

Record Number: 1180

Author, Monographic: Lachance, M.//Bobée, B.//Desjardins, M.//Bertrand, R.//Ricbourg, J. P.

Author Role:

Title, Monographic: Étude comparative de deux méthodes de mesure d'empoussiérage dans l'air - Phase 2

Translated Title:

Reprint Status:

Edition:

Author, Subsidiary:

Author Role:

Place of Publication: Québec

Publisher Name: INRS-Eau

Date of Publication: 1979

Original Publication Date: Octobre 1979

Volume Identification:

Extent of Work: v, 54

Packaging Method: pages et 4 annexes

Series Editor:

Series Editor Role:

Series Title: INRS-Eau, Rapport de recherche

Series Volume ID: 117

Location/URL:

ISBN: 2-89146-114-2

Notes: Rapport annuel 1979-1980

Abstract: Rapport rédigé pour la Société Asbestos Corporation
15.00\$

Call Number: R000117

Keywords: rapport/ ok/ dl

ETUDE COMPARATIVE DE DEUX METHODES
DE MESURE D'EMPOUSSIERAGE DANS L'AIR

PHASE 2

Rapport scientifique No 117

Soumis à la société Asbestos Corporation

par

M. Lachance et B. Bobée
INRS-Eau

et

M. Desjardins, R. Bertrand et J.P. Ricbourg
INRS-Pétrole

Octobre 1979

INSTITUT NATIONAL DE LA RECHERCHE SCIENTIFIQUE
Université du Québec
C.P. 7500, Sainte-Foy
Québec, GIV 4C7

TABLE DES MATIERES

	PAGE
1. INTRODUCTION.....	1
1.1 Résumé de l'étude précédente.....	2
1.2 Données considérées.....	2
1.3 Approche utilisée.....	4
2. Résultats de l'étude statistique.....	5
2.1 Description des séries de mesures.....	5
2.2 Comparaison globale des trois séries de mesures...	14
2.3 Détermination de groupes de stations.....	22
2.4 Relations linéaires entre les séries de mesures...	31
2.4.1 Coefficient de corrélation.....	31
2.4.2 Linéarité et dispersion des relations.....	38
2.4.3 Régressions linéaires et intervalles de confiance associés.....	43
Conclusion.....	51
Références bibliographiques.....	54
Annexe A: Données.....	A-1
Annexe B: Classification ascendante hiérarchique.....	B-1
Annexe C Graphiques.....	C-1
Annexe D Etude morphologique au microscope électronique à balayage	D-1

LISTE DES TABLEAUX

PAGE

Tableau 2.1.	Description de la distribution des valeurs obtenues pour les séries A, B et C.....	6
Tableau 2.2.	Caractéristiques statistiques des 3 séries de mesures.....	12
Tableau 2.3.	Séparation des données en 5 classes: seuils et effectif de chaque classe.....	15
Tableau 2.4.	Coordonnées et contributions relatives sur les 4 premiers axes factoriels des points représen- tatifs des classes de valeurs pour chacune des séries de mesures.....	18
Tableau 2.5.	Tableaux de contingence obtenus en croisant deux à deux les séries A, B et C.....	23
Tableau 2.6.	Composition des groupes de stations déterminés par la classification ascendante hiérarchique.....	26
Tableau 2.7.	Caractéristiques statistiques de chacun des groupes de stations déterminés par la clas- sification ascendante hiérarchique.....	29
Tableau 2.8.	Coefficients de corrélation entre les séries A, B et C pour différents intervalles des valeurs de A.....	33
Tableau 2.9.	Coefficients de corrélation entre les séries A, B et C pour chaque groupe de stations.....	36

LISTE DES TABLEAUX (SUITE)

PAGE

Tableau 2.10. Nombre, moyenne et écart-type des valeurs de B groupées en 10 classes équiprobables, moyenne et écart-type des valeurs de A cor- respondantes.....	39
Tableau 2.11. Nombre, moyenne et écart-type des valeurs de C groupées en 10 classes équiprobables, moyenne et écart-type des valeurs de A correspondantes.....	40
Tableau 2.12. Equations de régression et coefficients de détermination pour l'ensemble des stations et pour un groupe de stations.....	45

LISTE DES FIGURES

	PAGE
Figure 2.1	Distribution des valeurs obtenues pour la série A... 7
Figure 2.2	Distribution des valeurs obtenues pour la série B... 8
Figure 2.3	Distribution des valeurs obtenues pour la série C... 9
Figure 2.4	Représentation dans le plan des deux premiers axes principaux des points paramètres (3 séries, 5 classes) et des centres de gravité des groupes de points observations (stations) classés hiérarchiquement.....17
Figure 2.5	Représentation des distances euclidiennes calculées dans l'espace factoriel (4 dimensions) entre les points représentant les 5 classes des trois séries de mesures A, B et C.....20
Figure 2.6	Partie supérieure de l'arbre montrant la hiérarchie des 5 groupes de stations déterminés à partir des coordonnées sur les 4 premiers axes factoriels.....25
Figure 2.7	Relation entre la série B et la série A: étude de la linéarité et de la dispersion.....41
Figure 2.8	Relation entre la série C et la série A: étude de la linéarité et de la dispersion.....42
Figure 2.9	Régression linéaire simple pour l'ensemble des mesures entre la méthode optique (A) et la méthode du RDM (B) et intervalles de confiance associés.....46

LISTE DES FIGURES (SUITE)

PAGE

Figure 2.10	Régression linéaire pour l'ensemble des mesures entre la méthode optique (A) et la méthode du RDM (C) et intervalles de confiance associés.....	47
Figure 2.11	Régression linéaire entre les séries de mesures au RDM à la fin de l'échantillonnage (C) et au début de l'échantillonnage (B) et intervalles de confiance associés.....	48

INTRODUCTION

Dans cette étude, on vise à:

- poursuivre l'étude entreprise précédemment pour la Société Asbestos Corporation (DESJARDINS et al., 1979), concernant l'existence d'une relation entre les mesures optiques (comptage de fibres d'amiante au microscopie) et les mesures physiques (mesure au RDM de l'absorption des rayons β par une masse de poussière);
- étudier l'existence potentielle d'une relation entre ces deux méthodes de mesure avec un nombre plus réduit et plus homogène de stations d'échantillonnage;
- vérifier la représentativité des échantillons recueillis avec les deux systèmes de mesure en comparant au microscope électronique à balayage, la composition morphologique des deux types d'échantillonnage.

La poursuite de cette étude a comme objectif de tenter de répondre à la question qui avait été soulevée dans l'étude précédente à savoir si les deux méthodes mesurent la même chose du point de vue de la dimension et de la nature des particules. Les données qui avaient été considérées dans la première étude ne permettent pas de répondre à cette question, à cause de la trop grande disparité dans les postes de mesure.

Dans ce qui suit, on rappelle brièvement les résultats de l'étude déjà effectuée (DESJARDINS et al., 1979) et après avoir caractérisé les données que l'on considère, on donne les grandes lignes de l'approche statistique qui sera suivie.

1.1 Résumé de l'étude précédente

L'étude précédente (phase I) a été faite sur 900 mesures effectuées par la Société Asbestos Corporation, lors d'une campagne de mesure d'empoussiérage dans l'air à divers endroits des usines et des mines pendant la période du 13 février au 21 avril 1978. Cette étude a fait ressortir:

- une très grande dispersion entre les valeurs obtenues pour chacune des mesures;
- la présence d'une certaine relation entre les deux types de mesures surtout évidentes pour des comptages supérieurs à 6 fibres/cm³;
- un intérêt réduit dans l'utilisation de cette relation en raison de la grande dispersion des valeurs et conséquemment de la grande imprécision dans la prédiction d'une mesure à partir de l'autre mesure;
- une différence non significative statistiquement dans les mesures du RDM effectuées au début et à la fin d'une expérience.

1.2 Données considérées

Les données à analyser pour l'étude statistique consistent en trois

séries de mesures. Comme pour l'étude précédente, la série des mesures optiques en fibres/cm³ est notée A, tandis que les deux autres séries correspondant aux mesures physiques sont notées B et C; la série notée B correspond aux mesures sur des échantillons d'air prélevés au cours des 4 premières minutes de l'échantillonnage tandis que la série notée C est obtenue sur des échantillons d'air prélevés pendant les 4 dernières minutes de l'échantillonnage, c'est-à-dire 82 minutes plus tard.

Chacune des mesures provient de 174 postes d'échantillonnage localisés à l'intérieur de trois usines de transformation de fibres d'amiante. Ces trois usines sont appelées British-Canadian # 1 (BC1), British-Canadian #2 (BC2) et Normandie (N). Chacun des postes d'échantillonnage est situé dans le circuit de transformation et de concentration de la fibre d'amiante. On dispose donc d'un ensemble de mesures provenant de stations d'échantillonnage possédant une certaine homogénéité. Ces mesures ont été effectuées pendant les périodes suivantes:

6 - 8 mars 1978,	moulin Normandie
24 février - 13 mars 1978,	moulin British-Canadian #1
13 - 16 février 1978,	moulin British-Canadian #2

Les données pour chacune des trois séries de mesures obtenues au cours de cette campagne apparaissent dans l'annexe A.

1.3 Approche utilisée

Pour l'étude mathématique, l'approche qui a été utilisée dans la comparaison des trois séries est essentiellement la même que celle utilisée dans la phase I de l'étude. Elle se divise en trois parties:

- on décrit l'ensemble des valeurs pour chacune des séries en présentant des histogrammes de fréquence de distribution et en calculant des caractéristiques statistiques globales;
- on procède à une comparaison globale des trois séries de mesures au moyen de l'analyse factorielle des correspondances. Dans cette phase de l'étude, on pousse plus loin l'analyse en délimitant un certain nombre de groupes de postes d'échantillonnage qui ont une similitude de comportement; ceci est effectué au moyen d'une méthode de classification appliquée sur les résultats obtenus de l'analyse factorielle des correspondances;
- on calcule ensuite des coefficients de corrélation entre chacune des séries de mesures prises deux à deux pour chacun des groupes de stations et pour différentes gammes de variation des valeurs obtenues au microscope optique.

Pour compléter l'analyse, on fait ressortir la linéarité et la dispersion dans la relation des séries B et C avec A et on calcule entre chacune des séries prises deux à deux, des équations de régression et leurs intervalles de confiance.

Les méthodes statistiques utilisées sont sensiblement les mêmes que celles employées dans la phase I de l'étude. Dans cette étude, on utilise en outre une méthode de classification hiérarchique (JAMBU, 1972; 1976) qui sera décrite sommairement en annexe B et on calcule des régressions linéaires simples avec leurs intervalles de confiance.

2. RESULTATS DE L'ETUDE STATISTIQUE

Les résultats de l'étude statistique sont décrits dans cette partie et la méthodologie utilisée est décrite brièvement lorsqu'elle diffère de celle élaborée dans la phase I de l'étude.

2.1 Description des séries de mesures

Pour obtenir les histogrammes de fréquence de distribution des données, on a encore séparé l'intervalle de variation de chacune des séries en 14 petits intervalles égaux (classes) correspondant à 1 fibre/cm³ pour la série A et à 0.1 mg/m³ pour les séries B et C. On a considéré dans deux classes séparées, d'une part les valeurs nulles et d'autre part les valeurs supérieures à 14 fibres/cm³ pour la série A et à 1.4 mg/m³ pour les séries B et C. On a ensuite compté le nombre de fois qu'on rencontre des mesures comprises dans ces classes. Les résultats de ce comptage apparaissent au tableau 2.1 et les histogrammes de fréquence absolue des distributions sont visualisés, dans les figures 2.1, 2.2 et 2.3 respectivement pour les séries A, B et C. Ces figures donnent les fréquences absolues pour chaque classe; la fréquence relative peut s'en déduire en divisant la fréquence observée par le nombre total d'observations (N = 174).

Tableau 2.1 Description de la distribution des valeurs obtenues pour les séries A, B et C.

classe	Série A				Série B				Série C			
	intervalle fibres/cm ³	fréquence absolue	fréquence relative (%)	fréquence cumulée (%)	intervalle (mg/m ³)	fréquence absolue	fréquence relative (%)	fréquence cumulée (%)	intervalle (mg/m ³)	fréquence absolue	fréquence relative (%)	fréquence cumulée (%)
1	= 0.0	0	0.0	0.0	= .00	16	9.2	9.2	= .00	11	6.3	6.3
2	0.0 - 1.0	26	14.9	14.9	.00 - .10	16	9.2	18.4	.00 - .10	19	10.9	17.2
3	1.0 - 2.0	43	24.7	39.7	.10 - .20	17	9.8	28.2	.10 - .20	19	10.9	28.2
4	2.0 - 3.0	48	27.6	67.2	.20 - .30	31	17.8	46.0	.20 - .30	25	14.4	42.5
5	3.0 - 4.0	23	13.2	80.5	.30 - .40	38	21.8	67.8	.30 - .40	35	20.1	62.6
6	4.0 - 5.0	14	8.0	88.5	.40 - .50	15	8.6	76.4	.40 - .50	19	10.9	73.6
7	5.0 - 6.0	6	3.4	92.0	.50 - .60	11	6.3	82.8	.50 - .60	12	6.9	80.5
8	6.0 - 7.0	8	4.6	96.6	.60 - .70	17	9.8	92.5	.60 - .70	13	7.5	87.9
9	7.0 - 8.0	1	.6	97.1	.70 - .80	1	.6	93.1	.70 - .80	7	4.0	92.0
10	8.0 - 9.0	0	.0	97.1	.80 - .90	4	2.3	95.4	.80 - .90	3	1.7	93.7
11	9.0 - 10.0	1	.6	97.7	.90 - 1.00	3	1.7	97.1	.90 - 1.00	2	1.1	94.8
12	10.0 - 11.0	1	.6	98.3	1.00 - 1.10	1	.6	97.7	1.00 - 1.10	2	1.1	96.0
13	11.0 - 12.0	1	.6	98.9	1.10 - 1.20	0	.0	97.7	1.10 - 1.20	3	1.7	97.7
14	12.0 - 13.0	0	.0	98.9	1.20 - 1.30	0	.0	97.7	1.20 - 1.30	0	.0	97.7
15	13.0 - 14.0	0	.0	98.9	1.30 - 1.40	1	.6	98.3	1.30 - 1.40	0	.0	97.7
16	> 14.0	2	1.1	100.0	> 1.40	3	1.7	100.0	> 1.40	4	2.3	100.0

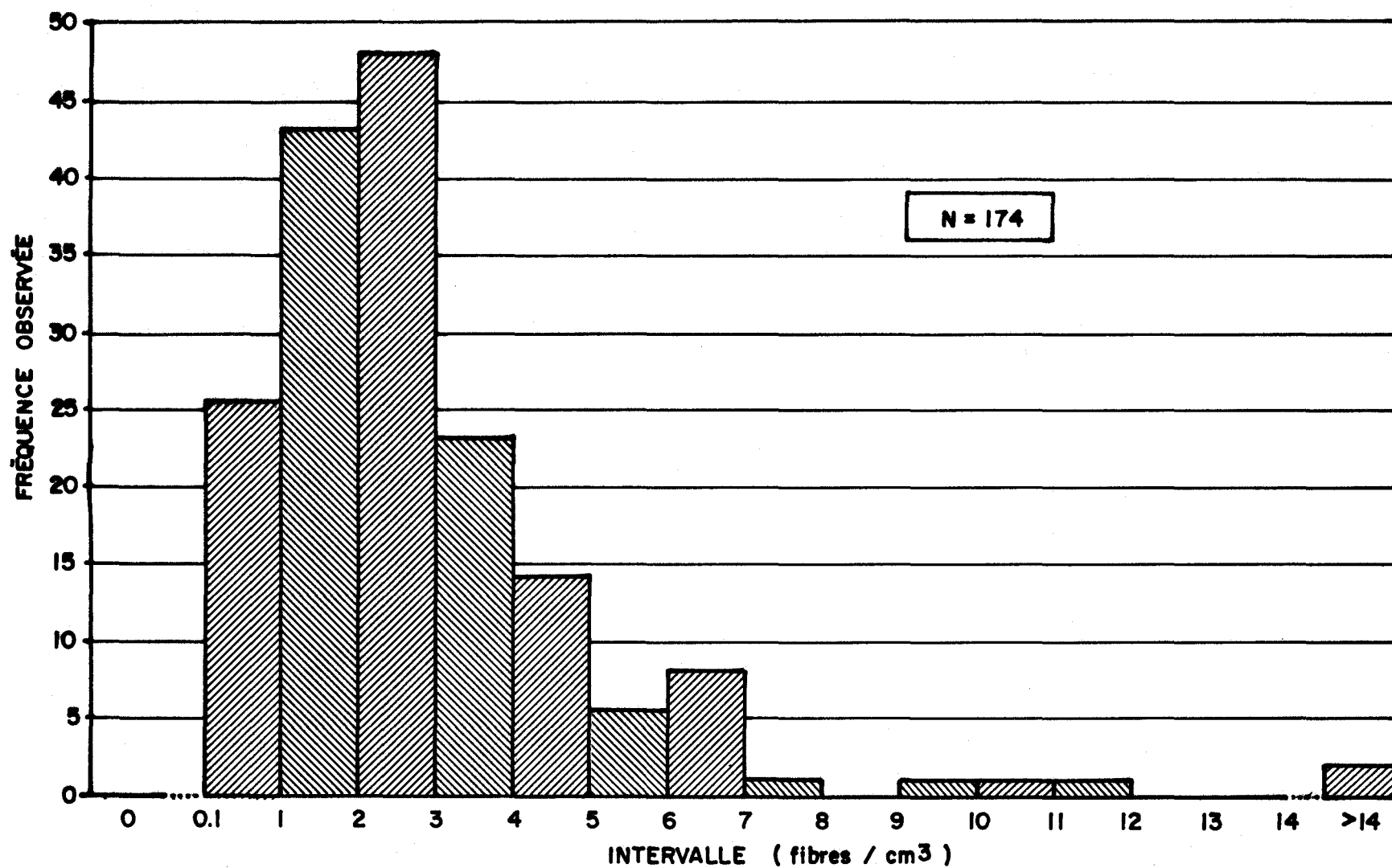


Figure 2.1. Distribution des valeurs obtenues pour la série A .

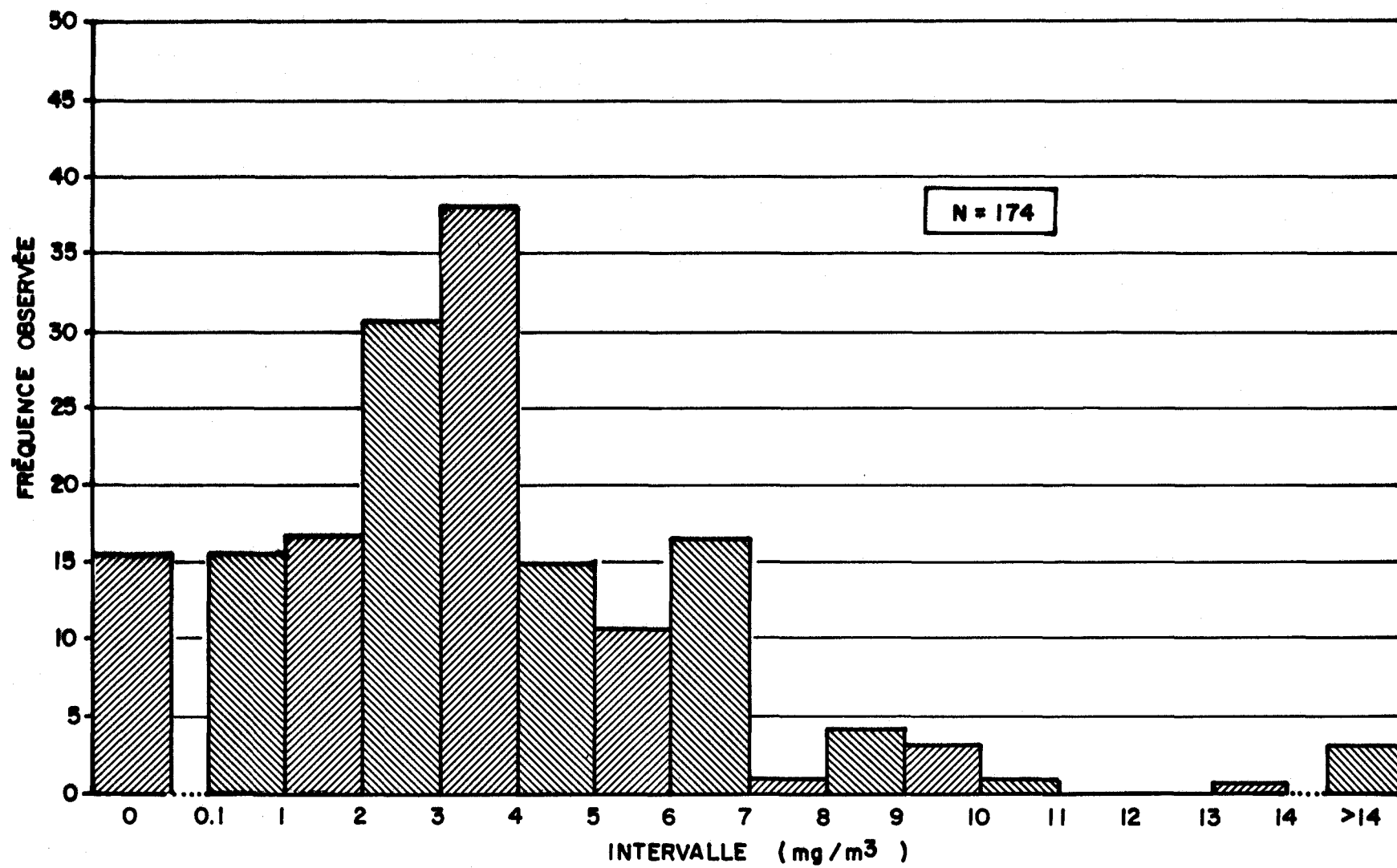


Figure 2.2. Distribution des valeurs obtenues pour la série B .

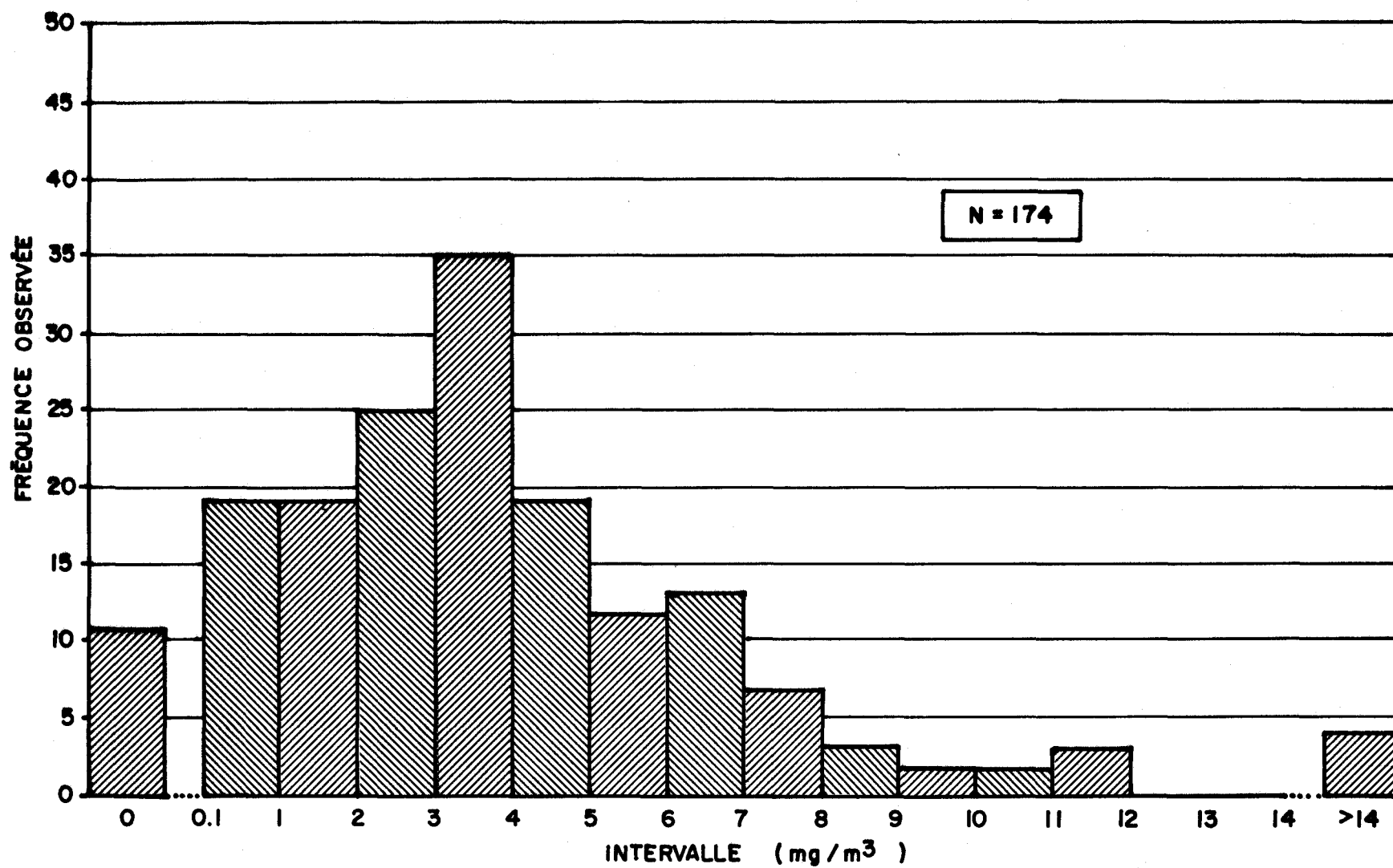


Figure 2.3. Distribution des valeurs obtenues pour la série C .

Pour les trois séries, on observe une distribution avec une légère asymétrie positive, c'est-à-dire avec une queue vers la droite. La fréquence maximum (mode) pour la série A se situe entre 2 et 3 fibres/cm³ et pour les séries B et C entre 0.3 et 0.4 mg/m³.

L'examen du tableau 2.1 montre qu'environ 75% des valeurs de B et C sont inférieures à 0.5 mg/m³. Quant aux valeurs de la série A, on note que 67% et 89% des valeurs sont inférieures à 3 et à 5 fibres/cm³ respectivement. Par conséquent, la majorité des mesures est constituée de valeurs plutôt faibles. En examinant les données brutes (annexe A), on retrouve rapidement les deux stations d'échantillonnage qui ont des valeurs nettement supérieures et qui apparaissent dans la dernière classe. Ces valeurs extrêmes peuvent être dues à des conditions très particulières non représentatives de l'ensemble de l'échantillonnage.

	A	B	C
BC1- 6072	51.7	2.27	2.71
BC1- 6084	83.2	2.82	3.42

Les séries B et C ont une distribution semblable; la série A a une distribution légèrement différente des deux autres, car elle apparaît étalée sur un nombre plus réduit de classes. La distribution différente pour la série A peut être influencée jusqu'à un certain point par le choix des intervalles délimitant les classes. On verra plus loin qu'il est possible de préciser cette observation en calculant les coefficients d'asymétrie et d'aplatissement pour chacune des séries.

Si on compare les distributions obtenues dans cette étude avec celles obtenues dans la phase I de l'étude avec 900 postes de mesures, on constate que:

- la distribution des valeurs de chacune des séries considérées dans cette étude a une allure tendant davantage vers une distribution normale (asymétrie plus faible). L'homogénéité des stations d'échantillonnage peut expliquer ce fait;
- la dispersion des valeurs pour chacune des séries est plus faible. Un nombre relativement faible apparaît dans la classe des valeurs supérieures. Cette plus faible dispersion est aussi indicatrice d'une plus grande homogénéité dans les stations d'échantillonnage retenues.

Afin de préciser les observations déduites de l'examen des histogrammes de fréquence de distribution, on a calculé au tableau 2.2 plusieurs caractéristiques statistiques générales des trois séries de mesures telles que la moyenne (m), l'écart-type (s), le minimum, le maximum, le coefficient de variation (s/m), le coefficient d'asymétrie (C_s) et le coefficient d'aplatissement (C_k). Pour les deux derniers, on a employé les formules suivantes:

$$C_s = \frac{1}{N} \left[\sum_{i=1}^N (x_i - m)^3 / s^3 \right]$$

$$C_k = \frac{1}{N} \left[\sum_{i=1}^N (x_i - m)^4 / s^4 \right] - 3$$

où N est le nombre de mesures et x_i est la i^{e} observation de la série considérée.

Tableau 2.2 Caractéristiques statistiques des 3 séries de mesures.

caractéristiques	SERIES		
	A	B	C
moyenne (m)	3.49 (2.75) *	.37 (.35)	.41 (.38)
écart-type (s)	7.36 (1.89)	.36 (.27)	.43 (.32)
minimum	.02	.00	.00
maximum	83.2 (11.2)	2.82 (1.77)	3.42 (2.17)
coefficient de variation (s/m)	2.1 (.69)	1.0 (.78)	1.0 (.83)
coefficient d'asymétrie (C_s)	9.1 (1.4)	3.3 (1.5)	3.6 (2.1)
coefficient d'aplatissement (C_k)	90.1 (3.3)	17.3 (4.8)	19.5 (8.2)

La moyenne, l'écart-type, le minimum et le maximum sont donnés en fibres/cm³ pour la série A et en mg/m³ pour les séries B et C.

* Les valeurs entre parenthèses sont obtenues en éliminant les valeurs extrêmes observées à BC1-6072 et à BC1-6084.

Rappelons que le coefficient d'asymétrie mesure le degré d'asymétrie d'une distribution de valeurs par rapport à une distribution normale (asymétrie nulle). Il est positif lorsque la distribution a une forte densité de faibles valeurs avec une longue queue de distribution pour les fortes valeurs et négatif dans le cas contraire. Le coefficient d'applatissage mesure le degré d'étroitesse de la distribution par rapport à une distribution normale (applatissage nul). Ce coefficient est négatif lorsque la distribution est plus étalée, et positif lorsqu'elle est plus étroite que la distribution normale.

Les caractéristiques statistiques ont été également calculées en éliminant les valeurs obtenues aux deux stations identifiées hétérogènes; ces valeurs apparaissent entre parenthèses au tableau 2.2. On voit que l'élimination de ces deux valeurs diminue considérablement l'écart-type, le coefficient de variation, le coefficient d'asymétrie et le coefficient d'applatissage, principalement pour la série A et donne des valeurs plus représentatives de l'ensemble des mesures.

En comparant ces caractéristiques avec celles obtenues pour l'ensemble des 900 stations d'échantillonnage, on voit que la moyenne et surtout l'écart-type des valeurs pour les trois séries est plus faible, confirmant l'observation que les données de cette étude sont moins dispersées et proviennent de stations possédant une plus grande homogénéité. La distribution des valeurs pour les trois séries se rapproche plus de la normalité à cause des coefficients d'asymétrie plus faibles (de 1.4 à 2.1 par rapport à l'intervalle observé dans l'étude précédente qui était de 3.9 à 5.0).

Si on compare les trois séries de mesures entre elles, on voit que les coefficients de variation, d'asymétrie et d'aplatissement pour la série A sont plus faibles que pour les séries B et C confirmant que la distribution des valeurs de A a une distribution plus proche de la loi normale et moins dispersée que pour les deux autres séries. Ces caractéristiques légèrement différentes donnent une indication de la nature physique différente dans les mesures effectuées au microscope optique et au RDM.

Enfin, les faibles coefficients d'asymétrie calculés pour chacune des séries, ne justifient pas d'effectuer une transformation (logarithmique par exemple) pour étudier les relations entre chacune des séries; il est en effet nécessaire pour effectuer des études régressives, de disposer de séries provenant d'une loi normale ou ne s'en éloignant pas trop.

2.2 Comparaison globale des trois séries de mesures

Cette analyse faite au moyen de l'analyse factorielle des correspondances vise à comparer globalement les trois séries de mesures et à faire ressortir leurs interrelations. Comme nous le verrons plus loin, cette analyse permettra par l'application d'une méthode de classification, de classer les stations d'échantillonnage en un certain nombre de groupes.

La méthode d'analyse factorielle des correspondances a déjà été décrite dans l'étude précédente (phase I) et la méthodologie d'application est essentiellement la même ici. Dans une première étape, les données relatives aux trois séries de mesures ont été séparées en 5 classes d'effectifs approximativement égaux. Les seuils déterminés pour obtenir ces classes et le nombre de valeurs comprises dans ces classes apparaissent au tableau 2.3. Les

Tableau 2.3 Séparation des données en 5 classes: seuils et effectif de chaque classe.

séries	classe	Borne inférieure (A: fibres/cm ³) (B et C mg/m ³)	Borne supérieure (A: fibres/cm ³) (B et C mg/m ³)	effectif
A	1	.02	1.35	34
	2	1.35	2.05	35
	3	2.05	2.85	36
	4	2.85	4.05	35
	5	4.05	100.0	34
B	1	.000	.135	35
	2	.135	.275	35
	3	.275	.355	35
	4	.355	.565	35
	5	.565	5.0	34
C	1	.000	.145	37
	2	.145	.295	34
	3	.295	.390	34
	4	.390	.600	35
	5	.600	5.0	34

points représentatifs de chaque classe sont notés A_i , B_i et C_i ($i = 1...5$) et l'on utilise un codage binaire comme cela a été fait dans l'étude précédente.

L'analyse* a permis de déterminer les axes principaux d'un nuage de 174 points (stations) dans un espace à 15 dimensions (trois séries séparées en 5 classes). Les premiers axes principaux expliquent les pourcentages suivants de variabilité totale du nuage de points:

axe	%	% cumulé
1	21.2	21.2
2	18.1	39.3
3	13.7	53.0
4	10.5	63.5
5	7.6	71.1
6	7.2	78.4
7	6.5	84.9

Après 4 axes, le pourcentage de variabilité expliquée devient faible indiquant que le nuage des 174 points est suffisamment bien représenté dans un espace à 4 dimensions avec 63.5% de variabilité expliquée cumulée.

Les points représentant les 5 classes de chacune des séries dans le plan des deux premiers axes principaux sont visualisés à la figure 2.4 et les coordonnées de ces points sur les 4 premiers axes principaux sont données au tableau 2.4. Les contributions relatives des points sur chacun des axes sont représentées dans ce tableau par des astérisques suivant leur intensité. La

* Il s'agit d'une analyse de même type que celle réalisée dans la phase I, à laquelle on peut se reporter pour la description de la méthode et de ses caractéristiques.

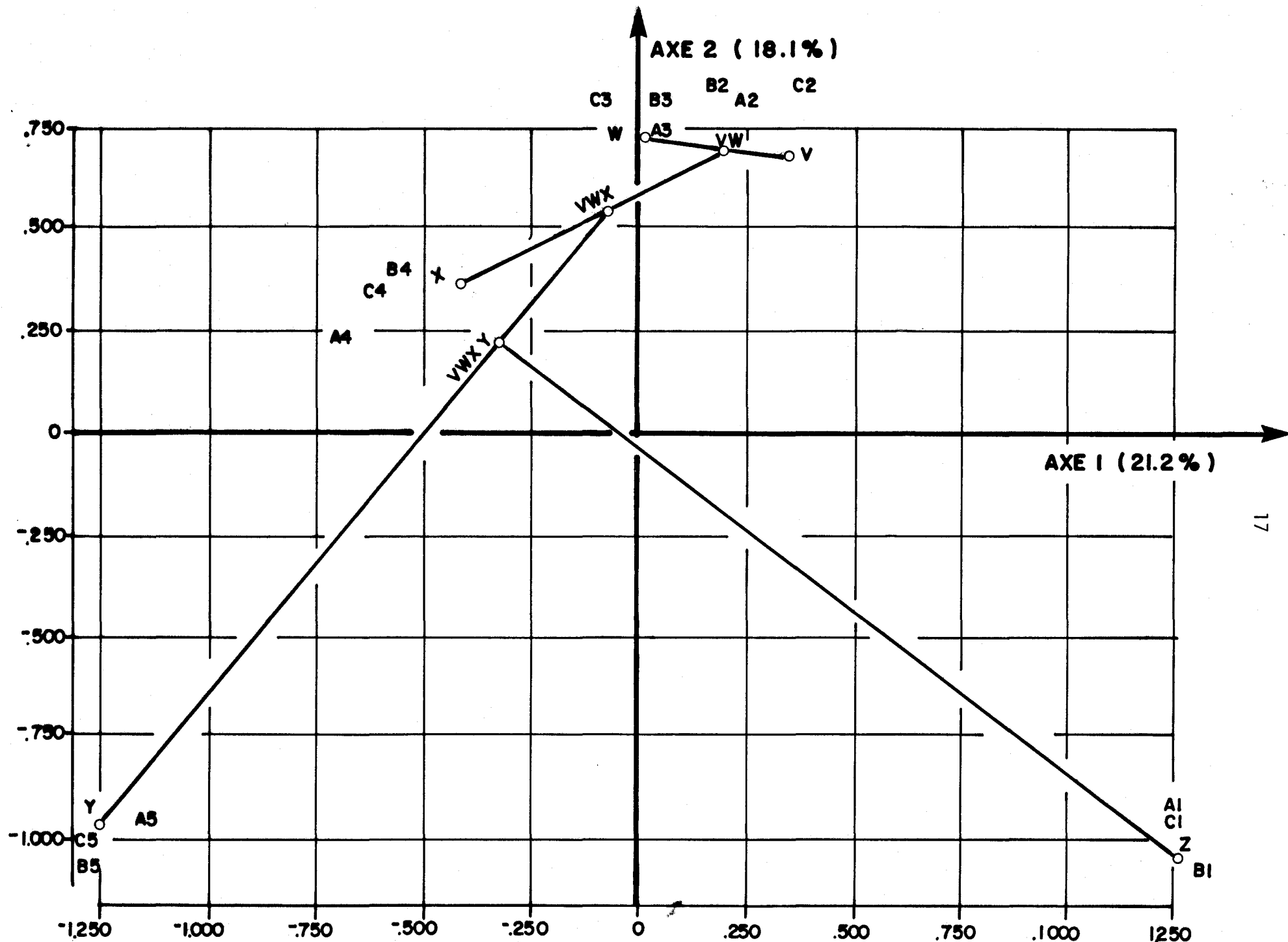


Figure 2 .4. Représentation dans le plan des deux premiers axes principaux des points paramètres (3 séries , 5 classes) et des centres de gravité des groupes de points observations (stations) classés hiérarchiquement .

Tableau 2.4 Coordonnées et contributions relatives sur les 4 premiers axes factoriels des points représentatifs des classes de valeurs pour chacune des séries de mesures.

point variable	coordonnées sur les axes factoriels				contribution relative cumulée
	1	2	3	4	
A1	1.43 ***	- .94 *	.19	- .04	.72
B1	1.48 ***	-1.10 **	.23	.06	.87
C1	1.39 ***	- .99 **	.16	- .02	.79
A2	.31	.85 *	- .29	.90 *	.43
B2	.26	.91 *	-1.11 **	.65 *	.65
C2	.45	.91 *	-1.13 **	.65 *	.66
A3	.10	.75 *	- .51	.98 **	.47
B3	.13	.85 *	.22	-1.14 **	.53
C3	.01	.82 *	.34	-1.21 **	.55
A4	-.65 *	.22	1.11 **	.14	.43
B4	-.53	.38	1.21 **	.53	.55
C4	-.62	.36	1.20 **	.73 *	.63
A5	-1.19 **	-.96 *	-.49	.01	.62
B5	-1.38 ***	-1.08 **	-.57	-.11	.83
C5	-1.33 ***	-1.02 **	-.63	-.16	.78

Contribution relative (CR)
indiquée par:

***	.40 < CR ≤ .55
**	.25 < CR ≤ .40
*	.10 < CR ≤ .25

contribution relative cumulée après 4 axes, qui exprime la qualité de la représentation dans un espace à 4 dimensions, est plus élevée pour les classes 1 et 5 (.62-.87) que pour les classes intermédiaires (.43-.66) indiquant que les points représentant les classes extrêmes sont un peu mieux représentés.

L'examen des points paramètres (3 séries, 5 classes) dans le plan des deux premiers axes principaux (figure 2.4) montre que les points représentant la même classe pour chacune des 3 séries sont pratiquement voisins. Les points représentant les classes extrêmes s'opposent sur l'axe 1; l'axe 2 oppose les deux classes extrêmes aux classes intermédiaires. Cette proximité sur ce plan entre points représentant la même classe indique que globalement les 3 séries de mesures ont une certaine similitude de variation; notons cependant que cette proximité est renforcée par le codage binaire effectué.

On constate que les coordonnées des points représentant la même classe commencent à différer sur les axes 3 et 4 spécialement en ce qui concerne les classes 2, 3 et 4. Afin de mieux visualiser cette différence, on a calculé les distances euclidiennes (somme des carrés des distances projetées sur chacun des axes) entre les points représentant la même classe pour chacune des séries prises deux à deux (distance A1-B1, A1-C1, B1-C1, A2-B2...). Les résultats de ces calculs apparaissant à la figure 2.5 font ressortir les points suivants:

- Les séries B et C sont beaucoup plus reliées entre elles qu'avec la série A, spécialement pour les classes 2 à 5. En effet, les

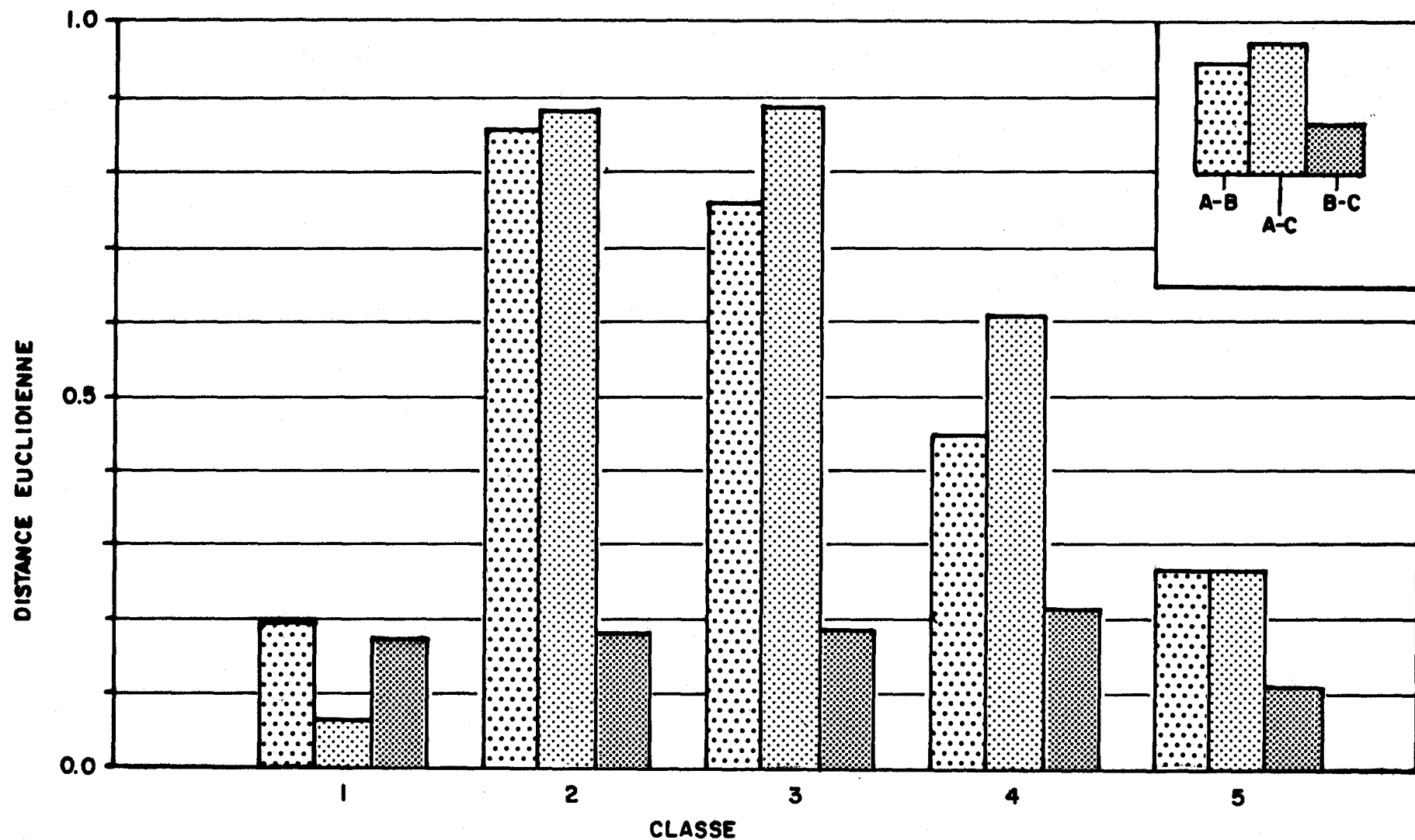


Figure 2.5. Représentation des distances euclidiennes calculées dans l'espace factoriel (4 dimensions) entre les points représentant les 5 classes des trois séries de mesures A, B et C.

distances B2-C2, B3-C3, B4-C4 et B5-C5 sont beaucoup plus faibles que les distances A2-B2, A2-C2, A3-B3... Cette constatation avait déjà été effectuée lors de l'étude précédente. Les mesures au RDM au début et à la fin de la période d'échantillonnage sont beaucoup plus corrélées entre elles qu'avec la mesure au microscope optique;

- il existe encore autant de différence dans les relations entre B et C qu'entre A et B et entre A et C pour les très faibles valeurs. La distance B1-C1 est plus grande que A1-C1 et presque égale à A1-B1. Il est normal de constater que pour des niveaux d'empoussiérage très faibles, la mesure au début et à la fin de la période d'échantillonnage peut varier à cause d'une certaine imprécision dans la mesure, d'une fluctuation attribuée au bruit de fond naturel ou encore d'un changement de niveau d'empoussiérage au cours de la période d'échantillonnage;
- comme lors de l'étude précédente, les relations entre A et B et entre A et C apparaissent moins fortes pour les valeurs intermédiaires (classes 2, 3 et 4) que pour les valeurs très fortes ou très faibles. En effet, les distances A2-B2, A2-C2, A3-B3... sont beaucoup plus grandes que A1-B1, A1-C1, A5-B5 et A5-C5. Cette observation d'une faible relation entre les mesures optiques et les mesures au RDM pour les valeurs intermédiaires, nécessite d'être confirmée par le calcul de coefficients de corrélation puisque les classes intermédiaires pour les trois séries (points paramètres) sont moins bien représentées dans l'espace à quatre dimensions que les classes extrêmes (tableau 2.4).

En résumé, on peut retenir qu'il existe encore une forte relation entre les séries B et C qui correspondent aux mesures prises avec la même technique à des temps différents, alors que la série A (méthode optique) est moins bien reliée avec B et C. Ceci confirme les résultats obtenus lors de la phase I.

On a recalculé les tableaux de contingence donnant la répartition des 174 échantillons selon les 5 classes de chacune des séries prises 2 à 2 (tableau 2.5). Ce tableau permet de confirmer les observations notées à partir des calculs des distances entre les points-paramètres (figure 2.5). En effet, la présence d'un nombre important de valeurs sur la diagonale du tableau, indicatrice d'une forte relation, se traduit par une faible distance entre les points. On remarque que les éléments diagonaux des tableaux de contingence sont plus importants entre B et C qu'entre A et B et entre A et C. Pour les classes extrêmes, le nombre d'éléments sur la diagonale pour les trois séries de mesures est plus important que pour les classes intermédiaires confirmant que la relation est plus faible pour les valeurs intermédiaires. On retrouve, par exemple, presque autant d'éléments A3-B2 que d'éléments A2-B2.

2.3 Détermination de groupes de stations

Les données à analyser proviennent d'une campagne d'échantillonnage à 174 postes localisés dans 3 usines de transformation des fibres d'amiante. Malgré le fait que l'ensemble des stations est localisé uniquement dans le circuit de transformation de la fibre, on doit s'attendre à avoir une certaine hétérogénéité dans les postes d'échantillonnage. C'est pourquoi, la recherche de groupes de stations homogènes du point de vue physique nécessite non seulement la connaissance de la localisation précise

Tableau 2.5 Tableaux de contingence obtenus en croisant 2 à 2 les séries A, B et C.

	B1	B2	B3	B4	B5	TOTAL
A1	26	2	4	2	0	34
A2	5	14	9	7	0	35
A3	2	13	13	5	3	36
A4	1	4	8	14	8	35
A5	1	2	1	7	23	34
TOTAL	35	35	35	35	34	174

	C1	C2	C3	C4	C5	TOTAL
A1	24	7	3	0	0	34
A2	5	14	8	8	0	35
A3	5	11	12	4	4	36
A4	1	1	8	16	9	35
A5	2	1	3	7	21	34
TOTAL	37	34	34	35	34	174

	C1	C2	C3	C4	C5	TOTAL
B1	30	3	2	0	0	35
B2	4	20	7	2	2	35
B3	3	9	15	8	0	35
B4	0	2	10	18	5	35
B5	0	0	0	7	27	34
TOTAL	37	34	34	35	34	174

du poste d'échantillonnage, mais également la connaissance de l'étape de transformation de la fibre et des conditions d'empoussiérage de tout le milieu ambiant. Dans cette section, on vise uniquement à déterminer mathématiquement un certain nombre de groupes de stations qui possèdent une homogénéité par rapport aux trois types de mesures et à laquelle on peut attribuer un sens physique. La détermination de ces groupes de stations permet de vérifier la présence de meilleures relations entre les deux types de mesures (méthode optique et RDM) et comporte l'intérêt de classer les stations d'échantillonnage en groupes et de caractériser ces groupes par leur niveau d'empoussiérage.

La méthode de regroupement utilisée est la classification ascendante hiérarchique (JAMBU, 1972; 1976) décrite en annexe B. Cette technique permet d'effectuer de manière automatique et selon un critère d'agrégation choisi, des regroupements en classes et d'ordonner les groupes de stations de façon à visualiser la hiérarchie du classement obtenu.

L'analyse factorielle des correspondances permet de représenter simultanément dans un espace factoriel de dimensions données les 15 points paramètres (3 séries, 5 classes) et les 174 points stations. Par conséquent, on peut utiliser les coordonnées des points stations sur les 4 premiers axes principaux pour calculer entre chaque paire de points un tableau de distances à partir duquel la technique de classification procède. La classification obtenue par cette technique est visualisée à la figure 2.6 qui montre la partie supérieure de l'arbre hiérarchique avec 5 groupes de stations, soient les groupes V, W, X, Y et Z.

La composition de chacun des groupes déterminés à partir de cette méthode de classification est donnée au tableau 2.6. Le nombre de stations

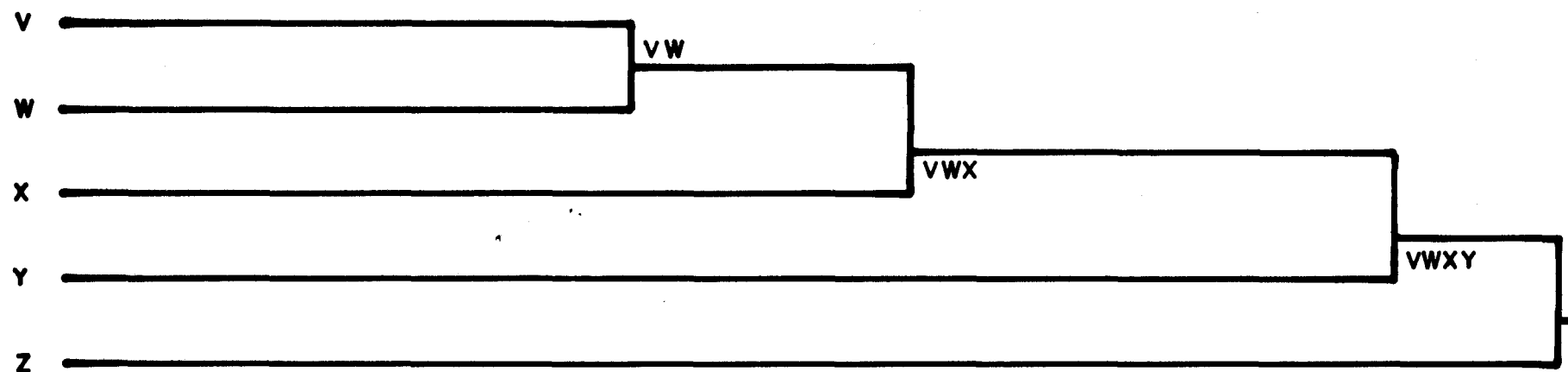


Figure 2.6 . Partie supérieure de l'arbre montrant la hiérarchie des 5 groupes de stations déterminés à partir des coordonnées sur les 4 premiers axes factoriels .

Tableau 2.6 Composition des groupes de stations déterminés par la classification ascendante hiérarchique.

(V)

1	BC1-6051-A	30	BC1-6075-B	121	N-6014-A	162	N-6067-D
2	BC1-6051-B	44	BC2-6046	122	N-6014-B	163	N-6001-A
3	BC1-6039	45	BC2-6047	136	N-6041-A	164	N-6001-B
4	BC1-6024-A	65	BC2-6022-B	137	N-6041-B	165	N-6001-C
5	BC1-6024-B	66	BC2-6022-C	138	N-6042-A	166	N-6002-A
13	BC1-6033	69	BC2-6033	141	N-6043-B	167	N-6002-B
16	BC1-6057	71	BC2-6031-A	146	N-6079-A	168	N-6002-C
17	BC1-6059	72	BC2-6031-B	148	N-6039-A	169	N-6003
21	BC1-6022-B	83	BC2-6069-A	149	N-6039-B	170	N-6004-A
24	BC1-6023-A	84	BC2-6069-B	150	N-6054-A	171	N-6004-B
27	BC1-6069-B	86	BC2-6061	157	N-6009-A		
29	BC1-6075-A	88	BC2-6064-B	159	N-6067-A		

(W)

19	BC1-6021	67	BC2-6058-A	105	N-6025-B	140	N-6043-A
50	BC2-6010	68	BC2-6058-B	119	N-6060-A	142	N-6044-A
52	BC2-6030-A	73	BC2-6020-A	131	N-6076-A	147	N-6079-B
57	BC2-6024-B	79	BC2-6017-A	132	N-6076-B		
63	BC2-6023-C	85	BC2-6037	133	N-6086		
64	BC2-6022-A	103	N-6024-B	139	N-6042-B		

(X)

6	BC1-6029-A	54	BC2-6030-C	82	BC2-6060	144	N-6044-C
11	BC1-6038	55	BC2-6025-A	102	N-6024-A	145	N-6044-D
20	BC1-6022-A	59	BC2-6024-D	104	N-6025-A	151	N-6006-A
22	BC1-6073-A	60	BC2-6067-A	106	N-6050-A	152	N-6006-B
23	BC1-6073-B	61	BC2-6067-B	107	N-6050-B	153	N-6007-A
25	BC1-6023-B	70	BC2-6032	109	N-6082-B	154	N-6007-B
46	BC2-6011	74	BC2-6020-B	125	N-6017-A	155	N-6008-A
48	BC2-6013	75	BC2-6068	129	N-6046-A	156	N-6008-B
49	BC2-6014	77	BC2-6018-B	130	N-6046-B	158	N-6009-B
51	BC2-6026-A	78	BC2-6018-C	135	N-6085-B	160	N-6067-B
53	BC2-6030-B	80	BC2-6017-B	143	N-6044-B	161	N-6067-C

Tableau 2.6 (suite). Composition des groupes de stations déterminés par la classification ascendante hiérarchique.

Y	7	BC1-6029-B	56	BC2-6024-A	111	N-6020-B	120	N-6060-B
	8	BC1-6037-A	58	BC2-6024-C	112	N-6020-C	123	N-6016-A
	9	BC1-6037-B	62	BC2-6067-C	113	N-6021-A	124	N-6016-B
	10	BC1-6037-C	76	BC2-6018-A	114	N-6021-B	126	N-6017-B
	12	BC1-6060	81	BC2-6017-C	115	N-6022-A	127	N-6045-A
	14	BC1-6072	97	N-6069	116	N-6022-B	128	N-6045-B
	15	BC1-6084	108	N-6082-A	117	N-6047-A	134	N-6085-A
	47	BC2-6012	110	N-6020-A	118	N-6047-B		

Z	18	BC1-6086	36	BC1-6045-C	87	BC2-6064-A	96	N-6027
	26	BC1-6069-A	37	BC1-6078	89	BC2-6064-C	98	N-6057
	28	BC1-6069-C	38	BC1-6077	90	BC2-6048	99	N-6071
	31	BC1-6070-A	39	BC1-6076	91	BC2-6049	100	N-6028
	32	BC1-6070-B	40	BC1-6079	92	BC2-6050	101	N-6058
	33	BC1-6071	41	BC1-6080	93	BC2-6051	172	N-6075-A
	34	BC1-6045-A	42	BC1-6081	94	BC2-6052	173	N-6075-B
	35	BC1-6045-B	43	BC1-6082	95	BC2-6053	174	N-6075-C

diffère d'un groupe à un autre; il est égal à 46, 21, 44, 31 et 32 pour les groupes V, W, X, Y et Z respectivement. A partir de l'arbre hiérarchique présenté à la figure 2.6, on peut décider de considérer uniquement 4 groupes ou 3 groupes de stations en réunissant en un seul groupe les groupes V et W ou les groupes V, W et X.

On peut calculer les centres de gravité de chacun des groupes de stations dans l'espace factoriel puisqu'on connaît les coordonnées des points sur chacun des axes factoriels. On a reporté sur le plan des deux premiers axes principaux, les centres de gravité de chacun des groupes de stations (figure 2.4). On a réuni par des droites les centres de gravité des groupes voisins, par exemple, V et W sont réunis par une droite et le point VW est le centre de gravité du groupe résultant de la réunion de ces groupes.

Grâce aux propriétés de l'analyse factorielle des correspondances, on peut associer à chacun des groupes de stations les points paramètres caractéristiques de chacun des groupes en examinant les proximités entre les points paramètres et groupes de stations. Ainsi (cf. fig. 2.4), on peut associer au groupe Y les valeurs très fortes pour les trois séries de mesures (A5, B5, C5), et au groupe Z leurs valeurs très faibles (A1, B1, C1); les groupes V et W sont plus ou moins associés aux classes 2 et 3 des trois séries de mesures alors que le groupe X est plutôt associé aux points A4, B4 et C4.

Afin de mieux caractériser chacun des groupes de stations, on a calculé les caractéristiques statistiques de chacun des groupes de stations à partir des valeurs de chaque série de mesures (tableau 2.7). On

Tableau 2.7. Caractéristiques statistiques de chacun des groupes de stations déterminés par la classification ascendante hiérarchique.

SERIE A (fibres/cm ³)							
caractéristiques	G R O U P E S						
	V	W	X	Y	Z	VW	VWX
nombre	46	21	44	31 (29)*	32	67	111
moyenne	2.12	2.61	3.22	9.17 (5.15)	.90	2.27	2.65
écart-type	1.03	.77	1.63	16.18 (1.84)	1.07	.98	1.35
minimum	.60	.70	1.40	2.10 (2.10)	.02	.60	.60
maximum	5.70	4.20	11.20	83.20 (10.20)	5.00	5.70	11.20
coef. variation	.48	.29	.51	1.76 (.36)	1.19	.43	.51

SERIE B (mg/m ³)							
caractéristiques	G R O U P E S						
	V	W	X	Y	Z	VW	VWX
nombre	46	21	44	31 (29)*	32	67	111
moyenne	.23	.33	.42	.89 (.78)	.03	.26	.33
écart-type	.07	.05	.11	.51 (.26)	.04	.08	.12
minimum	.00	.28	.21	.56 (.56)	.00	.00	.00
maximum	.45	.50	.70	2.82 (1.77)	.13	.50	.70
coef. variation	.32	.15	.26	.58 (.34)	1.38	.31	.37

SERIE C (mg/m ³)							
caractéristiques	G R O U P E S						
	V	W	X	Y	Z	VW	VWX
nombre	46	21	44	31 (29)*	32	67	111
moyenne	.28	.33	.46	.97 (.82)	.04	.30	.36
écart-type	.28	.08	.11	.65 (.32)	.07	.23	.21
minimum	.03	.10	.26	.41 (.41)	.00	.03	.03
maximum	1.97	.45	.74	3.42 (2.17)	.36	1.97	1.97
coef. variation	.98	.24	.24	.67 (.39)	1.72	.79	.58

*Les valeurs entre parenthèses ont été calculées en éliminant les valeurs extrêmes observées à BC1-6072 et BC1-6084.

peut vérifier ainsi que les groupes Z, V, W, X et Y correspondent dans cet ordre à des valeurs généralement croissantes pour chacun des types de mesures. En éliminant les deux stations non représentatives comprises dans le groupe Y (BC1-6072 et BC1-6084) pour lesquelles on observe des valeurs extrêmes, on constate que les caractéristiques statistiques présentent une moins grande dispersion. En examinant les minimums et les maximums de chacun des groupes, on voit que les intervalles de variation se chevauchent. On a calculé en outre, les caractéristiques statistiques pour le groupe VW et le groupe VWX qui sont associées à des valeurs intermédiaires.

Dans cette section, on a montré qu'il était possible de déterminer mathématiquement un certain nombre de groupes de stations relativement homogènes par rapport aux trois séries de mesures et de caractériser chacun de ces groupes en fonction des classes de mesures de A, B et C. Une étude sommaire de la répartition géographique de chacun des groupes à l'intérieur des trois usines a montré que ces groupes déterminés mathématiquement correspondent à une certaine réalité physique. On retrouve en effet sur un plancher plusieurs stations appartenant au même groupe ou à un groupe dont les caractéristiques sont voisines. Cependant, cette répartition des groupes à l'intérieur des trois usines n'est pas présentée ici. En effet, d'autres facteurs importants tels que la structure du milieu ambiant, les diverses fuites de poussières, la mise en circulation par ventilation ou circulation etc., viennent s'ajouter au facteur géographique. Il est impossible compte tenu des informations disponibles, d'interpréter l'influence de tous ces facteurs sur la formation des groupes.

La détermination de ces groupes homogènes par rapport aux 3 séries de mesures peut permettre de vérifier s'il existe à l'intérieur de ces groupes de stations des relations plus fortes entre les mesures optiques et physiques (RDM), que celles assez faibles existant pour l'ensemble des données (phase I).

2.4 Relations linéaires entre les séries de mesures

L'étude des relations linéaires entre chacune des séries prises deux à deux est faite en calculant dans une première étape des coefficients de corrélation:

- pour différents intervalles de variation des valeurs de A;
- pour chacun des groupes de stations déterminés dans la section précédente.

Comme dans l'étude précédente, on vérifie la linéarité des relations entre chacune des séries et on identifie la dispersion des points observés. Enfin, on calcule, selon les résultats obtenus avec les coefficients de corrélation, des équations de régression linéaire entre chacune des séries et leurs intervalles de confiance.

2.4.1 Coefficients de corrélation

Pour calculer des coefficients de corrélation, les séries de valeurs doivent suivre une distribution approximativement normale. La

présence d'une forte asymétrie dans la distribution des valeurs peut en effet, enlever tout sens physique à la valeur du coefficient de corrélation calculé et nécessiter d'effectuer une transformation (logarithmique, par exemple) sur les données. On a vu précédemment (section 2.1) lors de l'étude des histogrammes de fréquence, que la distribution de l'ensemble des valeurs avait une légère asymétrie positive, mais ne justifiait pas d'effectuer une telle transformation. On a vérifié cependant si tous les coefficients de corrélation calculés avaient un sens physique en portant en graphique les valeurs d'une série par rapport à une autre. Ces graphiques sont montrés à l'annexe C, pour des valeurs de A inférieures à 10 fibres/cm^3 et pour des valeurs de B et C inférieures à 1.5 mg/m^3 . Les points superposés sont indiqués par un chiffre variant de 1 à 9.

Afin de mieux comparer les relations entre chacune des séries avec les 174 stations d'échantillonnage par rapport à celles calculées dans l'étude précédente pour l'ensemble des 900 stations d'échantillonnage, on a calculé (tableau 2.8) les coefficients de corrélation pour l'ensemble des valeurs et pour différents intervalles de variation des valeurs de A:

intervalle	1	$A < 12$	fibres/cm^3
intervalle	2	$A < 6$	fibres/cm^3
intervalle	3	$6 \leq A < 12$	fibres/cm^3
intervalle	4	$3 \leq A < 6$	fibres/cm^3
intervalle	5	$1 \leq A < 3$	fibres/cm^3
intervalle	6	$A < 1$	fibres/cm^3

Tableau 2.8. Coefficients de corrélation entre les séries A, B et C pour différents intervalles des valeurs de A.

séries correlées	coefficient de corrélation	nombre de valeurs	intervalles de A (fibres/cm ³)
A-B A-C B-C	** .77 ** .77 ** .90	174 174 174	global global global
A-B A-C B-C	** .72 ** .66 ** .83 (** .91) ⁺	172 172 172	A < 12.0 A < 12.0 A < 12.0
A-B A-C B-C	** .65 ** .60 ** .75 (** .87) ⁺	160 160 160	A < 6.0 A < 6.0 A < 6.0
A-B A-C B-C	.30 .32 ** .94	12 12 12	6.0 ≤ A < 12.0 6.0 ≤ A < 12.0 6.0 ≤ A < 12.0
A-B A-C B-C	* .27 .13 ** .46 (** .76) ⁺	53 53 53	3.0 ≤ A < 6.0 3.0 ≤ A < 6.0 3.0 ≤ A < 6.0
A-B A-C B-C	* .19 * .18 ** .79	82 82 82	1.0 ≤ A < 3.0 1.0 ≤ A < 3.0 1.0 ≤ A < 3.0
A-B A-C B-C	** .51 ** .51 ** .91	25 25 25	A < 1.0 A < 1.0 A < 1.0

** significativement différents de zéro au niveau de 1%

* significativement différents de zéro au niveau de 5%

⁺ coefficient de corrélation calculé en éliminant les valeurs obtenues à la station BC1-6051-B

Les résultats des calculs de coefficients de corrélation qui apparaissent au tableau 2.8, font ressortir les points suivants:

- de façon générale, la relation entre B et C est généralement très bonne si l'on considère l'ensemble des valeurs ou différents intervalles de variation de A. La relation apparaissait un peu plus faible ($r = .46$) pour les valeurs de A comprises entre 3 et 6 fibres/cm³, mais l'analyse du graphique correspondant a montré que la valeur observée à la station BC1-6051-B (annexe A) était responsable de cette plus faible relation;
- la relation entre A et B et entre A et C est bonne ($.60 \leq r < .72$) lorsqu'on considère un grand intervalle de variation ($A < 12$ ou $A < 6$ fibres/cm³);
- la relation entre les deux types de mesures (A-B et A-C) est bonne ($r = 0.51$) pour des valeurs de A très faibles ($A < 1$ fibre/cm³);
- la relation entre A et B et entre A et C est très faible pour les intervalles 3,4 et 5 de variation de A ($6 \leq A < 12$, $3 \leq A < 6$ et $1 \leq A < 3$).

On donne ici un tableau comparatif des résultats des corrélations calculées dans l'étude précédente (phase I) et dans cette étude (phase II).

intervalle de A	A - B		A - C		B - C	
	phase I	phase II	phase I	phase II	phase I	phase II
A < 6	.62	.65	.56	.60	.66	.87
3 ≤ A < 6	.28	.27	.25	.13	.58	.76
1 ≤ A < 3	.32	.19	.28	.18	.65	.79
A < 1	.34	.51	.11	.51	.29	.91

De façon générale, on remarque que la relation entre B et C est meilleure que dans l'étude précédente, particulièrement pour les valeurs de A très faibles ($A < 1$ fibre/cm³). La relation entre A et B et entre A et C n'est pas meilleure que dans l'étude précédente sauf pour les valeurs de A inférieures à 1 fibre/cm³ où elle est meilleure. Il existe encore de façon générale, une certaine dispersion qui diminue la valeur du coefficient de corrélation entre A et B et entre A et C.

On a ensuite effectué des calculs de coefficients de corrélation pour chacun des groupes de stations homogènes déterminés mathématiquement par la méthode de classification. Les résultats de ces calculs apparaissent au tableau 2.9. On observe les points suivants:

- la relation entre A et B et entre A et C est non significative pour l'ensemble des groupes, sauf pour le groupe Y associé aux valeurs

Tableau 2.9. Coefficients de corrélation entre les séries A, B et C pour chaque groupe de stations.

séries corrélées	G R O U P E S				
	V (N = 46)	W (N = 21)	X (N = 44)	Y (N = 31)	Z (N = 32)
A - B	-**.28	-.14	.17	**.89(**.61) ^[2]	-.15
A - C	.24	.05	.01	**.90(**.54)	-.06
B - C	-.01(.20) ^[1]	.02	**.37	**.97(**.87)	**.66

séries corrélées	G R O U P E S	
	VW (N = 67)	VWX (N = 111)
A - B	-.07	**.27
A - C	*.23	**.25
B - C	-.04(**.39) ^[1]	**.34(**.66) ^[1]

les coefficients de corrélation indiqués par:

** sont significativement différents

* sont significativement différents

[1] la valeur entre parenthèse a été obtenue en éliminant BCl-6051-A et BCl-6051-B

[2] les valeurs entre parenthèses ont été obtenues en éliminant BCl-6072 et BCl-6084

très fortes où le coefficient de corrélation est généralement bon ($r \approx 0.6$) en éliminant les deux stations non représentatives, BC1-6072 et BC1-6084;

- la relation entre B et C est excellente pour le groupe Y associé aux valeurs très fortes, bonne pour le groupe Z associé aux valeurs très faibles, faible pour le groupe X associé aux valeurs intermédiaires élevées et non significative pour les deux autres groupes associés aux valeurs intermédiaires plutôt faibles.

La détermination de 5 groupes de stations homogènes n'a pas permis, comme on pouvait l'espérer, d'améliorer la relation entre les mesures optiques et les mesures au RDM. On a trouvé un seul groupe (Y), qui contient 29 stations, où la relation A-B et A-C est encore bonne. L'élimination des 2 valeurs aberrantes (cf. tab. 2.9) diminue la corrélation qui était fictive en raison des valeurs extrêmes. De plus, on a observé des coefficients de corrélation non significatifs entre B et C pour les groupes V et W, ce qui apparaît contradictoire avec les résultats présentés au tableau 2.8. En éliminant, par exemple, après examens graphiques, les valeurs aberrantes pour B et C aux stations BC1-6051-A et BC1-6051-B, le coefficient de corrélation entre B et C pour le groupe V passe de .01 à 0.20. Les faibles relations observées indiquent donc que la dispersion des valeurs est trop grande et peut masquer les relations lorsqu'on considère des groupes trop petits. Cette hypothèse est vérifiée en obtenant un coefficient de corrélation plus élevé entre B et C en réunissant en un seul groupe les groupes V et W ou les groupes V, W et X (tableau 2.9).

Rappelons que la comparaison globale des trois séries de mesures a été faite sur des données codées et que le regroupement des stations est obtenu en fonction des classes des valeurs. Lorsque l'on revient aux valeurs brutes pour le calcul des coefficients de corrélation, la dispersion des valeurs observées masque les relations. Les faibles relations observées pour chacun des groupes V, W, X et Z ne montrent pas nécessairement que les groupes de stations ne correspondent à aucune réalité physique. La relation entre les séries A, B et C n'existerait pas au niveau local à cause des nombreux facteurs qui peuvent affecter les résultats, mais elle existe au niveau global. Cette constatation observée qualitativement à partir des résultats de l'analyse des correspondances (figure 2.5) sera confirmée dans la section suivante.

2.4.2 Linéarité et dispersion des relations

Pour vérifier la linéarité de la relation globale et en même temps, faire ressortir la dispersion dans la relation entre les mesures optiques et celles du RDM, on a utilisé la méthodologie élaborée dans la phase I de l'étude. On a séparé l'intervalle de variation des 172 valeurs de B et de C en 10 classes comprenant entre 15 et 20 valeurs. On a calculé ensuite la moyenne et l'écart-type des valeurs de B et de C pour chacune des classes, ainsi que la moyenne et l'écart-type des valeurs de A correspondantes. Les résultats de ces calculs pour les valeurs de B et celles de C sont donnés aux tableaux 2.10 et 2.11. Sur les figures 2.7 et 2.8, on a porté en graphique les moyennes des classes de B et de C en fonction des moyennes et des écart-types des valeurs de A correspondantes. Ces deux figures font ressortir les points suivants:

Tableau 2.10. Nombre, moyenne et écart-type des valeurs de B groupées en 10 classes équiprobables, moyenne et écart-type des valeurs de A correspondantes.

classe	intervalle de B (mg/m ³)	nombre	moyenne de B	écart-type de B	moyenne de A	écart-type de A
1	.000 - .001	16	.000	.000	1.38	1.40
2	.001 - .135	19	.057	.042	.65	.51
3	.135 - .225	19	.188	.022	2.45	1.39
4	.225 - .275	16	.253	.015	2.28	.52
5	.275 - .315	19	.297	.013	2.51	.77
6	.315 - .355	16	.331	.011	2.21	.94
7	.355 - .435	18	.389	.024	2.86	1.23
8	.435 - .565	17	.497	.044	3.81	2.38
9	.565 - .665	15	.619	.030	4.22	1.10
10	.665 - 1.80	17	.901	.287	5.44	2.24

Tableau 2.11. Nombre, moyenne et écart-type des valeurs de C groupées en 10 classes équiprobables, moyenne et écart-type des valeurs de A correspondantes.

classe	intervalle de C (mg/m ³)	nombre	moyenne de C	écart-type de C	moyenne de A	écart-type de A
1	.000 - .025	21	.008	.009	1.03	1.28
2	.025 - .145	16	.091	.043	1.50	1.43
3	.145 - .235	17	.189	.029	2.14	1.09
4	.235 - .295	17	.266	.017	1.84	.74
5	.295 - .345	19	.323	.013	2.54	.88
6	.345 - .390	15	.364	.010	2.55	1.03
7	.390 - .475	18	.427	.021	3.59	2.35
8	.475 - .600	17	.534	.041	3.09	.91
9	.600 - .760	16	.671	.042	4.34	1.28
10	.760 - 2.20	16	1.074	.412	5.39	2.34

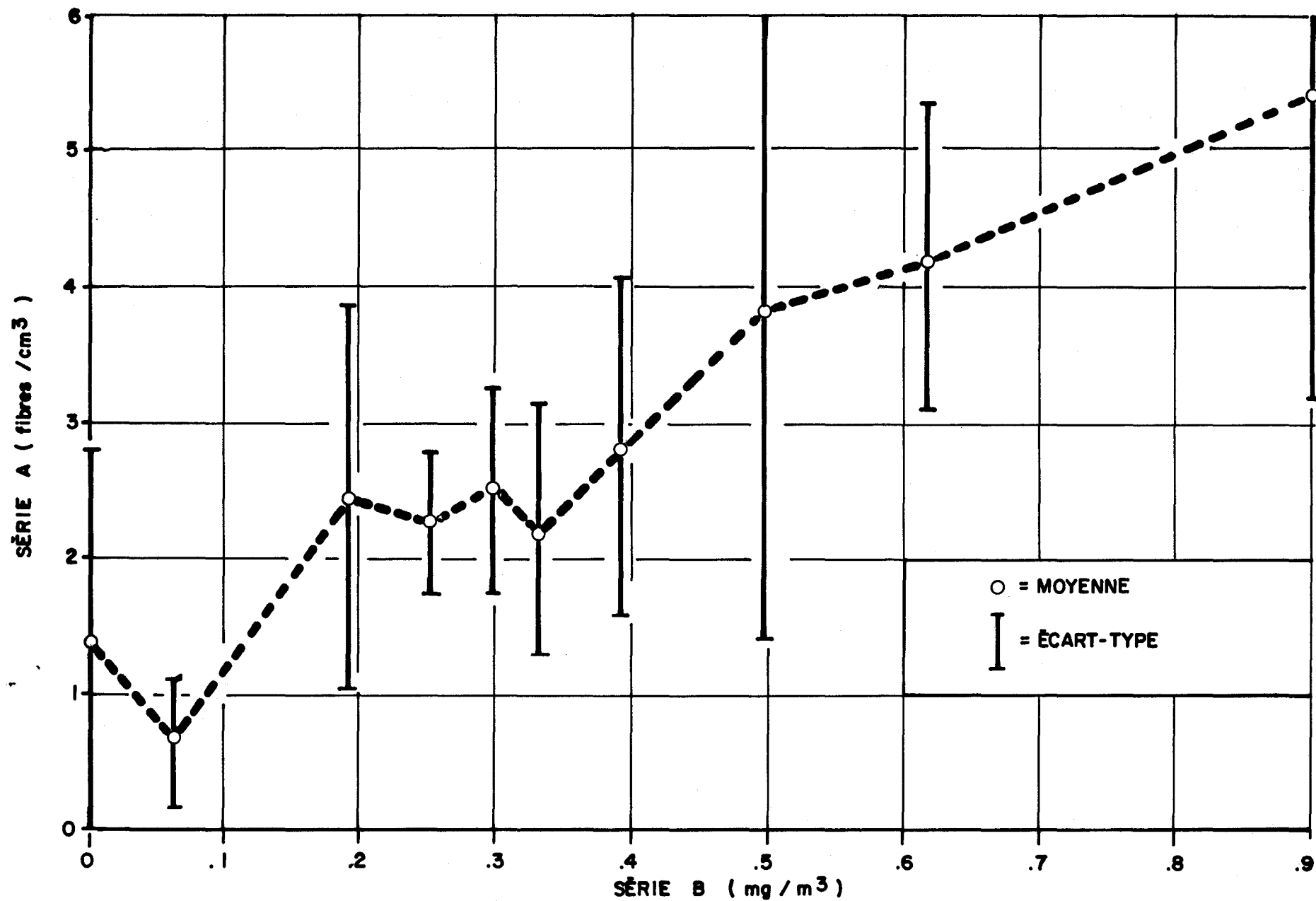


Figure 2.7 . Relation entre la série B et la série A : Étude de la linéarité et de la dispersion

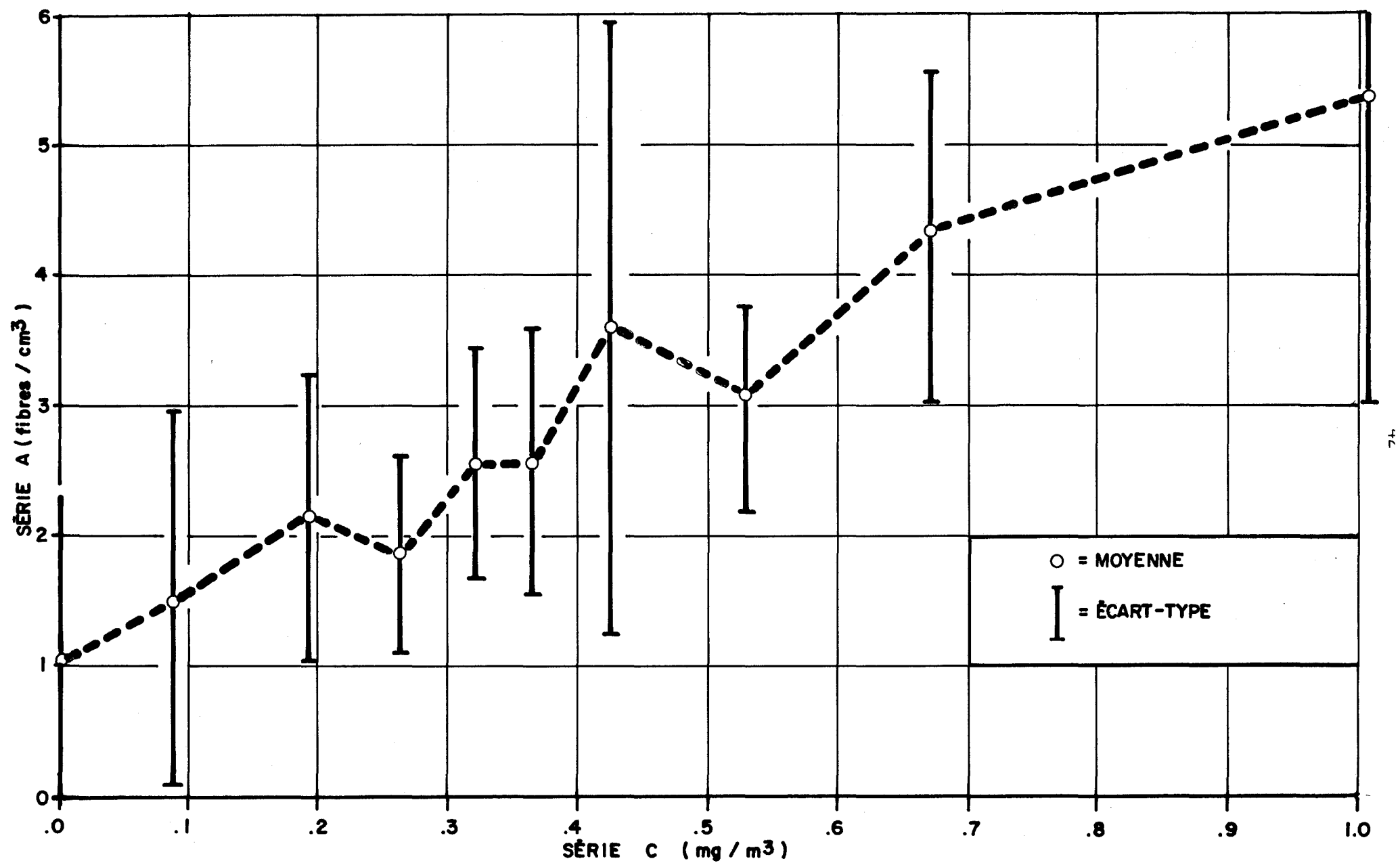


Figure 2.8 . Relation entre la série C et la série A : Étude de la linéarité et de la dispersion .

- la relation est approximativement linéaire entre les moyennes des classes de A et B et de A et C;
- pour chaque classe de B et de C, il existe une grande dispersion autour de la moyenne de A;
- on retrouve dans la classe 1 une valeur moyenne de A supérieure à 1 fibre/cm³, alors que les valeurs de B et de C sont pratiquement nulles;
- l'écart-type des valeurs de A dans la classe 1 est supérieur à la moyenne .

En résumé, cette méthode simple montre que la relation entre la méthode optique et la méthode physique est linéaire en moyenne et justifie le calcul d'équations de régression linéaire simple. La dispersion très grande des valeurs observées rend les relations moins intéressantes pour la prédiction individuelle et nécessite le calcul des intervalles de confiance sur les valeurs prédites par les équations de régression. Enfin, on montre qu'il existe un certain problème de justesse de mesures entre le RDM et la méthode optique pour des niveaux d'empoussiérage très faibles.

2.4.3 Régressions linéaires et intervalles de confiance associés

Malgré la dispersion observée dans les relations entre chacune des séries, le coefficient de corrélation au niveau global est bon et justifie le calcul d'équations de régression linéaire simple entre chacune des

séries. Les intervalles de confiance à 95% et 90% montreront l'utilité de ces équations de régression.

On a calculé des équations de régression permettant de prédire en premier lieu les valeurs du RDM à partir des mesures optiques et ensuite les valeurs du RDM à la fin de l'échantillonnage à partir de celles du début de l'échantillonnage. Ces équations sont données au tableau 2.12 pour l'ensemble des 172 stations et pour le groupe de stations Y déterminé par la méthode de classification. On donne également le coefficient de détermination r^2 (carré du coefficient de corrélation) qui exprime le pourcentage de variance expliquée par l'équation de régression. Les figures 2.9, 2.10 et 2.11 montrent les 3 droites de régression obtenues pour l'ensemble des valeurs et les intervalles de confiance à 95% et à 90% des valeurs prédites par ces équations. Ainsi, pour une valeur de B au RDM égale à $.50 \text{ mg/m}^3$, on obtient une valeur prédite de A égale à 3.52 fibres/cm^3 . On peut prédire avec 95% de confiance que pour cette valeur de B, la valeur de A est comprise entre $.91$ et 6.13 fibres/cm^3 ; ce large intervalle est dû à la dispersion des valeurs autour de la régression et empêche des prédictions précises.

Ces équations de régression permettent les observations suivantes:

- à partir des valeurs nulles au RDM, on prédit des mesures optiques de l'ordre de 1 fibre/cm^3 . Cette observation avait été soulignée dans la section précédente. Ceci impliquerait qu'à certaines stations, l'appareil RDM ne serait pas suffisamment sensible à cause de la nature des fibres ou que la mesure optique est surestimée pour les faibles valeurs;

Tableau 2.12. Equations de régression et coefficients de détermination pour l'ensemble des stations et pour un groupe de stations.

ENSEMBLE (172 valeurs)	$A = 5.04 B \pm 1.00$; $r^2 = .51$ cf. fig. 2.9 $A = 3.90 C \pm 1.26$; $r^2 = .43$ cf. fig. 2.10 $C = .98 B \pm .04$; $r^2 = .69$ cf. fig. 2.11
GROUPE DE STATIONS Y (29 valeurs)	$A = 4.27 B \pm 1.83$; $r^2 = .37$ $A = 3.12 C \pm 2.56$; $r^2 = .30$ $C = 1.06 B$; $r^2 = .75$

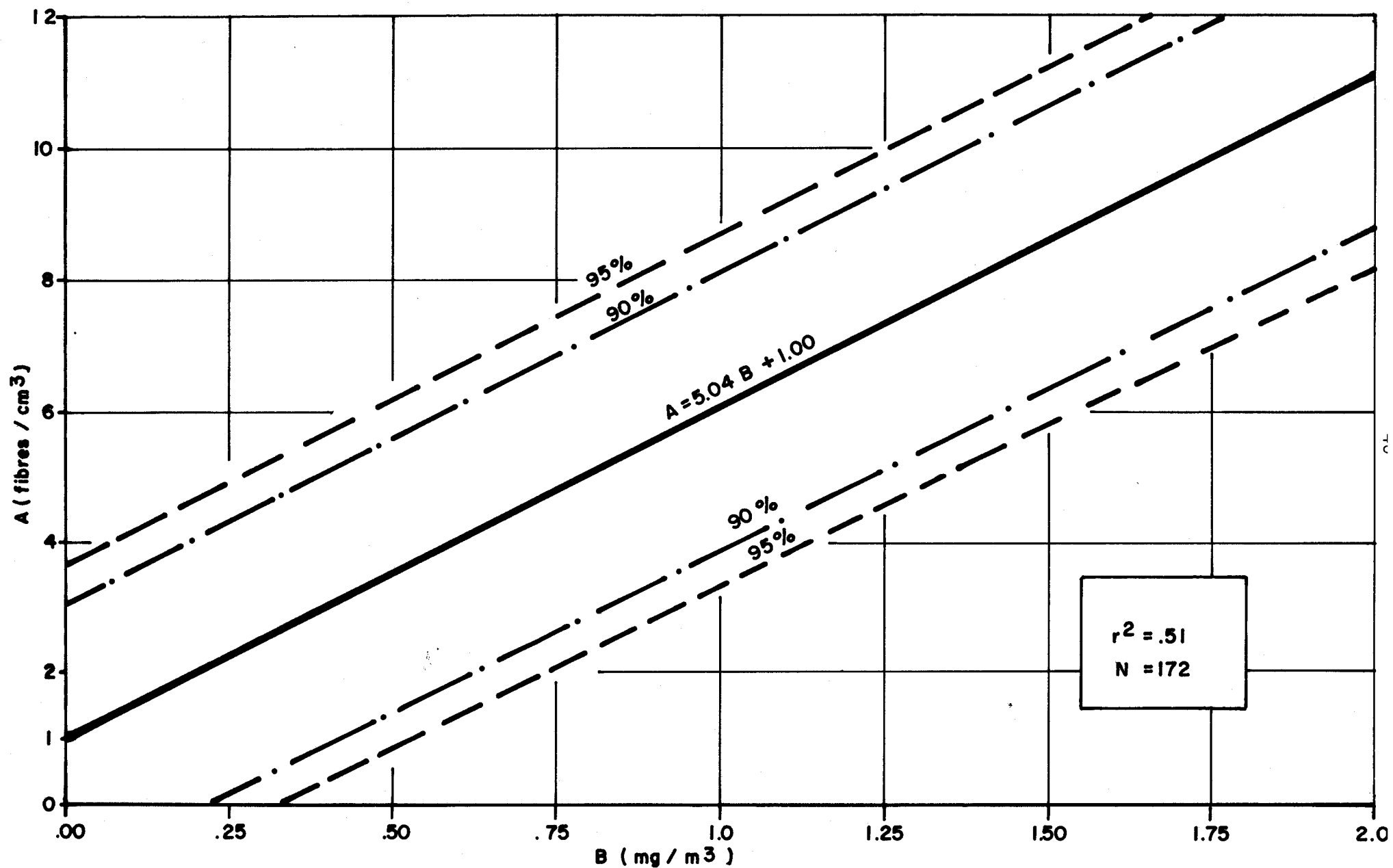


Figure 2.9 . Régression linéaire simple pour l'ensemble des mesures entre la méthode optique (A) et la méthode du R.D.M (B) et intervalles de confiances associés .

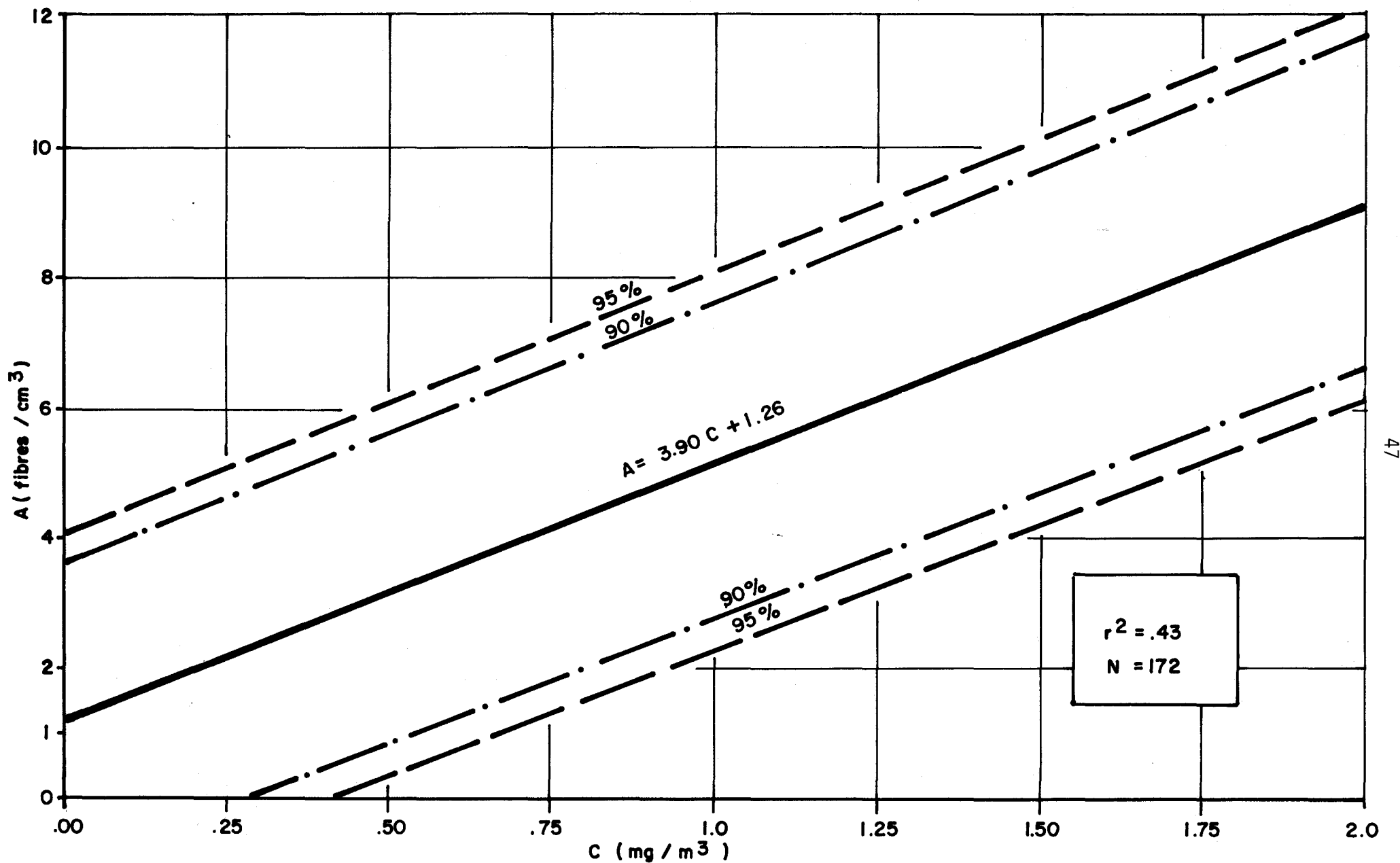


Figure 2.10 . Régression linéaire simple pour l'ensemble des mesures entre la méthode optique (A) et la méthode du R.D.M (C) et intervalles de confiances associés ;

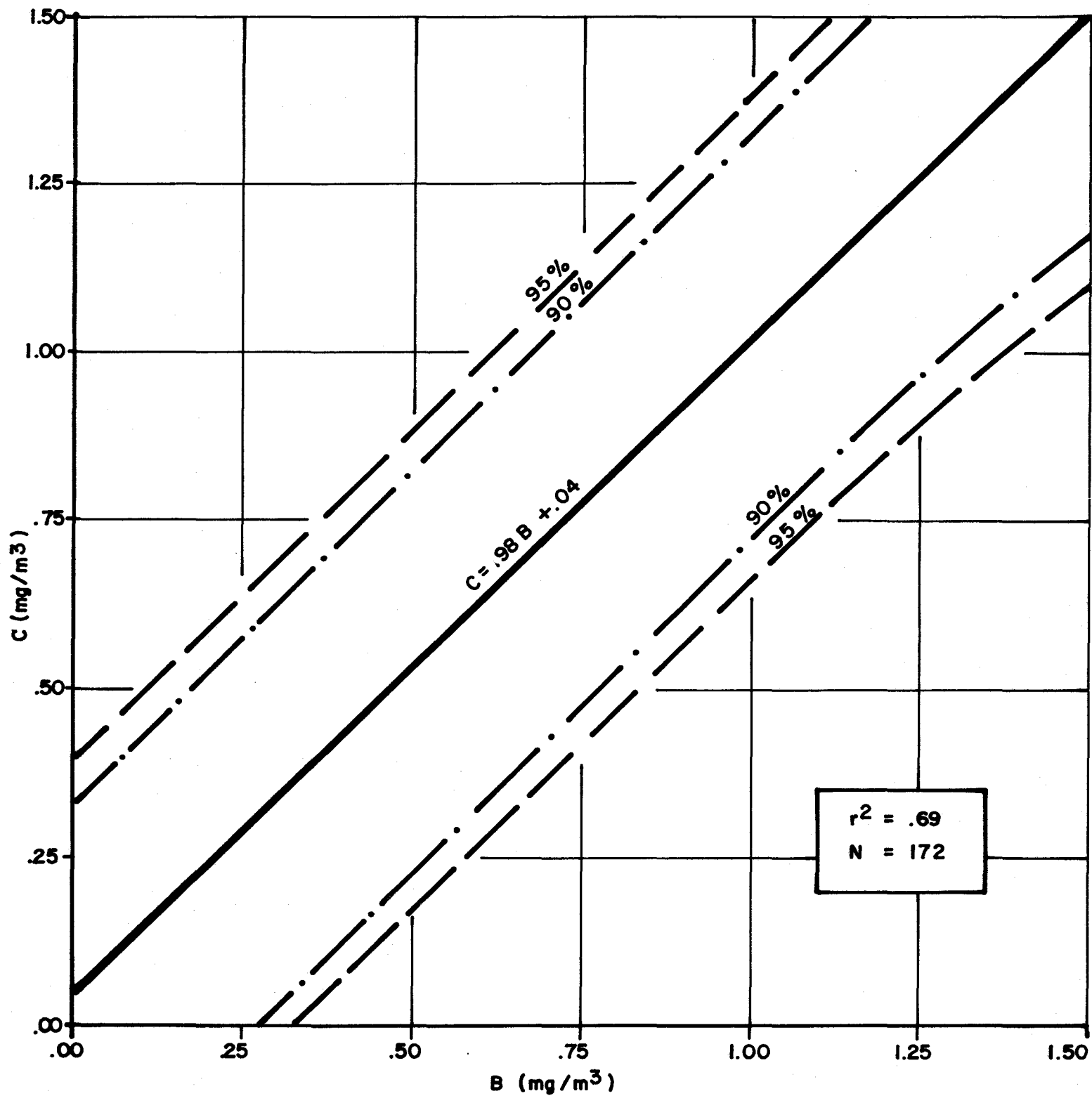


Figure 2. II . Régression linéaire entre les séries de mesures du R.D.M. à la fin de l'échantillonnage (C) et au début de l'échantillonnage (B) et intervalles de confiance associés .

- les intervalles de confiance autour des droites de régression obtenues entre A et B et entre A et C sont très grands, en raison de la dispersion des valeurs observées, et rendent la relation trouvée, moins utile pour prédire avec précision une mesure à partir d'une autre mesure;
- ces intervalles de confiance permettent de prédire des valeurs en fibres/cm³ inférieures à des normes avec un certain degré de confiance. Par exemple, on peut prédire (figure 2.9) qu'avec une valeur du RDM (B) de .27 mg/m³, il y a 95% de chances d'obtenir une valeur de A inférieure ou égale à 5 fibres/cm³.
- l'équation de régression obtenue entre les deux mesures au RDM (figure 2.11) permet de prédire avec une bonne précision une mesure à partir d'une autre. Pour une mesure de B de 1 mg/m³, on peut prédire que la valeur de C sera comprise entre .62 et 1.33 mg/m³ avec une probabilité de 95%; cet intervalle est relativement réduit en raison de la corrélation assez forte entre B et C. Il est de plus possible de montrer à l'aide de tests sur les coefficients de régression que les équations du tableau 2.12 correspondent au modèle théorique $B = C$.
- les équations de régression linéaire obtenues pour le groupe de stations Y sont légèrement différentes de celles obtenues pour l'ensemble des valeurs. La pente et l'ordonnée à l'origine sont diffé-

rentes parce que les stations du groupe Y comportent des valeurs élevées ($21. < A < 10.2 \text{ fibres/cm}^3$; voir tableau 2.7) et que la relation A-B et A-C n'est pas complètement linéaire sur tout l'intervalle de variation (figure 2.7 et 2.8). Ces équations de régression valables pour un seul groupe de stations présentent ici un intérêt réduit et ne justifient pas le calcul d'intervalles de confiance.

CONCLUSION

Cette étude mathématique est la suite de l'étude entreprise précédemment (phase I) concernant l'existence d'une relation entre les mesures optiques (comptage au microscope) et les mesures physiques (RDM). On a étudié l'existence potentielle d'une relation plus forte avec un nombre plus réduit (174 par rapport à 900) et surtout plus homogène, de stations d'échantillonnage. Chacun des postes d'échantillonnage est situé dans le circuit de transformation et de concentration de la fibre d'amiante. L'approche utilisée dans la comparaison des méthodes était essentiellement la même. On a, en outre, déterminé un regroupement de stations en fonction des classes des valeurs mesurées afin de rechercher la présence de relations locales plus fortes. De cette étude, on peut tirer les conclusions suivantes:

- les valeurs obtenues pour chacune des méthodes de mesure ont une dispersion moins grande que dans l'étude précédente et une distribution qui se rapproche d'une loi normale, confirmant la plus grande homogénéité entre les stations.
- la distribution des mesures optiques est quelque peu différente (asymétrie et aplatissement différents) de celle obtenue pour les mesures au RDM et peut être expliquée par la précision différente des deux méthodes et probablement par la nature différente du matériel utilisé;

- on trouve une relation linéaire au niveau global entre les mesures optiques et les mesures au RDM;
- au niveau local, on ne trouve pas de relation intéressante entre les mesures optiques et les mesures au RDM sauf pour groupe de stations où l'on enregistre des valeurs élevées;
- la relation globale entre les deux méthodes n'est pas excellente et ne permet pas une prédiction précise des valeurs individuelles;
- il existe une bonne relation entre les mesures du RDM effectuées au début (B) et à la fin (C) d'une expérience

En résumé, les conclusions qui se dégagent de cette étude confirment et précisent celles de l'étude précédente (phase I). Cependant la phase II qui avait été entreprise en considérant l'hypothèse que des données plus homogènes provenant uniquement du circuit de transformation et de concentration de la fibre, pourraient conduire à de meilleures équations de régression, a été partiellement décevante. En effet, au niveau local, que ce soit en considérant des tranches de variation de A ou des groupes homogènes issus de l'analyse des correspondances, les valeurs observées montrent toujours une grande dispersion autour des droites de régression. On peut en conclure que:

- les 2 types de mesures (méthode optique et RDM) sont relatives à des caractéristiques différentes du niveau d'empoussiérage en

fibres d'amiante. Ceci est confirmé par une comparaison morphologique des 2 types d'échantillonnage au microscope électronique à balayage (voir annexe D)

- les mesures de niveaux d'empoussiérage sont très sensibles aux conditions ambiantes et posent le problème de la représentativité de l'échantillonnage en un point donné.

REFERENCES BIBLIOGRAPHIQUES

DESJARDINS, M., R. BERTRAND, M. LACHANCE et B. BOBEE.

Etude comparative de deux méthodes de mesure d'empoussiérage dans l'air. INRS-Eau, rapport scientifique No. 100, 63 p.

JAMBU, M. (1972).

Techniques de classification appliquées à des données de sciences humaines. Thèse de 3^e cycle. L.S.M. - ISUP, Université de Paris VI, 272 p.

JAMBU, M. (1976).

Sur l'interprétation mutuelle d'une classification hiérarchique et d'une analyse des correspondances. Rev. Statist. Appl. 24(2), p. 45-73.

ANNEXE A

DONNEES

NO	STATION	A	B	C	NO	STATION	A	B	C
1	BC1-6051-A	3.00	.17	.80	51	BC2-6026-A	3.10	.43	.58
2	BC1-6051-B	4.00	.22	1.97	52	BC2-6030-A	2.70	.29	.36
3	BC1-6039	1.50	.18	.29	53	BC2-6030-B	1.90	.32	.48
4	BC1-6024-A	2.50	.18	.22	54	BC2-6030-C	1.90	.31	.40
5	BC1-6024-B	1.40	.19	.22	55	BC2-6025-A	3.70	.52	.63
6	BC1-6029-A	4.00	.70	.48	56	BC2-6024-A	6.90	.91	1.01
7	BC1-6029-B	7.10	1.38	1.08	57	BC2-6024-H	3.90	.35	.33
8	BC1-6037-A	2.10	.91	.86	58	BC2-6024-C	4.50	.63	.41
9	BC1-6037-B	9.20	1.77	2.17	59	BC2-6024-D	3.00	.67	.59
10	BC1-6037-C	6.10	.57	.67	60	BC2-6067-A	3.30	.36	.32
11	BC1-6038	3.30	.31	.51	61	BC2-6067-B	3.50	.50	.52
12	BC1-6060	3.20	.57	.82	62	BC2-6067-C	4.20	.63	.51
13	BC1-6033	1.40	.28	.03	63	BC2-6023-C	2.20	.50	.32
14	BC1-6072	51.70	2.27	2.71	64	BC2-6022-A	2.50	.34	.36
15	BC1-6084	83.20	2.82	3.42	65	BC2-6022-B	1.10	.31	.27
16	BC1-6057	1.60	.12	.28	66	BC2-6022-C	1.40	.32	.27
17	BC1-6059	.60	.18	.25	67	BC2-6058-A	2.00	.32	.32
18	BC1-6086	.60	.04	0.00	68	BC2-6058-B	2.20	.31	.30
19	BC1-6021	2.50	.32	.10	69	BC2-6033	2.30	.29	.20
20	BC1-6022-A	2.70	.44	.70	70	BC2-6032	3.20	.28	.26
21	BC1-6022-B	1.70	.17	.14	71	BC2-6031-A	1.70	.27	.36
22	BC1-6073-A	3.60	.39	.74	72	BC2-6031-B	1.90	.22	.27
23	BC1-6073-B	1.90	.21	.40	73	BC2-6020-A	2.20	.32	.45
24	BC1-6023-A	1.10	.36	.35	74	BC2-6020-B	2.20	.47	.65
25	BC1-6023-B	1.90	.37	.34	75	BC2-6068	2.10	.38	.42
26	BC1-6069-A	.80	.13	.03	76	BC2-6018-A	2.60	.82	.87
27	BC1-6069-B	1.60	.13	.36	77	BC2-6018-B	1.50	.56	.53
28	BC1-6069-C	1.10	.09	.36	78	BC2-6018-C	1.60	.39	.54
29	BC1-6075-A	5.70	.14	.14	79	BC2-6017-A	.70	.34	.32
30	BC1-6075-B	5.70	.19	.19	80	BC2-6017-B	1.70	.37	.36
31	BC1-6070-A	1.10	.03	.13	81	BC2-6017-C	4.30	.66	1.14
32	BC1-6070-B	.90	.08	.18	82	BC2-6060	3.00	.29	.40
33	BC1-6071	.50	0.00	.02	83	BC2-6069-A	.80	.32	.24
34	BC1-6045-A	.20	.01	.03	84	BC2-6069-B	1.40	.19	.20
35	BC1-6045-B	.10	0.00	.01	85	BC2-6037	1.80	.31	.34
36	BC1-6045-C	.20	.03	.02	86	BC2-6061	1.50	.25	.30
37	BC1-6078	.02	0.00	0.00	87	BC2-6064-A	1.70	0.00	0.00
38	BC1-6077	.20	.01	.01	88	BC2-6064-B	2.40	0.00	.15
39	BC1-6076	.04	0.00	.01	89	BC2-6064-C	2.70	0.00	0.00
40	BC1-6079	.60	.01	.02	90	BC2-6048	.02	.03	.04
41	BC1-6080	1.50	0.00	.01	91	BC2-6049	.08	.02	.02
42	BC1-6081	.20	.01	0.00	92	BC2-6050	.12	0.00	0.00
43	BC1-6082	.90	0.00	.02	93	BC2-6051	.04	.09	.09
44	BC2-6046	2.40	.18	.16	94	BC2-6052	1.20	.08	.10
45	BC2-6047	2.20	.17	.12	95	BC2-6053	.03	0.00	.02
46	BC2-6011	4.10	.36	.34	96	N-6027	5.00	0.00	0.00
47	BC2-6012	4.30	.59	.66	97	N-6069	3.20	.63	.65
48	BC2-6013	2.00	.40	.40	98	N-6057	.70	0.00	0.00
49	BC2-6014	3.90	.46	.59	99	N-6071	1.30	0.00	0.00
50	BC2-6010	4.20	.33	.38	100	N-6028	2.00	0.00	0.00

NU	STATION	A	R	C	NO	STATION	A	B	C
101	N-605A	3.10	0.00	0.00	151	N-6006-A	3.20	.51	.44
102	N-6024-A	3.00	.5A	.43	152	N-6006-B	1.40	.55	.44
103	N-6024-B	3.30	.29	.37	153	N-6007-A	2.00	.33	.4A
104	N-6025-A	3.00	.25	.31	154	N-6007-B	2.50	.63	.5A
105	N-6025-B	3.80	.2A	.33	155	N-6008-A	2.90	.27	.45
106	N-6050-A	3.70	.4A	.59	156	N-6008-B	3.00	.37	.54
107	N-6050-B	3.00	.40	.42	157	N-6009-A	2.80	.27	.34
108	N-6082-A	6.10	.56	.64	158	N-6009-B	3.40	.31	.42
109	N-6082-B	6.30	.54	.46	159	N-6067-A	2.70	.20	.27
110	N-6020-A	7.00	.87	.79	160	N-6067-B	2.90	.38	.32
111	N-6020-B	5.30	.62	.77	161	N-6067-C	1.90	.42	.26
112	N-6020-C	4.80	.66	.69	162	N-6067-D	1.30	.45	.23
113	N-6021-A	6.80	.98	.92	163	N-6001-A	1.60	.25	.30
114	N-6021-B	3.70	.67	.69	164	N-6001-B	1.90	.27	.29
115	N-6022-A	5.00	.87	1.14	165	N-6001-C	2.80	.23	.28
116	N-6022-B	10.20	1.04	1.13	166	N-6002-A	2.40	.25	.11
117	N-6047-A	3.00	.65	.61	167	N-6002-B	2.50	.26	.15
118	N-6047-B	4.20	.61	.56	168	N-6002-C	1.90	.23	.15
119	N-6060-A	2.70	.43	.32	169	N-6003	.80	.18	.18
120	N-6060-B	5.60	.88	.91	170	N-6004-A	1.20	.31	.27
121	N-6014-A	2.40	.26	.25	171	N-6004-B	1.50	.34	.29
122	N-6014-B	2.40	.22	.38	172	N-6075-A	.50	.03	.05
123	N-6016-A	5.20	.69	.61	173	N-6075-B	.50	.0A	.13
124	N-6016-B	4.60	.62	.68	174	N-6075-C	1.00	.07	.08
125	N-6017-A	11.20	.48	.42					
126	N-6017-B	6.10	.63	.64					
127	N-6045-A	3.90	.70	.80					
128	N-6045-B	4.30	.75	.75					
129	N-6046-A	3.80	.44	.43					
130	N-6046-B	3.80	.55	.38					
131	N-6076-A	2.50	.35	.36					
132	N-6076-B	2.40	.29	.37					
133	N-6086	2.80	.31	.32					
134	N-6085-A	5.90	.75	.72					
135	N-6085-B	6.40	.37	.47					
136	N-6041-A	1.90	.16	.24					
137	N-6041-B	2.80	.26	.33					
138	N-6042-A	2.80	.22	.24					
139	N-6042-B	3.00	.34	.36					
140	N-6043-A	2.20	.32	.33					
141	N-6043-B	2.60	.28	.23					
142	N-6044-A	2.50	.31	.43					
143	N-6044-B	3.00	.41	.50					
144	N-6044-C	4.10	.41	.50					
145	N-6044-D	4.30	.44	.35					
146	N-6079-A	2.70	.26	.15					
147	N-6079-B	2.80	.29	.14					
148	N-6039-A	2.00	.33	.19					
149	N-6039-B	1.80	.24	.22					
150	N-6054-A	1.80	.23	.19					

ANNEXE B

CLASSIFICATION ASCENDANTE

HIERARCHIQUE

1. Classification ascendante hiérarchique (C.A.H.)

La méthode de classification ascendante hiérarchique (Jambu, 1972) est une technique qui permet de regrouper de manière automatique et selon des critères d'agrégation choisis, des données en classes et de les ordonner de façon à visualiser la hiérarchie du classement obtenu.

On reconnaît deux types d'utilisation de cette méthode qui permettent d'interpréter les résultats de l'A.F.C. (Jambu, 1976a). On peut, soit utiliser la C.A.H. sur le même tableau de données traité par l'A.F.C., soit l'utiliser sur le tableau des facteurs obtenus de l'analyse des correspondances, à condition de définir la formule de distance appropriée.

Pour résumer la méthode de C.A.H., on choisit de se placer dans l'espace des paramètres à p dimension R^p dans lequel se retrouvent les n points observations. La C.A.H. procède à partir d'un tableau de $n(n-1)/2$ distances entre chaque paire possible de points. Selon les deux types d'utilisation, on a

$$a) \quad d^2(i, i') = \sum_{j=1}^p \left(\frac{p_{ij}}{p_i} - \frac{p_{i'j}}{p_i} \right)^2 / p_j$$

$$b) \quad d^2(i, i') = \sum_{\alpha} \left((F_{\alpha}(i) - F_{\alpha}(i')) \right)^2$$

La première relation définit la distance dite du χ^2 entre deux points observations. Nous avons vu qu'elle est équivalente à une distance euclidienne entre deux points dont les coordonnées du point i seraient $p_{ij} / p_i \sqrt{p_j}$ ($j=1, 2, \dots, p$). Dans la deuxième relation, la coordonnée du point i est $F_{\alpha}(i)$ ($\alpha = 1, 2, \dots, r < p$).

La méthode peut se résumer ainsi:

- on recherche la distance minimum entre l'ensemble des paires de points
- on forme une première classe avec ces deux points ($S=i'Ui''$)

et on calcule son centre de gravité:

$$a) \quad \frac{P_{sj}}{P_s \sqrt{P_j}} = \frac{P_{i'j} + P_{i''j}}{(P_{i'} + P_{i''}) \sqrt{P_j}}$$

$$b) \quad F_{\alpha}(s) = \frac{P_{i'} F_{\alpha}(i') + P_{i''} F_{\alpha}(i'')}{P_{i'} + P_{i''}}$$

- on attribue à cette première classe un indice de diamètre $\tau(s)$ qui est un moment centré d'ordre 2

on calcule $\tau(s)$ par cette formule:

$$\tau(s) = \frac{P_{i'} P_{i''}}{P_{i'} + P_{i''}} d^2(i', i'')$$

- Après avoir remplacé les deux points constituant la première classe par leur centre de gravité, on recalcule un nouveau tableau de distances de dimension $(n-1)(n-2)/2$.
- on recherche une autre paire de points dont la distance est minimum.

- on recommence (n - 1) fois toutes les opérations décrites jusqu'au moment où il n'y a plus que deux parties à agréger.

Critère d'agrégation choisi

Il existe plusieurs formules de calcul de distances entre classes conduisant à des regroupements de classes plus ou moins différents. La formule de distance utilisée dans le programme de calcul (Jambu, (1974) permet de maximiser le moment centré d'ordre 2 d'une partition. En effet, supposons l'étape de rang (n-1) réalisée et la partition Q_{n-1} obtenue sur R^p . On cherche à agréger deux parties de R^p , (soit deux classes s et t de Q_{n-1}), dont la réunion maximisera le moment centré d'ordre 2 du nuage associé à la partition Q_n , soit $M^2(N(Q_n))$. Q_n est la partition constituée de Q_{n-1} à laquelle on retire s et t et à laquelle on ajoute sUt:

$$M^2(N(Q_n)) = M^2(N(Q_{n-1})) - M^2(sUt) + M^2(s) + M^2(t)$$

$$\text{avec } \tau(n) = \tau(sUt) = M^2(N(Q_{n-1})) - M^2(N(Q_n))$$

Contribution des paramètres j à la formation des classes d'observations sur R^p

L'inertie totale du nuage des points i dans R^p peut être exprimée par la relation:

$$\text{inertie totale} = \sum_q \tau(q) = \sum_q p_q \rho^2(q)$$

dans laquelle $\rho^2(q)$ est la distance du centre de gravité de la classe q au centre de gravité du nuage des points dans R^p ; $\rho^2(q)$ est défini selon le type d'utilisation:

$$a) \quad \rho^2(q) = \sum_{j=1}^p \left(\frac{p_{qj}}{p_q} - p_j \right)^2 / p_j$$

$$b) \quad \rho^2(q) = \sum_{\alpha} F_{\alpha}^2(q)$$

La quantité $\rho^2(q)$ permet d'apprécier si la classe q est au centre (dans ce cas, la classe q se rapproche d'un profil moyen) ou très éloigné du centre. On pose

$$c_{qj} = \left(\frac{p_{qj}}{p_q} - p_j \right)^2 / p_j \quad \text{et } S \text{ est le signe de } (p_{qj} - p_j p_q)$$

c_{qj} est la contribution absolue du point j dans R^n à la distance de la classe q au centre du nuage dans R^p . En étudiant la quantité $S \cdot c_{qj} / \rho^2(q)$, on connaît les points j qui sont discriminants pour la classe q , c'est-à-dire qui fournissent les plus fortes contributions et le signe de S indique si p_{qj} est très faible ou très élevé. (Si S est > 0 , p_{qj} est très élevé et si S est < 0 , p_{qj} est très faible). Ces contributions peuvent également être calculées sur un tableau de données reconstituées à partir d'un nombre réduit d'axes factoriels. Dans ce cas, on peut caractériser facilement les groupes d'observations déterminés sur les plans factoriels par les paramètres responsables du regroupement.

Interprétation des classes produites par la C.A.H. sur les axes factoriels.

Il est intéressant de comparer les indices des noeuds de la hiérarchie aux valeurs propres associées à l'A.F.C. On a en effet:

$$\text{Inertie totale} = \sum_q \tau(q) = \sum_{\alpha} \lambda_{\alpha}$$

Soient s et t deux classes qui sont les successeurs immédiats de la classe q. On a :

$$\tau(q) = \frac{p_t \cdot p_s}{p_t + p_s} \sum_{\alpha} \left((F_{\alpha}(s) - F_{\alpha}(t))^2 \right)$$

On pose

$$I(\alpha, q) = \frac{p_t \cdot p_s}{p_t + p_s} \left((F_{\alpha}(s) - F_{\alpha}(t))^2 \right)$$

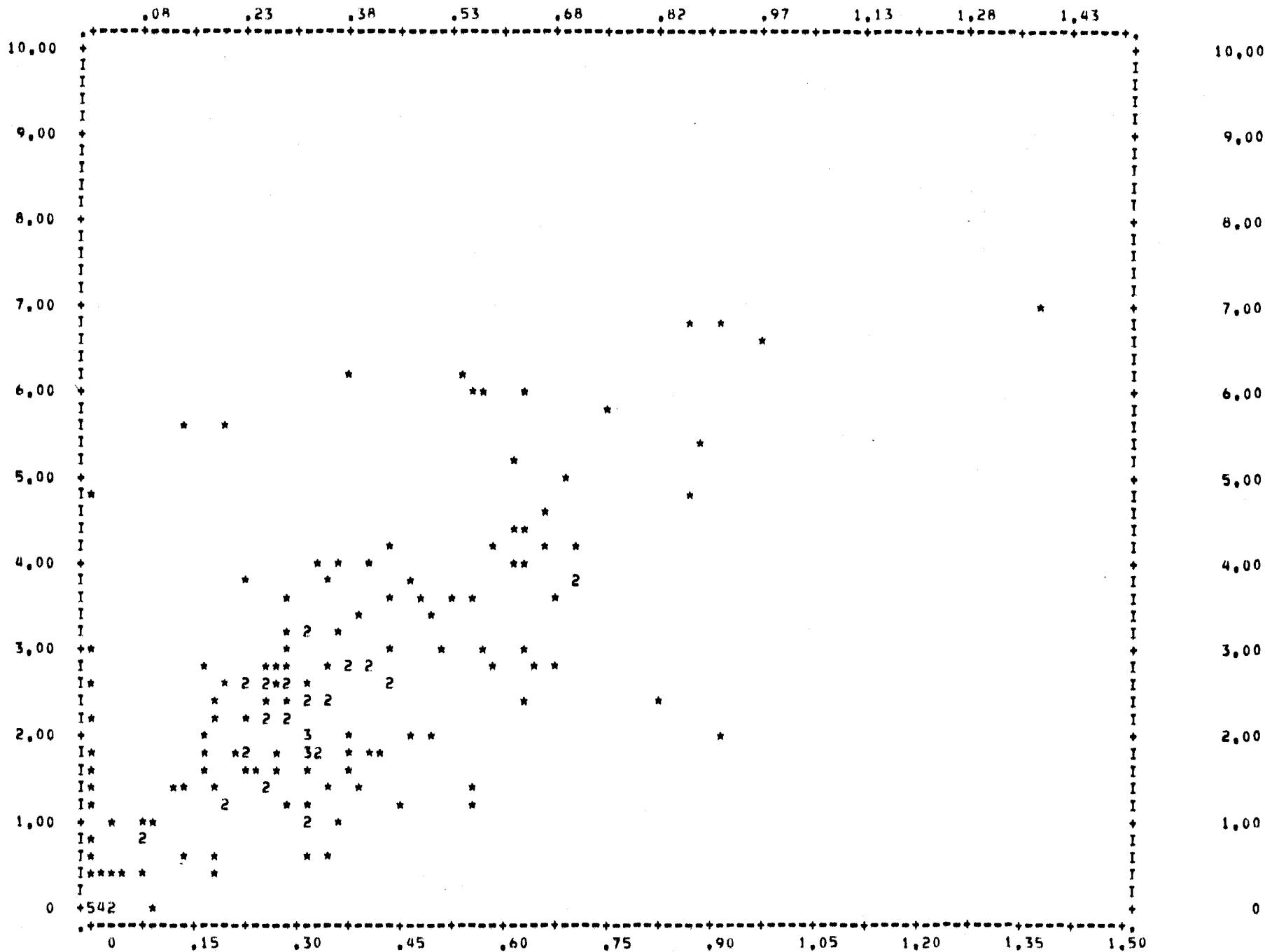
La quantité $I(\alpha, q)$ représente la contribution absolue de la paire (α, q) à l'inertie totale et peut s'exprimer indifféremment comme la contribution absolue de la classe q à l'axe α ou de l'axe α au noeud q. En pratique, on considère les quantités $I(\alpha, q) / \tau(q)$ et $I(\alpha, q) / \lambda_{\alpha}$. En particulier la quantité $I(\alpha, q) / \tau(q)$ représente la contribution relative de l'axe α à la classe q. Si cette valeur se rapproche de 1, c'est que la séparation entre classes pour ce noeud se fait presque exclusivement dans la direction de l'axe α (Jambu, 1976b).

ANNEXE C

GRAPHIQUES

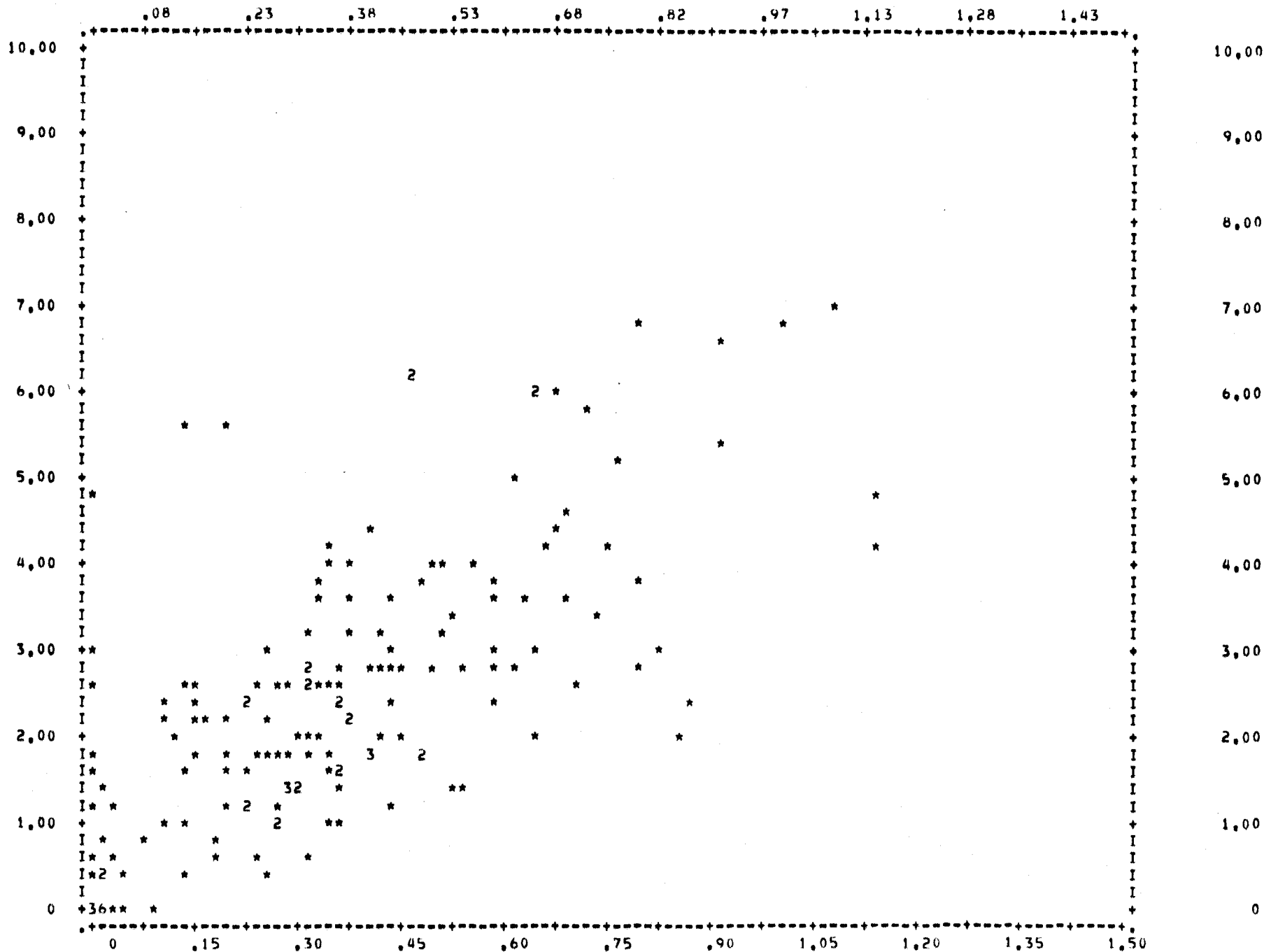
SCATTERGRAM OF (DOWN) A

(ACROSS) B



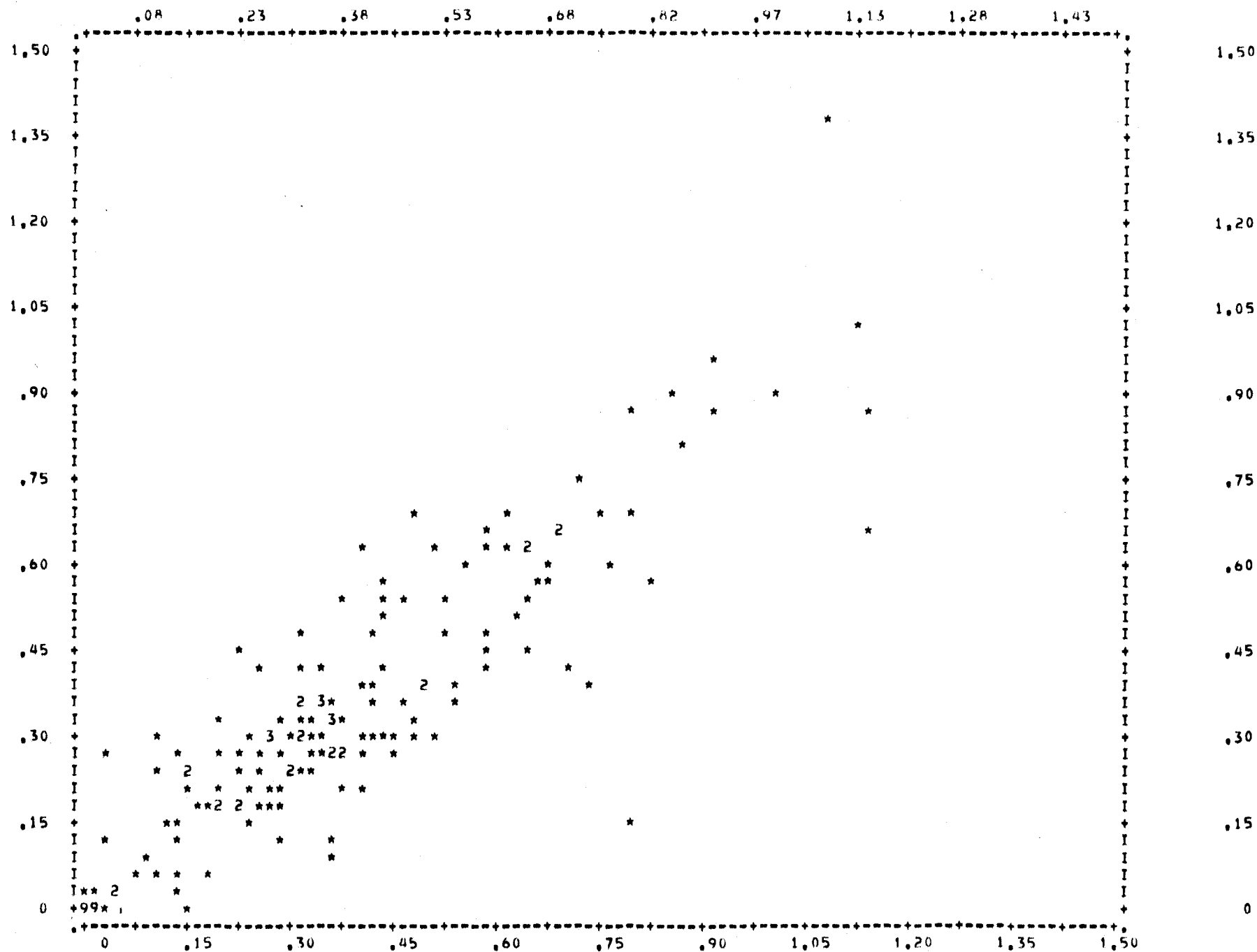
SCATTERGRAM OF (DOWN) A

(ACROSS) C



SCATTERGRAM OF (DOWN) B

(ACROSS) C



ANNEXE D

ETUDE MORPHOLOGIQUE

AU

MICROSCOPE ELECTRONIQUE A BALAYAGE.

Une étude morphologique comparative au microscope électronique à balayage sur des échantillons d'empoussièrage d'amiante obtenus par deux techniques différentes de mesures (RDM et filtre à membrane) a été effectuée afin de trouver une explication dans le manque de corrélation des résultats de l'étude (Etude comparative de deux méthodes de mesure d'empoussièrage dans l'air) effectuée par l'INRS.

Echantillonnage.

La Société Asbestos Corporation a fourni des échantillons provenant de deux postes différents:

POSTE A

- A-1 Filtre à membrane ordinaire 90 mn, 2L/mn numéro 8410;
- A-2 Filtre à membrane avec collecteur 90 mn, 2L/mn à cyclone de 10 microns numéro 8548;
- A-3 Lectures gravimétriques au RDM muni d'un collecteur à cyclone de 10 microns.

POSTE B

- B-1 Filtre à membrane ordinaire 30 mn, 2L/mn numéro 8403;
- B-2 Filtre à membrane avec collecteur à cyclone de 10 microns 30 mn, 2L/mn numéro 8549;

B-3 Lectures gravimétriques au RDM muni d'un collecteur à cyclone de 10 microns.

Préparation des échantillons pour le microscope électronique.

a) RDM.

La totalité des dépôts a été prélevée sur le disque, passée aux ultrasons pendant une minute dans le chloroforme afin d'éliminer la couche de vaseline enrobant les particules puis filtrée sur millipore.

Un quartier du filtre a été découpé, déposé sur un support puis recouvert d'une couche d'or palladium de 200 Å pour observations au microscope.

b) Filtre à membrane.

Un quartier du filtre a été découpé, déposé sur un support puis recouvert d'une couche d'or palladium de 200 Å pour observations au microscope.

Conditions de travail.

L'étude a été faite dans les conditions normales d'utilisation du microscope à balayage à un grossissement de 400 fois (gros-

sissement utilisé dans la méthode optique de comptage sur membrane).

L'observation d'une vingtaine de plages a montré une très bonne homogénéité des échantillons.

Observations et conclusions.

Une photographie représentative de chaque échantillon a été prise, la plage observée étant de 170 microns x 170 microns (voir pages D-5 et D-6).

Les photographies IA et 3A (IB et 3B) montrent la morphologie des échantillons ayant servi à l'étude de corrélation de l'INRS. On voit immédiatement que les deux méthodes ne mesurent pas la même chose.

Dans le cas de la membrane seules les fibres et particules inférieures à 10 microns sont recueillies alors qu'avec le RDM, malgré l'utilisation d'un collecteur à cyclone de 10 microns, des particules pouvant aller jusqu'à 20 microns sont recueillies et mesurées.

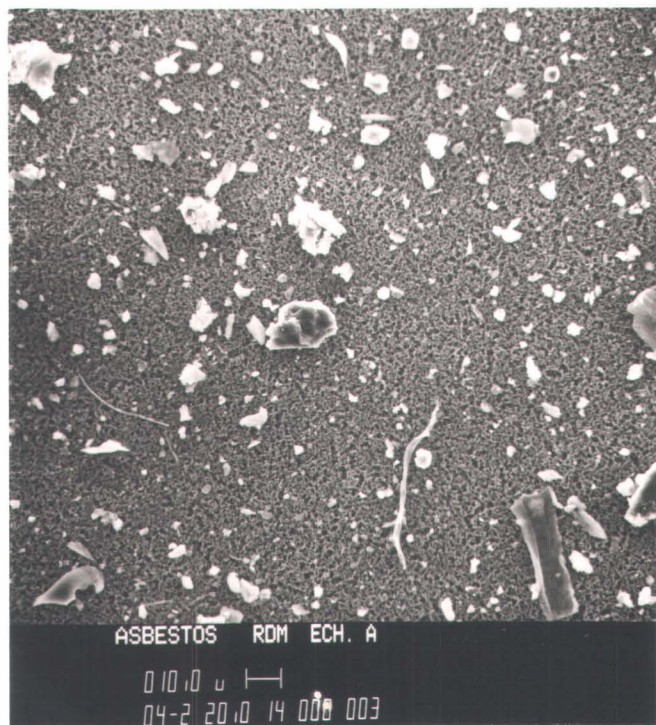
Ceci pourrait expliquer le manque de corrélation des résultats des deux méthodes car le RDM mesurant des poids, la présence d'une ou plusieurs particules de grosse taille influence énormément

la variation de poids du prélèvement.

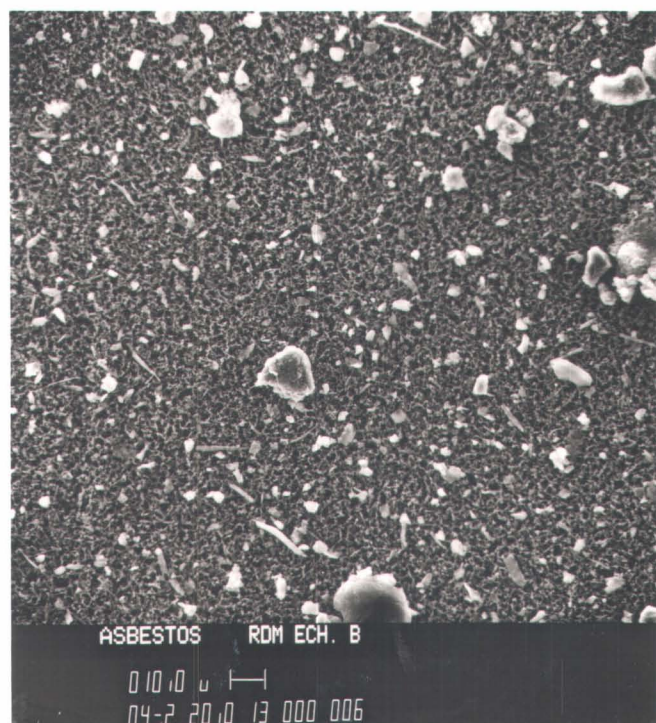
Par contre les photographies 1A et 2A (1B et 2B) montrent une morphologie semblable ce qui laisse à penser que le collecteur à cyclone de 10 microns du RDM ne semble pas jouer son rôle dans l'élimination des particules supérieures à 10 microns.

Remarque: Aucune étude méthodique n'a été effectuée par micro-analyse afin de vérifier que toutes les particules observées sont de l'amiante, cette étude dépassant le cadre du projet. De toute façon le RDM ne faisant pas la différence entre des fibres et des particules autres que de l'amiante, l'étude par microanalyse n'aurait apporté aucune amélioration des résultats et la conclusion serait restée la même.

ECHANTILLONNAGE AU RDM

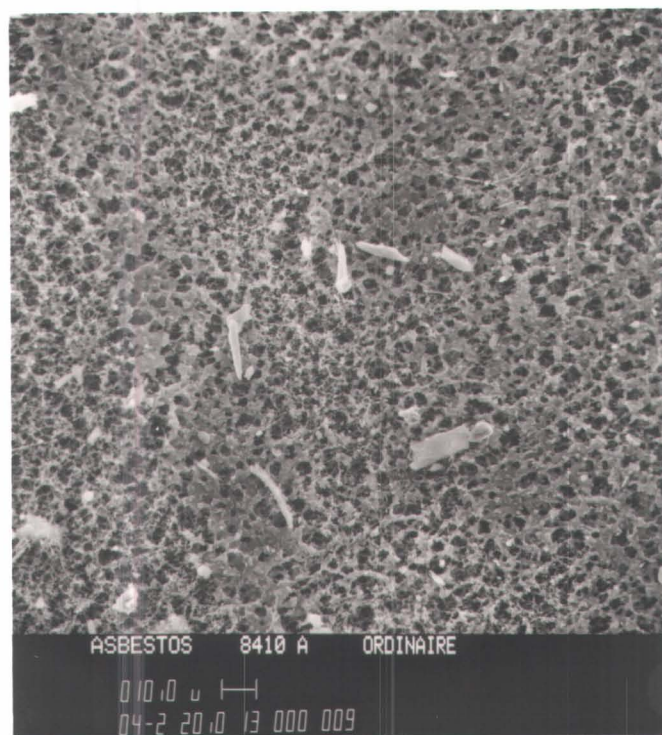


IA

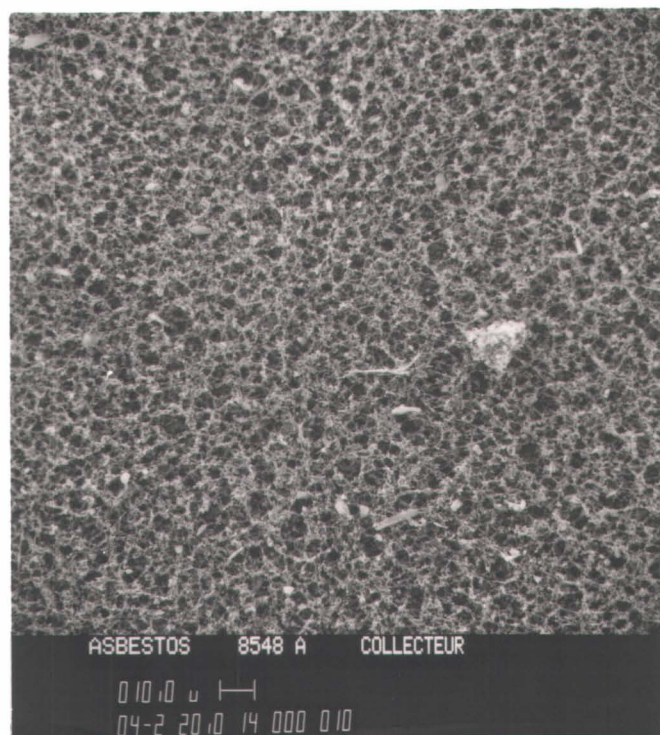


IB

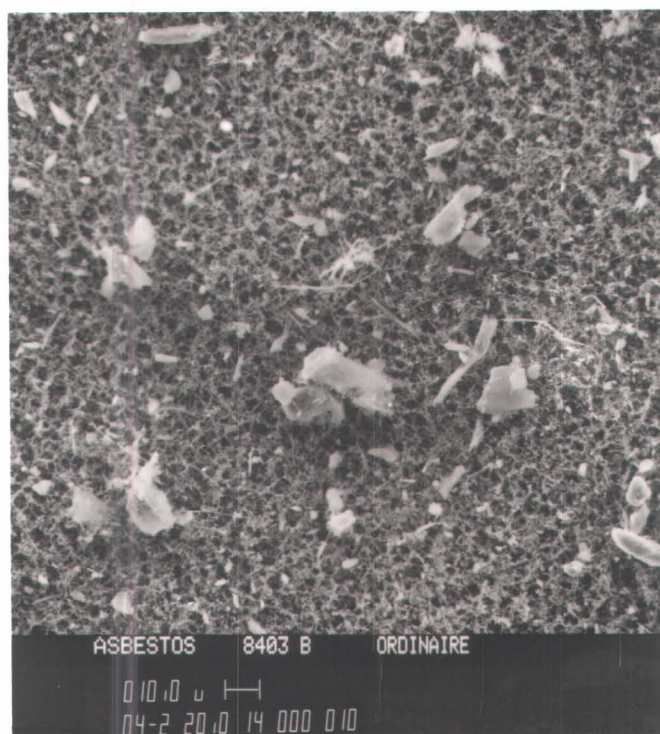
ECHANTILLONNAGE SUR MEMBRANE



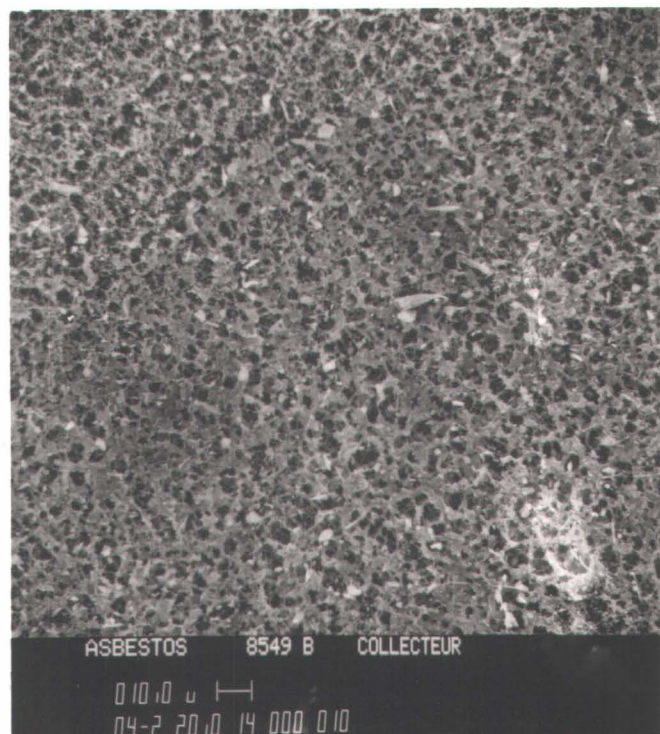
2A



3A



2B



3B