57	<u>2.46</u>	
12									.42		.92			.50	
13		1.06				<u>2.07</u>	<u>1.82</u>	.50	.26	<u>1.43</u>		<u>1.89</u>	.64	<u>2.33</u>	
14		.99				<u>2.35</u>	<u>3.22</u>					.93		<u>1.69</u>	
15		.75		<u>1.73</u>							<u>1.59</u>	<u>1.36</u>		.93	
16	.90	.07		.72	<u>3.44</u>	1.11	1.16	.35	.57	.15	<u>1.51</u>	.82	<u>2.15</u>	.95	
17		.10	.95			<u>1.42</u>	1.55		.52	-.15			<u>1.44</u>	1.07	
18	.22	-.20	.89			<u>1.94</u>	<u>3.40</u>	-.96	.93	-.17	1.04	<u>1.26</u>	1.06	<u>2.28</u>	
19		.29	<u>1.37</u>	<u>3.35</u>		.51	<u>1.42</u>			.36	<u>3.39</u>		.38	<u>2.45</u>	

— $\alpha = 0.01$ très significativement différent de zéro - $1.54 < C_s < 1.54$

--- $\alpha = 0.05$ significativement différent de zéro - $1.17 < C_s < 1.17$

TABLEAU 4.14; Tests de normalité,

	SiO ₂	Cond.	N _T	NH ₃	NO ₃	P _T	P _h	Ca	Fe	Mg	Mn	Na	K	Cl	SO ₄
1	.97				.92				.98		<u>.82</u>				
2		.96			.94			.95		.96		.96		<u>.90</u>	
3					.94				.94						
4	.93		.97		.92	<u>.74</u>	<u>.68</u>	.94		.95	<u>.80</u>				.92
5	.97	.98	.95	<u>.86</u>	<u>.81</u>			.98				.94		.94	
6		.93	.91	<u>.77</u>	.96			<u>.90</u>	.93	.93		<u>.90</u>	<u>.46</u>		
7			.92	<u>.89</u>		.96	<u>.85</u>				<u>.59</u>	.91	.97	.93	
8		.95	.92	<u>.90</u>	.97	.96	<u>.89</u>	.94	.97	.96	<u>.74</u>	<u>.84</u>	.98	.92	
9					.97	.99	.96	.95	.98	.93	<u>.81</u>				
10		<u>.85</u>	.95	<u>.73</u>		.91	<u>.78</u>	<u>.90</u>	.94	.96	<u>.81</u>		.96	<u>.72</u>	
11		.94	<u>.90</u>	<u>.78</u>		<u>.82</u>	<u>.88</u>	.91	.96	.94	.91	<u>.84</u>	.96	<u>.74</u>	
12									.96		.93			.95	
13		<u>.90</u>				<u>.76</u>	<u>.61</u>	.94	.98	<u>.90</u>		<u>.83</u>	.93	<u>.75</u>	
14		.94				<u>.67</u>	<u>.53</u>					.91		<u>.85</u>	
15		.92		<u>.72</u>							<u>.84</u>	<u>.81</u>		<u>.89</u>	
16	.92	.97		.91	<u>.61</u>	<u>.84</u>	<u>.83</u>	.95	.93	.94	<u>.84</u>	.92	<u>.81</u>	.91	
17		.97	.95			<u>.84</u>	<u>.83</u>		.96	.97			<u>.82</u>	<u>.90</u>	
18	.97	.97	<u>.90</u>			<u>.70</u>	<u>.50</u>	.92	.91	.98	.92	<u>.89</u>	<u>.89</u>	<u>.70</u>	
19		.97	<u>.87</u>	<u>.58</u>		.93	<u>.85</u>			.97	<u>.55</u>		.92	<u>.74</u>	

La probabilité que l'échantillon considéré provienne d'une loi normale est inférieure:

— à 1% seuil: 0.868
 --- à 5% seuil: 0.905

faibles asymétries pour l'ensemble des stations considérées. Par contre, les paramètres manganèse, phosphore hydrolysable, phosphore total et azote ammoniacal paraissent fortement asymétriques.

test de W: L'hypothèse de normalité est acceptable pour le fer et le magnésium ainsi que, dans la majorité des cas, pour la conductivité, le calcium et l'azote total. Elle est à rejeter dans l'ensemble, pour le manganèse, le phosphore hydrolysable, le phosphore total et l'azote ammoniacal.

Pour les autres paramètres, on ne peut tirer de conclusion générale. Pour la silice et les sulfates, on ne dispose pas de résultats puisque ces paramètres étaient fortement auto-correlés dans l'ensemble. Le test de W est confirmé par le test sur le coefficient d'asymétrie qui est cependant un peu moins puissant que le test W de normalité. La normalité du paramètre conductivité peut s'expliquer par le fait qu'il intègre l'effet de plusieurs autres paramètres (les ions majeurs).

4.3.3.2 Ajustement de loi

On considère le cas de la conductivité aux stations 16 et 17. On ajuste à ce paramètre une loi normale et une loi gamma (cf. 2.2.4.4). Une étude graphique permettra d'identifier le meilleur des deux ajustements (cf. 2.2.4.3). D'après le paragraphe précédent et le tableau 4.14, la loi normale devrait pouvoir s'appliquer. Cependant, la loi gamma est plus souple et peut conduire à un ajustement meilleur. Son emploi a, d'autre part, un plus grand sens physique puisque son domaine de variation est $(0, +\infty)$ alors que celui de la loi normale est $(-\infty, +\infty)$. Le tableau 4.15 donne les valeurs de la conductivité

TABLEAU 4.15: Valeurs observées classées aux stations 16 et 17 du paramètre conductivité et probabilité empirique correspondante.

Conductivité (µmhos/cm) Valeurs classées		
F	Station 16	Station 17
.048	150	86
.095	152	105
.143	191	116
.190	209	117
.238	214	120
.286	217	128
.333	231	134
.381	270	140
.429	288	143
.476	295	144
.524	304	160
.571	308	168
.619	314	181
.666	327	181
.714	361	184
.762	363	190
.810	363	195
.857	375	211
.905	420	218
.952	460	228

(probabilité empirique au non dépassement)

$$F = \frac{k}{n+1}$$

classées par ordre croissant; F est la probabilité au dépassement empirique (plotting position) calculée par la formule:

$$F = \frac{k}{N + 1}$$

où k est le numéro d'ordre de la valeur classée;

N est la taille de l'échantillon.

Les lois normale et gamma sont ajustées par la méthode du maximum de vraisemblance; les valeurs prises par les paramètres de ces lois sont:

Station	Loi normale	Loi gamma
16	$\mu = 290.6$	$\alpha = 0.0384$
	$\sigma = 86.2$	$\lambda = 11.1476$
17	$\mu = 46.9$	$\alpha = 0.0988$
	$\sigma = 13.1$	$\lambda = 15.5616$

Les valeurs de la loi gamma et les intervalles de confiance correspondant au niveau 95% sont donnés, pour diverses probabilités au dépassement, aux tableaux 4.16 (station 16) et 4.17 (station 17). Ces calculs sont effectués suivant la technique décrite par BOBEE et ROBITAILLE (1976). Les lois sont tracées aux figures 4.2 (station 16) et 4.3 (station 17).

On constate, sur ces figures, que l'ajustement des lois est, dans l'ensemble, peu différent. Cependant à la station 16 (figure 4.2) la loi gamma semble donner un meilleur ajustement pour les valeurs élevées, tandis que la loi normale paraît sous-estimer les événements de faible probabilité au dépassement. A la station 17, l'emploi de la loi normale, plus simple que celui

TABLEAU 4.16: Ajustement de la loi Gamma par la méthode du maximum de vraisemblance (station 16, Conductivité).

Valeurs des paramètres de la loi:

- . paramètre d'échelle: $\alpha = .0384$
- . paramètre de forme: $\lambda = 11.1476$

Caractéristiques de la population:

- . Moyenne: $\mu = 290.60$
- . Ecart-type: $\sigma = 87.04$
- . Coeff. asymétrie: $C = 0.60$
- . Coeff. variation: $C_v^s = 0.30$

Probabilité au dépassement	Evénement X_T	Ecart-type de X_T	Intervalle de Confiance					
			50%		80%		95%	
.001	635.	133.	545.	724.	465.	805.	375.	895.
.005	563.	93.	500.	626.	443.	683.	380.	746.
.010	532.	78.	479.	584.	432.	631.	379.	684.
.020	496.	63.	454.	538.	416.	576.	373.	619.
.050	447.	45.	417.	477.	389.	505.	359.	535.
.100	406.	34.	383.	429.	363.	450.	340.	473.
.150	380.	29.	361.	400.	343.	417.	323.	437.
.200	360.	26.	343.	378.	327.	394.	309.	412.
.500	282.	22.	267.	297.	254.	310.	240.	324.
.800	216.	19.	203.	229.	192.	240.	179.	253.
.850	202.	19.	189.	215.	178.	227.	165.	240.
.900	186.	20.	172.	199.	160.	212.	147.	225.
.950	163.	24.	147.	180.	132.	195.	115.	211.
.980	141.	33.	118.	163.	98.	183.	75.	206.
.990	127.	41.	99.	155.	74.	180.	47.	207.
.995	115.	49.	82.	148.	52.	178.	19.	212.
.999	94.	68.	48.	140.	7.	182.	-40.	228.

TABLEAU 4.17: Ajustement de la loi Gamma par la méthode du maximum de vraisemblance (station 17, Conductivité).

Valeurs des paramètres de la loi:

- . paramètre d'échelle: $\alpha = .0988$
- . paramètre de forme: $\lambda = 15.5616$

Caractéristiques de la population:

- . Moyenne: $\mu = 157.45$
- . Ecart-type: $\sigma = 39.91$
- . Coeff. asymétrie: $C^s = 0.51$
- . Coeff. variation: $C^v = 0.25$

Probabilité au dépassement	Evénement X_T	Ecart-type de X_T	Intervalle de Confiance					
			50%		80%		95%	
.001	310.	56.	272.	348.	238.	382.	200.	420.
.005	279.	40.	252.	306.	228.	330.	201.	357.
.010	265.	33.	243.	288.	222.	308.	200.	331.
.020	250.	27.	232.	268.	215.	285.	197.	303.
.050	228.	20.	215.	242.	203.	254.	190.	267.
.100	210.	15.	200.	220.	191.	230.	181.	240.
.150	199.	13.	190.	207.	182.	215.	173.	224.
.200	190.	12.	182.	198.	175.	205.	167.	213.
.500	154.	10.	148.	161.	142.	167.	135.	173.
.800	123.	9.	117.	129.	112.	135.	106.	141.
.850	117.	9.	111.	123.	105.	128.	99.	135.
.900	109.	10.	102.	115.	97.	121.	90.	128.
.950	98.	12.	90.	106.	83.	113.	75.	121.
.980	87.	15.	76.	97.	67.	106.	56.	117.
.990	80.	19.	67.	92.	55.	104.	43.	117.
.995	74.	23.	58.	89.	45.	103.	30.	118.
.999	63.	31.	42.	84.	23.	103.	1.	124.

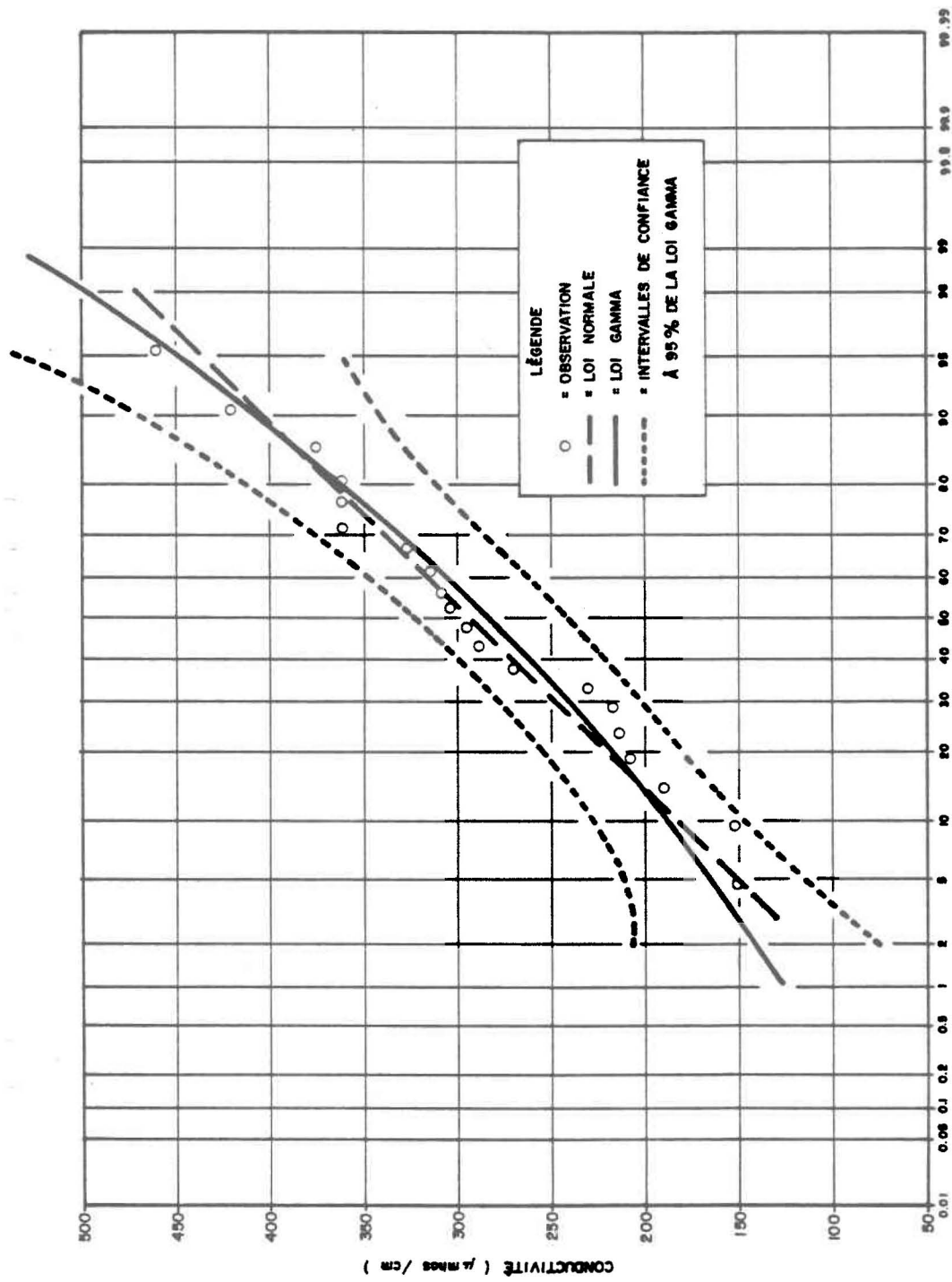


Figure 4.2 . Ajustement des lois normale et gamma (Conductivité , station 16)

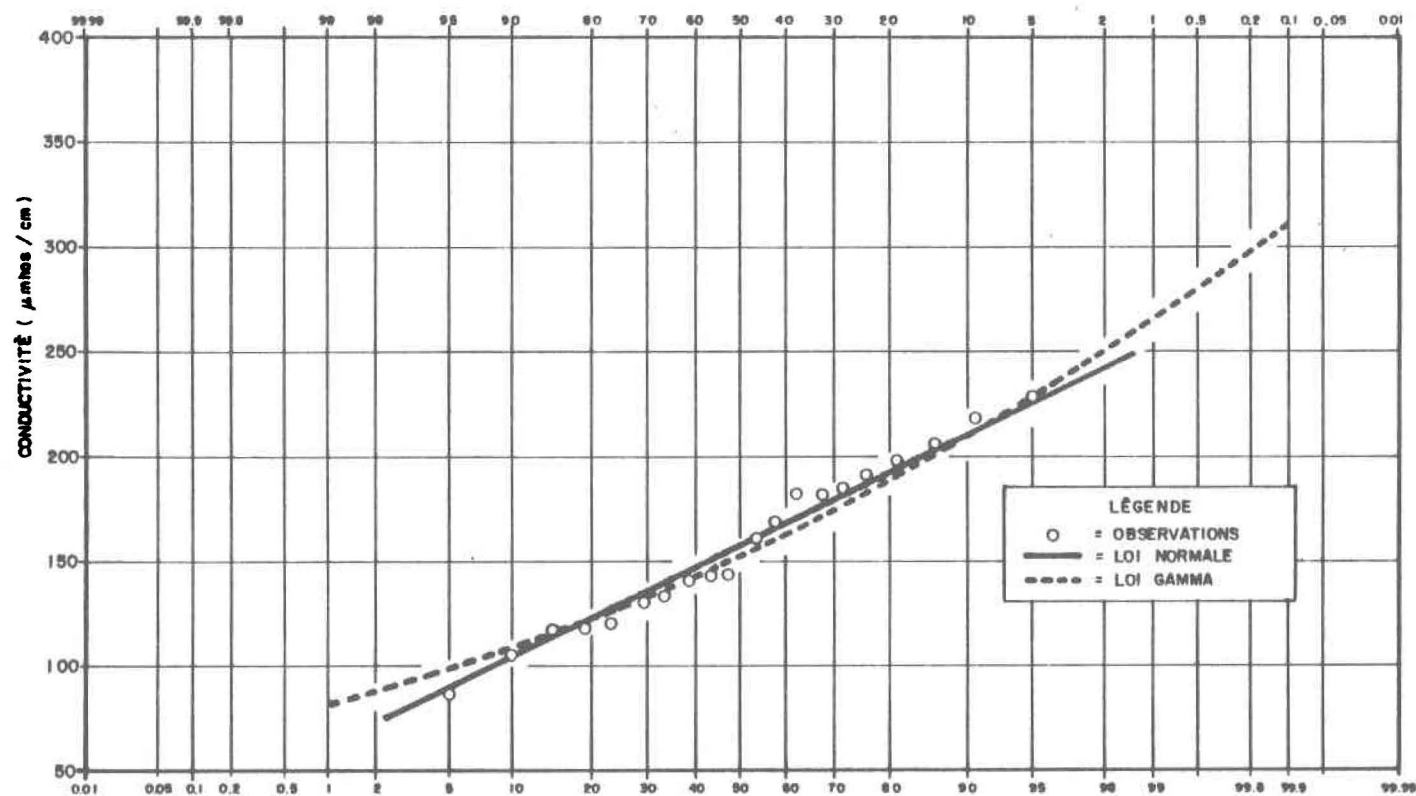


Figure 4.3 . Ajustement des lois normale et gamma - Conductivité , station 17 .

de la loi gamma, est adéquat. On est maintenant en mesure de déterminer la valeur de la conductivité qui correspond à une probabilité au dépassement donnée ainsi que l'intervalle de confiance de cette valeur. Ainsi, à la station 16, l'événement de probabilité au dépassement 0.1 (10%) est:

$$Q_{90} = 405 \text{ } \mu\text{mhos/cm}$$

L'intervalle de confiance à 95% de Q_{90} est:

$$340 < Q_{90} < 475 \text{ } (\mu\text{mhos/cm})$$

Réciproquement, si on se fixe la valeur de 450 $\mu\text{mhos/cm}$ pour la conductivité, la probabilité de dépassement de cette valeur à la station 16 est: 95%.

4.3.3.3 Conclusion

Certains paramètres de qualité comme le fer, le calcium, la conductivité, l'azote total ou le magnésium peuvent, à certaines stations, en première approximation, être étudiés par l'intermédiaire d'une loi normale. Pour plus de précision, et si on dispose des moyens de calcul nécessaires, on obtiendra de meilleurs résultats par l'emploi d'autres lois (gamma, Pearson type III, etc...) plus souples mais dont l'ajustement est un peu plus long et les propriétés moins connues. L'étude de paramètres tels que le manganèse, le phosphore hydrolysable, le phosphore total et l'azote ammoniacal nécessite l'emploi de lois asymétriques.

4.4 Etude de la variabilité spatiale du réseau

On a réalisé deux types d'étude:

- une analyse des inter-relations entre stations, pour chaque paramètre de qualité, par le calcul des coefficients de corrélation.

La détermination des régressions linéaires et du contenu d'information de la moyenne régionale que permet cette analyse, vise à réduire la densité du réseau;

- une analyse globale, pour l'ensemble du réseau, des moyennes des paramètres de qualité du groupe I (ions majeurs, silice et conductivité). On sera ainsi en mesure de caractériser par rapport à ces paramètres différentes zones du bassin.

4.4.1 Relations entre stations

4.4.1.1 Etude des coefficients de corrélation

Les résultats du calcul des coefficients de corrélation entre les stations prises deux à deux, pour tous les paramètres de qualité, sont regroupés dans l'annexe A1.

Ayant constaté que les coefficients de corrélation relatifs aux paramètres fer et manganèse sont faibles, on n'a considéré dans l'étude que les deux groupes de paramètres: ions majeurs (groupe I) et substances nutritives (groupe II).

On repère, pour chaque groupe, les forts coefficients de corrélation ce qui conduit aux tableaux 4.18 et 4.19; ces coefficients sont très significativement différents de zéro (au niveau de signification 0.01) et prennent des valeurs supérieures ou égales à 0.80.

On a synthétisé ces deux tableaux de façon à faire ressortir les couples de stations les plus fortement corrélées, pour les paramètres ions majeurs (groupe I) et substances nutritives (groupe II), chaque groupe étant considéré globalement (tableau 4.20).

TABLEAU 4.18: Coefficients de corrélation significatifs entre stations pour les paramètres du groupe "ions majeurs".

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1		17	<u>1 2 3 4</u> 5 6 7 8		8	<u>2 3 4</u> 8	<u>2 3 4</u> 6 8	1 8	<u>2 3 4 5</u> 5 7 8	8	8		8				7 8		87
2			8	6 3 8	2 3 5	1 8	8	8	8	8				8					
3					6 8	<u>1 2 3 4</u> 8 7	<u>1 2 3 4</u> 6 8	<u>1 8 2</u> 3	<u>2 3 4 5</u> <u>6 7 8</u>	8	5 8	8	8	<u>8</u>	5		7 8	8	7 8
4					8				8										
5									8										
6							<u>1 3 4 8</u> <u>2 5 7</u>	<u>1 4 8</u> 2 7	<u>1 2 3 4</u> 5 7 8	<u>1 8</u>	8	8	8	5 8	8		8 7	8	8
7								<u>1 2 3 4</u> 5 7 8	<u>1 2 3 4 8</u> 5 6 7 8					5 8	5 7				
8									<u>2 3 4</u> 8	8	8	8	8	5 8	5				
9									<u>8</u>	<u>8</u>			8	8			5 8 7	8	8
10										6 8	8	8	<u>1 3 8</u>	<u>8 6</u>	3	3	8 3	8	<u>1 3 4</u> 8
11											8	8		<u>1 2 3 4</u> 5 6 7 8	4 5	4 2	5 8	8	<u>1 2 3 4</u> 5 8
12														8	<u>1 3 8 6</u>		8	8	
13														4 8				7 8	7
14															2 3 5 6	2 3 4	1 5 8	1 8	<u>1 2 3 8</u>
15																2 3 4	3 4		<u>2 3 4</u>
16																			<u>2 3</u>
17																		<u>1 2 3 5</u> 7 8	<u>1 3 5 7</u> 8
18																			<u>1 2 7 8</u>
19																			

Légende: 1 SiO₂ 3 Ca 5 Na 7 Cl
2 Cond 4 Mg 6 K 8 SO₄

- à un numéro non souligné correspond un coefficient de corrélation calculé compris entre 0.80 et 0.90.

- à un numéro souligné correspond un coefficient de corrélation compris entre 0.90 et 1.

- Pour les numéros des paramètres, voir tableau 3.2.

- Pour les numéros des stations, voir figure 3.2.

TABLEAU 4.19: Coefficients de corrélation significatifs entre stations pour les paramètres du groupe "substances nutritives".

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1		2	<u>1</u> <u>2</u> <u>3</u> <u>4</u> <u>5</u>	2		<u>1</u> <u>2</u> <u>3</u> <u>4</u> <u>5</u>	<u>1</u> <u>2</u> <u>3</u> <u>4</u> <u>5</u>	1	<u>1</u> <u>2</u>	2								2	
2								1	<u>2</u>	2							2	<u>2</u>	2
3						<u>1</u> <u>2</u> <u>3</u> <u>4</u> <u>5</u>	<u>1</u> <u>2</u> <u>4</u> <u>5</u>	<u>1</u> <u>2</u> <u>4</u> <u>5</u>	<u>1</u> <u>2</u>								2		
4																			
5																			
6							<u>1</u> <u>2</u> <u>3</u> <u>4</u> <u>5</u>	<u>2</u> <u>3</u> <u>4</u> <u>5</u>	<u>1</u> <u>2</u> <u>3</u>	2	2	4						5	
7								<u>1</u> <u>2</u> <u>3</u> <u>4</u> <u>5</u>	<u>1</u> <u>2</u> <u>4</u>	2	2							5	
8									1 3	2	2							5	
9										2								2	
10											2 <u>3</u> <u>4</u>			4 5				<u>2</u>	2
11														<u>3</u> 4					
12																			
13														3				2	2
14															3				
15																			
16																			
17																		2 3 1	1 2 3
18																		<u>2</u>	
19																			

Légende: 1 N_T 3 NO₃ 5 P_h

2 NH₃ 4 P_T

- Pour les numéros des paramètres, voir tableau 3.2.

- à un numéro non souligné correspond un coefficient de corrélation calculé compris entre 0.80 et 0.90.

- à un numéro souligné correspond un coefficient de corrélation calculé compris entre 0.90 et 1.

- Pour les numéros des stations, voir figure 3.2.

TABLEAU 4.20: Couples de stations à forts coefficients de corrélation.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1			X O			X O	X O		X O										
2																			
3						X O	X O	X O	X O										
4																			
5																			
6							X O	X O	X O										
7								X O	X O										
8									X										
9																			
10											O (P)			O (P)					
11														X					X
12																			
13																			
14																			X
15																			X
16																			X
17																		O X (N)	O (N) X
18																			
19																			

Légende: X: ions majeurs (groupe I)
O: substances nutritives (groupe II)

plus spécifiquement:
O (N): formes azotées
O (P): formes phosphorées

Pour les numéros des stations, voir figure 3.2.

Le tableau 4.20 permet d'établir une distinction très nette entre deux groupes de stations:

- stations 1 à 9, en aval du bassin
- stations 10 à 19, en amont du bassin.

Il existe des corrélations importantes entre stations d'un même groupe mais les corrélations entre stations de groupes différents sont faibles. Il est intéressant de constater que ces deux groupes de stations correspondent à ceux qu'on avait identifiés en 3.2.2: le groupe de stations 10 à 19 (groupe I) était échantillonné systématiquement une semaine plus tôt que le groupe de stations 1 à 9 (groupe II). Un délai d'une semaine est suffisant pour que des variations importantes des concentrations des paramètres chimiques aient pu se produire (par exemple, à la suite d'événements météorologiques). Cela explique qu'on ne puisse pas, pour un paramètre donné, dans ces conditions, détecter de corrélations entre deux stations appartenant chacune à l'une des deux zones d'échantillonnage, même si elles sont proches géographiquement (sur le même cours d'eau). C'est ainsi que la station 8 de la Yamaska est correlée, pour la majorité des paramètres, avec les stations 3, 6, 7 qui sont en aval, mais ne l'est pas avec la station 10, pourtant située immédiatement en amont, mais échantillonnée une semaine plus tard. On observe aussi cette coupure entre les stations 9 et 17 de la rivière Noire.

On constate ainsi très nettement l'influence des délais d'échantillonnage sur l'interprétation des résultats: il est donc souhaitable que ces délais soient réduits au maximum, compte-tenu des contraintes matérielles.

Considérons séparément les deux groupes de stations envisagés:

- stations 1 à 9 (groupe II)

Les tableaux 4.18 et 4.19 nous permettent de dégager les points suivants:

- . Les stations sont corrélées entre elles le long de la rivière Yamaska. On remarque plus spécialement le couple (1, 3) pour l'ensemble des paramètres (ions majeurs et substances nutritives), ainsi que le couple (6, 7) pour l'ensemble des paramètres (sauf potassium). Ces stations sont corrélées avec la station 9 de la rivière Noire. Plus généralement, les couples suivants:

(1, 3) (1, 6) (1, 7) (1, 9)
 (3, 6) (3, 7) (3, 9)
 (6, 7) (6, 9)
 (7, 9)

sont bien corrélés en ce qui concerne:

- les ions majeurs (sauf parfois le potassium)
 - les substances nutritives, pour les couples formés des stations 1, 3, 6, 7
 - les substances azotées (azote total et ammoniacal) pour les couples formés avec la station 9.
- . Les stations 1, 3, 6 sont mieux corrélées avec la station 9 de la rivière Noire qu'avec la station 8 de la Yamaska, située en amont du confluent des deux rivières. Cependant, la station 7, située sur la Yamaska, en aval du confluent, est également corrélée avec les stations 8 et 9. On ne peut donc pas attribuer

par cette étude, d'influence prépondérante à la rivière Noire ou au cours amont de la Yamaska sur le cours inférieur de la Yamaska.

- Remarquons enfin l'isolement caractéristique des stations 2, 4, 5 (surtout 4 et 5) qui ne sont corrélées ni entre elles, ni avec d'autres stations du réseau (à l'exception des sulfates et de la conductivité pour (2, 5)). On peut en conclure que les rivières David, Salvail et Chibouet (respectivement stations 2, 4 et 5) drainent des sous-bassins aux caractéristiques propres dont l'influence se perd probablement par dilution (faible débit relatif) quand ces rivières se jettent dans le cours principal de la Yamaska. Notons que ces sous-bassins sont à vocation agricole.

- Stations 10 à 19 (groupe I)

- Les corrélations sont plus faibles sur cette portion du bassin, ce qui s'explique en partie par la plus grande ramification du réseau.
- Les coefficients de corrélation relatifs aux substances nutritives sont en général assez faibles, sauf pour les couples de stations suivants:
 - surtout (10, 11) pour le phosphore;
 - ainsi que (10, 14) (17, 18) et (17, 19) pour l'azote;
 - et (18, 19) pour l'azote ammoniacal.
- Pour les ions majeurs, les couples de stations suivants sont fortement corrélés:

- surtout (11, 14) pour l'ensemble des ions majeurs;
- ainsi que (17, 18) et (17, 19) pour les ions majeurs (sauf magnésium et potassium) et (18, 19) pour la silice, la conductivité, les chlorures et les sulfates;
- (10, 19), (11, 19), (14, 19), (15, 19), (16, 19) pour certains ions majeurs comme le calcium, les sulfates et la conductivité.

En considérant les cours d'eau dans leur ensemble, l'analyse précédente montre qu'il existe des corrélations entre stations le long des cours d'eau suivants:

- Yamaska (région II)
- Noire (région I)
- Yamaska Sud-Est (région I)

On constate, par contre, la faiblesse relative des corrélations entre les stations de la rivière Yamaska-Nord pour lesquelles cependant on observe des valeurs élevées des paramètres considérés (cf. 4.2.1 et 4.2.3). Dans le cas des substances nutritives, on peut peut-être expliquer ce fait par leur origine: rejets ponctuels d'égouts ou d'industries qui sont spécialement importants sur ce cours d'eau (ville de Granby, industrie laitière, élevage porcin, textile etc..., cf. 3.1.5). Ce type de rejet influence plus spécifiquement la station de mesure la plus proche; les variations s'y font sentir rapidement et de façon aigüe, les transformations chimiques et biologiques au sein du cours d'eau (pouvoir auto-épurateur, par exemple) n'ont pas eu le temps d'être effectives, etc...: les stations du cours d'eau sont ainsi individualisées.

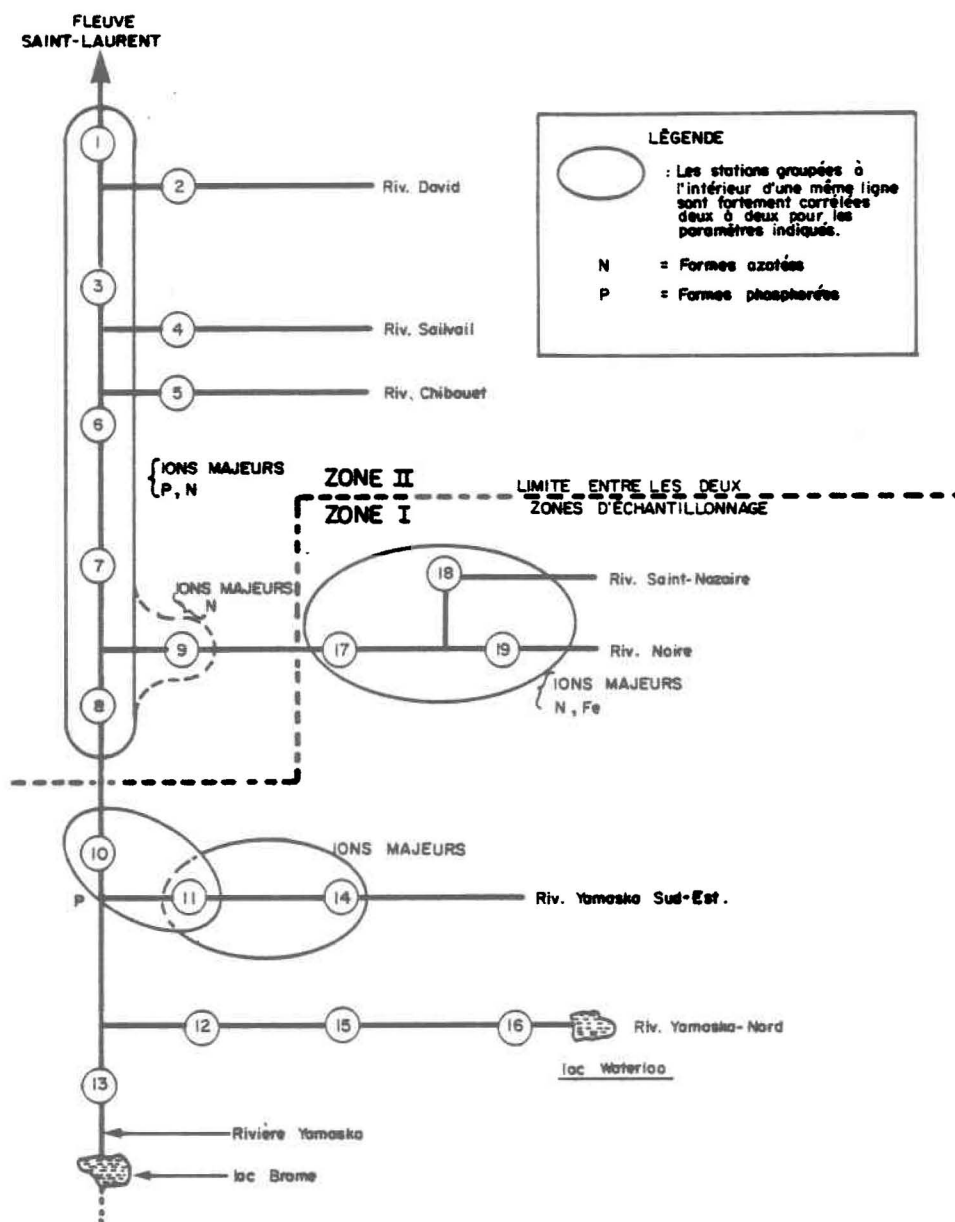


Figure 4.4. Schéma des groupes des stations fortement corrélées

Notons qu'un paramètre se distingue des autres quand on considère l'ensemble des stations: les sulfates sont, en effet, assez régulièrement corrélés pour tous les couples de stations du bassin (sauf stations 16, 4, 5 et, à un degré moindre, 2). Cela caractérise une certaine homogénéité spatiale des sulfates qui peut être explicable par une cause commune telle que l'origine atmosphérique des précipitations.

Le fer et le manganèse n'ont pas été pris en compte dans cette étude. Dans le cas de ces paramètres, on ne constate des coefficients de corrélations importants que pour les couples de stations suivants:

fer	(17, 18) et (17, 19) rivière Noire
manganèse	(12, 15) Yamaska-Nord et (7, 8) Yamaska.

Les résultats de cette étude sont résumés de façon globale sur la figure 4.4. L'analyse des coefficients de corrélation débouche sur le calcul de régressions linéaires et du contenu d'information de la moyenne régionale.

4.4.1.2 Régressions linéaires

Un fort coefficient de corrélation est indicatif d'une bonne liaison linéaire mais des liaisons fictives peuvent exister (cf. 2.3.1). Pour s'assurer de la réalité de la liaison, il est important de tracer les valeurs observées. On peut ensuite, éventuellement, calculer l'équation de régression, déterminer l'intervalle de confiance à un niveau donné de la valeur reconstituée, et évaluer la précision de cette reconstitution.

L'étude a été réalisée pour le couple des stations 3 et 1 et pour l'ensemble des paramètres de qualité, exceptés le fer et le manganèse. On reconstitue les valeurs de la station 1, située en aval du cours de la Yamaska, à par-

tir de celles de la station 3. On a choisi ces deux stations car elles présentent, dans l'ensemble, les plus forts coefficients de corrélation pour chaque paramètre.

L'annexe A-2 présente les graphiques réalisés pour chaque paramètre et comportant:

- . les points d'observation (symbole o);
- . la droite de régression (symbole I) dont l'équation est donnée sous chaque graphique;
- . les intervalles de confiance à 95% de la valeur reconstituée (symbole X).

L'annexe A-3 présente les valeurs reconstituées à la station 1 avec les intervalles de confiance et la précision relative, en pourcentage, de chaque reconstitution (cf. 2.3.1.2). On a reconstitué 20 valeurs, avec un pas constant, sur l'intervalle de variation du paramètre de qualité considéré à la station 3, de la façon suivante:

Soit $Y = ax + b$ la droite de régression

x_m minimum observé à la station 3

x_M maximum observé à la station 3

Y_i valeur reconstituée est telle que:

$$Y_i = a \left(\frac{x_M - x_m}{19} \cdot i + x_m \right) + b \quad i = 0, 19$$

Les graphiques de l'annexe A-2 et les précisions données à l'annexe A-3 permettent de constater que les régressions linéaires présentent peu d'intérêt pour les substances nutritives (sauf l'azote total) et les sulfates. Elles sont, par contre, utilisables pour tous les autres paramètres. Dans ce cas, on

observe que la précision de la reconstitution est meilleure pour les valeurs élevées que pour les faibles valeurs, ce qui est normal puisque l'on considère la précision relative (à la valeur reconstituée) alors que l'erreur absolue de la reconstitution varie peu, qu'il s'agisse de faibles ou de fortes valeurs reconstituées.

On a distingué deux groupes de paramètres, en fonction de la précision de la reconstitution.

Paramètre	Intervalle de variation de la précision de la reconstitution d'une valeur sur la gamme des valeurs reconstituées (%)
Conductivité	11 à 21
Calcium	12.5 à 22
Magnésium	12 à 29
Potassium	10 à 34
Azote total	15 à 75
Sodium	17 à 63
Chlorures	19 à 73
Silice	20 à 72

On pourra donc envisager la suppression des observations de ces paramètres à la station 1 et les reconstituer, si besoin est, à partir des observations de la station 3, d'après les graphiques de l'annexe A-2. Cette étude peut être généralisée au bassin tout entier et, selon l'objectif visé, des réductions importantes du réseau peuvent être envisagées par cette méthode, pour les paramètres considérés.

4.4.1.3 Contenu d'information de la moyenne régionale

L'existence de corrélations entre les diverses stations du bassin peut être interprétée comme une redondance d'information et il est possible de calculer (cf. 2.3.1.3) le nombre p' de stations indépendantes qui serait suffisant pour obtenir une information sur la moyenne régionale d'un paramètre de qualité identique à celle donnée par le réseau des p stations actuelles. Ces p' stations devront être indépendantes (c'est-à-dire non corrélées entre elles) pour que l'information qu'elles apportent soit maximale; en pratique il est cependant difficile de trouver, parmi celles qui existent déjà, des stations qui réalisent complètement cette condition. L'intérêt de cette étude est donc de donner des indications sur la réduction possible du réseau. On a effectué une application dans le cas du fer.

Dans le cas du fer, on obtient d'après le tableau A.1.9 de l'annexe A-1, en utilisant l'équation (2.64):

$$\bar{\rho} = 0.29 \quad (\text{coefficient de corrélation moyen})$$

$$I = 0.16 \quad (\text{contenu en information})$$

Le nombre de stations indépendantes qui donneront la même information sur la moyenne régionale du fer sera, d'après l'équation (2.66):

$$p' = 4$$

Au regard du tableau A.1.9 et en tenant compte des divers sous-bassins hydrographiques et de leur importance, on a choisi les quatre stations suivantes:

station 3 rivière Yamaska (proche de l'exutoire)

station 9 rivière Noire (exutoire)

station 11 rivière Yamaska Sud-Est (exutoire)

station 16 rivière Yamaska Nord (amont)

Les coefficients de corrélation sont les suivants:

	3	9	11	16
3		.26	-.02	-.08
9			-.13	-.03
11				.36
16				

Ces coefficients de corrélation échantillonnés ne sont pas exactement nuls, mais dans aucun cas la valeur théorique correspondante de la population n'est significativement différente de zéro au niveau 5%. La moyenne régionale obtenue à partir de ces quatre stations est de 0.12 mg Fe/l alors que la moyenne régionale obtenue à partir des dix-neuf stations du réseau actuel est de 0.13 mg Fe/l. Il serait donc, suivant l'objectif visé, possible d'utiliser ces quatre stations pour avoir une connaissance sommaire de la concentration régionale moyenne du fer sur le bassin.

4.4.1.4 Conclusion

Cette étude a mis en évidence les points suivants:

- Un délai d'une semaine séparant les dates d'échantillonnage des zones amont et aval du bassin a une influence importante sur l'étude des relations qui existent entre les stations. En fait, de tels délais sont nuisibles pour toute étude spatiale nécessitant la comparaison d'observations réalisées à des stations différentes mais aux mêmes dates. On recommande en conséquence, de réaliser l'échantillonnage d'un bassin versant dans le délai minimum, compte tenu des contraintes physiques, pour disposer de données approximativement synchrones.

- On a identifié des groupements de stations fortement corrélées entre elles pour certains types de paramètres (cf. figure 4.4). L'intérêt principal de ces corrélations est de conduire à l'établissement d'équations de prédiction qui permettent de réduire la densité du réseau (exemple pour les stations 1 et 3).
- Enfin, on a réalisé, dans le cas du fer, un calcul indicatif permettant d'évaluer la redondance d'information du réseau pour ce paramètre et, en conséquence, d'envisager une réduction possible en ce qui concerne la connaissance de la moyenne régionale du fer.

4.4.2 Etude globale des ions majeurs

On envisage dans cette section, une étude spatiale globale de tous les paramètres du groupe I (ions majeurs, silice et conductivité). L'analyse des correspondances (cf. 2.3.2) permet de réaliser une étude synthétique complète des dix-neuf stations du réseau par rapport à ces huit paramètres.

4.4.2.1 Réalisation de l'analyse des correspondances

Le tableau initial est composé des moyennes des huit paramètres de qualité aux dix-neuf stations du réseau (tableau 19 x 8). Les paramètres sont codifiés de la manière suivante:

1	silice	3	calcium	5	sodium	7	chlorures
2	conductivité	4	magnésium	6	potassium	8	sulfates

Ce tableau subit les transformations suivantes:

- pour chaque paramètre de qualité, on détermine deux seuils qui divisent l'ensemble des dix-neuf moyennes relatives au paramètre

en trois groupes approximativement équiprobables: valeurs élevées (E), valeurs moyennes (M), valeurs faibles (F);

- chaque paramètre de qualité est ainsi subdivisé en trois classes et la dimension du nouveau tableau d'étude sera 19 x 24. On le constitue en associant à chaque moyenne la valeur codée binaire 1 ou 0, selon qu'elle appartient ou non à une classe considérée.

C'est sur ce tableau, formé uniquement des valeurs 1 ou 0, que l'analyse des correspondances est effectuée. Les transformations précédentes ont pour but:

- de donner un poids égal dans l'analyse à chaque paramètre de qualité;
- de rendre l'analyse des correspondances plus discriminante.

Il est cependant clair que le codage a un effet de pondération puisque l'on remplace les valeurs brutes par des 0 ou des 1.

4.4.2.2 Résultats

Le programme de calcul utilisé nous a donné les résultats suivants:

Axe factoriel	Valeurs propres λ_i	% de variance expliquée	% cumulé
1	.75	37.45	37.45
2	.49	24.60	62.05
3	.26	12.95	75.00
4	.18	8.91	83.91
5	.09	4.52	88.43
6	.06	3.13	91.56

λ_i est la valeur propre associée à l'axe factoriel i correspondant. Elle représente la variance du nuage de points expliquée par cet axe. Le tableau s'arrête au sixième axe factoriel, la variance résiduelle étant alors inférieure à 9%. En pratique, nous s'utiliserons pour l'interprétation que les trois premiers axes factoriels qui expliquent 75% de la variance totale.

Les figures 4.5 et 4.6 donnent les projections des points "stations" (observations) et des points "paramètres" (variables) dans le plan des deux premiers axes factoriels qui contribuent à 62% de la variance totale. Ces figures mettent en évidence, respectivement, les groupements de points "paramètres" (figure 4.5) et les groupements de points "stations" (figure 4.6). Sur ces figures le point représenté, par exemple, par le symbole 3E caractérise les valeurs élevées du calcium.

4.4.2.3 Etude des points "paramètres"

Chaque point paramètre de la figure 4.5 est repéré sur le plan des deux premiers axes factoriels par ses composantes sur ces axes. Les points d'un même groupe ont des composantes voisines sur les axes factoriels et sont situés dans une même région du plan. Pour déterminer plus précisément les groupements, on a indiqué à côté de chaque point paramètre la valeur de la composante sur le troisième axe factoriel. On établit ainsi des sous-groupes de points en fonction de leur position par rapport à cet axe. On distingue trois groupes de points distincts:

- le groupe (I), (2F, 3F, 4F, 5F, 6F, 7F, 8F) est constitué des valeurs faibles des moyennes des paramètres (sauf la silice). Les deux premiers facteurs contribuent fortement à expliquer la variance des points de ce groupe. D'après le troisième facteur, on peut établir

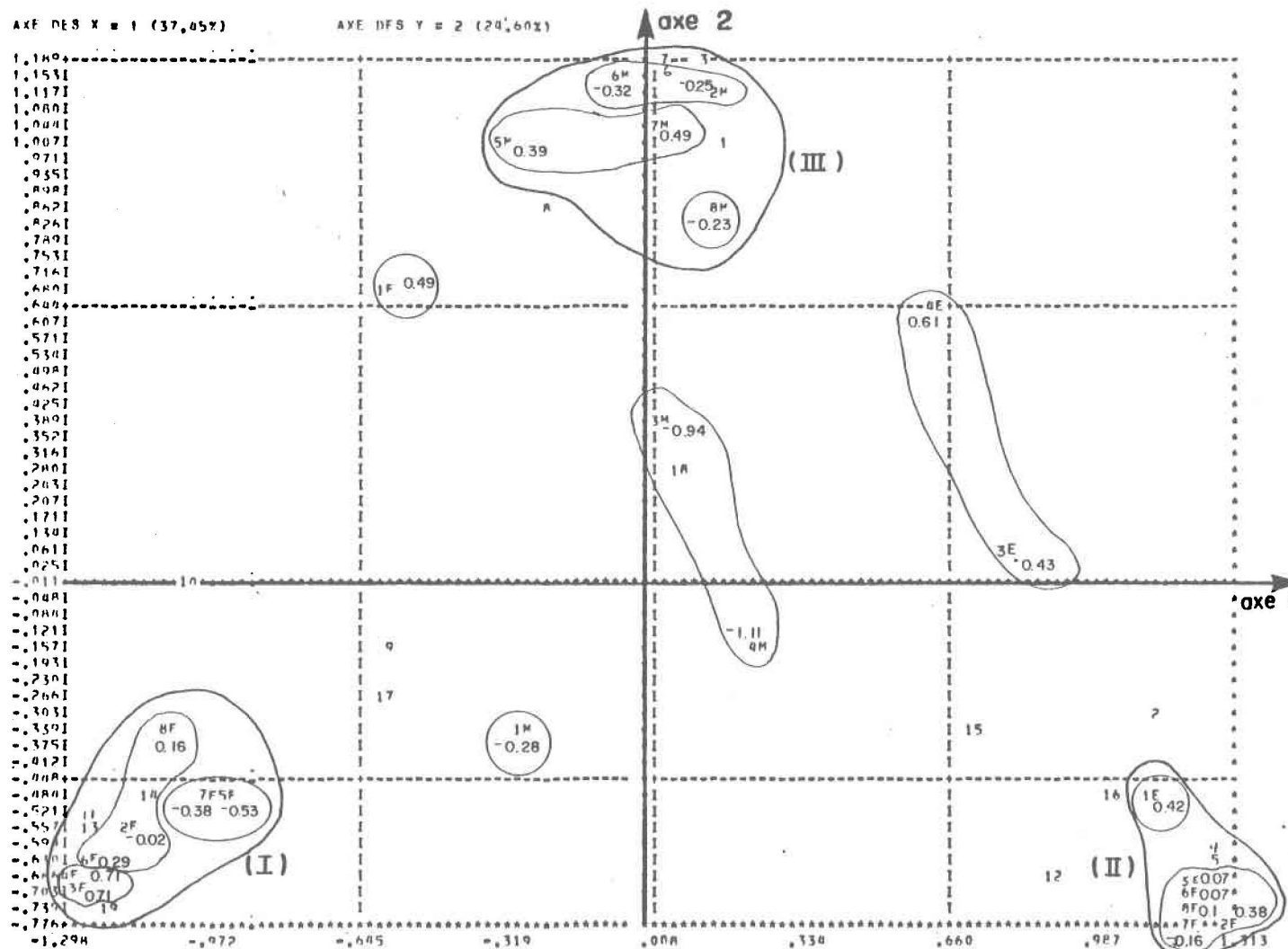


Figure 4.5 . Analyse factorielle des ions majeurs dans le plan des axes 1 et 2 - Étude des points paramètres .

POINTS DOUBLES (4, 5) (6, 7) (11, 13) (3F, 4F) (5F, 6F)

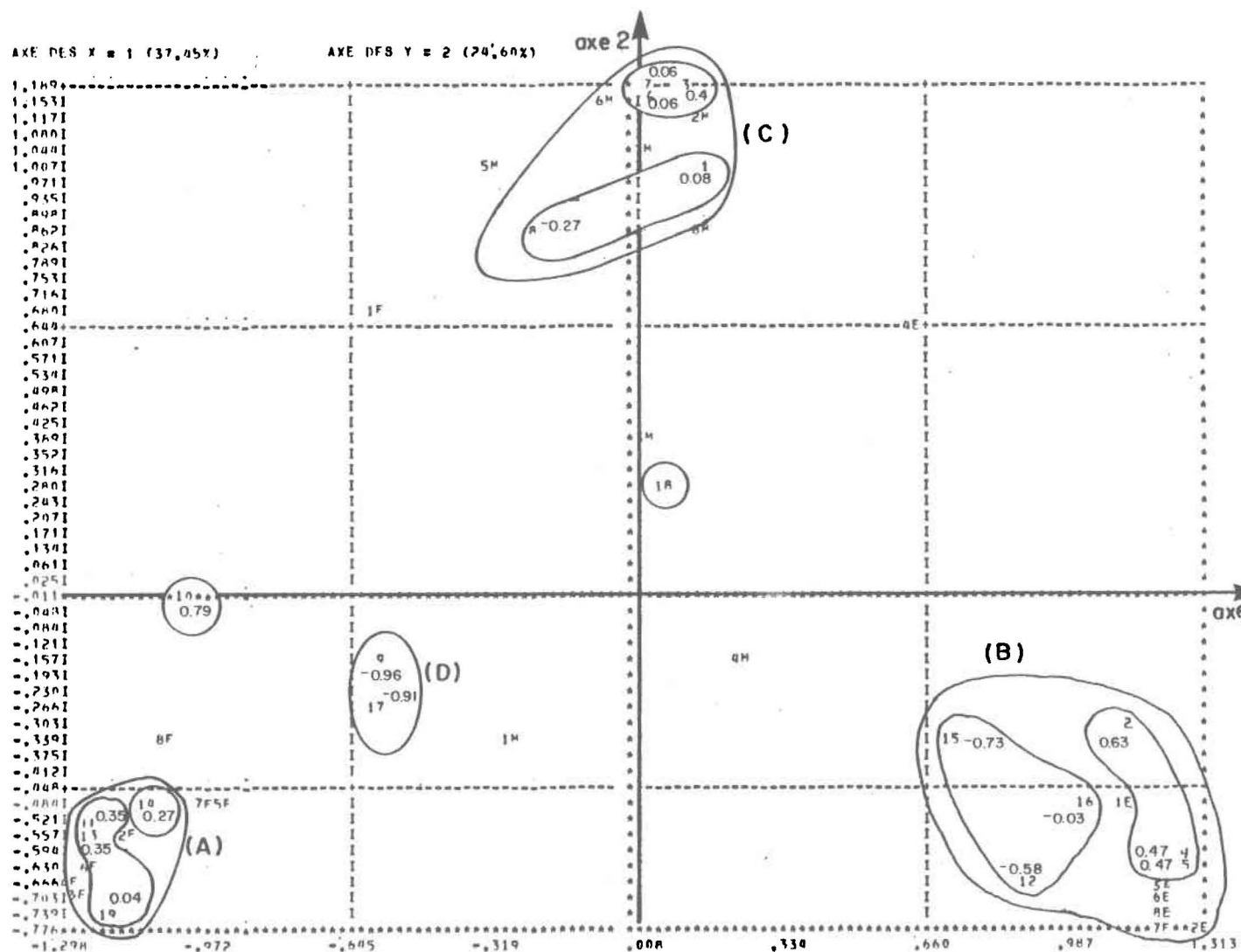


Figure 4.6. Analyse factorielle des ions majeurs dans le plan des axes 1 et 2. Etude des points stations.

les sous-groupes suivants:

- . 3F, 4F (calcium, magnésium)
- . 5F, 7F (sodium, chlorures)
- . 6F, 2F, 8F (conductivité, potassium, sulfates).

- Le groupe (II) (1E, 2E, 5E, 6E, 7E, 8E) est constitué des valeurs élevées des moyennes des paramètres (sauf le calcium et le magnésium). Les deux premiers facteurs contribuent bien à expliquer la variance des points de ce groupe. Le quatrième facteur contribue en partie à expliquer la variance de la silice. D'après le troisième facteur, on établit les sous-groupes suivants:

- . 2E, 5E, 6E, 7E, 8E (conductivité, sodium, potassium, chlorures, sulfates)
- . 1E (silice)

- Le groupe (III) (2M, 5M, 6M, 7M, 8M) est constitué des valeurs intermédiaires des moyennes des paramètres (sauf la silice, le calcium et le magnésium). Le deuxième facteur contribue fortement à la variance totale mais le premier facteur y contribue peu. Le quatrième facteur contribue un peu à la variance des sulfates. D'après le troisième axe factoriel, on établit les sous-groupes suivants:

- . 2M, 6M, 8M (conductivité, potassium, sulfates)
- . 5M, 7M (sodium, chlorures)

- Les points 1F, 1M, 3M, 6M, 3E, 4E ne peuvent être inclus dans les groupes précédents. Les deux premiers facteurs expliquent une partie de leur variabilité. D'après l'étude des facteurs d'ordre su-

périeur, on déduit les facteurs influant: les troisième et quatrième facteurs pour 3M, 4M, 3E et 4E et le cinquième facteur pour 1M et 1F. On établit les sous-groupes suivants:

- . 3E, 4E (calcium et magnésium)
- . 3M, 4M (calcium et magnésium)
- . 1F, 1M sont laissés à part.

On peut dire que dans l'ensemble (sauf pour les paramètres silice, calcium et magnésium):

- . le premier axe factoriel oppose les faibles moyennes aux moyennes élevées (axe d'échelle);
- . le deuxième axe factoriel oppose les moyennes intermédiaires aux moyennes extrêmes (faibles ou élevées).

En conclusion:

- trois paramètres ont un comportement différent de celui des autres: les valeurs faibles et intermédiaires de la silice et les valeurs intermédiaires et élevées du calcium et du magnésium;
- les cinq autres paramètres, conductivité, sodium, potassium, chlorures, sulfates ont un comportement assez semblable pour l'ensemble des moyennes observées.
- on a constaté des similitudes de comportement plus spécifiques pour les groupes de paramètres suivants:
 - . sodium et chlorures
 - . conductivité, potassium, sulfates
 - . calcium et magnésium.

4.4.2.4 Etude des points stations

D'après la figure 4.6, on distingue les groupements suivants:

- groupe A: stations 11, 13, 14, 19. Les deux premiers facteurs contribuent fortement à la variance des stations 11, 13 et 19, alors que le cinquième facteur contribue en partie à la variance de la station 14. On distingue donc deux sous-groupes:
 - . stations 11, 13, 19
 - . station 14

- groupe B: stations 2, 4, 5, 12, 15, 16. Les deux premiers facteurs contribuent fortement à la variance des stations 4 et 5. Le troisième facteur contribue en partie à la variance de la station 2. Les troisième et quatrième facteurs contribuent à la variance des stations 12 et surtout 15. En remarquant que les groupes de stations 2, 4, 5 d'une part, et 12, 15, 16 d'autre part, ont tendance à s'opposer sur le troisième axe factoriel, on peut déterminer les sous-groupes suivants:
 - . stations 2, 4, 5 (rivières David, Salvail, Chibouet)
 - . stations 12, 15, 16 (rivière Yamaska-Nord)

- groupe C: stations 1, 3, 6, 7, 8 (rivière Yamaska, en aval). Les deux premiers facteurs contribuent fortement à la variance des stations 3, 6 et 7, alors que le cinquième facteur contribue, pour une bonne part, à la variance des stations 1 et 8. Ainsi, on obtient les sous-groupes:
 - . stations 3, 6, 7
 - . stations 1, 8

Le premier axe factoriel oppose le groupe A au groupe B. Le deuxième axe factoriel oppose le groupe C aux groupes A et B.

- Les stations 9, 10, 17 et 18 ne peuvent pas être classées dans les groupes précédents. Le troisième facteur contribue de façon complémentaire à la variance des stations 9, 10 et 17. Les stations 9 et 17 s'opposent à la station 10, le long de cet axe. La station 18 présente un comportement très particulier: le quatrième facteur contribue presque entièrement à sa variance totale. On obtient donc les trois groupes suivants:

- . groupe D, stations 9, 17 (rivière Noire)
- . station 10
- . station 18 (rivière St-Nazaire)

En conclusion, on distingue trois grands groupes bien distincts A (11, 13, 14, 19), B (2, 4, 5, 12, 15, 16), C (1, 3, 6, 7, 8), un groupe autonome D (9, 17) et deux stations individualisées 10 et 18, cette dernière station ayant un comportement nettement particulier.

Les stations de chacun des groupes identifiés ont un comportement homogène, pour l'ensemble des paramètres considérés; il serait donc possible, en choisissant une station représentative de chaque groupe, d'avoir une vue d'ensemble du comportement de toutes les stations. On peut remarquer qu'en considérant les stations 3, 9, 11 et 16, on a une station représentant chacun des groupes A, B, C et D; or, on a montré (cf. 4.4.1.3) que dans le cas du fer (paramètre qui n'est pas considéré ici), ces quatre stations étaient "indépendantes" et pouvaient représenter l'ensemble du réseau pour le calcul de la moyenne régionale.

4.4.2.5 Etude simultanée des points paramètres et des points stations

- Le groupe de stations A est proche (cf. figures 4.5 et 4.6) du groupe de paramètres (I) et éloigné des autres groupes de paramètres. On peut donc associer le groupe de stations A aux faibles valeurs moyennes des ions majeurs (sauf silice).
- On peut de même associer les stations du groupe B aux moyennes élevées des paramètres du groupe (II) (ions majeurs, sauf calcium et magnésium).
- Les stations du groupe C sont à rapprocher du groupe (III) (ions majeurs, sauf silice, calcium et magnésium).
- Les stations 17 et 19 du groupe D ainsi que la station 10 pourraient être associées aux paramètres du groupe (I), mais de façon moins nette que celles du groupe A.
- On ne peut rien conclure quant à la station 18.

Le tableau 4.21 résume les résultats de cette étude et les associations mises en évidence.

La comparaison des tableaux 4.4 et 4.21 montre l'avantage de l'analyse des correspondances. L'étude empirique du tableau des moyennes 4.1 (cf. 4.2.1) ne permet que de réaliser un classement global des stations vis-à-vis des ions majeurs pris dans leur ensemble. L'analyse des correspondances, par contre, caractérise toutes les stations vis-à-vis de chacun des ions majeurs, en faisant ressortir les stations et les paramètres aux comportements excentriques. Toute l'information du tableau 4.1 est ainsi utilisée, ce qu'une simple étude directe ne saurait réaliser, en raison du grand nombre de données et de la difficulté

TABLEAU 4.21: Analyse factorielle sur les ions majeurs: interprétation des résultats.

Caractéristiques	Paramètres influents	Stations	Localisation
Moyennes élevées	Silice, Conductivité, Sodium Potassium, Chlorures, Sulfates	2, 4, 5, 12, 15, 16	rivières David, Salvail, Chibouet; rivière Yamaska-Nord
Moyennes intermédiaires	Conductivité, Sodium, Potassium, Chlorures, Sulfates	1, 3, 6, 7, 8	rivière Yamaska (de l'embouchure jusqu'à la station 8, à 4 kilo- mètres en amont du confluent avec la rivière Noire)
Moyennes faibles	Conductivité, Calcium, Magnésium, Sodium, Potassium, Chlorures, Sulfates	11, 13, 14, 19	rivière Yamaska Sud-Est (11, 14); amont de la rivière Noire (station 19 à 2.8 km de Roxton Falls); amont de la rivière Yamaska (station 13, à 5.1 km en aval d'Adamsville)
Moyennes plutôt faibles	Conductivité, Calcium, Magnésium, Sodium, Potassium, Chlorures, Sulfates	10, 17, 9	Station 10: rivière Yamaska à 1.6 km en aval de Farnham après la confluence des rivières Ya- maska Sud-Est et Yamaska-Nord. Stations 9, 17: rivière Noire de l'embouchure jusqu'au pont- route de St-Hector de Bagot.
Pas d'interprétation		18	rivière St-Nazaire, affluent de la rivière Noire.

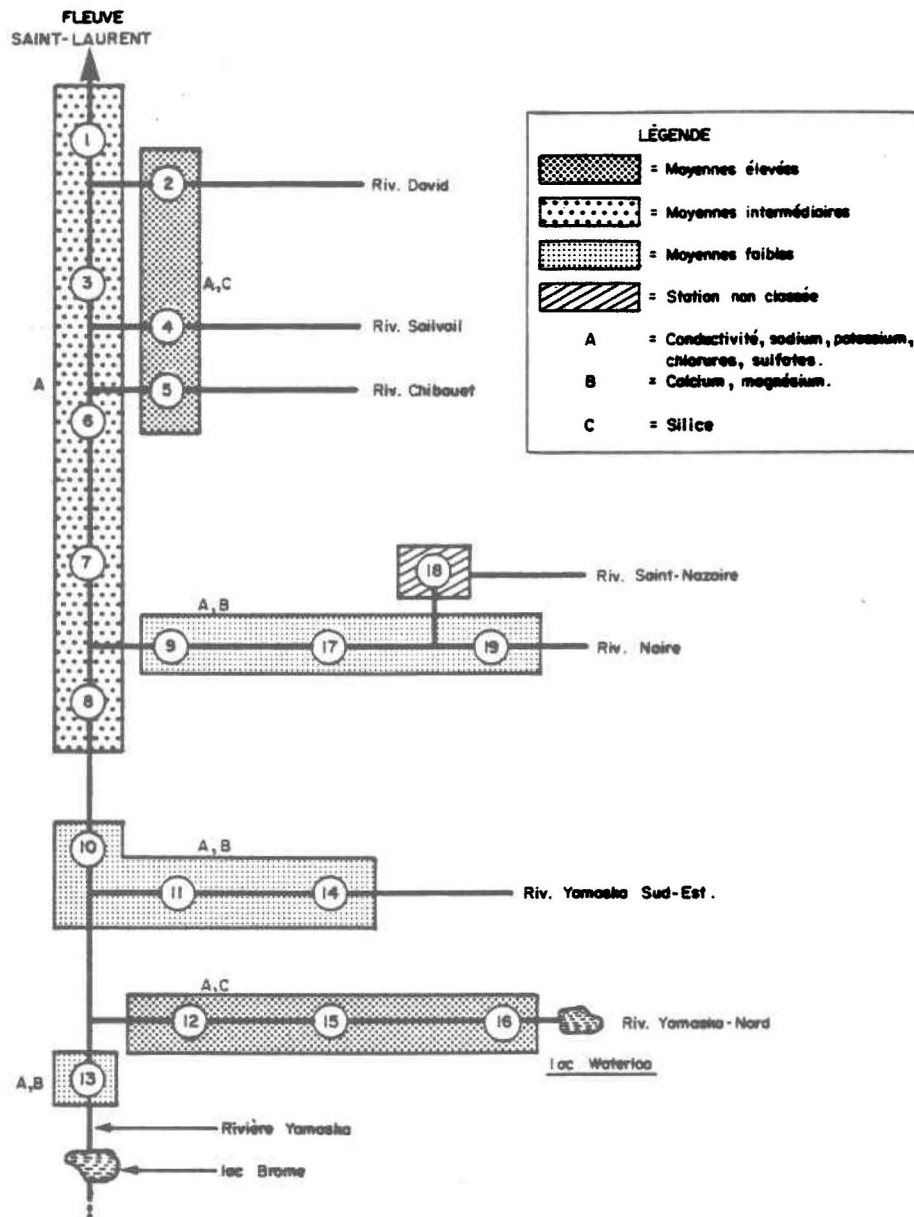


Figure 4.7. Visualisation des résultats de l'analyse des correspondances appliquée aux ions majeurs

de les discriminer.

On a établi à la figure 4.7 un zonage du bassin versant d'après les résultats du tableau 4.21.

4.5 Conclusion

Ce chapitre constitue une étude du réseau de mesure du bassin versant de la rivière Yamaska, de ses performances et des résultats qu'il fournit. Deux conclusions importantes sont à rappeler:

- On a constaté un enrichissement excessif des eaux en substances nutritives à l'ensemble des stations du bassin (sauf à la station 13) et plus particulièrement aux stations de la rivière Yamaska-Nord.
- Des problèmes dus au fer et surtout au manganèse se posent pour l'alimentation en eau potable, à bon nombre de stations et, plus particulièrement, aux stations de la rivière Yamaska-Nord.

On a, par ailleurs, acquis divers autres résultats concernant.

- certaines caractéristiques statistiques des paramètres de qualité (moyennes, écarts-types, maxima, ...);
- la synthèse de l'information disponible sur les paramètres de qualité (lois statistiques);
- les performances du réseau (précisions atteintes sur les moyennes);
- la variabilité spatiale de certains paramètres (analyse des correspondances sur les ions majeurs);
- les possibilités de réduire le réseau (régressions linéaires, contenu de l'information).

Il n'est cependant pas possible de réaliser une véritable synthèse de tous ces résultats qui permette de donner des réponses suffisamment complètes et précises aux problèmes:

- a) de la qualité de l'eau pour l'alimentation et la vie aquatique;
 - b) des performances du réseau;
 - c) de l'utilité des traitements statistiques mis en oeuvre.
- a) On ne donne pas de réponse correcte au problème de la qualité de l'eau pour la vie aquatique et l'alimentation, pour plusieurs raisons:
- des paramètres de qualité importants comme les paramètres bactériologiques (coliformes totaux, coliformes fécaux, streptocoques fécaux, ...) ou comme des paramètres chimiques tels que le mercure (dont la présence sur le bassin est prouvée), les pesticides, les détergents, etc... n'ont pas été mesurés par le réseau. Par contre, des paramètres dont la connaissance est beaucoup moins utile, comme les ions majeurs, sont abondamment échantillonnés. On ne peut, dans ces conditions, espérer qu'une réponse très partielle aux problèmes de qualité envisagés. En ce qui concerne plus particulièrement la vie aquatique, certains indicateurs biologiques tels que la présence concomitante de deux espèces de poissons ou la teneur en métaux des écailles de poissons pourraient être considérés. Ces données, en raison de leur nature, devraient cependant être acquises sur des bases différentes du réseau actuel.
- En ce qui concerne les paramètres mesurés, les caractéristiques statistiques étudiées sont surtout les moyennes, les écarts-types et les maxima. Mais on ne sait pas si le choix de ces caracté-

ristiques est pertinent: la connaissance de certains quantiles aurait peut-être donné plus de renseignements utiles. On a constaté l'existence de dépassements de critères de qualité mais on n'a pas calculé, faute d'indications précises, la probabilité réelle associée à ces dépassements (ce qui demande une étude particulière assez longue, cf. chapitre 5).

b) On ne peut pas juger des performances du réseau, en effet:

- on ne sait pas si les précisions atteintes sur les moyennes sont suffisantes; d'autre part, on ne connaît pas les précisions atteintes sur d'autres caractéristiques statistiques qui n'ont pas été calculées, telles que les quantiles. On ne peut donc pas envisager d'éventuelles modifications des fréquences de mesure (augmentation ou réduction à partir de ces résultats).
- Les études spatiales (régressions, calculs de contenu d'information, étude spatiale des moyennes par l'analyse des correspondances) ne peuvent pas, pour les mêmes raisons, déboucher sur des recommandations précises: telle régression est-elle acceptable, telle réduction du réseau est-elle envisageable? On ne peut fournir de réponses sans critères de jugement.

c) Enfin, on a envisagé, au cours de ce chapitre, diverses méthodes statistiques relatives à la synthèse de l'information (distributions statistiques), au traitement global d'un grand nombre de données (analyse des correspondances), à la réduction du réseau de mesures (régressions linéaires, contenu de l'information). Mais ces méthodes sont souvent restées au stade méthodologique, faute d'un cadre pour leurs applications et d'objectifs de précision fixés à priori.

Il est clair que tous les problèmes soulevés précédemment ont une cause unique: le manque de définition initiale précise des objectifs du réseau. La nécessité de la démarche rationnelle exposée dans le premier chapitre de ce travail se fait ici sentir de façon cruciale. En effet, si une définition précise des objectifs du réseau avait présidé à la planification, on aurait:

- mesuré les paramètres de qualité utiles;
- identifié à priori les caractéristiques statistiques pertinentes;
- utilisé les méthodes statistiques adéquates.

On saurait alors:

- si les précisions désirées sur les caractéristiques statistiques sont atteintes par le réseau et comment il faut, en conséquence, modifier les fréquences de mesure;
- s'il est possible de réduire la densité du réseau: pour quels paramètres de qualité et à quelles stations.

En bref, on saurait si le réseau atteint les objectifs précis qu'on lui a fixés dans la connaissance de la qualité de l'eau pour l'alimentation et la vie aquatique et, s'il ne les atteint pas, comment le modifier.

C'est l'application de cette démarche rationnelle que nous allons mettre en oeuvre au chapitre suivant, dans le cadre de l'étude de la qualité de l'eau d'alimentation, à certaines stations spécialement affectées du bassin.

CHAPITRE 5

ETUDE D'UN OBJECTIF: ALIMENTATION EN EAU POTABLE

Ce chapitre tente d'illustrer les principes méthodologiques de rationalisation exposés dans le premier chapitre.

On définit de façon précise certains objectifs que le réseau doit atteindre; on en déduit le type d'information nécessaire ainsi que la précision requise sur cette information pour évaluer la qualité de l'eau pour l'objectif défini. Puis on étudie les performances du réseau. Si le réseau n'atteint pas les objectifs qui lui sont fixés, on propose les modifications nécessaires.

Il ne s'agit pas, dans ce chapitre, de résoudre l'ensemble des problèmes de potabilité de l'eau sur le bassin de la rivière Yamaska mais simplement d'illustrer une démarche méthodologique.

5.1 Définition des objectifs

5.1.1 Choix des objectifs

Des villes importantes, en pleine expansion démographique, sont situées sur le bassin versant de la rivière Yamaska. Leur approvisionnement en eau potable est un objectif primordial. Pour évaluer la qualité de l'eau du bassin, en vue de cette utilisation, la connaissance d'un grand nombre de paramètres physico-chimiques et bactériologiques est requise. Nous nous limitons ici à l'étude des deux seuls paramètres, pertinents pour cet usage de l'eau, dont nous disposons: le fer et le manganèse. Les résultats préliminaires obtenus

au chapitre 4 prouvent la présence excessive de fer et de manganèse à bon nombre de stations du bassin et montrent l'importance de l'étude de ces paramètres. Le fer et le manganèse donnent, quand ils sont en trop grande quantité, un goût métallique désagréable à l'eau de boisson et provoquent une certaine turbidité. Ils sont gênants au niveau de la cuisson des aliments, du lavage, des appareils de filtration et des systèmes de distribution.

Pour situer l'ampleur du problème sur le bassin de la rivière Yamaska signalons que les valeurs extrêmes observées aux dix-neuf stations du réseau dépassent toutes la norme prescrite de 0.05 mg/l en ce qui concerne le manganèse. On observe des concentrations maximales de .52, .50 et .67 mg Mn/l aux stations 12, 15 et 16 de la rivière Yamaska-Nord. A ces trois stations les concentrations moyennes dépassent la norme (respectivement .21, .17 et .23 mg Mn/l). La norme de .3 mg/l prescrite sur le fer est dépassée à sept stations du réseau. Aux trois stations 12, 15 et 16 on observe des concentrations maximales de 0.50, 0.55 et 0.32 mg Fe/l. La rivière Yamaska-Nord semble la plus touchée, dans l'ensemble du bassin, par les problèmes de fer et de manganèse. De plus la ville de Granby, située sur son cours, est le plus grand consommateur d'eau de tout le bassin de la Yamaska. La rivière Saint-Nazaire, affluent de la rivière Noire dans une zone agricole, supporte aussi des concentrations élevées de fer. Nous localisons donc l'étude sur ces deux sous-bassins:

- . rivière Yamaska-Nord
- . rivière Saint-Nazaire.

5.1.2 Définition des caractéristiques statistiques d'intérêt

Dans le cadre de l'objectif général d'étude "eau potable" nous avons fait le choix des paramètres de qualité fer et manganèse (cf. 5.1.1). Pour quantifier le problème dans l'esprit du chapitre 1 il faut déterminer les carac-

téristiques statistiques d'intérêt, les précisions sur les caractéristiques et le niveau de confiance requis (cf. 1.3.1). Ce choix est ici arbitraire étant donné le caractère méthodologique de l'étude. Il devrait, de préférence, reposer sur des considérations pratiques reliées à l'usage considéré.

Les caractéristiques statistiques d'étude sont: moyennes, quantiles et probabilités; les précisions requises sont fixées à:

- . 20% sur les moyennes
- . 25% sur le quantile 90

Les critères de qualité à respecter sont les normes canadiennes pour l'eau d'alimentation soit: 0.3 mg/l pour le fer et 0.05 mg/l pour le manganèse. La localisation de l'étude est: - stations 12, 15, 16 rivière Yamaska-Nord; - station 18 rivière Saint-Nazaire.

Les renseignements précédents suffisent à définir de façon précise les objectifs du réseau. Pour mettre en application la méthodologie décrite au premier chapitre, l'étude de l'utilisation eau potable est effectuée en considérant les objectifs concrets suivants:

- . connaissance de la moyenne des concentrations en fer et en manganèse aux stations des rivières Yamaska-Nord et Saint-Nazaire avec une précision de 20% au niveau de confiance 95%;
- . connaissance de la probabilité de dépassement des critères de qualité à ces stations pour l'objectif visé;
- . connaissance du quantile 90 des distributions de fer et de manganèse à ces stations avec une précision de 25% au niveau de confiance 95%.

5.2 Réponses du réseau actuel

Dans cette section, nous étudions successivement les résultats obtenus sur la moyenne, sur le quantile 90 et sur la probabilité de dépassement des critères de qualité aux stations 12, 15, 16 et 18 pour les paramètres fer et manganèse. On cherche aussi les précisions obtenues sur la moyenne et le quantile 90 à l'aide des données acquises. L'étude du quantile 90 et de la précision atteinte sur ce quantile nécessite l'ajustement de lois statistiques aux séries d'observations et la détermination des intervalles de confiance au niveau requis (cf. 5.2.2).

5.2.1 Connaissance de la moyenne

D'après les tableaux 4.1 et 4.6, on déduit le tableau suivant:

Station	Fer		Manganèse	
	moyenne (mg/l)	Précision (%) niveau 95%	moyenne (mg/l)	Précision (%) niveau 95%
12	.25	24	.21	30
15	.22	33	.17	32
16	.15	23	.22	37
18	.21	25	.08	22

On remarque que la précision souhaitée de 20% sur la moyenne, au niveau de confiance 95%, n'est pas atteinte. Il faudra donc augmenter la fréquence de mesure (cf. 5.3.1).

5.2.2 Ajustements des lois statistiques

Le programme de calcul utilisé effectue l'ajustement de plusieurs lois statistiques asymétriques à chaque série d'observations étudiée (cf. 4.4.3.2).

Un choix de lois est, en effet, à priori très difficile à réaliser sur une base théorique. De plus, l'ajustement de lois est encore peu employé en qualité de l'eau et il n'y a pas de précédent sur lequel se baser. On a donc réalisé les ajustements suivants (cf. 2.2.4):

- loi Gamma, méthode du maximum de vraisemblance;
- loi Pearson type III, méthode des moments (correction d'asymétrie $(C_g)_3$, cf. 2.1.4);
- loi Log-gamma, méthode des moments appliquée à la série des valeurs observées;
- loi Log-Pearson type III, méthode des moments appliquée à la série des valeurs observées.

On donne un exemple des résultats obtenus dans le cas du manganèse à la station 12 (tableaux 5.1 à 5.6). On a tracé les lois théoriques ajustées et les observations avec la probabilité empirique $P_k = \frac{k}{N+1}$ dans les cas suivants:

- manganèse stations 12, 15, 16 (figures 5.1 à 5.6)
- fer stations 12, 18 (figures 5.7 à 5.10).

Notons que l'hypothèse d'indépendance des observations, qui est un préliminaire requis avant d'ajuster une loi, est acceptée dans tous les cas, sauf pour le fer à la station 15 (cf. test d'ANDERSON 2.1.5). Le choix des lois est effectué par une étude visuelle des graphiques; en effet les différents tests (Chi-carré KOLMOGOROV-SMIRNOV) n'étant pas assez discriminants, ils ne sont pas utilisés ici (cf. 2.2.4.2). La sélection retenue est résumée au tableau suivant:

TABLEAU 5.1: Valeurs classées et probabilités empiriques
(station 12, Manganèse).

La probabilité empirique au dépassement est: $P_k = \frac{k}{(N + 1)}$

Valeurs classées mg Mn/l	Probabilité empirique au non-dépassement
25.00	.04762
50.00	.09524
70.00	.14286
100.00	.19048
100.00	.23810
110.00	.28571
140.00	.33333
150.00	.38095
160.00	.42857
160.00	.47619
170.00	.52381
230.00	.57143
240.00	.61905
240.00	.66667
250.00	.71429
250.00	.76190
320.00	.80952
390.00	.85714
430.00	.90476
520.00	.95238

TABLEAU 5.2: Caractéristiques de l'échantillon. Test d'indépendance des observations (station 12, Manganèse).

Caractéristiques de l'échantillon
des valeurs observées

Taille	:	20
Moyenne	:	205.25
Ecart-type	:	130.32
Coeff. d'asymétrie	:	0.92
Coeff. de variation	:	0.63

Caractéristiques de l'échantillon
des logarithmes des valeurs observées

Taille	:	20
Moyenne	:	2.22
Ecart-type	:	0.32
Coeff. d'asymétrie	:	- 0.79
Coeff. de variation	:	0.15

- Résultat du test de Wald-Wolfowitz pour l'indépendance: $u = 0.216$
- On accepte l'hypothèse d'indépendance au niveau de signification 5%

TABLEAU 5.3: Ajustement de la loi Gamma par la méthode
du maximum de vraisemblance (station 12, Manganèse).

Valeurs des paramètres de la loi:

paramètre d'échelle: $\alpha = .0117$
paramètre de forme: $\lambda = 2.3953$

Caractéristiques de la population:

Moyenne: $\mu = 205.25$
Ecart-type: $\sigma = 132.62$
Coeff. asymétrie: $C = 1.29$
Coeff. variation: $C_v^s = 0.65$

Probabilité au dépassement	Evénement X_T	Ecart-type de X_T	Intervalle de Confiance					
			50%		80%		95%	
.001	861.	345.	628.	1094.	418.	1304.	184.	1538.
.005	701.	228.	548.	854.	409.	993.	255.	1147.
.010	630.	182.	508.	753.	398.	863.	275.	986.
.020	558.	139.	465.	652.	380.	737.	286.	831.
.050	460.	91.	399.	522.	344.	577.	282.	638.
.100	383.	63.	340.	425.	302.	464.	259.	507.
.150	336.	52.	300.	371.	269.	402.	234.	438.
.200	301.	47.	269.	332.	241.	361.	209.	392.
.500	177.	36.	153.	202.	131.	224.	106.	248.
.800	94.	22.	79.	109.	66.	122.	52.	136.
.850	80.	22.	65.	95.	52.	108.	37.	123.
.900	64.	28.	46.	83.	29.	100.	10.	118.
.950	45.	43.	16.	75.	-10.	101.	-40.	130.
.980	29.	66.	-15.	74.	-55.	113.	-99.	158.
.990	21.	81.	-34.	76.	-83.	125.	-138.	180.
.995	15.	96.	-49.	80.	-107.	138.	-172.	202.
.999	6.	123.	-77.	89.	-151.	164.	-235.	247.

TABLEAU 5.4: Ajustement de la loi Pearson type 3 par la méthode des moments avec la correction d'asymétrie.

$$(C_s)_3 = C_s \left[\left(1 + \frac{6.51}{N} + \frac{20.20}{N^2} \right) + \left(\frac{1.48}{N} + \frac{6.77}{N^2} \right) C_s^2 \right]$$

Valeurs des paramètres de la loi:

- . paramètre d'échelle: $\alpha = .0126$
- . paramètre de forme: $\lambda = 2.6892$
- . paramètre de position: $M = 8.4554$

Caractéristiques de la population:

- . Moyenne: $\mu = 205.25$
- . Ecart-type: $\sigma = 130.32$
- . Coeff. asymétrie: $C = 1.22$
- . Coeff. variation: $C_v^s = 0.63$

Probabilité au dépassement	Evénement X_T	Ecart-type de X_T	Intervalle de Confiance					
			50%		80%		95%	
.001	836.	322.	619.	1053.	423.	1249.	205.	1468.
.005	684.	214.	540.	829.	410.	959.	265.	1104.
.010	617.	171.	502.	733.	398.	837.	281.	953.
.020	549.	132.	459.	638.	379.	718.	290.	807.
.050	455.	87.	396.	513.	343.	566.	284.	625.
.100	380.	61.	339.	421.	302.	458.	260.	499.
.150	334.	50.	300.	368.	270.	399.	236.	433.
.200	300.	45.	270.	331.	243.	358.	212.	389.
.500	179.	35.	156.	203.	134.	224.	111.	248.
.800	95.	22.	81.	110.	67.	123.	52.	138.
.850	81.	22.	66.	95.	53.	109.	38.	124.
.900	64.	27.	47.	82.	30.	98.	13.	116.
.950	45.	41.	17.	72.	-8.	97.	-35.	124.
.980	27.	62.	-14.	69.	-52.	106.	-94.	148.
.990	18.	77.	-34.	70.	-80.	117.	-133.	169.
.995	11.	91.	-50.	73.	-105.	128.	-167.	190.
.999	1.	119.	-79.	81.	-152.	154.	-232.	234.

TABLEAU 5.5: Ajustement de la loi Log-Gamma par la méthode des moments
appliquée à la série des valeurs observées.

Valeurs des paramètres de la loi:

paramètre d'échelle: $\alpha = 41.2866$
paramètre de forme: $\lambda = 92.7788$

Caractéristiques de la population:

Moyenne: $\mu = 2.25$
Ecart-type: $\sigma = 0.23$
Coeff. asymétrie: $C = 0.21$
Coeff. variation: $C^s_v = 0.10$

Probabilité au dépassement	Evénement X_T	Ecart-type de Log (X_T)	Intervalle de Confiance					
			50%		80%		95%	
.001	1091.	.258	731.	1627.	510.	2334.	341.	3490.
.005	783.	.187	585.	1046.	451.	1359.	336.	1820.
.010	669.	.158	523.	855.	419.	1067.	328.	1366.
.020	565.	.131	461.	692.	384.	831.	313.	1020.
.050	441.	.099	378.	514.	329.	590.	282.	688.
.100	356.	.078	315.	402.	282.	448.	250.	507.
.150	308.	.069	277.	343.	251.	378.	226.	422.
.200	276.	.064	250.	305.	228.	334.	206.	369.
.500	173.	.057	159.	189.	147.	205.	134.	224.
.800	112.	.058	102.	122.	94.	133.	86.	145.
.850	101.	.060	92.	111.	85.	121.	77.	133.
.900	90.	.065	81.	99.	74.	109.	67.	121.
.950	75.	.078	67.	85.	60.	95.	53.	107.
.980	62.	.102	53.	73.	46.	84.	39.	98.
.990	55.	.122	45.	66.	38.	79.	32.	95.
.995	49.	.143	39.	61.	32.	75.	26.	94.
.999	39.	.196	29.	53.	22.	70.	16.	95.

TABIEAU 5.6: Ajustement de la loi Log-Pearson type 3 par la méthode des moments sur la série des valeurs observées.

Valeurs des paramètres de la loi:

paramètre d'échelle: $\alpha = -6.9744$
 paramètre de forme: $\lambda = 5.0993$
 paramètre de position: $M = 2.9441$

Caractéristiques de la population:

Moyenne: $\mu = 2.21$
 Ecart-type: $\sigma = 0.32$
 Coeff. asymétrie: $C_s = -0.89$
 Coeff. variation: $C_v = 0.15$

Probabilité au dépassement	Evénement X_T	Ecart-type de Log (X_T)	Intervalle de Confiance					
			50%		80%		95%	
.001	680.	.270	447.	1033.	306.	1507.	201.	2296.
.005	606.	.196	447.	821.	340.	1081.	250.	1467.
.010	566.	.163	440.	729.	350.	915.	272.	1180.
.020	521.	.129	426.	637.	355.	763.	290.	934.
.050	449.	.089	392.	516.	346.	584.	301.	670.
.100	385.	.067	347.	427.	316.	469.	285.	520.
.150	342.	.062	311.	377.	285.	411.	259.	452.
.200	309.	.063	280.	341.	256.	372.	232.	411.
.500	182.	.083	160.	207.	143.	232.	125.	264.
.800	92.	.104	78.	108.	68.	125.	57.	147.
.850	77.	.116	64.	92.	54.	108.	45.	129.
.900	60.	.138	49.	75.	40.	91.	32.	112.
.950	41.	.190	30.	55.	23.	72.	17.	97.
.980	25.	.276	17.	39.	11.	57.	7.	88.
.990	18.	.348	11.	31.	6.	51.	4.	87.
.995	13.	.426	7.	25.	4.	46.	2.	89.
.999	6.	.622	2.	17.	1.	39.	0.	104.

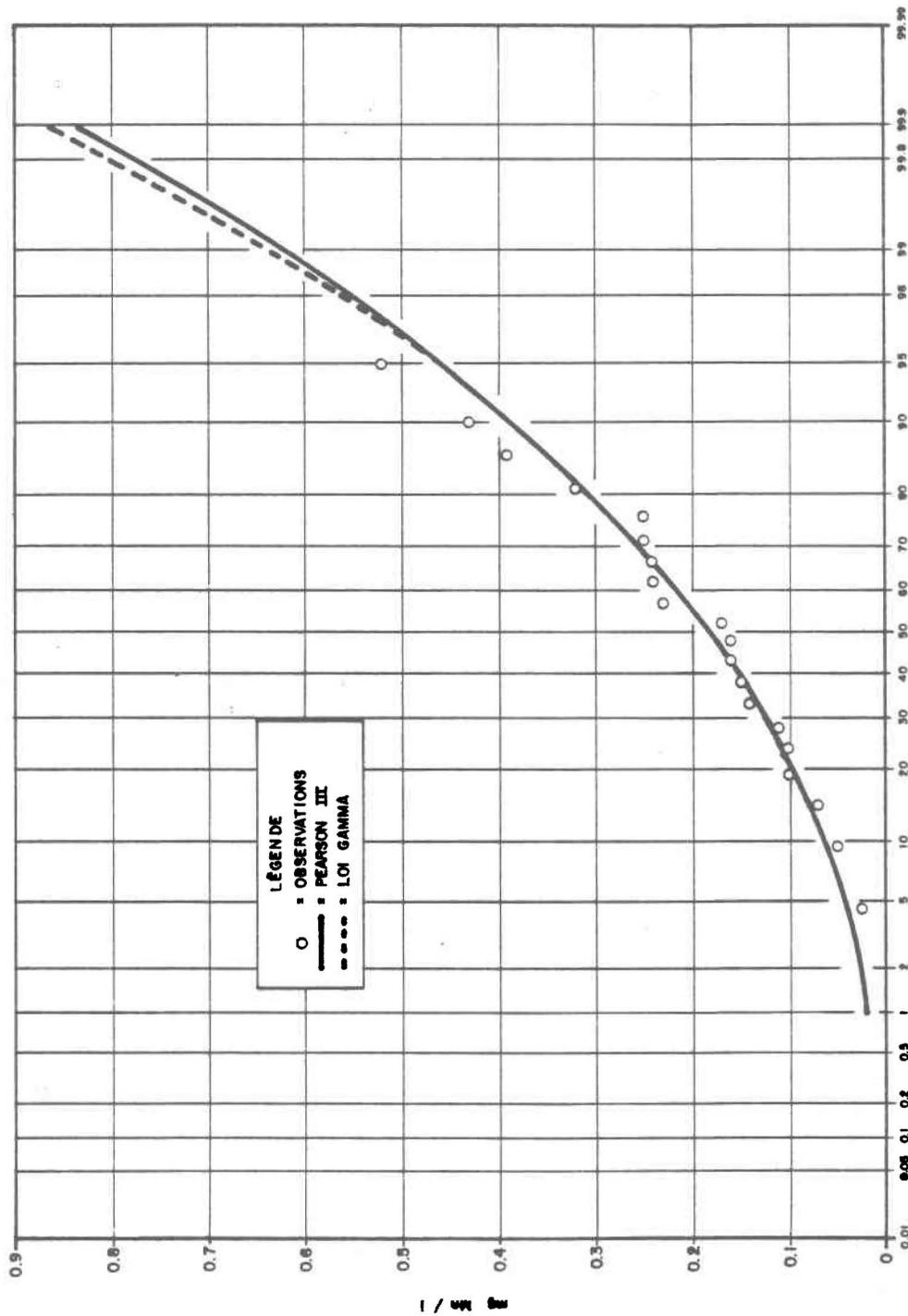


Figure 5.1. Ajustement, lois gamma et Pearson type III (Manganèse, station 12).

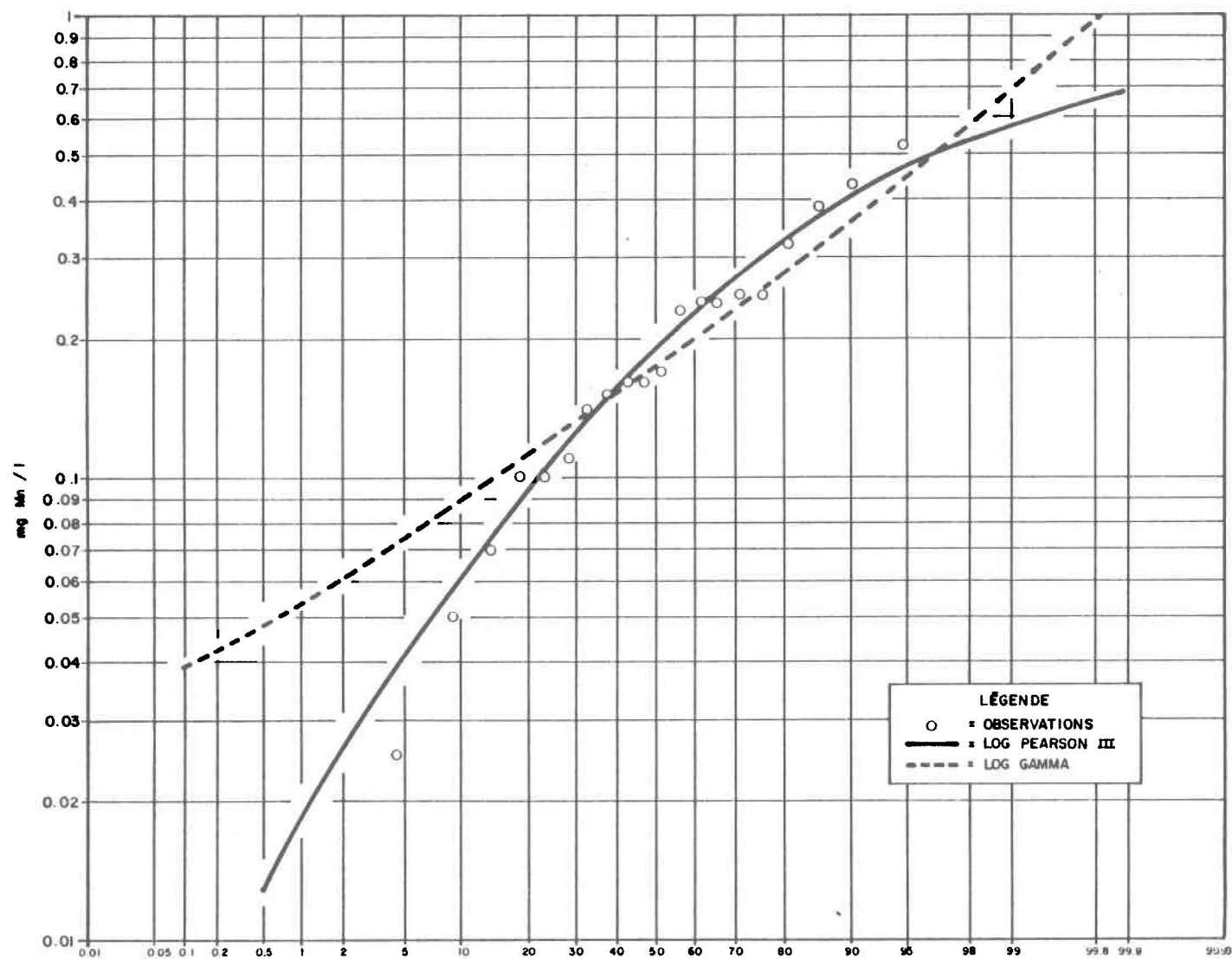


Figure 5.2. Ajustement : lois log gamma et log Pearson III (Manganèse station I2)

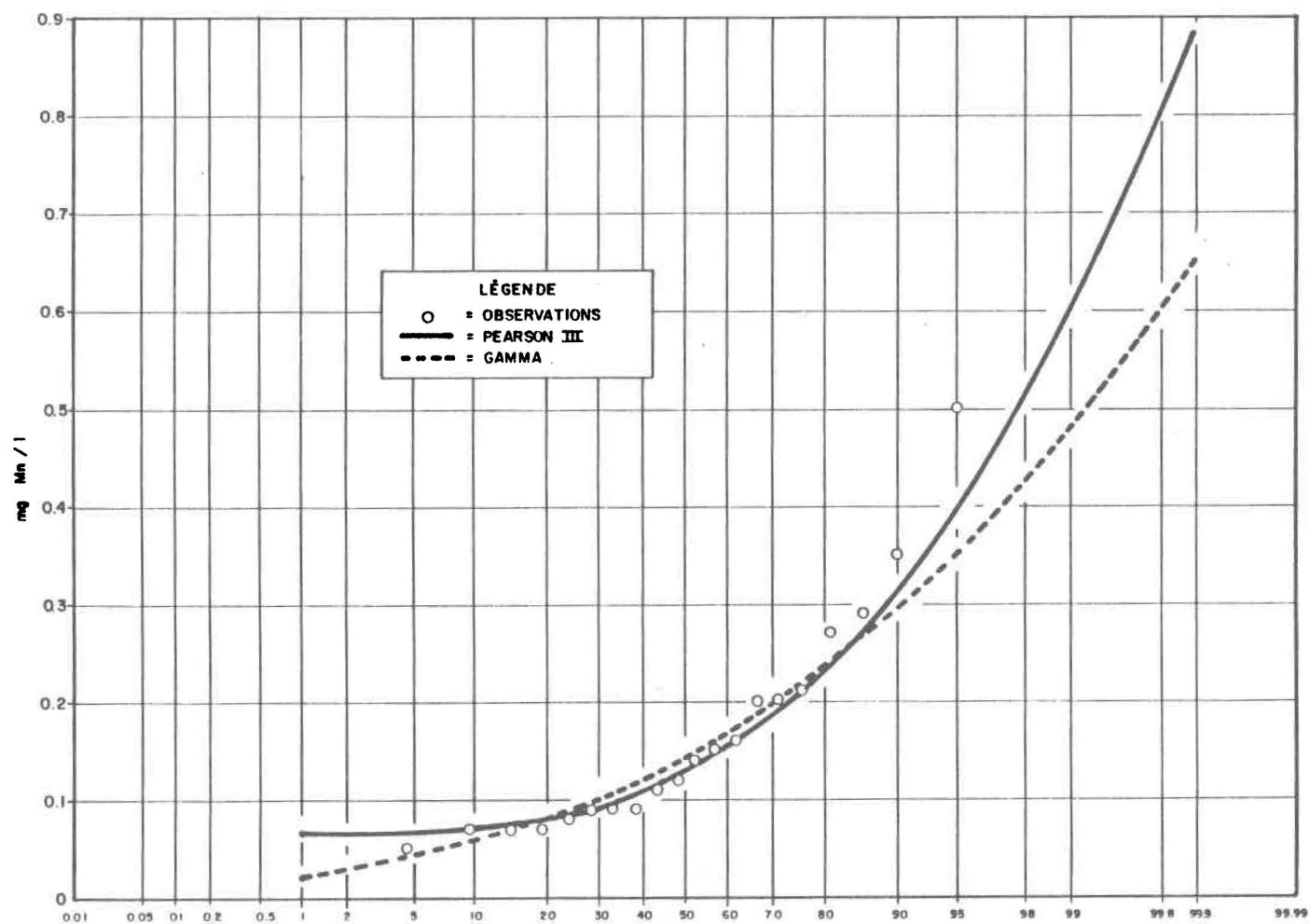


Figure 5.3. Ajustement lois gamma et Pearson type III (Manganèse ,station 15).

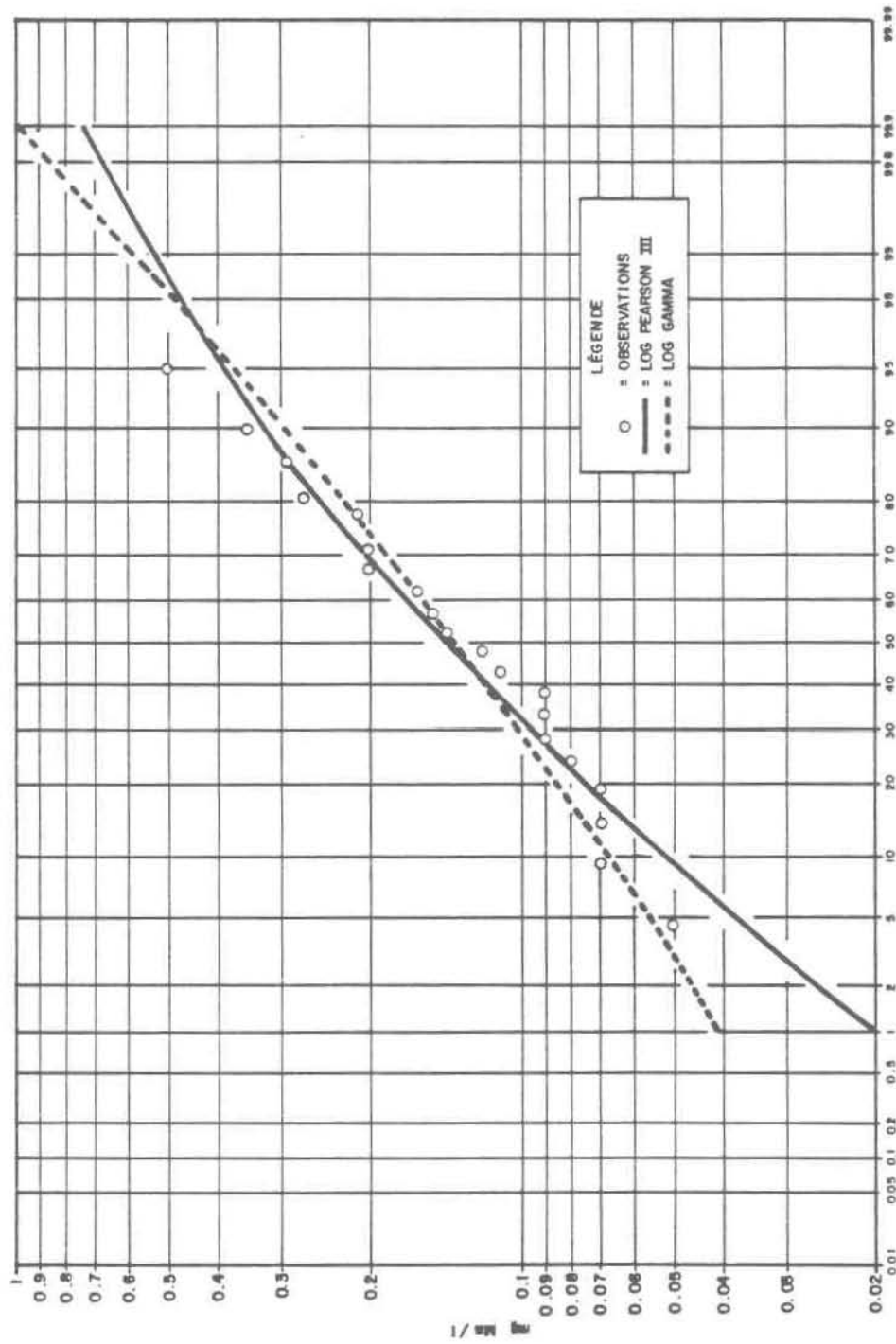


Figure 5.4 . Ajustement : lois log gamma et log Pearson III (Manganèse , station 15) .

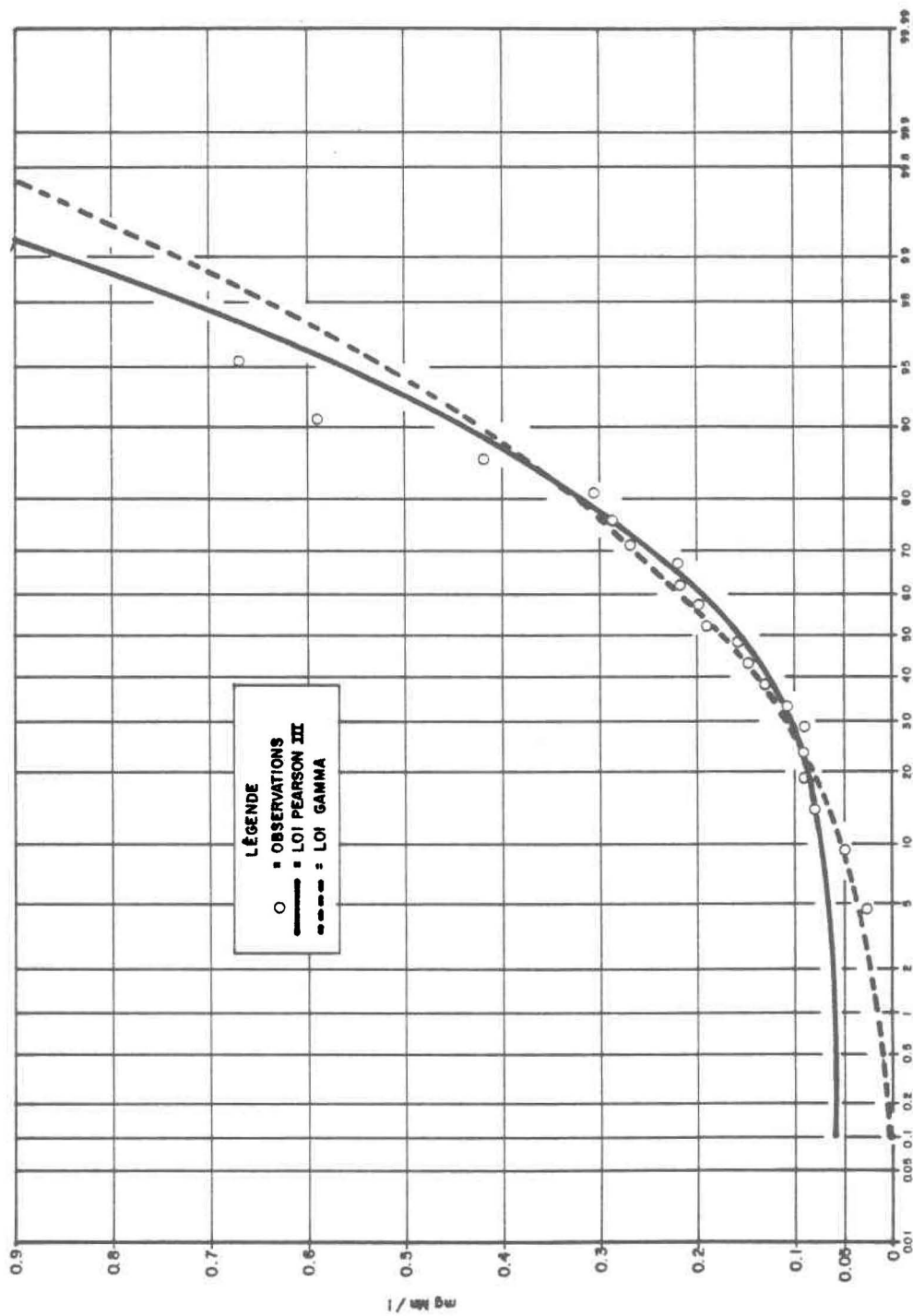


Figure 5.5. Ajustement : lois gamma et Pearson type III (Manganèse , station 16).

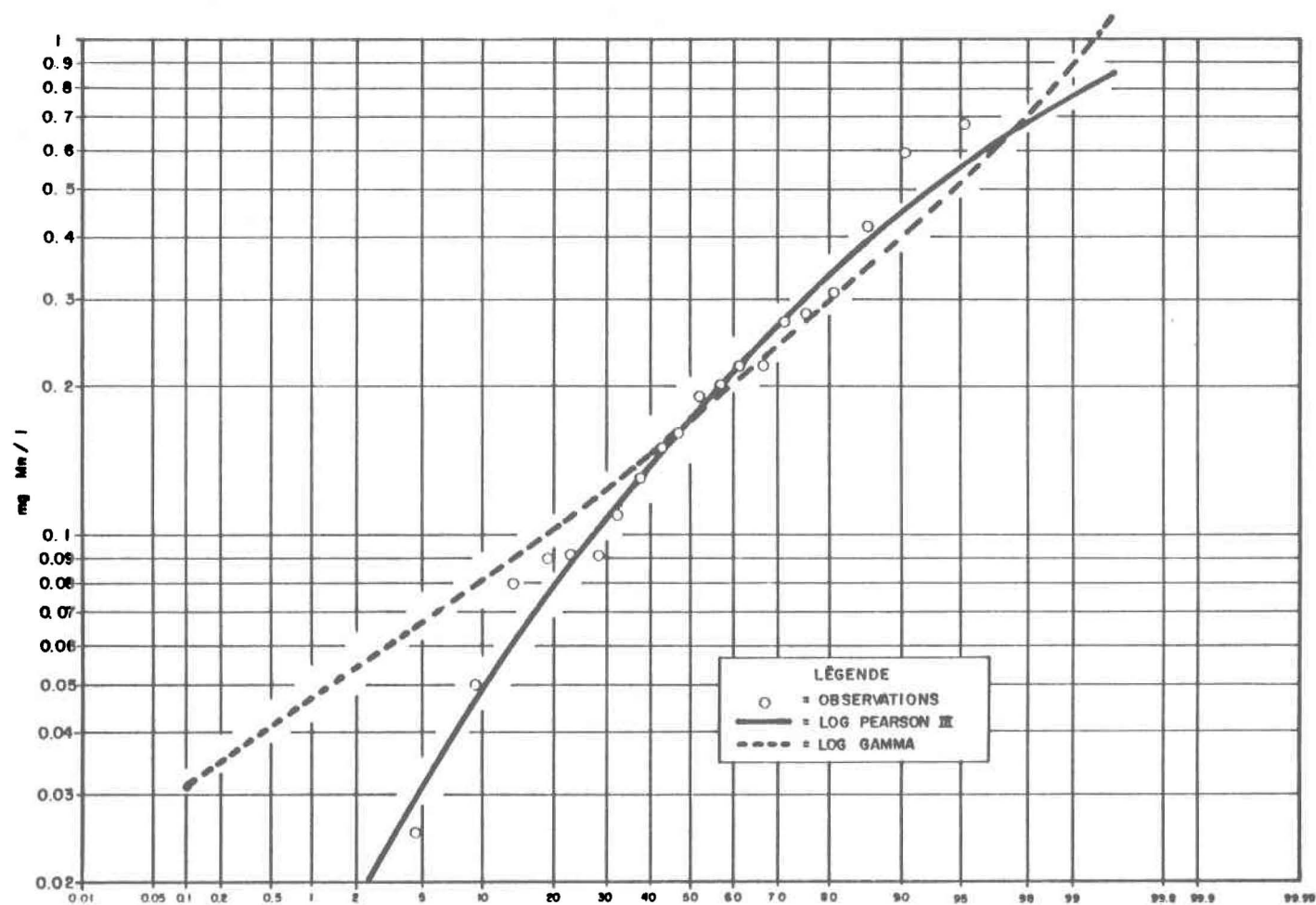


Figure 5.6 . Ajustement : lois log gamma et log-Pearson type III (Manganèse, station 16) .

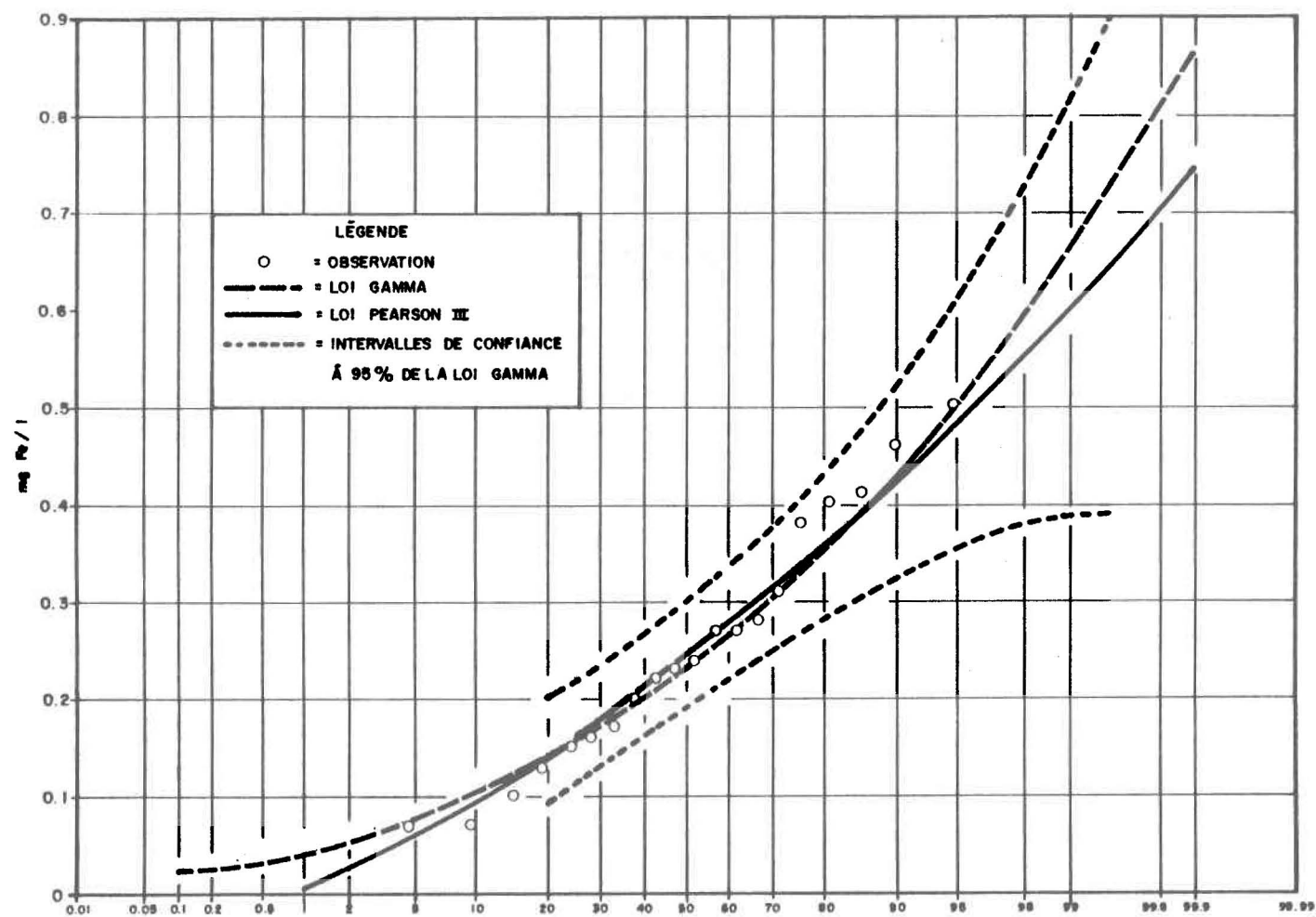


Figure 5.7. Ajustement : loi gamma et Pearson type III (Fer, station I2).

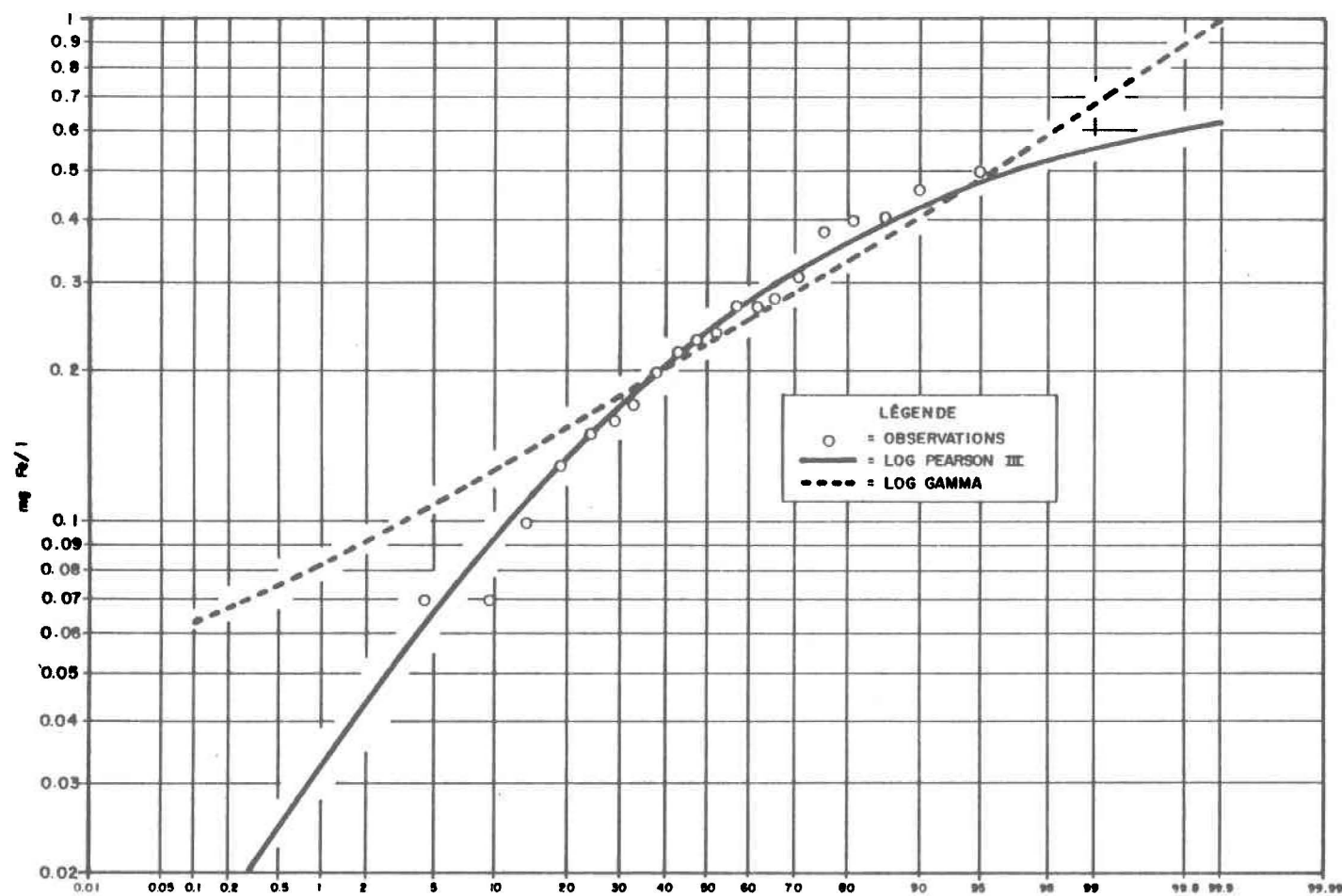


Figure 5.8 . Ajustement , lois log gamma et log Pearson III (Fer , station 12) .

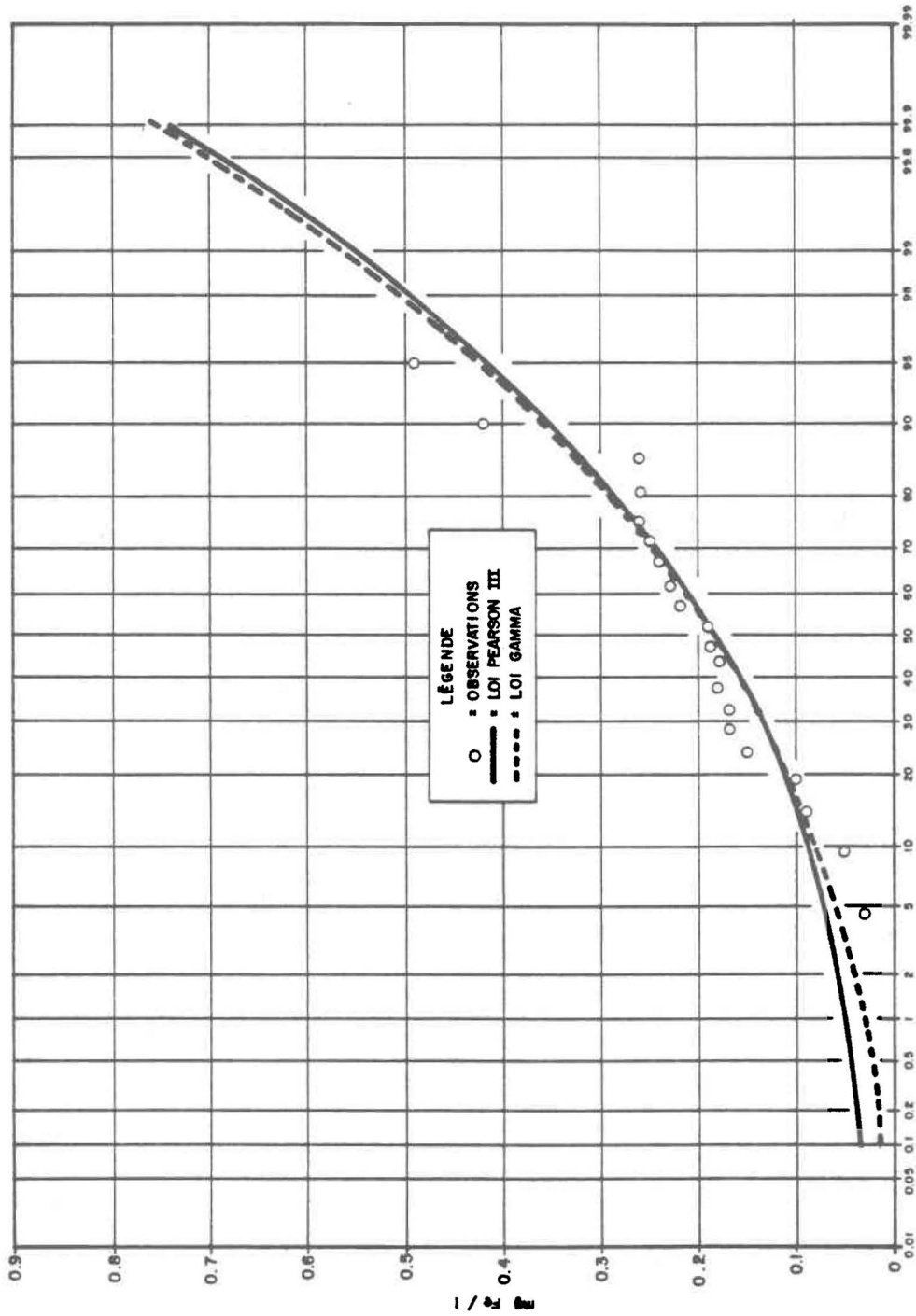


Figure 5.9. Ajustement : lois gamma et Pearson III (Fer , station 18)

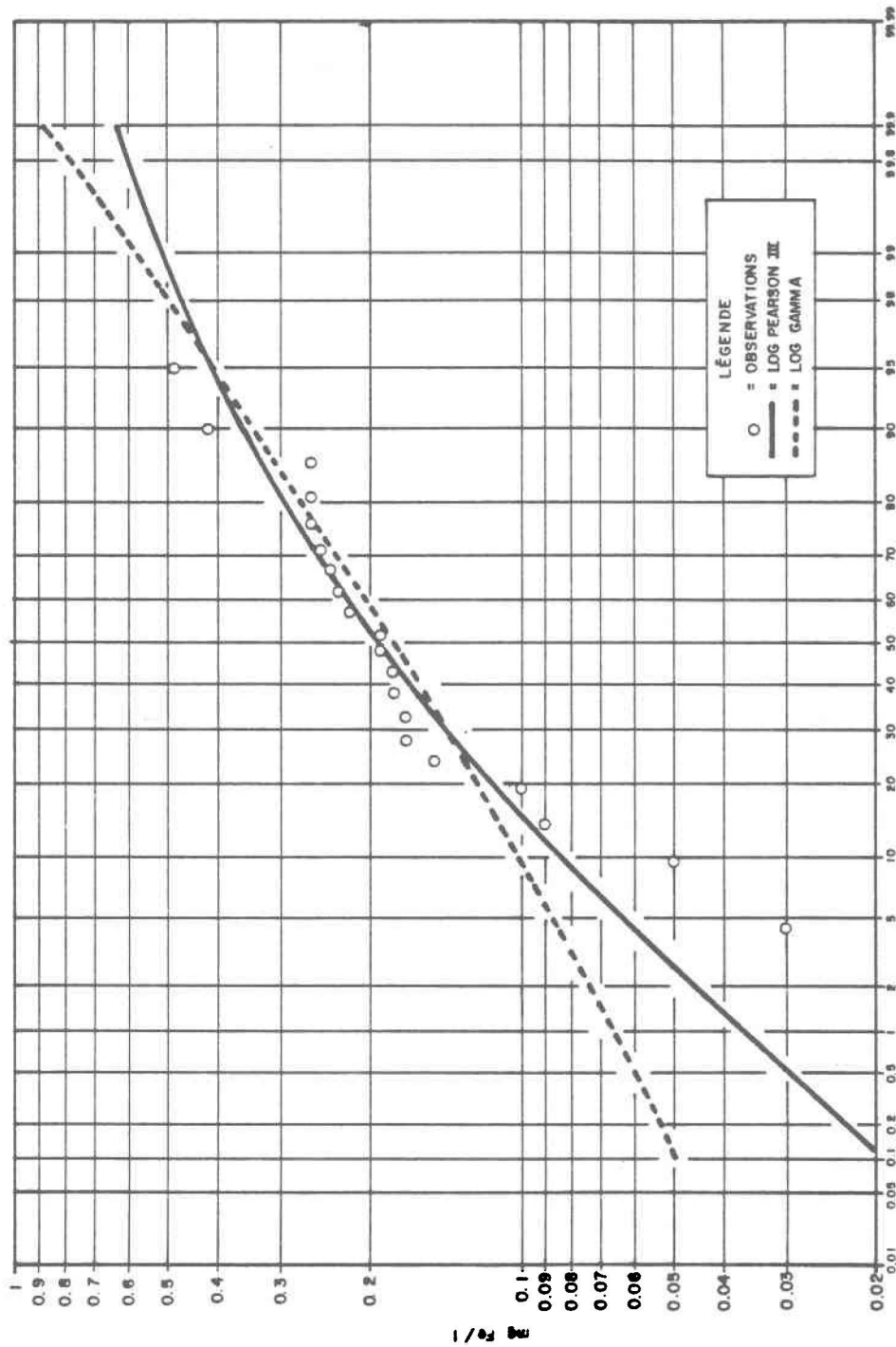


Figure 5. 10. Ajustement, lois log gamma et log Pearson type III (Fer , station 18) .

Paramètre	Station	Loi	Caractéristiques de la loi		
			m	α	λ
Manganèse	12	Gamma	0	0.0117	2.3953
	15	Pearson 3	66.74	0.0076	0.7475
	16	Log-Pearson 3	3.2572	-7.1070	7.5502
Fer	12	Log-Pearson 3	2.8295	-7.0946	3.5214
	18	Gamma	0	0.0155	3.2104

5.2.3 Intervalles de confiance

On dispose de deux méthodes pour déterminer les intervalles de confiance (cf. 2.2.4.4). La méthode 1 est celle utilisée par le programme (cf. 2.2.4.4), la méthode 2 utilise les résultats établis sur l'intervalle de confiance d'un quantile (cf. 2.1.3.1). Dans la méthode 1 l'intervalle de confiance est calculé d'après l'équation 2.49; la méthode 2 utilise l'équation 2.15.

On a comparé les intervalles de confiance obtenus par ces deux méthodes au niveau de confiance 95% pour le manganèse à la station 12 (cf. tableau 5.7). La loi ajustée est une loi gamma. Le tableau 5.7 appelle deux remarques:

- l'intervalle de confiance obtenu par la méthode 2 est légèrement plus large que celui obtenu par la méthode 1 mais cette différence est faible (de l'ordre de 4 à 7%). On pourrait donc, éventuellement, envisager de généraliser le calcul d'intervalles de confiance par la méthode 2 (cf. 2.2.4.4).
- Quelle que soit la méthode de calcul, les intervalles de confiance calculés sont relativement larges. C'est la petite taille des échantillons ($N = 20$) qui en est la cause.

TABLEAU 5.7: Intervalles de confiance de la loi Gamma à 95%
Manganèse, station 12 (méthodes 1 et 2)

Loi Gamma $\alpha = 0.0117$ $\lambda = 2.3953$

Probabilité au non dépassement	Quantile µg Mn/l ou ppb	Intervalles de confiance à 95%			
		Méthode 1		Méthode 2	
.10	64	13	116	20	108
.20	95	52	138	46	143
.50	117	106	248	110	244
.80	301	209	392	193	409
.85	336	234	438	211	461
.90	383	259	507	232	534
.95	460	282	638	251	669
.98	558	286	831	234	882
.99	630	275	986	180	1080

La fonction densité de probabilité de la loi Gamma est:

$$f(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)} e^{-\alpha x} x^{\lambda-1}$$

$$\Gamma(\lambda) \approx \lambda^\lambda e^{-\lambda} \sqrt{\frac{2\pi}{\lambda}} \left[1 + \frac{1}{12\lambda} + \dots \right]$$

$$\alpha^\lambda = 0.0000236$$

$$\Gamma(\lambda) = 1.2369$$

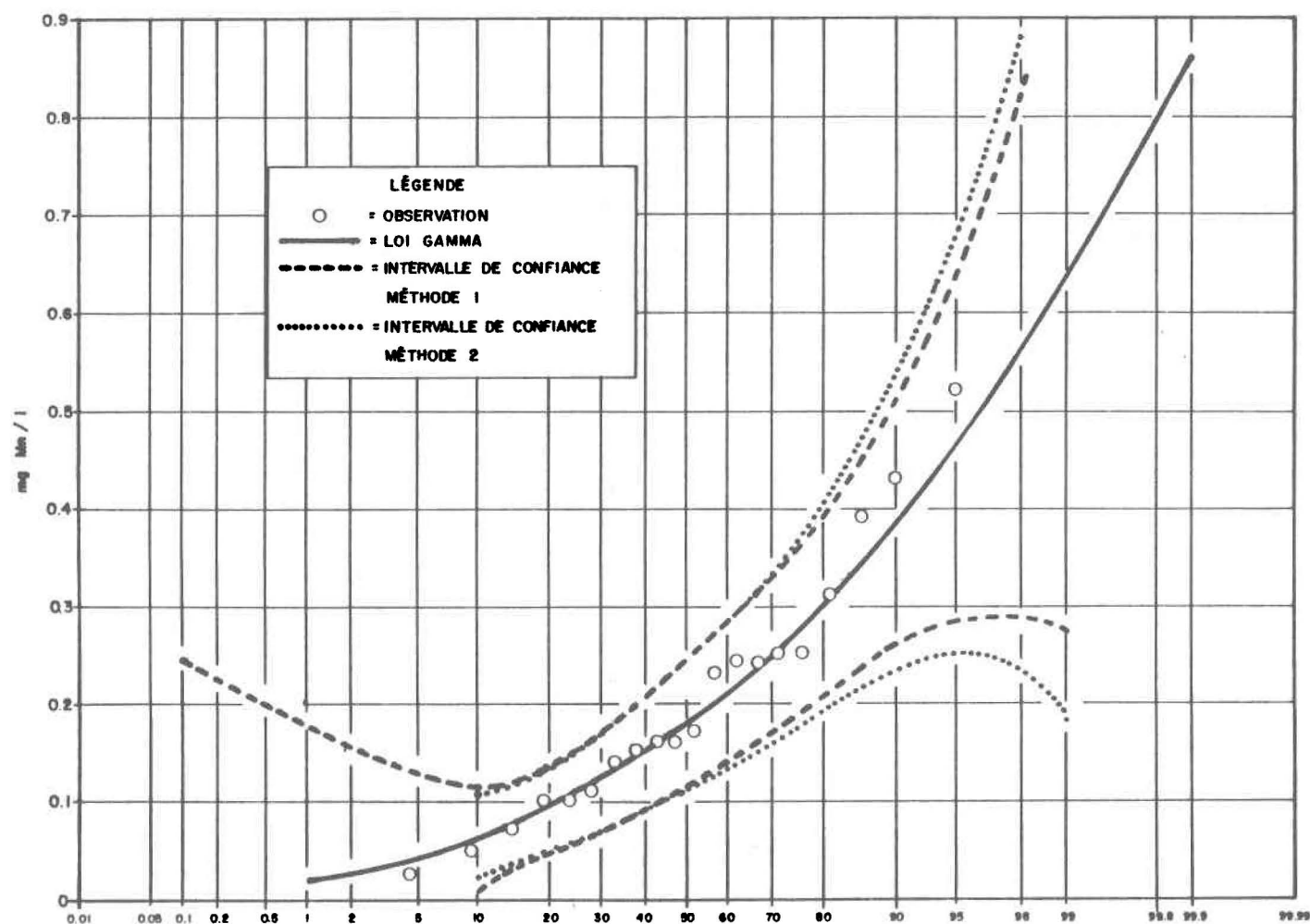


Figure 5. 11. Comparaison des intervalles de confiance de la loi gamma (Manganèse , station 12) méthode 1 et 2 .

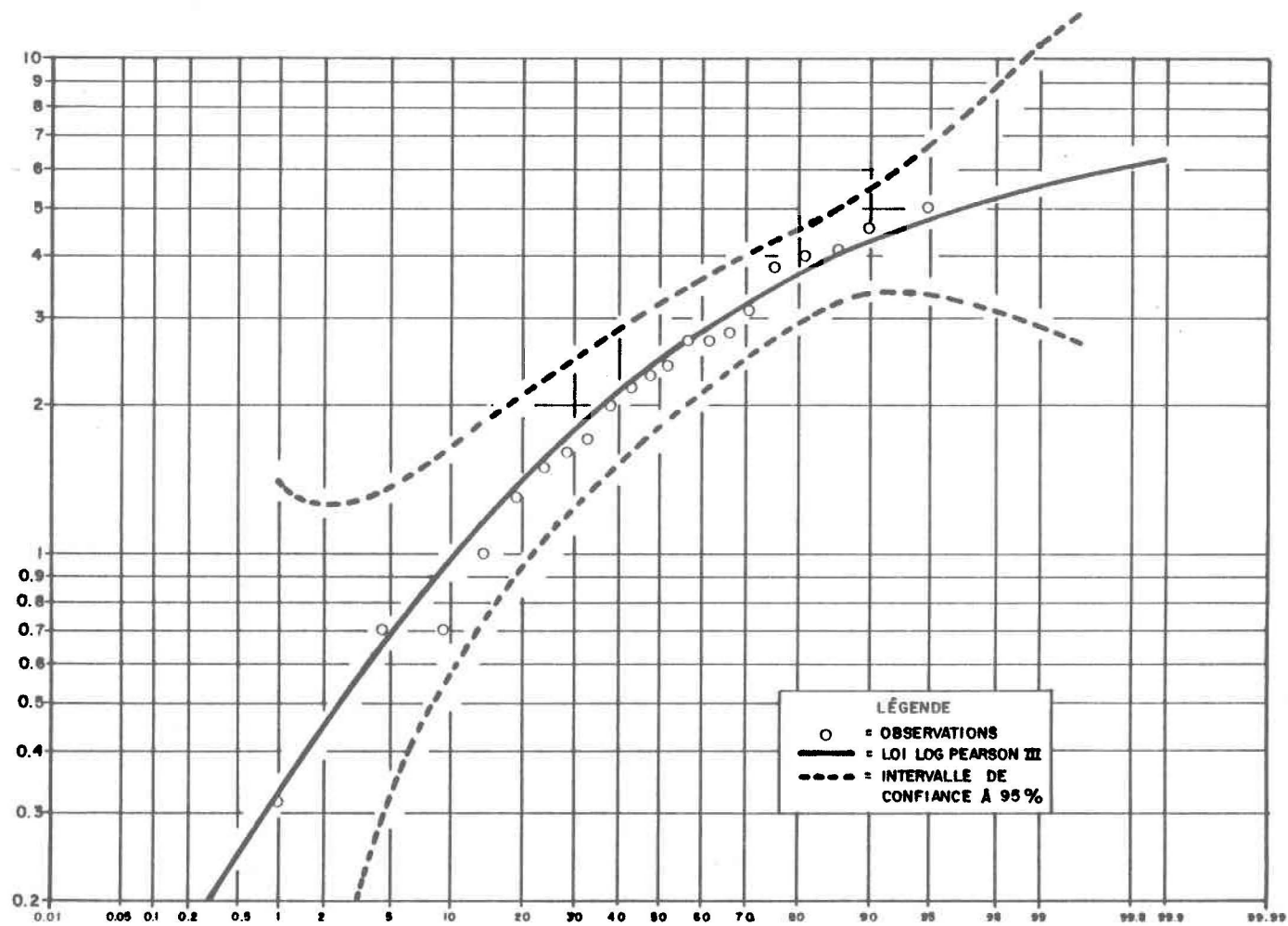


Figure 5.12. Intervalles de confiance à 95 % de la loi log Pearson type III (Fer, station 12).

On a tracé dans le cas du manganèse, station 12, les intervalles de confiance à 95% de la loi gamma par les deux méthodes d'après le tableau 5.7 (figure 5.11). On donne un autre exemple d'intervalle de confiance à 95% dans le cas d'une loi log-Pearson type III pour le fer à la station 12 (calcul par la méthode 1) (figure 5.12).

La connaissance de ces intervalles de confiance au niveau 95% va nous permettre (cf. 5.2.5) d'évaluer les précisions atteintes sur le quantile 90 étudié au niveau 95% (ou sur tout autre quantile désiré).

Avec le tableau 5.7 on peut par exemple affirmer que la concentration de manganèse à la station 12 qui est dépassée 10% des fois (probabilité au dépassement 0.10) sera comprise dans l'intervalle 0.26, 0.51 mg Mn/l, au niveau de confiance 95%, c'est-à-dire dans 95% des cas. On peut ainsi affirmer que le quantile 90 dépasse largement la norme de qualité prescrite (0.05 mg Mn/l).

5.2.4 Probabilité de dépassement de critère de qualité

A partir des courbes (figures 5.1 à 5.10) représentatives des lois statistiques choisies (cf. 5.2.2) on peut déterminer la probabilité de dépassement des critères de qualité imposés pour le manganèse (stations 12, 15, 16) et le fer (stations 12, 18). On obtient le tableau 5.8.

Paramètre	Station	Critère de qualité (eau de boisson)	Probabilité de dépassement (%)
Manganèse	12	.05 mg Mn/l	95
	15		93
	16		90
Fer	12	.3 mg Fe/l	34
	18		18

TABLEAU 5.8: Probabilités de dépassement de critères de qualité
(manganèse, stations 12, 15, 16; fer, stations 12, 18)

On remarque sur ce tableau que la probabilité de dépassement de la norme canadienne pour l'eau de boisson est extrêmement élevée dans le cas du manganèse (supérieure ou égale à 0.90) à toutes les stations de la rivière Yamaska-Nord. Il nous est possible d'affirmer ici, qu'en référence à la norme canadienne sur l'eau de boisson, de sérieux problèmes sont à craindre en l'absence de traitement, en raison d'une présence excessive de manganèse, pour l'eau d'alimentation de la rivière Yamaska-Nord. Les critères relatifs au fer sont aussi dépassés: les probabilités de dépassement sont assez élevées, surtout sur la Yamaska-Nord (station 12).

L'origine du fer et du manganèse peut être géologique ou industrielle. Il semble que dans le cas du bassin de la rivière Yamaska, ces deux éléments aient une origine géologique. Cette hypothèse est confirmée par deux faits:

- . Une étude de l'INRS-Eau (1976, vol. 1) sur les débits massiques a montré une croissance de ces débits au printemps (époque de crue) et une décroissance en été et en hiver (époques d'étiage), ce qui dénote une origine diffuse, donc possiblement géologique, et non ponctuelle, comme ce serait, par exemple, le cas d'un rejet d'usine.
- . Les géologues qui ont étudié la région des cantons de l'Est confirment la présence de pyrite dans les roches consolidées de tous les cantons de l'Est. On note d'ailleurs la présence de schistes gris ferrugineux dans la région de Granby. On peut citer aussi la présence de pierres de construction de matière magnésienne siliceuse et ferrugineuse dans des carrières situées entre St-Hyacinthe et St-Dominique (cf. CLARK, rapport géologique No. 101, Comtés de Bagot, St-Hyacinthe, Shefford).

Le manganèse est souvent associé au fer dans la pyrite. Après oxydation puis désagrégation et altération des roches une partie de ces éléments se retrouve dans les cours d'eau. Pour remédier à ces problèmes il faudra mettre en place les procédés adéquats de déferrisation et de démantanisation dans les stations de traitement de l'eau d'alimentation.

5.2.5 Précision atteinte sur le quantile 90 au niveau de confiance 95%

La précision sur le quantile 90 (probabilité au dépassement 0.1) peut être obtenue par l'une des deux méthodes décrites précédemment (cf. 5.2.3). Dans le cas de la méthode 2, d'après les équations 2.17 et 2.19, on a calculé la précision ϵ sur le quantile 90, pour les cas envisagés (manganèse, 12, 15, 16 et fer 12, 18), par la formule:

$$\epsilon = 13.15 \times 10^{-2} \frac{1}{f(Q_{90})} \cdot \frac{1}{Q_{90}} \quad (5.1)$$

Q_{90} est déduit, pour les lois choisies (cf. 2.2.5), des courbes des figures 5.1 à 5.10.

Paramètre	station	Précision relative sur le quantile 90 au niveau de confiance 95%	
		Méthode 1 (%)	Méthode 2 (%)
Manganèse	12	32.3	39.5
	15	43.0	37.4
	16	39.0	45.8
Fer	12	24.2	24.5
	18	28.5	34.5

TABEAU 5.9: Précision sur le quantile 90 au niveau de confiance (95%)
(manganèse, stations 12, 15, 16; fer, stations 12, 18).

On remarque qu'il y a peu de différence entre les résultats donnés par les deux méthodes. Il semble que la méthode 2 donne des valeurs plus élevées que la méthode 1 (sauf pour le manganèse à la station 15).

La précision désirée dans l'objectif du réseau (25%) n'est atteinte que dans le cas du fer à la station 12. Il faudra donc modifier la fréquence de mesures en conséquence dans les autres cas (cf. 5.3.2).

5.3 Correction du réseau pour atteindre les objectifs visés

Il s'agit de corriger les fréquences de mesure aux stations où les précisions sur la moyenne ou le quantile 90 sont insuffisantes.

5.3.1 Correction de la fréquence de mesure: cas de la moyenne

On calcule le nombre de mesures indépendantes nécessaires à l'obtention d'une précision de 20% au niveau de confiance 95% par la formule 2.10 (cf 2.1.1.2). On obtient le tableau 5.10.

Paramètre	Station	Nombre de mesures
Fer	12	28
	15	47
	16	24
	18	29
Manganèse	12	27
	15	40
	16	43
	18	60

TABLEAU 5.10: Nombre de mesures indépendantes nécessaires à l'obtention d'une précision de 20% au niveau de confiance 95% sur la moyenne des paramètres fer et manganèse aux stations 12, 15, 16 et 18.

Ces nombres de mesures peuvent se traduire en fréquences de mesures. Ainsi, pour le fer à la station 15, 47 observations peuvent se répartir en quatre ans d'observations mensuelles. On pourrait aussi envisager deux ans d'observations bimensuelles mais il faudrait alors s'assurer que ces observations sont indépendantes.

5.3.2 Correction de la fréquence de mesure: cas du quantile 90

On calcule le nombre de mesures nécessaires à l'obtention d'une précision de 25% au niveau de confiance 95% sur le quantile 90 au moyen de la formule 2.21 (cf. 2.1.3.2). Dans notre cas:

$$N_1 = 20$$

$$\epsilon_2 = 0.25$$

$$N_2 = 20 \left(\frac{\epsilon_1}{0.25} \right)^2 \quad (5.2)$$

On donne au tableau 5.11, le nombre N_2 obtenu à partir du tableau 5.9 (méthode 2) et de l'équation 5.2.

Paramètre	Station	Nombre de mesures
Manganèse	12	50
	15	45
	16	67
Fer	12	20
	18	37

TABLEAU 5.11: Nombre de mesures indépendantes nécessaires à l'obtention d'une précision de 25% sur le quantile 90, au niveau de confiance 95% (fer station 12, 18; manganèse stations 12, 15, 16).

Il faudra augmenter de façon importante le nombre de mesures relatives au manganèse pour atteindre la précision désirée. On pourrait envisager pour le manganèse à la station 15 une fréquence d'observations mensuelles durant une période de quatre ans. Sur deux années une fréquence bimensuelle serait nécessaire mais il faudrait s'assurer de l'indépendance des observations.

5.4 Conclusions

Deux types de conclusions sont à tirer de ce chapitre:

- d'un point de vue méthodologique
 - . on a montré qu'une définition précise des objectifs auxquels le réseau doit répondre permet d'évaluer clairement les performances de ce réseau, c'est-à-dire ici les précisions atteintes sur les moyennes et les quantiles étudiés, et d'envisager les modifications à apporter quand ces précisions sont insuffisantes. On a ainsi donné un exemple d'application de la méthodologie développée au premier chapitre dans le cas particulier de l'alimentation en eau potable.
 - . Pour obtenir les réponses désirées sur la qualité de l'eau d'alimentation il a fallu mettre en oeuvre certains types de méthodes statistiques comme l'ajustement de lois et le calcul des intervalles de confiance de ces lois. On a montré ainsi qu'il était tout à fait concevable d'utiliser pour les études de qualité de l'eau l'ajustement de lois statistiques qui permettent de donner des réponses rigoureuses aux problèmes de dépassements de critères, de calcul de quantiles et de précision atteinte sur ces quantiles.

- d'un point de vue pratique:

- . La faible taille des échantillons étudiés ($N = 20$) est cause des ajustements peu précis de lois, des larges intervalles de confiance et des précisions assez médiocres rencontrées. Mais on s'est surtout attaché ici à illustrer la démarche méthodologique.
- . On a, cependant, démontré de façon rigoureuse la présence excessive de fer aux stations 12 et 18 et de manganèse aux stations 12, 15 et 16 de la rivière Yamaska-Nord avec une probabilité de dépassement, à ces trois stations, de la norme canadienne pour l'eau de boisson (0.05 mg Mn/l) supérieure ou égale à 0.90. Il serait intéressant dans des études futures d'associer la probabilité au dépassement d'un critère au risque économique correspondant comme cela se pratique en hydrologie (construction de barrages, etc...).

CONCLUSION

Ce travail confirme les problèmes déjà constatés d'un enrichissement excessif en substances nutritives des eaux du bassin et de la présence de fer et surtout de manganèse à des concentrations dépassant les normes canadiennes prescrites pour l'eau d'alimentation. Ces problèmes sont plus cruciaux sur la rivière Yamaska-Nord.

Nous avons envisagé diverses méthodes statistiques d'étude du réseau de mesure visant à sa réduction ou à son amélioration et nous avons constaté que des recommandations précises dans ce sens ne pouvaient intervenir qu'à la suite d'une définition précise des objectifs de ce réseau. C'est cet aspect méthodologique que ce travail a essentiellement cherché à illustrer.

Nous avons tenté de montrer d'une façon théorique, au cours du premier chapitre, que la définition d'objectifs précis permettait de concevoir une planification rationnelle de l'acquisition de données. Cette planification a pour but d'obtenir au moindre coût et avec le maximum d'efficacité la connaissance désirée sur la qualité de l'eau. L'analyse du réseau du bassin versant de la rivière Yamaska, après deux années de fonctionnement, prouve la nécessité d'une telle démarche rationnelle si l'on veut obtenir des résultats pertinents et si l'on veut évaluer correctement les performances du réseau. Au cours du dernier chapitre nous avons tenté, en guise d'exemple, de mettre en application cette démarche dans le cadre de l'alimentation en eau potable.

La difficulté de la planification d'un réseau de mesures provient de l'interaction d'un grand nombre de contraintes très diverses, relatives:

- . à la chimie, à la biologie, à la toxicologie de l'eau;
- . aux techniques de prélèvement, de conservation et d'analyse des échantillons;
- . à la connaissance de la géologie, de la physiographie, de l'hydrologie, de l'utilisation du territoire, de l'urbanisation et de l'industrialisation du bassin versant;
- . aux méthodes de traitement des données;
- . à la détermination des critères de qualité, etc...

Mais c'est, en dernière analyse, de la prise en compte de toutes ces contraintes au sein d'un plan rationnel d'acquisition de données, soumis à l'orientation d'objectifs précis, que dépend la connaissance que l'on veut acquérir sur la qualité de l'eau.

REFERENCES

- ANDERSON, R.L. (1941).
 "Distribution of the serial correlation coefficient". Ann. Math. Stat., 8(1): 1-13.
- BENZECRI, . (1973).
 "L'analyse des correspondances". T. 2, Dunod, Paris, 619 p.
- BOBEE, B. (1973).
 "Sample Error of T-Years events computed by fitting a Pearson type III distribution"., Wat. Res. Res., 9(5): 1264-1270.
- BOBEE, B. et G. MORIN (1973).
 "Détermination des intervalles de confiance de la loi Pearson type III par les statistiques d'ordre". Jour. Hydrol., 20: 137-153.
- BOBEE, B. (1975).
 "The Log-Pearson type 3 distribution and its application in hydrology". Wat. Res. Res., 11(5): 681-689.
- BOBEE, B. et R. ROBITAILLE (1975).
 "Correction of bias in the estimation of the coefficient of skewness". Wat. Res. Res., 11(6): 851-854.
- BOBEE, B. et R. ROBITAILLE (1977).
 "The use of the Pearson type III and Log-Pearson type III distributions revisited". Wat. Res. Res., 13(2): 427-443.
- BREMOND, R. et R. VUICHARD (1972).
 "Les paramètres de la qualité de l'eau". Ministère de la protection de la nature et de l'environnement. Secrétariat permanent pour l'étude des problèmes de l'eau.
- CAILLE, A., P.G. CAMPBELL, D. CLUIS, P. COUTURE, L. TALBOT (1975).
 "Impact de l'agriculture sur la qualité des eaux de surface du bassin de la rivière Yamaska". INRS-Eau, rapport scientifique No. 50, 99 p.
- FISHER, R.A. (1931).
 "The moments of the distribution for normal samples of measures of departure from normality". Proc. Roy. Soc. London(A), 130: 16-28.
- FOURGEAUD, C. et A. FUCHS. (1972).
 "Statistique". 2ième édition, Dunod, Paris, 334 p.
- HAHN, G.J. et S.S. SHAPIRO (1967).
 "Statistical Models in engineering". John Wiley, New York, 353 p.

REFERENCES (SUITE)

- HARNED, H.S. et B.B. OWEN (1958).
 "The physical chemistry of electrolytic solutions". 3rd edition, Reinhold Publishing Corp., New York.
- LAPOINTE, R. et J.L. SASSEVILLE (1976).
 "Indice de danger environnemental sur les bassins des rivières Yamaska et Saint-François". INRS-Eau, rapport scientifique No. 65, 59 p.
- LEBART, L. et J.P. FENELON (1973).
 "Statistique et Informatique appliquées". 2ième édition, Dunod, Paris.
- MALO-LALANDE, D. (1976).
 "L'impact des activités agricoles sur la qualité de la rivière Yamaska". INRS-Eau, thèse de maîtrise.
- MANN, H.B. et D.R. WHITNEY (1947).
 "On the test of whether one of two random variables is stochastically larger than other". Ann. Math. Stat., 8: 50-60.
- MATALAS, N.C. et M.A. BENSON (1968).
 "Note of the standard error of the coefficient of skewness". Wat. Res. Res., 4(1):
- MATALAS, N.C. et W.B. LANGBEIN (1962).
 "Information content of the mean". Jour. Geophys. Res. 67(9):
- SHERWANI, J.K. et D.H. MOREAU (1975).
 "Strategies for water quality monitoring". Water Resources Research Institute of the university of North Carolina.
- WALD, A. et J. WOLFOWITZ (1943).
 "An exact test for randomness in the non-parametric case based on serial correlation". Ann. Math. Stat., Baltimore XIV.
- WILSON, A.L. (1974).
 "The design of long-term programs for the measurement of river water quality", from: Design of environmental information systems, R.A. Deininger, Ann Arbor Science Publ. 1974, chap 9: 199-226.
- INRS-Eau (1976).
 "Etude intégrée de la qualité des eaux des bassins versants des rivières Saint-François et Yamaska". Vol. 1: Secteur des métaux traces, Rapport No. 51, 131 p. - Vol. 2: Secteur des substances nutritives, Rapport No. 52, 127 p.
- O.P.D.Q. (1973).
 "Utilisation des eaux du bassin versant de la rivière Yamaska". Office de la planification et de développement du Québec.

ANNEXE A-1

ANNEXE A-1

Tableaux des coefficients de corrélation entre
stations pour chaque paramètre de qualité.

Chaque tableau correspond à un paramètre de qualité. Le triangle supérieur droit du tableau donne la valeur des coefficients de corrélation (r) calculés selon la relation 3.54 au moyen des séries d'observations de chaque couple de stations. Chaque coefficient r est un estimateur de la vraie valeur ρ qu'on obtiendrait si on disposait des populations. On a testé si les vraies valeurs ρ étaient significativement différentes de zéro (cf. 2.3.1.1). Ces résultats sont indiqués dans le triangle inférieur du tableau selon la convention suivante:

<u>Symbole</u>	<u>Signification</u>
0.00	ρ n'est pas significativement différent de zéro (seuil $\alpha = 0.05$)
1.00	ρ est significativement différent de zéro (seuil $\alpha = 0.05$)
2.00	ρ est très significativement différent de zéro (seuil $\alpha = 0.01$).

Les valeurs critiques r_0 correspondant aux seuils envisagés sont d'après l'équation 3.56, pour $N = 20$:

$\alpha = 5\%$	$r_0 = 0.445$
$\alpha = 1\%$	$r_0 = 0.562$

NUMERO DU PARAMETRE : 1

SILICE

MATRICE DES COEFFICIENTS DE CORRELATION

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	.81	.92	.45	.68	.79	.76	.86	.69	.62	.71	.37	.51	.53	.30	.18	.31	.37	.38
2	2.00	1.00	.77	.48	.67	.82	.76	.76	.62	.62	.63	.20	.45	.59	.12	.11	.45	.45	.36
3	2.00	2.00	1.00	.50	.63	.85	.86	.87	.76	.69	.66	.40	.65	.53	.37	.03	.36	.35	.45
4	1.00	1.00	1.00	1.00	.60	.18	.30	.28	.06	.03	.10	.06	-.04	-.10	-.16	-.02	-.20	-.00	-.11
5	2.00	2.00	2.00	2.00	1.00	.52	.55	.54	.59	.23	.30	.29	.23	.11	.06	.23	.11	.19	.05
6	2.00	2.00	2.00	0.00	1.00	1.00	.95	.88	.88	.83	.70	.27	.77	.71	.29	-.02	.56	.53	.55
7	2.00	2.00	2.00	0.00	1.00	2.00	1.00	.89	.88	.78	.66	.26	.73	.61	.25	-.04	.55	.57	.59
8	2.00	2.00	2.00	0.00	1.00	2.00	2.00	1.00	.74	.71	.71	.25	.66	.52	.24	.05	.40	.35	.48
9	2.00	2.00	2.00	0.00	2.00	2.00	2.00	2.00	1.00	.72	.60	.42	.78	.60	.45	.09	.59	.57	.59
10	2.00	2.00	2.00	0.00	0.00	2.00	2.00	2.00	2.00	1.00	.87	.49	.90	.83	.52	.07	.75	.69	.81
11	2.00	2.00	2.00	0.00	0.00	2.00	2.00	2.00	2.00	2.00	1.00	.49	.68	.81	.46	.32	.72	.67	.81
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	.47	.55	.89	.30	.60	.55	.65
13	1.00	1.00	2.00	0.00	0.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	1.00	.62	.53	-.11	.61	.49	.67
14	1.00	2.00	1.00	0.00	0.00	2.00	2.00	1.00	2.00	2.00	2.00	1.00	2.00	1.00	.60	.23	.83	.83	.82
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	2.00	1.00	2.00	1.00	.25	.64	.55	.70
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	.23	.23	.25
17	0.00	1.00	0.00	0.00	0.00	2.00	1.00	0.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	0.00	1.00	.85	.92
18	0.00	1.00	0.00	0.00	0.00	1.00	2.00	0.00	2.00	2.00	2.00	1.00	1.00	2.00	1.00	0.00	2.00	1.00	.86
19	0.00	0.00	1.00	0.00	0.00	1.00	2.00	1.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	0.00	2.00	2.00	1.00

NUMERO DU PARAMETRE 1 2

CONDUCTIVITE

MATRICE DES COEFFICIENTS DE CORRELATION

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	.62	.96	.42	.69	.85	.89	.79	.94	.46	.49	.66	.32	.46	.55	.59	.71	.63	.69
2	2.00	1.00	.60	.68	.91	.54	.55	.52	.70	.40	.40	.37	.26	.34	.37	.43	.51	.55	.46
3	2.00	2.00	1.00	.53	.86	.86	.89	.81	.93	.48	.60	.62	.33	.54	.64	.64	.75	.65	.75
4	0.00	2.00	1.00	1.00	.76	.36	.38	.36	.57	.05	.25	.11	-.09	.14	.36	.18	.35	.29	.22
5	2.00	2.00	2.00	2.00	1.00	.62	.57	.54	.74	.24	.39	.41	.13	.35	.41	.42	.45	.52	.46
6	2.00	1.00	2.00	0.00	2.00	1.00	.84	.80	.83	.55	.59	.58	.23	.52	.47	.64	.55	.57	.69
7	2.00	1.00	2.00	0.00	2.00	2.00	1.00	.81	.88	.61	.53	.68	.24	.53	.67	.67	.76	.64	.71
8	2.00	1.00	2.00	0.00	1.00	2.00	2.00	1.00	.85	.65	.69	.65	.46	.69	.66	.72	.67	.60	.79
9	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	.46	.48	.53	.33	.43	.54	.51	.68	.60	.62
10	1.00	0.00	1.00	0.00	0.00	1.00	2.00	2.00	1.00	1.00	.59	.55	.49	.49	.51	.80	.68	.57	.72
11	1.00	0.00	2.00	0.00	0.00	2.00	1.00	2.00	1.00	2.00	1.00	.52	.51	.91	.74	.81	.58	.68	.89
12	2.00	0.00	2.00	0.00	0.00	2.00	2.00	2.00	1.00	1.00	1.00	1.00	.15	.61	.75	.72	.62	.40	.70
13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	1.00	0.00	1.00	.50	.36	.50	.47	.51	.60
14	1.00	0.00	1.00	0.00	0.00	1.00	1.00	2.00	0.00	1.00	2.00	2.00	1.00	1.00	.80	.83	.52	.64	.88
15	1.00	0.00	2.00	0.00	0.00	1.00	2.00	2.00	1.00	1.00	2.00	2.00	0.00	2.00	1.00	.80	.75	.57	.81
16	2.00	0.00	2.00	0.00	0.00	2.00	2.00	2.00	1.00	2.00	2.00	2.00	1.00	2.00	2.00	1.00	.72	.69	.94
17	2.00	1.00	2.00	0.00	1.00	1.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	1.00	2.00	2.00	1.00	.82	.79
18	2.00	1.00	2.00	0.00	1.00	2.00	2.00	2.00	2.00	2.00	0.00	1.00	1.00	2.00	2.00	2.00	2.00	1.00	.82
19	2.00	1.00	2.00	0.00	1.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00

210

NUMERO DU PARAMETRE : 3

AZOTE TOTAL

MATRICE DES COEFFICIENTS DE CORRELATION

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	.74	.96	.76	.48	.85	.85	.82	.90	.37	.55	-.16	.48	-.24	-.23	-.10	.51	.58	.39
2	2.00	1.00	.75	.49	.40	.66	.69	.81	.66	.49	.70	-.24	.67	-.11	-.26	-.10	.62	.53	.69
3	2.00	2.00	1.00	.69	.58	.91	.92	.89	.94	.46	.64	-.10	.54	-.16	-.13	-.02	.52	.59	.45
4	2.00	1.00	2.00	1.00	.23	.56	.62	.56	.68	-.01	.27	-.40	.21	-.47	-.51	-.31	.46	.43	.23
5	1.00	0.00	2.00	0.00	1.00	.65	.63	.45	.66	.32	.24	.30	.26	-.04	-.01	.24	.29	.35	.25
6	2.00	2.00	2.00	2.00	2.00	1.00	.97	.89	.91	.54	.64	.12	.45	.04	.03	.16	.55	.63	.52
7	2.00	2.00	2.00	2.00	2.00	2.00	1.00	.87	.95	.54	.60	-.05	.53	-.11	-.06	.09	.65	.70	.57
8	2.00	2.00	2.00	2.00	1.00	2.00	2.00	1.00	.85	.63	.79	-.15	.66	-.04	-.06	.03	.54	.57	.60
9	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	.47	.51	-.12	.56	-.24	-.16	-.01	.58	.65	.46
10	0.00	1.00	1.00	0.00	0.00	1.00	1.00	2.00	1.00	1.00	.72	.32	.68	.44	.25	.54	.54	.66	.79
11	1.00	2.00	2.00	0.00	0.00	2.00	2.00	2.00	1.00	2.00	1.00	.05	.63	.37	-.02	.26	.45	.52	.67
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	-.28	.74	.61	.73	-.23	-.03	.02
13	1.00	2.00	1.00	0.00	0.00	1.00	1.00	2.00	1.00	2.00	2.00	0.00	1.00	-.10	-.30	-.06	.65	.56	.70
14	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	0.00	1.00	.59	.62	-.19	-.09	.19
15	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	0.00	2.00	1.00	.39	-.25	-.22	-.05
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	2.00	0.00	2.00	0.00	1.00	.08	.40	.35
17	1.00	2.00	1.00	1.00	0.00	1.00	2.00	1.00	2.00	1.00	1.00	0.00	2.00	0.00	0.00	0.00	1.00	.84	.86
18	2.00	1.00	2.00	0.00	0.00	2.00	2.00	2.00	2.00	2.00	1.00	0.00	1.00	0.00	0.00	0.00	2.00	1.00	.79
19	0.00	2.00	1.00	0.00	0.00	1.00	2.00	2.00	1.00	2.00	2.00	0.00	2.00	0.00	0.00	0.00	2.00	2.00	1.00

NUMERO DU PARAMETRE : 4

AZOTE AMMONIACAL

MATRICE DES COEFFICIENTS DE CORRELATION

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	.86	.98	.81	.51	.88	.80	.73	.94	.81	.73	-.03	.60	-.29	.10	-.13	.77	.81	.70
2	2.00	1.00	.80	.75	.35	.73	.63	.65	.90	.84	.76	-.13	.78	-.33	.01	-.14	.81	.92	.89
3	2.00	2.00	1.00	.77	.56	.94	.88	.80	.92	.86	.78	.04	.56	-.24	.18	-.07	.72	.79	.69
4	2.00	2.00	2.00	1.00	.31	.60	.57	.54	.74	.67	.58	-.22	.58	-.42	-.06	-.29	.82	.71	.62
5	1.00	0.00	1.00	0.00	1.00	.52	.47	.38	.59	.46	.31	.27	.40	-.16	.33	.22	.31	.40	.35
6	2.00	2.00	2.00	2.00	1.00	1.00	.94	.86	.84	.85	.86	.18	.40	-.09	.29	-.04	.56	.71	.63
7	2.00	2.00	2.00	2.00	1.00	2.00	1.00	.91	.79	.81	.87	.15	.38	-.09	.33	-.11	.52	.63	.56
8	2.00	2.00	2.00	1.00	0.00	2.00	2.00	1.00	.77	.86	.86	.10	.52	-.04	.28	-.06	.51	.68	.59
9	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	.86	.77	-.07	.70	-.36	.11	-.20	.73	.85	.74
10	2.00	2.00	2.00	2.00	1.00	2.00	2.00	2.00	2.00	1.00	.88	-.03	.72	-.22	.08	-.06	.71	.92	.86
11	2.00	2.00	2.00	2.00	0.00	2.00	2.00	2.00	2.00	2.00	1.00	-.02	.49	-.17	.12	-.11	.63	.78	.75
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	-.20	.70	.79	.61	-.27	-.15	-.08
13	2.00	2.00	1.00	2.00	0.00	0.00	0.00	1.00	2.00	2.00	1.00	0.00	1.00	-.26	-.08	.03	.76	.84	.82
14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	0.00	1.00	.38	.61	-.41	-.34	-.24
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	0.00	0.00	1.00	.49	-.15	-.06	.04
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	0.00	2.00	1.00	1.00	-.15	-.06	.10
17	2.00	2.00	2.00	2.00	0.00	1.00	1.00	1.00	2.00	2.00	2.00	0.00	2.00	0.00	0.00	0.00	1.00	.86	.80
18	2.00	2.00	2.00	2.00	0.00	2.00	2.00	2.00	2.00	2.00	2.00	0.00	2.00	0.00	0.00	0.00	2.00	1.00	.94
19	2.00	2.00	2.00	2.00	0.00	2.00	2.00	2.00	2.00	2.00	2.00	0.00	2.00	0.00	0.00	0.00	2.00	2.00	1.00

NUMERO DU PARAMETRE : 5

NITRATES

MATRICE DES COEFFICIENTS DE CORRELATION

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	.47	.96	.65	.26	.89	.84	.71	.79	.56	.52	.09	.21	.38	.25	.16	.48	.50	.31
2	1.00	1.00	.51	.11	.30	.65	.70	.83	.65	.44	.75	.47	.66	.73	.65	.42	.59	.42	.71
3	2.00	1.00	1.00	.55	.35	.89	.78	.74	.79	.57	.59	.16	.25	.39	.22	.35	.40	.50	.28
4	2.00	0.00	1.00	1.00	.22	.46	.43	.31	.28	.07	.20	-.10	-.16	.05	-.03	-.13	.27	.40	.01
5	0.00	0.00	0.00	0.00	1.00	.54	.41	.40	.46	.08	.30	-.26	.15	.10	-.05	.18	.32	.46	.14
6	2.00	2.00	2.00	1.00	1.00	1.00	.96	.86	.91	.60	.62	.07	.47	.51	.41	.19	.84	.59	.48
7	2.00	2.00	2.00	0.00	0.00	2.00	1.00	.89	.91	.65	.65	.13	.58	.64	.59	.10	.74	.57	.61
8	2.00	2.00	2.00	0.00	0.00	2.00	2.00	1.00	.84	.64	.77	.32	.63	.74	.66	.33	.59	.45	.65
9	2.00	2.00	2.00	0.00	1.00	2.00	2.00	2.00	1.00	.58	.57	.14	.50	.57	.46	.24	.68	.54	.56
10	2.00	0.00	2.00	0.00	0.00	2.00	2.00	2.00	2.00	1.00	.67	.36	.63	.66	.54	.28	.54	.50	.66
11	1.00	2.00	2.00	0.00	0.00	2.00	2.00	2.00	2.00	2.00	1.00	.57	.75	.90	.70	.53	.58	.50	.70
12	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	1.00	.49	.53	.45	.66	.07	-.04	.33
13	0.00	2.00	0.00	0.00	0.00	1.00	2.00	2.00	1.00	2.00	2.00	1.00	1.00	.82	.79	.28	.66	.39	.83
14	0.00	2.00	0.00	0.00	0.00	1.00	2.00	2.00	2.00	2.00	2.00	1.00	2.00	1.00	.88	.38	.66	.45	.80
15	0.00	2.00	0.00	0.00	0.00	0.00	2.00	2.00	1.00	1.00	2.00	1.00	2.00	2.00	1.00	.08	.67	.36	.76
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	2.00	0.00	0.00	0.00	1.00	-.05	.05	.09
17	1.00	2.00	0.00	0.00	0.00	2.00	2.00	2.00	2.00	1.00	2.00	0.00	2.00	2.00	2.00	0.00	1.00	.85	.82
18	1.00	0.00	1.00	0.00	1.00	2.00	2.00	1.00	1.00	1.00	1.00	0.00	0.00	1.00	0.00	0.00	2.00	1.00	.60
19	0.00	2.00	0.00	0.00	0.00	1.00	2.00	2.00	2.00	2.00	2.00	0.00	2.00	2.00	2.00	0.00	2.00	2.00	1.00

NUMERO DU PARAMETRE : 6

PHOSPHORE TOTAL

MATRICE DES COEFFICIENTS DE CORRELATION

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	.60	.90	.07	.57	.82	.85	.72	.77	.52	.36	.56	-.19	.41	.29	.14	.02	.60	-.03
2	2.00	1.00	.38	.20	.09	.29	.47	.24	.45	.07	-.15	-.03	-.06	-.23	.04	.01	.17	.04	-.03
3	2.00	0.00	1.00	-.12	.74	.91	.89	.85	.62	.64	.60	.66	-.25	.60	.48	.27	-.05	.64	-.09
4	0.00	0.00	0.00	1.00	.03	-.15	-.02	-.17	.10	-.33	-.34	-.19	.11	-.19	-.39	-.20	-.11	-.03	-.32
5	2.00	0.00	2.00	0.00	1.00	.76	.59	.58	.30	.41	.45	.72	-.16	.52	.26	.55	-.19	.64	-.20
6	2.00	0.00	2.00	0.00	2.00	1.00	.84	.74	.64	.74	.65	.81	-.32	.68	.45	.43	.00	.66	.02
7	2.00	1.00	2.00	0.00	2.00	2.00	1.00	.82	.74	.72	.61	.58	-.31	.65	.43	.29	.08	.67	.14
8	2.00	0.00	2.00	0.00	2.00	2.00	2.00	1.00	.55	.61	.60	.62	-.33	.56	.51	.30	-.08	.42	-.07
9	2.00	1.00	2.00	0.00	0.00	2.00	2.00	1.00	1.00	.46	.19	.37	-.40	.42	.03	.09	.05	.59	.02
10	1.00	0.00	2.00	0.00	0.00	2.00	2.00	2.00	1.00	1.00	.90	.61	-.16	.83	.45	.32	.11	.56	.50
11	0.00	0.00	2.00	0.00	1.00	2.00	2.00	2.00	0.00	2.00	1.00	.60	-.12	.84	.64	.37	-.04	.49	.42
12	2.00	0.00	2.00	0.00	2.00	2.00	2.00	2.00	0.00	2.00	2.00	1.00	-.23	.58	.53	.55	-.09	.48	.06
13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	-.10	-.33	-.16	.44	.06	.45
14	0.00	0.00	2.00	0.00	1.00	2.00	2.00	1.00	0.00	2.00	2.00	2.00	0.00	1.00	.26	.32	.04	.74	.38
15	0.00	0.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	2.00	1.00	0.00	0.00	1.00	.34	-.11	.06	.04
16	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	-.04	.20	.20
17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	.12	.56
18	2.00	0.00	2.00	0.00	2.00	2.00	2.00	0.00	2.00	1.00	1.00	1.00	0.00	2.00	0.00	0.00	0.00	1.00	.26
19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	2.00	0.00	1.00

NUMERO DU PARAMETRE : 7

PHOSPHORE HYDROLYSABLE

MATRICE DES COEFFICIENTS DE CORRELATION

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	.64	.93	.12	.64	.65	.80	.75	.76	.61	.43	.55	-.08	.50	.22	.30	.27	.66	.32
2	2.00	1.00	.40	.35	.24	.42	.46	.38	.53	.24	.12	.13	.09	.01	.01	.13	.32	.16	.26
3	2.00	1.00	1.00	.05	.77	.93	.86	.86	.65	.65	.53	.66	-.17	.60	.34	.46	.15	.82	.21
4	0.00	0.00	0.00	1.00	.11	-.03	.08	.03	.13	-.16	-.23	-.03	.06	-.09	-.25	-.15	.02	.09	.01
5	2.00	0.00	2.00	0.00	1.00	.78	.60	.58	.23	.30	.26	.74	-.22	.34	.25	.66	-.17	.77	-.22
6	2.00	0.00	2.00	0.00	2.00	1.00	.87	.84	.68	.73	.60	.77	-.30	.70	.36	.54	.11	.80	.21
7	2.00	1.00	2.00	0.00	2.00	2.00	1.00	.92	.81	.67	.58	.49	-.25	.79	.25	.40	.12	.83	.30
8	2.00	0.00	2.00	0.00	2.00	2.00	2.00	1.00	.75	.68	.60	.58	-.37	.74	.33	.42	-.03	.80	.20
9	2.00	1.00	2.00	0.00	0.00	2.00	2.00	2.00	1.00	.63	.39	.27	-.28	.67	-.01	.04	.21	.51	.40
10	2.00	0.00	2.00	0.00	0.00	2.00	2.00	2.00	2.00	1.00	.90	.58	-.12	.83	.46	.25	.20	.50	.60
11	0.00	0.00	1.00	0.00	0.00	2.00	2.00	2.00	0.00	2.00	1.00	.60	-.19	.75	.68	.43	.14	.40	.57
12	1.00	0.00	2.00	0.00	2.00	2.00	1.00	2.00	0.00	2.00	2.00	1.00	-.35	.48	.50	.76	-.02	.53	.13
13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	-.21	-.26	-.29	.55	-.10	.40
14	1.00	0.00	2.00	0.00	0.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	0.00	1.00	.22	.24	.17	.73	.48
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	2.00	1.00	0.00	0.00	1.00	.46	-.06	.14	.21
16	0.00	0.00	1.00	0.00	2.00	1.00	0.00	0.00	0.00	0.00	0.00	2.00	0.00	0.00	1.00	1.00	-.12	.40	-.05
17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	.03	.73
18	2.00	0.00	2.00	0.00	2.00	2.00	2.00	2.00	1.00	1.00	0.00	1.00	0.00	2.00	0.00	0.00	0.00	1.00	.10
19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	2.00	0.00	0.00	1.00	0.00	0.00	2.00	0.00	1.00

NUMERO DU PARAMETRE : 8

CALCIUM

MATRICE DES COEFFICIENTS DE CORRELATION

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	.59	.93	.52	.69	.91	.95	.75	.85	.61	.39	.57	.69	.39	.58	.51	.63	.64	.54
2	2.00	1.00	.45	.83	.82	.54	.52	.58	.64	.16	.21	.32	.32	.23	.31	.22	.29	.42	.27
3	2.00	1.00	1.00	.42	.58	.91	.95	.82	.89	.64	.44	.55	.70	.43	.60	.56	.66	.63	.57
4	1.00	2.00	0.00	1.00	.87	.41	.45	.43	.55	.07	.10	.07	.20	.05	.16	.08	.28	.41	.11
5	2.00	2.00	2.00	2.00	1.00	.58	.60	.57	.67	.11	.14	.18	.30	.14	.20	.16	.22	.38	.16
6	2.00	1.00	2.00	0.00	2.00	1.00	.96	.81	.89	.62	.42	.53	.69	.42	.57	.57	.59	.58	.55
7	2.00	1.00	2.00	1.00	2.00	2.00	1.00	.83	.91	.71	.44	.55	.75	.43	.62	.61	.69	.66	.59
8	2.00	2.00	2.00	0.00	2.00	2.00	2.00	1.00	.88	.60	.55	.48	.75	.57	.54	.64	.50	.52	.61
9	2.00	2.00	2.00	1.00	2.00	2.00	2.00	2.00	1.00	.65	.51	.49	.75	.53	.62	.67	.63	.61	.62
10	2.00	0.00	2.00	0.00	0.00	2.00	2.00	2.00	2.00	1.00	.76	.70	.87	.72	.81	.83	.84	.75	.85
11	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	2.00	1.00	.56	.77	.94	.73	.83	.66	.76	.91
12	2.00	0.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00	2.00	2.00	1.00	.54	.71	.91	.69	.69	.53	.80
13	2.00	0.00	2.00	0.00	0.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	1.00	.71	.69	.77	.69	.71	.80
14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	1.00	2.00	2.00	2.00	2.00	1.00	.82	.89	.62	.62	.94
15	2.00	0.00	2.00	0.00	0.00	2.00	2.00	1.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	.86	.86	.70
16	1.00	0.00	2.00	0.00	0.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	.79	.65
17	2.00	0.00	2.00	0.00	0.00	2.00	2.00	1.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	.87
18	2.00	0.00	2.00	0.00	0.00	2.00	2.00	1.00	2.00	2.00	2.00	1.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00
19	1.00	0.00	2.00	0.00	0.00	1.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00

NUMERO DU PARAMETRE : 9

FER

MATRICE DES COEFFICIENTS DE CORRELATION

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	.44	.78	.40	.59	.60	.56	.24	.51	-.06	-.23	-.68	-.09	-.53	-.34	-.01	.08	.29	-.23
2	0.00	1.00	.42	.67	.74	.30	.00	.28	.08	.09	-.29	-.37	-.24	-.64	-.67	-.05	-.10	.04	-.29
3	2.00	0.00	1.00	.50	.56	.50	.20	.17	.26	.13	-.02	-.65	-.13	-.52	-.53	-.08	.16	.33	-.18
4	1.00	2.00	1.00	1.00	.57	.28	.26	-.09	-.04	-.06	-.29	-.36	-.47	-.33	-.40	.06	-.15	-.08	-.43
5	2.00	2.00	1.00	2.00	1.00	.53	.11	.13	.23	-.00	-.22	-.48	-.15	-.67	-.49	-.17	-.12	.06	-.41
6	2.00	0.00	1.00	0.00	1.00	1.00	.44	.43	.42	.15	.02	-.31	.04	-.43	-.41	.00	-.03	.09	-.25
7	1.00	0.00	0.00	0.00	0.00	0.00	1.00	.17	.54	-.16	-.39	-.13	-.03	-.14	.08	.11	.00	-.05	-.19
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	.37	.51	.27	-.00	.13	-.10	-.39	.38	.10	.14	.04
9	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	.39	-.13	-.13	.36	-.07	-.12	-.03	.48	.46	.38
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	.50	.25	.37	.24	-.35	.37	.69	.55	.62
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	.20	.21	.43	-.04	.36	.36	.41	.50
12	2.00	0.00	2.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	.17	.70	.44	.16	.08	-.07	.28
13	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	.08	.30	.10	.50	.37	.49
14	1.00	2.00	1.00	0.00	2.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	0.00	1.00	.58	.17	.15	.00	.41
15	0.00	2.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	1.00	-.01	-.14	-.17	.04
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	.25	.22	.09
17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	2.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	.89	.85
18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	1.00	.76
19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	1.00	0.00	1.00	0.00	0.00	0.00	2.00	2.00	1.00

NUMERO DU PARAMETRE : 10

MAGNESIUM

MATRICE DES COEFFICIENTS DE CORRELATION

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	.58	.96	.44	.58	.91	.93	.71	.87	.45	.33	.39	.40	.33	.48	.37	.46	.46	.25
2	2.00	1.00	.45	.44	.69	.62	.60	.63	.74	.53	.41	.34	.37	.36	.41	.34	.49	.69	.35
3	2.00	1.00	1.00	.36	.45	.91	.94	.72	.80	.41	.32	.30	.35	.29	.42	.32	.42	.45	.22
4	1.00	0.00	0.00	1.00	.67	.27	.28	.26	.49	-.06	.25	-.08	.17	.30	.24	.11	.10	.24	-.14
5	2.00	2.00	1.00	2.00	1.00	.55	.48	.59	.76	.32	.52	.36	.43	.64	.45	.49	.28	.39	.24
6	2.00	2.00	2.00	0.00	1.00	1.00	.97	.87	.88	.46	.34	.39	.37	.30	.38	.39	.37	.48	.23
7	2.00	2.00	2.00	0.00	1.00	2.00	1.00	.82	.85	.48	.36	.40	.42	.31	.41	.37	.39	.51	.25
8	2.00	2.00	2.00	0.00	2.00	2.00	2.00	1.00	.87	.50	.46	.32	.55	.46	.33	.44	.31	.51	.25
9	2.00	2.00	2.00	1.00	2.00	2.00	2.00	2.00	1.00	.56	.50	.48	.55	.53	.55	.54	.50	.60	.34
10	1.00	1.00	0.00	0.00	0.00	1.00	1.00	1.00	2.00	1.00	.73	.70	.76	.65	.77	.68	.84	.69	.87
11	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	1.00	2.00	1.00	.60	.75	.93	.82	.85	.60	.57	.81
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	2.00	2.00	1.00	.45	.63	.79	.72	.66	.42	.73
13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	2.00	2.00	1.00	1.00	.80	.59	.63	.42	.36	.63
14	0.00	0.00	0.00	0.00	2.00	0.00	0.00	1.00	1.00	2.00	2.00	2.00	2.00	1.00	.76	.84	.47	.40	.73
15	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	2.00	2.00	2.00	2.00	2.00	1.00	.85	.84	.60	.83
16	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	.64	.47	.73
17	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	2.00	2.00	2.00	0.00	1.00	2.00	2.00	1.00	.76	.78
18	1.00	2.00	1.00	0.00	0.00	1.00	1.00	1.00	2.00	2.00	2.00	0.00	0.00	0.00	2.00	1.00	2.00	1.00	.53
19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	1.00

NUMERO DU PARAMETRE : 11

MANGANESE

MATRICE DES COEFFICIENTS DE CORRELATION

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	.62	.79	.80	-.01	.33	.05	.05	.48	.30	-.05	-.14	-.17	-.29	-.12	-.09	.59	.15	-.19
2	2.00	1.00	.44	.52	-.10	.20	-.06	-.07	.19	.42	.01	-.17	-.12	-.37	-.11	-.15	.40	-.01	-.23
3	2.00	0.00	1.00	.67	.07	.41	.03	.09	.46	.15	-.23	-.21	-.17	-.21	-.22	-.06	.58	.14	-.21
4	2.00	1.00	2.00	1.00	.10	.40	.17	.25	.78	.16	-.10	-.27	-.02	-.21	-.22	-.04	.56	.12	-.26
5	0.00	0.00	0.00	0.00	1.00	.74	.54	.45	.31	-.16	-.07	.32	.72	.60	.25	.54	.24	.48	.55
6	0.00	0.00	0.00	0.00	2.00	1.00	.56	.59	.46	.07	-.14	.24	.53	.49	.23	.54	.21	.34	.22
7	0.00	0.00	0.00	0.00	1.00	1.00	1.00	.89	.35	-.03	-.22	.18	.59	.57	.20	.52	-.01	.01	-.12
8	0.00	0.00	0.00	0.00	1.00	2.00	2.00	1.00	.47	.05	-.22	.20	.40	.44	.24	.43	-.04	.10	-.19
9	1.00	0.00	1.00	2.00	0.00	1.00	0.00	1.00	1.00	-.27	-.39	-.37	.09	-.06	-.33	.05	.36	-.02	-.27
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	.69	.46	.02	-.06	.52	.08	.17	.27	.11
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	1.00	.43	.26	.05	.40	-.02	.14	.38	.53
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	.19	.64	.95	.70	-.03	.27	.33
13	0.00	0.00	0.00	0.00	2.00	1.00	2.00	0.00	0.00	0.00	0.00	0.00	1.00	.59	.19	.43	.18	.46	.58
14	0.00	0.00	0.00	0.00	2.00	1.00	2.00	0.00	0.00	0.00	0.00	2.00	2.00	1.00	.57	.91	-.18	.13	.34
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	2.00	0.00	2.00	1.00	.66	-.00	.27	.26
16	0.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	2.00	0.00	2.00	2.00	1.00	-.10	.04	.17
17	2.00	0.00	2.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	.52	.15
18	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	1.00	.71
19	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	2.00	0.00	0.00	0.00	0.00	2.00	1.00

NUMERO DU PARAMETRE : 12

SODIUM

MATRICE DES COEFFICIENTS DE CORRELATION

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	.76	.95	.68	.67	.77	.76	.58	.89	.49	.75	.69	.43	.67	.77	.47	.78	.59	.65
2	2.00	1.00	.76	.73	.80	.53	.59	.46	.70	.48	.65	.46	.41	.57	.68	.50	.64	.53	.55
3	2.00	2.00	1.00	.76	.70	.76	.78	.61	.86	.51	.87	.62	.39	.75	.86	.48	.78	.59	.67
4	2.00	2.00	2.00	1.00	.54	.38	.46	.29	.52	.17	.59	.27	-.03	.33	.60	.16	.36	.11	.19
5	2.00	2.00	2.00	1.00	1.00	.63	.63	.58	.65	.26	.64	.48	.23	.71	.73	.38	.59	.60	.54
6	2.00	1.00	2.00	0.00	2.00	1.00	.81	.72	.89	.62	.77	.74	.36	.86	.75	.50	.78	.72	.70
7	2.00	2.00	2.00	1.00	2.00	2.00	1.00	.92	.80	.59	.66	.70	.32	.85	.88	.53	.78	.56	.56
8	2.00	1.00	2.00	0.00	2.00	2.00	2.00	1.00	.65	.43	.54	.61	.23	.85	.83	.50	.65	.51	.47
9	2.00	2.00	2.00	1.00	2.00	2.00	2.00	2.00	1.00	.67	.79	.73	.56	.76	.74	.45	.85	.68	.75
10	1.00	1.00	1.00	0.00	0.00	2.00	2.00	0.00	2.00	1.00	.62	.54	.72	.61	.49	.63	.81	.66	.72
11	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	2.00	2.00	1.00	.54	.55	.83	.84	.58	.86	.76	.80
12	2.00	1.00	2.00	0.00	1.00	2.00	2.00	2.00	2.00	1.00	1.00	1.00	.29	.66	.61	.44	.72	.47	.72
13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	2.00	1.00	0.00	1.00	.44	.35	.61	.72	.71	.77
14	2.00	2.00	2.00	0.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	0.00	1.00	.88	.61	.84	.80	.75
15	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	2.00	2.00	0.00	2.00	1.00	.58	.81	.67	.64
16	1.00	1.00	1.00	0.00	0.00	1.00	1.00	1.00	1.00	2.00	2.00	0.00	2.00	2.00	2.00	1.00	.69	.62	.71
17	2.00	2.00	2.00	0.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	.84
18	2.00	1.00	2.00	0.00	2.00	2.00	2.00	1.00	2.00	2.00	2.00	1.00	2.00	2.00	2.00	2.00	2.00	1.00	.79
19	2.00	1.00	2.00	0.00	1.00	2.00	1.00	1.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00

NUMERO DU PARAMETRE : 13

POTASSIUM

MATRICE DES COEFFICIENTS DE CORRELATION

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	.76	.97	.68	.79	.23	.87	.55	.90	.72	.56	.58	.32	.61	.70	.54	.69	.68	.62
2	2.00	1.00	.71	.89	.67	.09	.67	.45	.57	.76	.71	.39	.44	.71	.62	.63	.56	.52	.53
3	2.00	2.00	1.00	.62	.82	.17	.86	.60	.91	.71	.61	.66	.36	.68	.74	.51	.63	.65	.63
4	2.00	2.00	2.00	1.00	.70	.12	.55	.42	.42	.71	.69	.29	.33	.72	.65	.77	.50	.41	.48
5	2.00	2.00	2.00	2.00	1.00	.10	.67	.48	.66	.73	.71	.78	.35	.79	.85	.57	.47	.68	.59
6	0.00	0.00	0.00	0.00	0.00	1.00	.15	.30	.16	.28	.21	.13	.04	.17	.10	.10	-.03	.10	.32
7	2.00	2.00	2.00	1.00	2.00	0.00	1.00	.73	.84	.73	.53	.54	.43	.57	.61	.40	.58	.59	.51
8	1.00	1.00	2.00	0.00	1.00	0.00	2.00	1.00	.65	.73	.66	.49	.37	.71	.63	.37	.36	.45	.57
9	2.00	2.00	2.00	0.00	2.00	0.00	2.00	2.00	1.00	.61	.43	.65	.20	.52	.71	.35	.65	.63	.58
10	2.00	2.00	2.00	2.00	2.00	0.00	2.00	2.00	2.00	1.00	.86	.53	.71	.80	.67	.55	.56	.72	.77
11	2.00	2.00	2.00	2.00	2.00	0.00	1.00	2.00	0.00	2.00	1.00	.53	.66	.94	.64	.51	.39	.63	.77
12	2.00	0.00	2.00	0.00	2.00	0.00	1.00	1.00	2.00	1.00	1.00	1.00	.18	.67	.84	.28	.38	.58	.57
13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	2.00	0.00	1.00	.46	.13	.06	.42	.47	.50
14	2.00	2.00	2.00	2.00	2.00	0.00	2.00	2.00	1.00	2.00	2.00	2.00	1.00	1.00	.80	.62	.40	.60	.72
15	2.00	2.00	2.00	2.00	2.00	0.00	2.00	2.00	2.00	2.00	2.00	2.00	0.00	2.00	1.00	.64	.46	.53	.63
16	1.00	2.00	1.00	2.00	2.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	2.00	2.00	1.00	.28	.27	.38
17	2.00	2.00	2.00	1.00	1.00	0.00	2.00	0.00	2.00	2.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	.65	.61
18	2.00	1.00	2.00	0.00	2.00	0.00	2.00	1.00	2.00	2.00	2.00	2.00	1.00	2.00	1.00	0.00	2.00	1.00	.77
19	2.00	1.00	2.00	1.00	2.00	0.00	1.00	2.00	2.00	2.00	2.00	2.00	1.00	2.00	2.00	0.00	2.00	2.00	1.00

NUMERO DU PARAMETRE : 14

CHLORE

MATRICE DES COEFFICIENTS DE CORRELATION

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	.82	.95	.25	.75	.78	.71	.58	.86	.35	.51	.46	.71	.55	.49	.66	.86	.67	.82
2	2.00	1.00	.73	.26	.77	.53	.44	.34	.69	.04	.29	.26	.64	.35	.15	.45	.74	.60	.75
3	2.00	2.00	1.00	.38	.73	.83	.74	.58	.80	.44	.66	.45	.67	.62	.61	.62	.88	.70	.80
4	0.00	0.00	0.00	1.00	.17	.17	.16	.19	.21	.18	.34	.04	.13	.09	.36	.01	.24	.01	.01
5	2.00	2.00	2.00	0.00	1.00	.70	.71	.63	.71	.14	.49	.33	.50	.64	.44	.49	.74	.70	.65
6	2.00	1.00	2.00	0.00	2.00	1.00	.83	.69	.82	.72	.65	.60	.64	.67	.78	.50	.83	.69	.69
7	2.00	0.00	2.00	0.00	2.00	2.00	1.00	.88	.82	.46	.44	.54	.39	.55	.82	.50	.71	.46	.56
8	2.00	0.00	2.00	0.00	2.00	2.00	2.00	1.00	.69	.38	.33	.52	.23	.53	.74	.36	.60	.26	.43
9	2.00	2.00	2.00	0.00	2.00	2.00	2.00	2.00	1.00	.46	.42	.46	.61	.46	.58	.55	.80	.58	.68
10	0.00	0.00	0.00	0.00	0.00	2.00	1.00	0.00	1.00	1.00	.45	.59	.25	.34	.68	.25	.31	.21	.20
11	1.00	0.00	2.00	0.00	1.00	2.00	0.00	0.00	0.00	1.00	1.00	.25	.44	.89	.62	.63	.57	.73	.57
12	1.00	0.00	1.00	0.00	0.00	2.00	1.00	1.00	1.00	2.00	0.00	1.00	.13	.33	.62	.30	.43	.16	.33
13	2.00	2.00	2.00	0.00	1.00	2.00	0.00	0.00	2.00	0.00	0.00	0.00	1.00	.46	.12	.52	.78	.80	.88
14	1.00	0.00	2.00	0.00	2.00	2.00	1.00	1.00	1.00	0.00	2.00	0.00	1.00	1.00	.61	.70	.60	.72	.66
15	1.00	0.00	2.00	0.00	0.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	0.00	2.00	1.00	.46	.53	.32	.34
16	2.00	1.00	2.00	0.00	1.00	1.00	1.00	0.00	1.00	0.00	2.00	0.00	1.00	2.00	1.00	1.00	.55	.64	.70
17	2.00	2.00	2.00	0.00	2.00	2.00	2.00	2.00	2.00	0.00	2.00	0.00	2.00	2.00	1.00	1.00	1.00	.80	.87
18	2.00	2.00	2.00	0.00	2.00	2.00	1.00	0.00	2.00	0.00	2.00	0.00	2.00	2.00	0.00	2.00	2.00	1.00	.83
19	2.00	2.00	2.00	0.00	2.00	2.00	1.00	0.00	2.00	0.00	2.00	0.00	2.00	2.00	0.00	2.00	2.00	2.00	1.00

NUMERO DU PARAMETRE : 15

MATRICE DES COEFFICIENTS DE CORRELATION

SULFATES

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00	.77	.89	.72	.85	.92	.80	.80	.92	.87	.87	.78	.85	.79	.67	<u>.51</u>	.90	.78	.88
2	2.00	1.00	.82	.83	.70	.83	.86	.80	.85	.80	.74	<u>.60</u>	<u>.67</u>	.81	.62	<u>.42</u>	<u>.66</u>	<u>.56</u>	<u>.68</u>
3	2.00	2.00	1.00	.78	.81	.97	.85	.94	.96	.91	.94	.84	.85	.91	.76	<u>.49</u>	.89	.82	.89
4	2.00	2.00	2.00	1.00	.84	.72	<u>.58</u>	<u>.69</u>	.82	<u>.67</u>	<u>.66</u>	<u>.56</u>	.70	.70	<u>.57</u>	<u>.30</u>	<u>.64</u>	<u>.64</u>	<u>.61</u>
5	2.00	2.00	2.00	2.00	1.00	.77	<u>.64</u>	<u>.67</u>	.88	.70	.75	<u>.58</u>	<u>.69</u>	<u>.64</u>	.53	<u>.16</u>	.79	<u>.68</u>	.79
6	2.00	2.00	2.00	2.00	2.00	1.00	.91	.94	.95	.92	.91	.88	.83	.92	.81	<u>.57</u>	.89	.82	.87
7	2.00	2.00	2.00	2.00	2.00	2.00	1.00	.84	.82	.84	.79	.70	<u>.61</u>	.84	.70	<u>.49</u>	.76	<u>.60</u>	.79
8	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	.90	.86	.87	.81	.80	.89	.78	<u>.47</u>	.79	.75	.75
9	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	.90	.91	.79	.84	.87	.75	<u>.41</u>	.87	.82	.87
10	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	.95	.82	.88	.90	.71	<u>.64</u>	.86	.82	.86
11	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	.65	.89	.91	.68	<u>.50</u>	.91	.88	.92
12	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	.78	.87	.88	.73	.84	.90	.77
13	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	.81	.58	<u>.59</u>	.77	.83	.77
14	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	.75	<u>.56</u>	.82	.83	.83
15	2.00	2.00	2.00	2.00	1.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	<u>.62</u>	.71	.78	.62
16	1.00	0.00	1.00	0.00	0.00	2.00	1.00	1.00	0.00	2.00	1.00	2.00	2.00	1.00	2.00	1.00	<u>.45</u>	<u>.53</u>	<u>.38</u>
17	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	1.00	.89	.95
18	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00	2.00	1.00	.83
19	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	0.00	2.00	2.00	1.00

ANNEXE A-2

ANNEXE A-2

Graphiques des régressions linéaires entre les stations

1 et 3 pour chaque paramètre de qualité

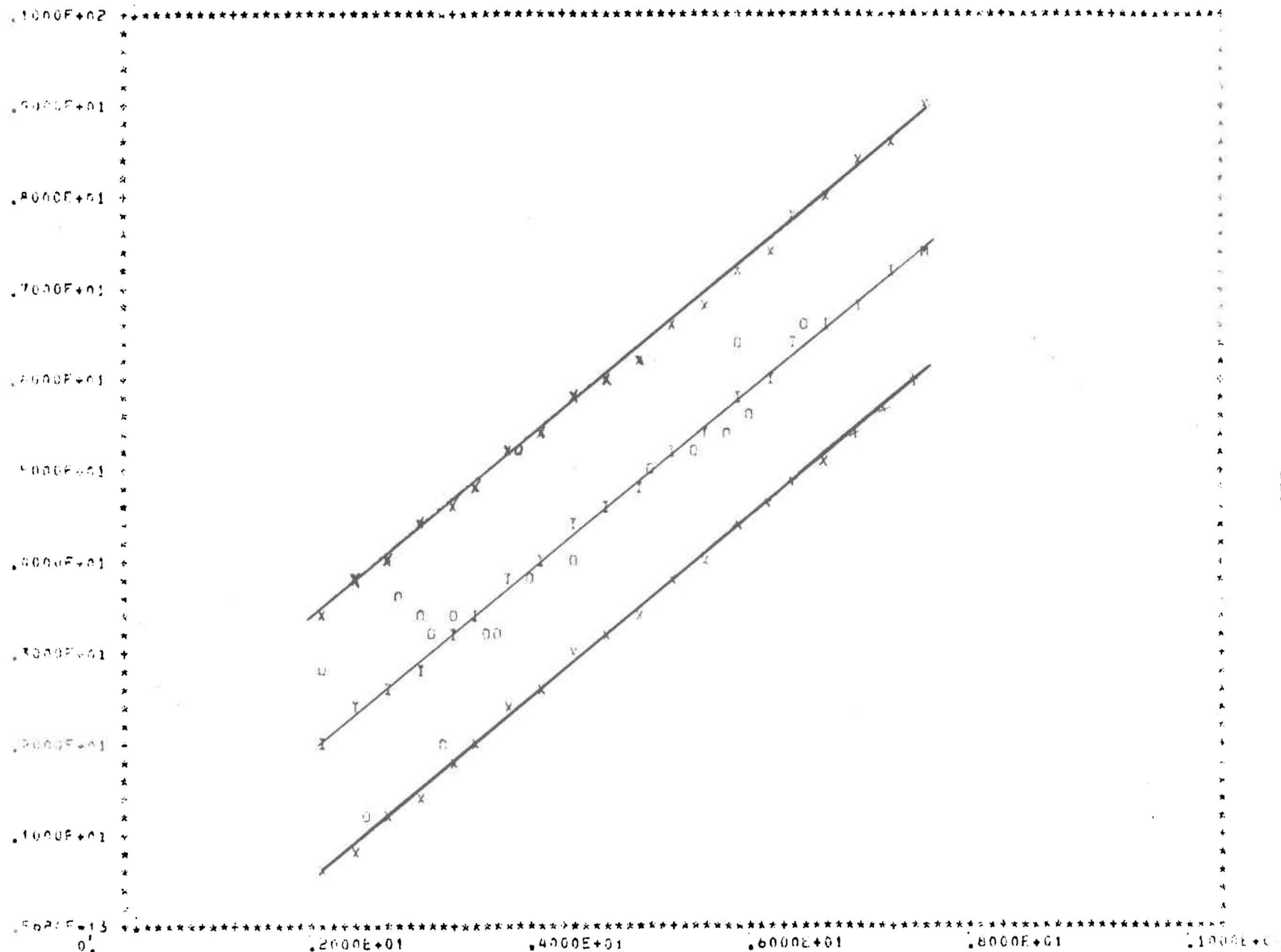
(sauf Fer et Manganèse).

La station 3 est la station de référence. On reconstitue les valeurs à la station 1.

<u>Symbole</u>	<u>Signification</u>
0	valeurs observées
I	droite de régression
X	intervalle de confiance à 95%
M	points superposés

On indique au-dessous de chaque graphique, l'équation de la droite de régression (y est la valeur reconstituée à la station 1), ainsi que les valeurs maximale et minimale du paramètre de qualité observées à la station 3 et dans l'intervalle desquelles l'équation de régression a été calculée.

Station 1



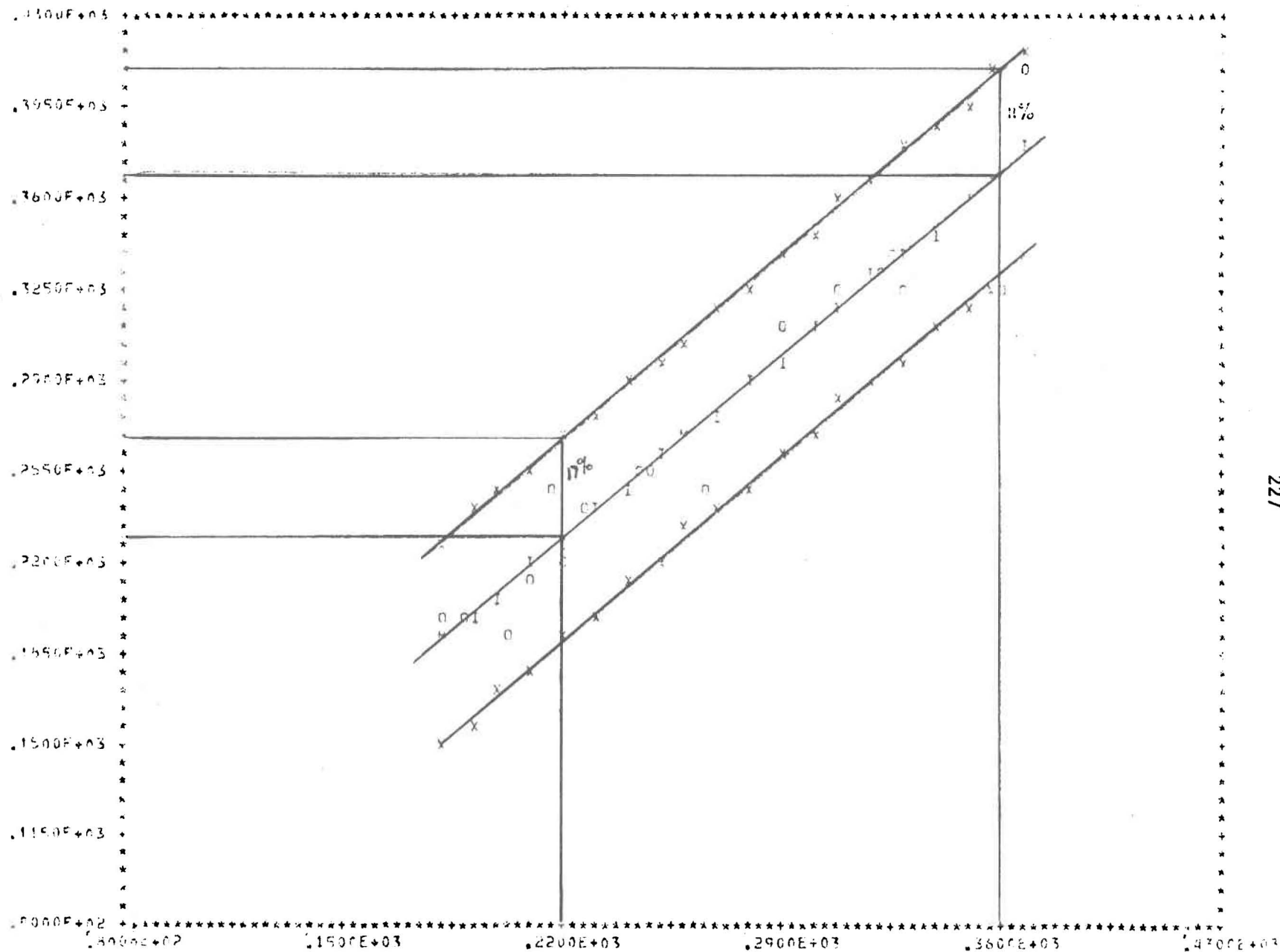
$$Y = 1.25 + .99X$$

MAXIMUM = 7.30

MINIMUM = 1.00

Silice (mg SiO₂/l)

Station 1



$$Y = 6.96 + 1.01X$$

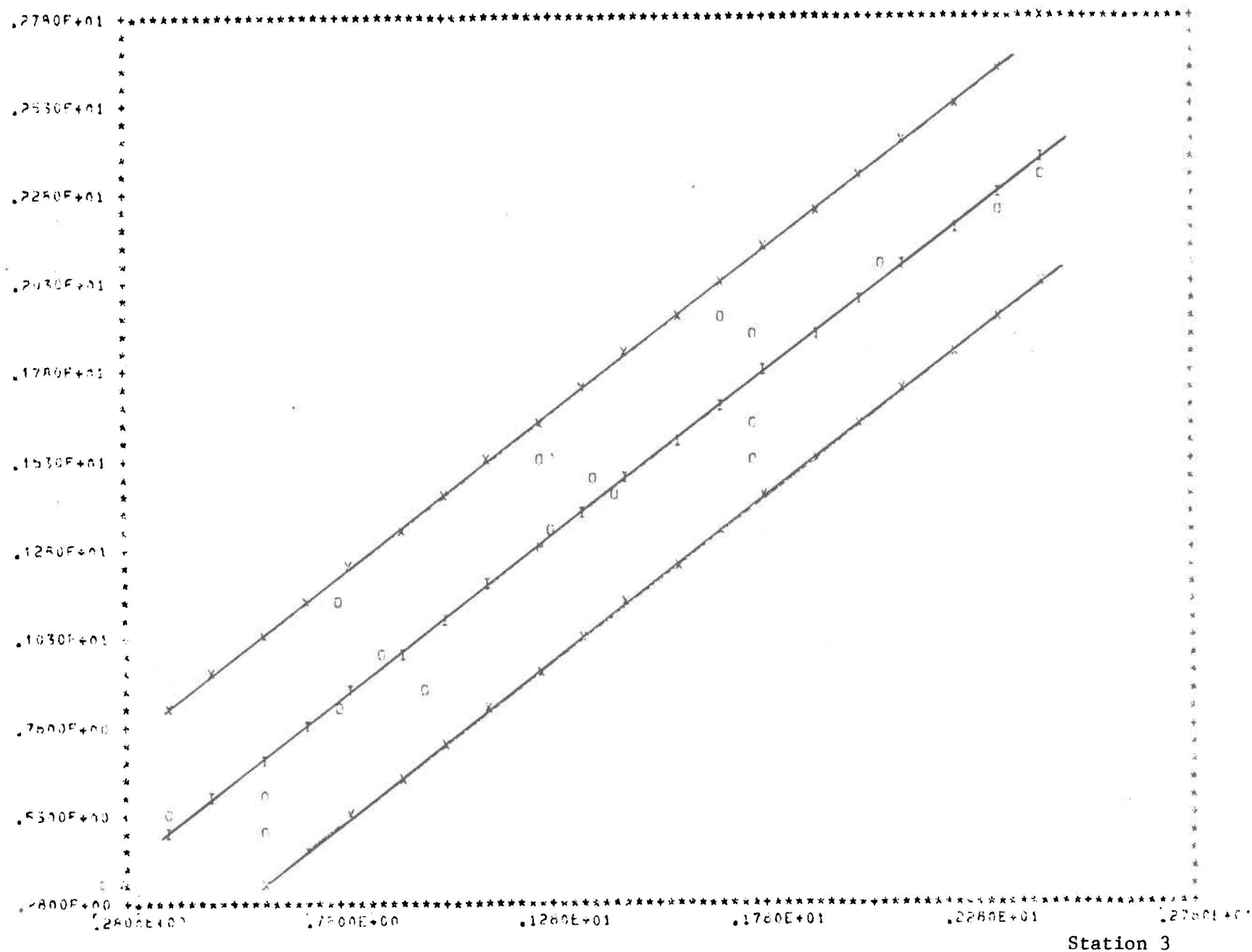
MAXIMUM = 368.00

MINIMUM = 181.00

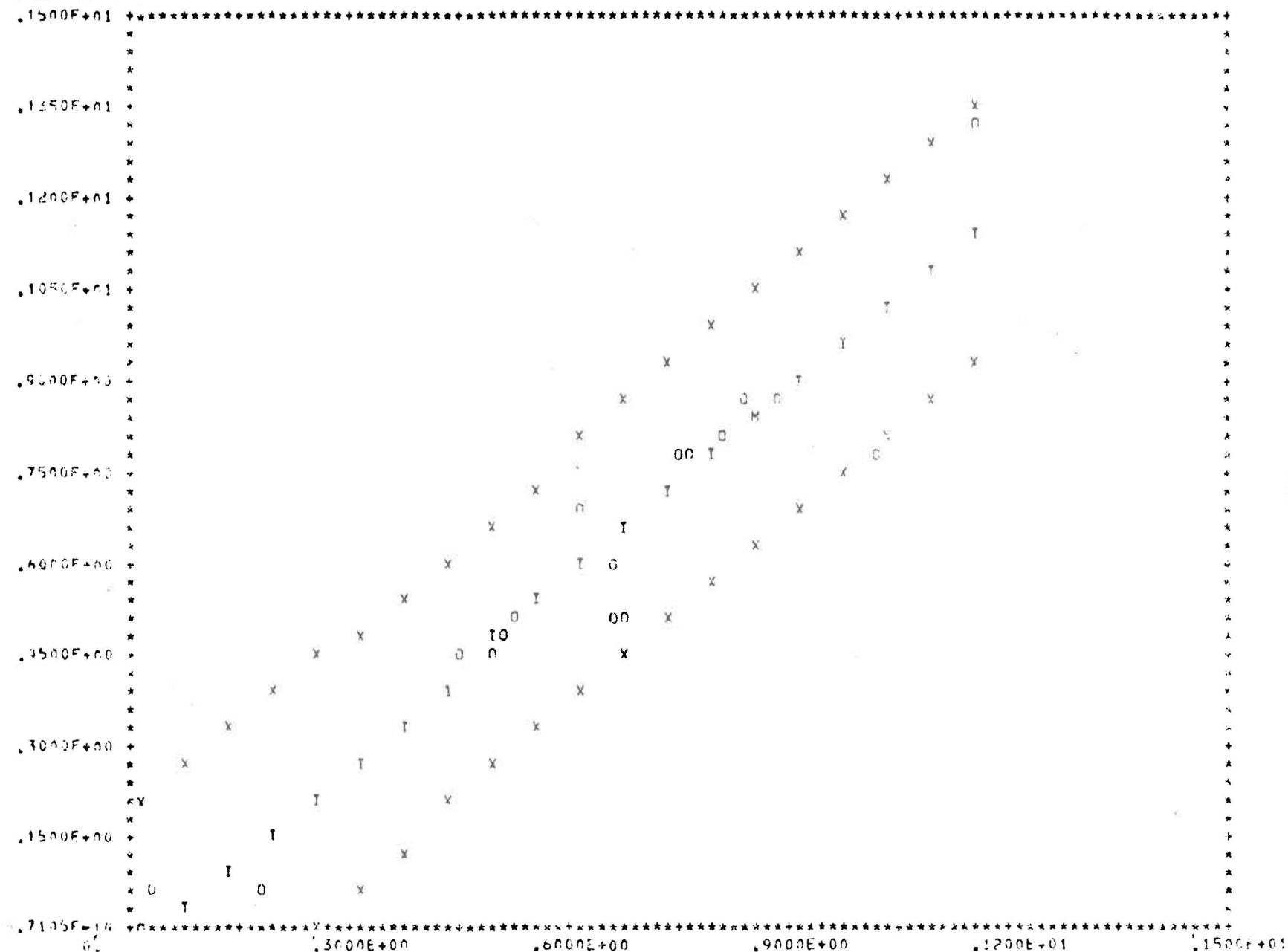
Conductivité (μmhos/cm)

Station 3

Station 1



Station 1



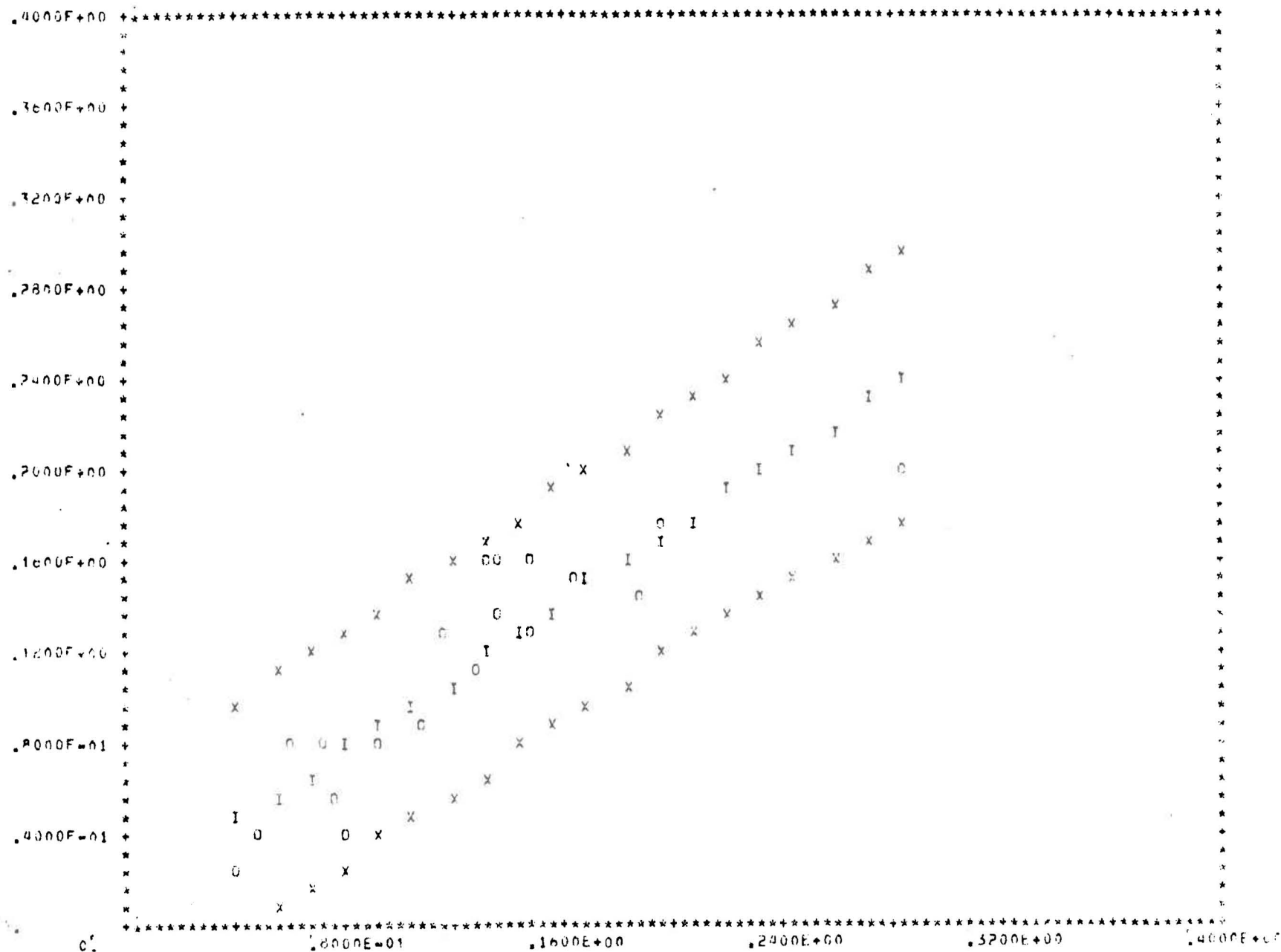
$$Y = -0.04 + 1.02X$$

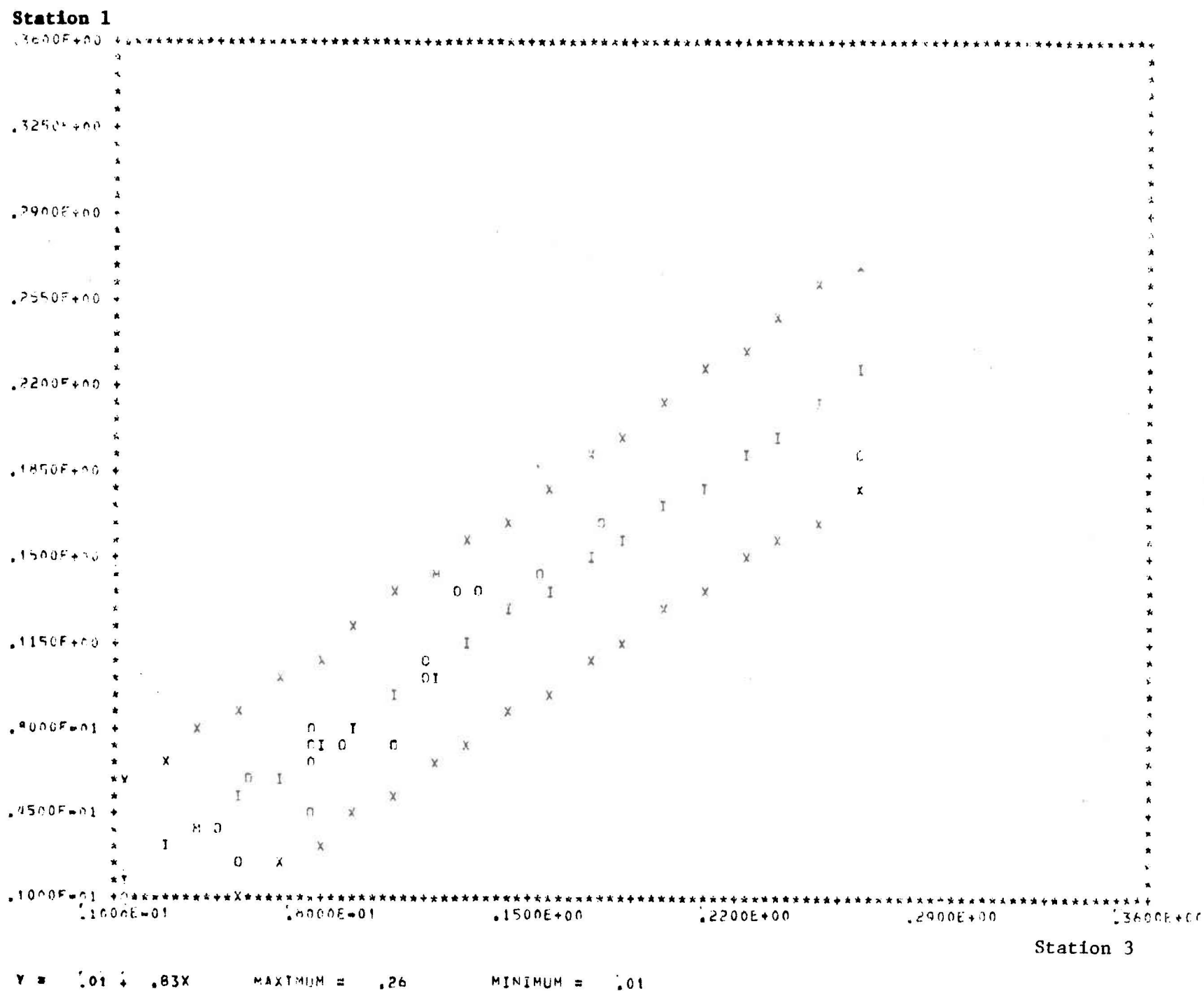
MAXIMUM = 1.15

MINIMUM = 0.01

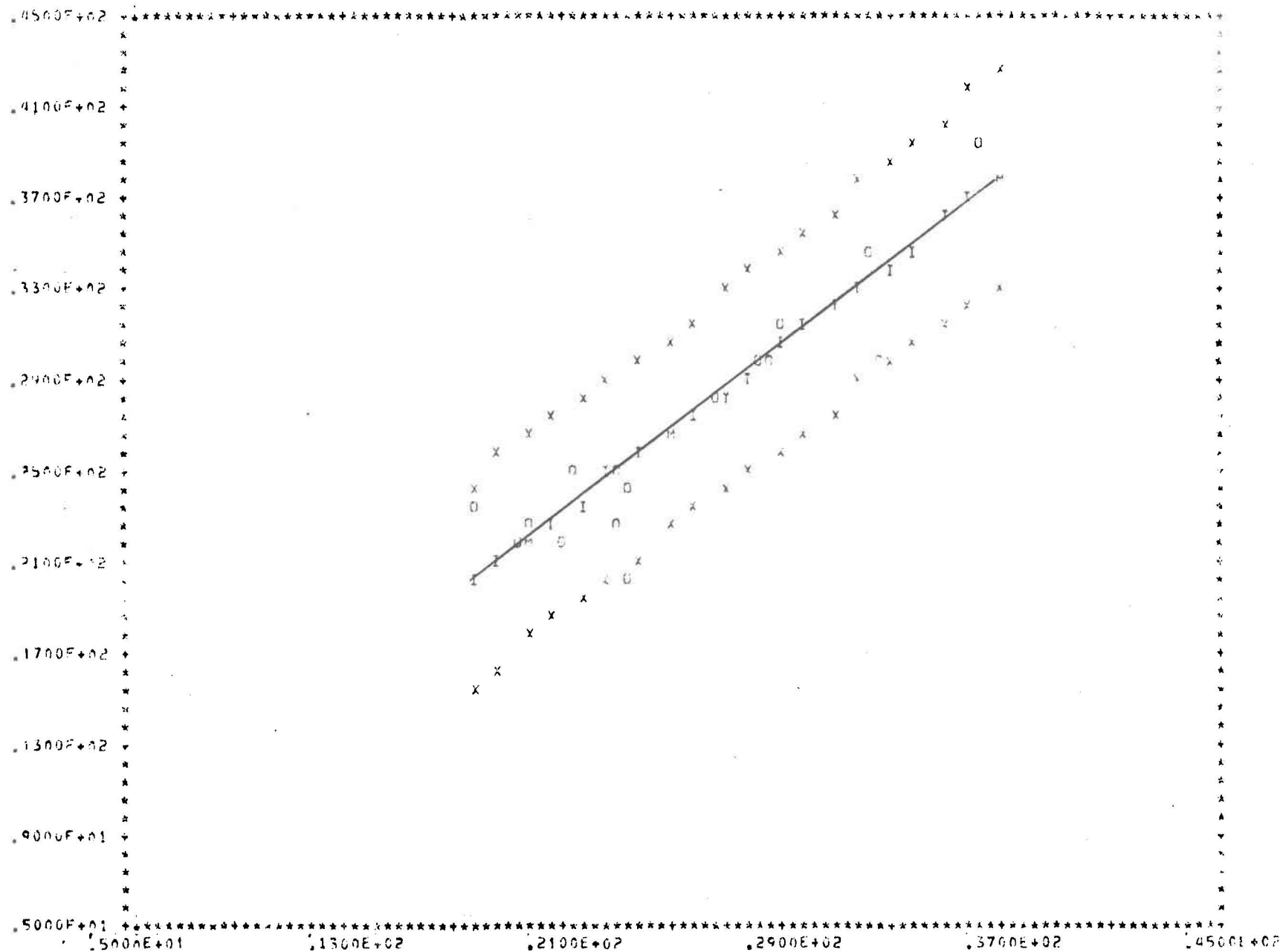
Nitrates (mg N/l)

Station 1





Station 1



232

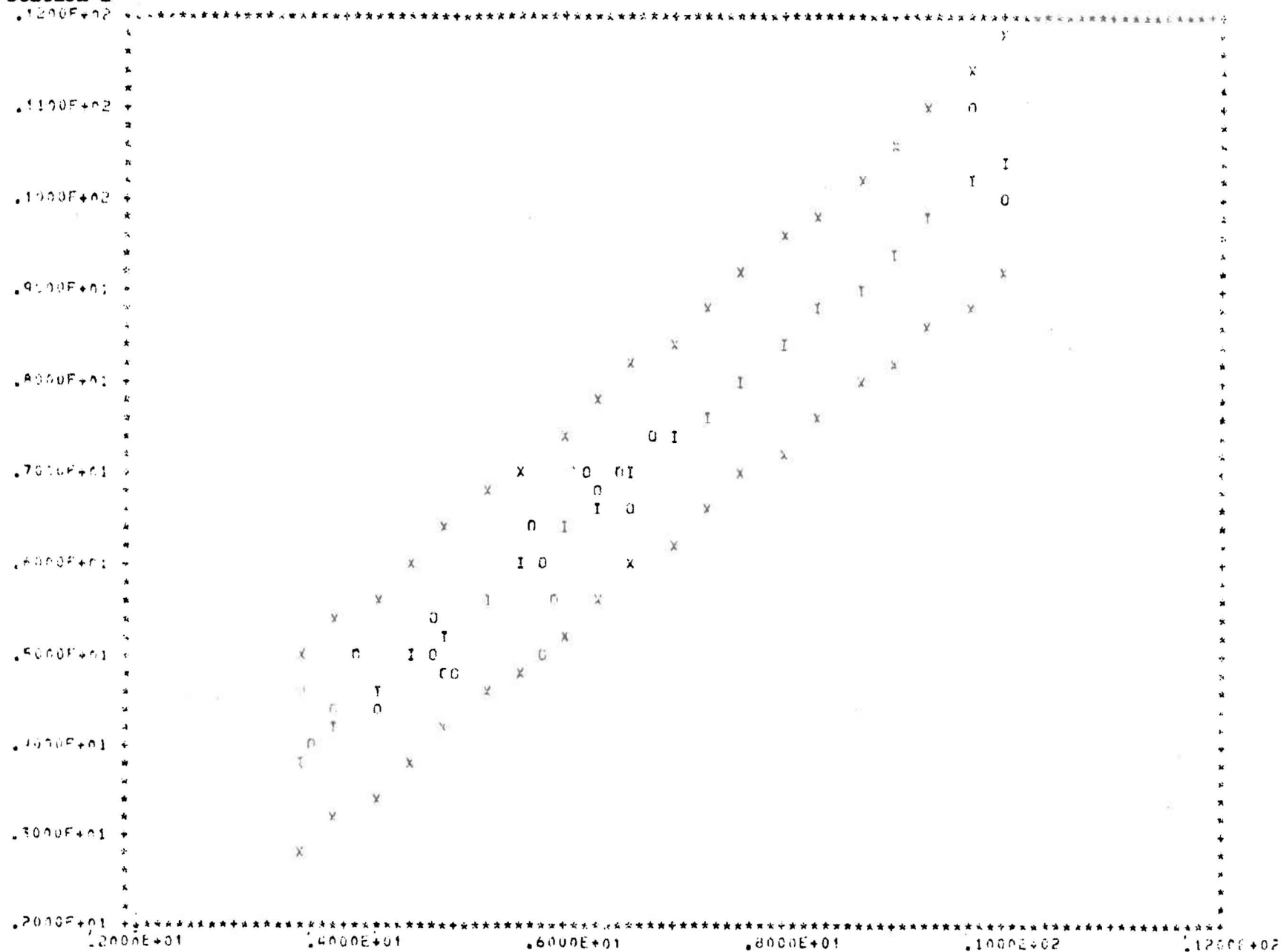
$r = 3.82 \pm .92X$

MAXIMUM = 37.00

MINIMUM = 17.70

Calcium (mg Ca/l)

Station 1



Station 3

Y = .20 + 1.03X MAXIMUM = 10.00 MINIMUM = 3.60

Magnésium (mg Mg/l)

Station 1

.4000E+02

.5400E+02

.4800E+02

.4200E+02

.3600E+02

.3000E+02

.2400E+02

.1800E+02

.1200E+02

.6000E+01

.2200E+01

Y = 2.46 ± .89X

MAXIMUM = 45.00

MINIMUM = 8.00

Sodium (mg Na/l)

Station 3

.6000E+02

.4800E+02

.3600E+02

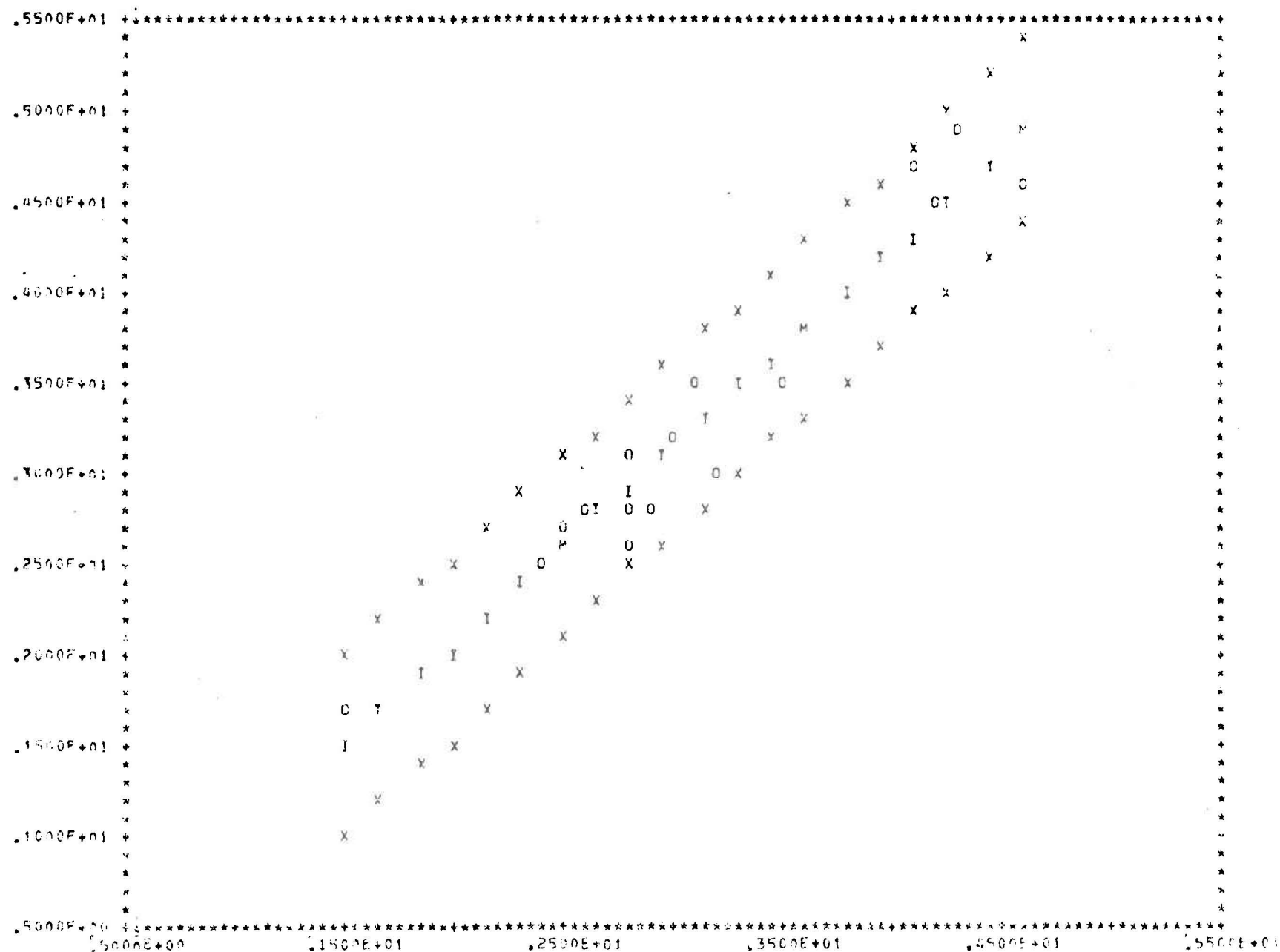
.2400E+02

.1200E+02

.6000E+01

.2200E+01

Station 1



Station 3

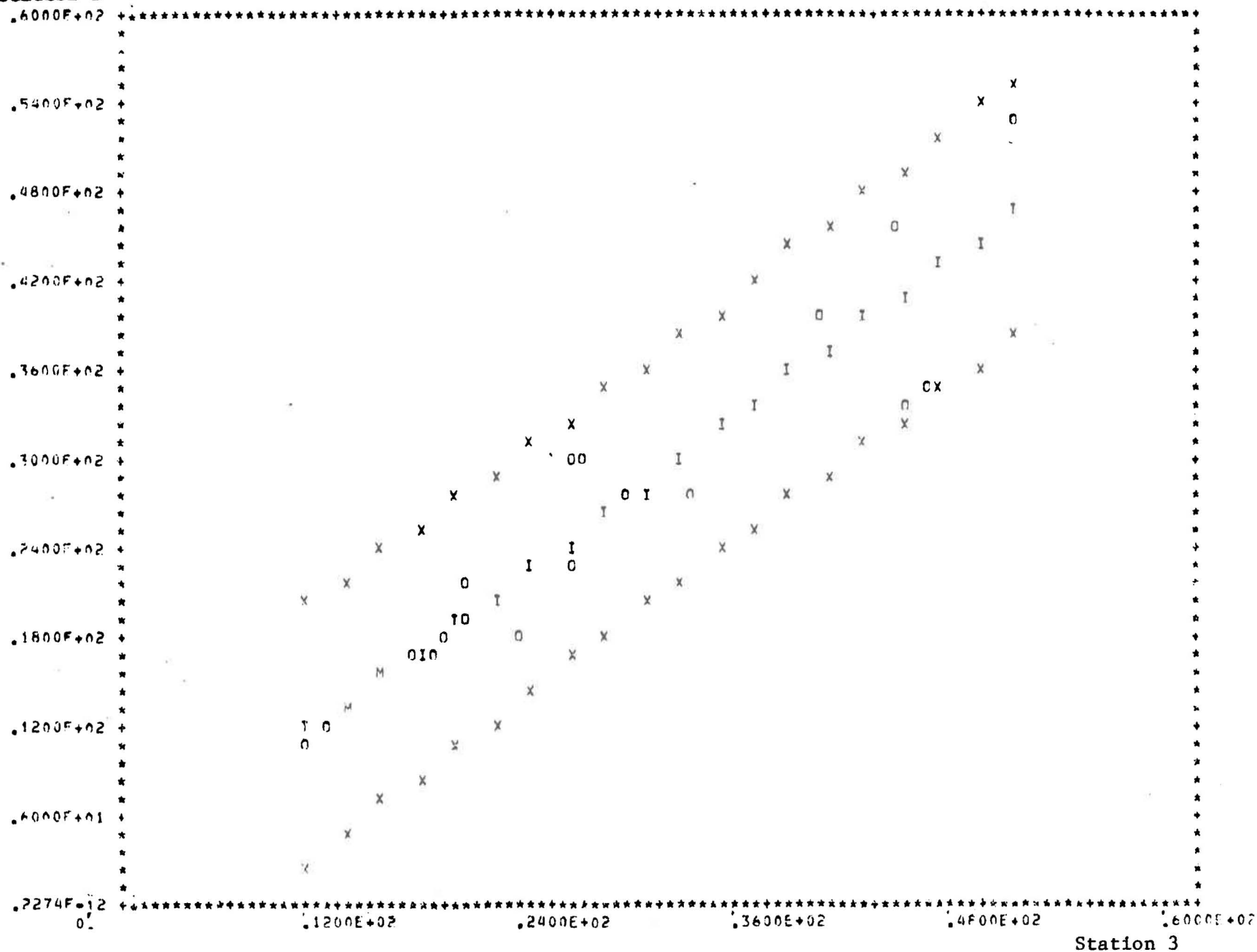
$$Y = -.11 + 1.08X$$

MAXIMUM = 4.60

MINIMUM = 1.50

Potassium (mg K/l)

Station 1



$$y = 2.17 + .89x$$

MAXIMUM = 50.00

MINIMUM = 10.40

Chlorures (mg Cl/l)

Station 1

.6000F+02

.5400F+02

.4800F+02

.4200F+02

.3600F+02

.3000F+02

.2400F+02

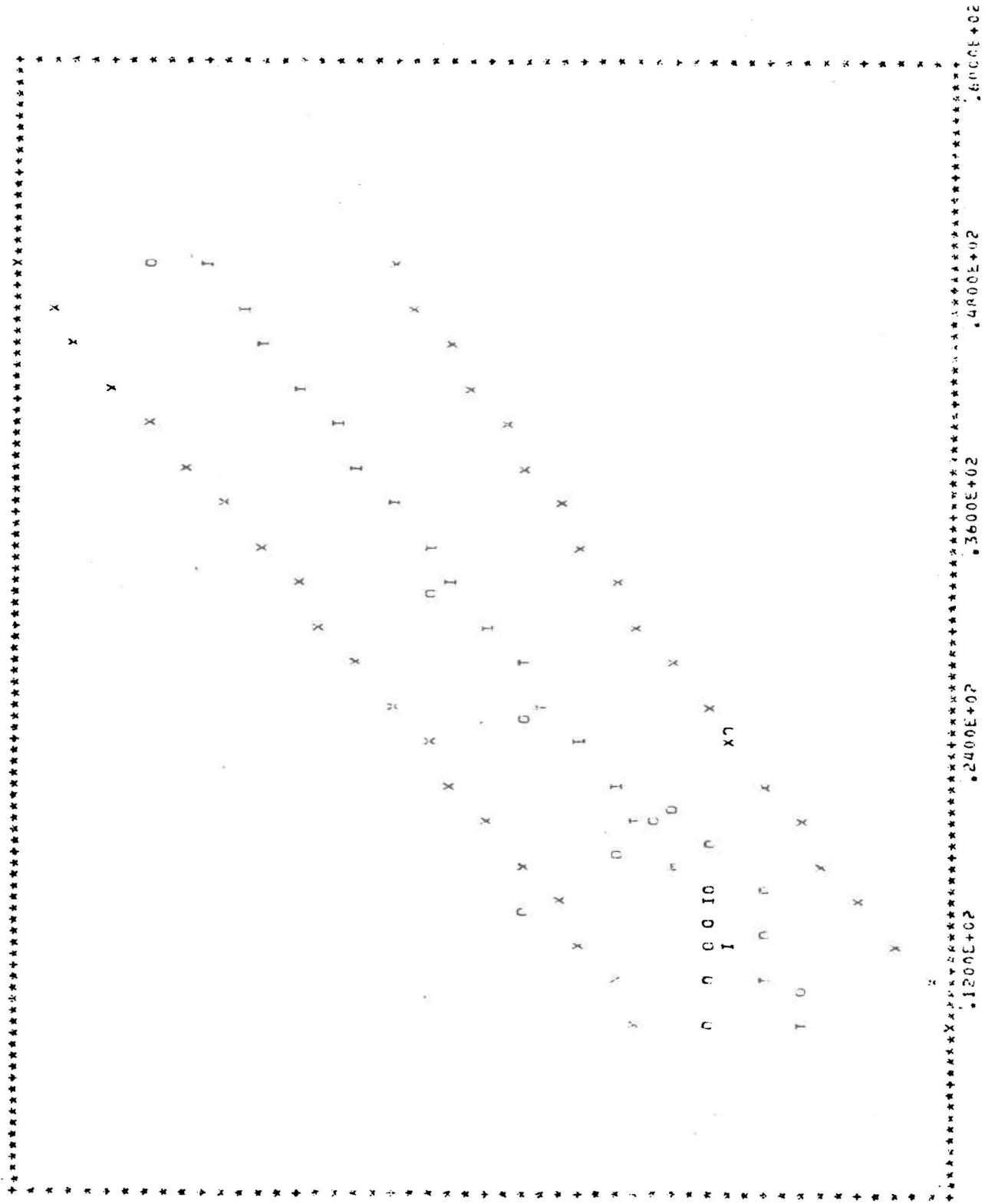
.1800F+02

.1200F+02

.6000F+01

.2270F+02

0.



Y = 1.18 + .96X

MAXIMUM = 49.00

MINIMUM = 9.00

Sulfates (mg SO₄/l)

ANNEXE A-3

ANNEXE A-3Régressions linéaires entre les stations1 et 3 pour chaque paramètre (sauf Fer et Manganèse):Reconstitution et précision.

Chaque tableau indique pour le paramètre considéré, les valeurs reconstituées à la station 1, à partir des valeurs à la station 3 (cf. 4.4.1.2), l'intervalle de confiance à 95% de la reconstitution et la précision relative de cette reconstitution en pourcentage.

Valeur reconstituée	Intervalle de confiance à 95%		Précision relative de la reconstitution (en %)
2.03	3.50	.56	72.56
2.31	3.76	.86	62.83
2.60	4.03	1.16	55.33
2.88	4.31	1.46	49.40
3.17	4.58	1.75	44.62
3.45	4.86	2.05	40.70
3.74	5.14	2.34	37.45
4.02	5.42	2.63	34.72
4.31	5.71	2.91	32.42
4.59	5.99	3.20	30.46
4.88	6.28	3.48	28.78
5.17	6.58	3.75	27.33
5.45	6.87	4.03	26.09
5.74	7.17	4.30	25.01
6.02	7.47	4.57	24.07
6.31	7.77	4.84	23.26
6.59	8.08	5.11	22.55
6.88	8.39	5.37	21.93
7.16	8.69	5.63	21.39
7.45	9.01	5.89	20.92

Silice (mg SiO₂ /l)

$$y = .99 x + .25$$

Valeur reconstituée	Intervalle de confiance à 95%		Précision relative de la reconstitution (en %)
189.55	229.28	149.81	20.96
199.47	238.82	160.13	19.72
209.40	248.41	170.40	18.63
219.33	258.04	180.62	17.65
229.26	267.72	190.80	16.78
239.19	277.46	200.92	16.00
249.12	287.24	210.99	15.31
259.04	297.08	221.00	14.68
268.97	306.98	230.97	14.13
278.90	316.92	240.88	13.63
288.83	326.92	250.74	13.19
298.76	336.97	260.54	12.79
308.68	347.07	270.29	12.44
318.61	357.23	280.00	12.12
328.54	367.43	289.65	11.84
338.47	377.69	299.25	11.59
348.40	387.99	308.80	11.37
358.33	398.34	318.31	11.17
368.25	408.74	327.77	10.99
378.18	419.18	337.18	10.84

Conductivité ($\mu\text{mhos/cm}$)

$$y = 1.01 x + 6.96$$

Valeur reconstituée	Intervalle de confiance à 95%		Précision relative de la reconstitution (en %)
.47 .57 .67 .77 .87 .97 1.08 1.18 1.28 1.38 1.48 1.58 1.68 1.79 1.89 1.99 2.09 2.19 2.29 2.39	.82 .92 1.02 1.11 1.21 1.31 1.41 1.51 1.61 1.71 1.81 1.92 2.02 2.12 2.23 2.33 2.44 2.54 2.65 2.76	.11 .22 .33 .43 .53 .64 .74 .84 .95 1.05 1.15 1.25 1.35 1.45 1.55 1.64 1.74 1.84 1.94 2.03	75.61 61.36 51.49 44.28 38.81 34.53 31.11 28.33 26.03 24.11 22.49 21.12 19.94 18.93 18.06 17.30 16.65 16.07 15.57 15.13

Azote total (mg N/l)

$$y = .94 x + .11$$

Valeur reconstituée	Intervalle de confiance à 95%		Précision relative de la reconstitution (en %)
.05	.25	- .14	362.87
.12	.31	- .07	158.67
.19	.38	- .00	101.24
.26	.45	.07	74.26
.32	.51	.13	58.65
.39	.58	.20	48.50
.46	.65	.27	41.42
.53	.72	.34	36.20
.59	.78	.40	32.23
.66	.85	.47	29.11
.73	.92	.53	26.60
.80	.99	.60	24.56
.86	1.06	.67	22.86
.93	1.13	.73	21.44
1.00	1.20	.80	20.24
1.07	1.27	.86	19.21
1.13	1.34	.93	18.33
1.20	1.41	.99	17.56
1.27	1.48	1.05	16.89
1.34	1.55	1.12	16.30

Azote ammoniacal (mg N/l)

$$y = .92 x + .05$$

Valeur reconstituée	Intervalle de confiance à 95%		Précision relative de la reconstitution (en %)
- .03	.20	- .25	- 909.02
.04	.26	- .19	613.95
.10	.32	- .12	225.01
.16	.38	- .06	136.30
.22	.44	.01	97.13
.28	.49	.07	75.15
.34	.55	.13	61.14
.40	.61	.20	51.48
.47	.67	.26	44.45
.53	.73	.32	39.13
.59	.80	.38	35.00
.65	.86	.44	31.70
.71	.92	.51	29.04
.77	.98	.57	26.85
.83	1.04	.63	25.03
.90	1.11	.69	23.51
.96	1.17	.74	22.22
1.02	1.23	.80	21.13
1.08	1.30	.86	20.19
1.14	1.36	.92	19.39

Nitrates (mg N/l)

$$y = 1.02 x - 0.4$$

Valeur reconstituée	Intervalle de confiance à 95%		Précision relative de la reconstitution (en %)
.05	.10	-.00	109.66
.06	.11	.01	89.45
.07	.12	.02	75.39
.08	.13	.03	65.10
.09	.14	.04	57.29
.10	.15	.05	51.20
.11	.16	.06	46.36
.12	.17	.07	42.43
.13	.18	.08	39.22
.14	.19	.09	36.56
.15	.20	.10	34.34
.16	.21	.11	32.47
.17	.22	.12	30.89
.18	.23	.13	29.54
.19	.24	.13	28.40
.20	.25	.14	27.41
.21	.26	.15	26.57
.22	.27	.16	25.85
.23	.29	.17	25.22
.24	.30	.18	24.68

Phosphore total (mg P/l)

$$y = .79 x + .01$$

Valeur reconstituée	Intervalle de confiance à 95%		Précision relative de la reconstitution (en %)
.02	.06	- .02	240.93
.03	.07	- .01	145.94
.04	.08	- .00	104.02
.05	.09	.01	80.55
.06	.10	.02	65.63
.07	.11	.03	55.39
.08	.12	.04	47.97
.09	.13	.05	42.39
.10	.14	.06	38.08
.12	.16	.08	34.68
.13	.17	.09	31.95
.14	.18	.10	29.72
.15	.19	.11	27.90
.16	.20	.12	26.38
.17	.21	.13	25.11
.18	.22	.14	24.04
.19	.24	.15	23.14
.20	.25	.16	22.37
.21	.26	.17	21.72
.22	.27	.18	21.15

Phosphore hydrolysable (mg P/l)

$$y = .83 x + .01$$

Valeur reconstituée	Intervalle de confiance à 95%		Précision relative de la reconstitution (en %)
20.04	24.57	15.50	22.63
20.97	25.45	16.48	21.39
21.90	26.34	17.46	20.27
22.83	27.23	18.43	19.28
23.76	28.13	19.39	18.39
24.69	29.04	20.34	17.60
25.62	29.95	21.29	16.90
26.55	30.87	22.23	16.27
27.48	31.80	23.17	15.70
28.41	32.73	24.09	15.20
29.34	33.68	25.01	14.76
30.28	34.63	25.93	14.37
31.21	35.58	26.83	14.02
32.14	36.55	27.73	13.72
33.07	37.52	28.62	13.45
34.00	38.49	29.51	13.22
34.93	39.48	30.38	13.01
35.86	40.46	31.26	12.84
36.79	41.46	32.12	12.69
37.72	42.46	32.98	12.56

Calcium (mg Ca/l)

$$y = .92 x + 3.82$$

Valeur reconstituée	Intervalle de confiance à 95%		Précision relative de la reconstitution (en %)
3.90	5.03	2.77	29.01
4.24	5.36	3.12	26.37
4.59	5.70	3.48	24.17
4.93	6.03	3.83	22.32
5.28	6.37	4.18	20.75
5.62	6.71	4.53	19.41
5.97	7.06	4.88	18.26
6.31	7.40	5.22	17.27
6.66	7.75	5.57	16.41
7.00	8.10	5.91	15.67
7.35	8.45	6.25	15.02
7.70	8.81	6.58	14.46
8.04	9.16	6.92	13.97
8.39	9.52	7.25	13.54
8.73	9.88	7.58	13.17
9.08	10.24	7.91	12.84
9.42	10.61	8.24	12.56
9.77	10.97	8.57	12.31
10.11	11.34	8.89	12.10
10.46	11.71	9.21	11.91

Magnésium (mg Mg/l)

$$y = 1.03 x + .20$$

Valeur reconstituée	Intervalle de confiance à 95%		Précision relative de la reconstitution (en %)
10.25	16.73	3.77	63.19
11.94	18.35	5.53	53.68
13.62	19.97	7.27	46.61
15.31	21.62	9.00	41.19
17.00	23.27	10.72	36.91
18.68	24.94	12.43	33.48
20.37	26.62	14.12	30.68
22.06	28.31	15.80	28.36
23.74	30.02	17.47	26.44
25.43	31.74	19.12	24.81
27.11	33.47	20.76	23.44
28.80	35.22	22.39	22.27
30.49	36.97	24.00	21.27
32.17	38.74	25.61	20.42
33.86	40.52	27.20	19.68
35.55	42.31	28.78	19.04
37.23	44.12	30.35	18.49
38.92	45.93	31.91	18.01
40.61	47.75	33.46	17.59
42.29	49.58	35.01	17.23

Sodium (mg Na/l)

$$y = .89 x + 2.46$$

Valeur reconstituée	Intervalle de confiance à 95%		Précision relative de la reconstitution (en %)
1.51 1.69 1.87 2.04 2.22 2.40 2.58 2.75 2.93 3.11 3.28 3.46 3.64 3.81 3.99 4.17 4.35 4.52 4.70 4.88	2.03 2.20 2.37 2.54 2.71 2.88 3.06 3.23 3.40 3.58 3.75 3.93 4.11 4.29 4.47 4.65 4.83 5.01 5.19 5.38	.99 1.18 1.37 1.55 1.73 1.91 2.10 2.28 2.46 2.64 2.81 2.99 3.17 3.34 3.52 3.69 3.86 4.03 4.20 4.37	34.29 30.20 26.92 24.24 22.03 20.17 18.61 17.28 16.14 15.15 14.31 13.57 12.94 12.38 11.90 11.48 11.12 10.80 10.53 10.29

Potassium (mg K/l)

$$y = 1.08 x - .11$$

Valeur reconstituée	Intervalle de confiance à 95%		Précision relative de la reconstitution (en %)
11.44	19.90	2.98	73.95
13.30	21.68	4.92	63.02
15.16	23.47	6.84	54.84
17.01	25.27	8.76	48.51
18.87	27.08	10.66	43.49
20.73	28.91	12.56	39.43
22.59	30.74	14.44	36.08
24.45	32.59	16.31	33.29
26.31	34.45	18.17	30.94
28.17	36.32	20.01	28.95
30.02	38.20	21.85	27.24
31.88	40.09	23.67	25.76
33.74	42.00	25.48	24.48
35.60	43.92	27.28	23.37
37.46	45.85	29.07	22.39
39.32	47.78	30.85	21.54
41.17	49.73	32.62	20.79
43.03	51.69	34.37	20.12
44.89	53.66	36.12	19.53
46.75	55.64	37.86	19.01

Chlorures (mg Cl/l)

$$y = .89 x + 2.17$$

Valeur reconstituée	Intervalle de confiance à 95%		Précision relative de la reconstitution (en %)
9.81 11.83 13.85 15.87 17.90 19.92 21.94 23.96 25.98 28.00 30.02 32.04 34.06 36.08 38.10 40.12 42.14 44.16 46.18 48.20	19.98 21.89 23.82 25.78 27.77 29.78 31.82 33.89 35.98 38.09 40.24 42.40 44.59 46.80 49.03 51.28 53.55 55.83 58.14 60.45	- .35 1.78 3.89 5.97 8.02 10.05 12.05 14.03 15.98 17.90 19.80 21.68 23.53 25.36 27.17 28.96 30.73 32.49 34.23 35.95	103.56 84.94 71.94 62.40 55.17 49.53 45.06 41.45 38.50 36.06 34.04 32.34 30.92 29.71 28.69 27.82 27.07 26.43 25.88 25.41

Sulfates (mg SO_4 /l)

$$y = .96 x + 1.18$$