

Etude comparative de deux méthodes de  
mesure d'empoussiérage dans l'air

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

RAPPORT SCIENTIFIQUE No 100  
1979

Rapport rédigé pour  
la Société Asbestos Corporation

par

M. Lachance, B. Bobée, M. Desjardins, R. Bertrand

ISBN 2-89146-097-9

DEPOT LEGAL 1979

Tous droits de traduction, de reproduction et d'adaptation réservés

© 1979 - Institut national de la recherche scientifique

Participation à la réalisation de l'étude

---

Bertrand, Rudolf	Agent de recherche	INRS-Pétrole
Bobée, Bernard	Professeur	INRS-Eau
Desjardins, Michel	Directeur	INRS-Pétrole
Labbé, Nicole	Secrétaire	INRS-Eau
Lachance, Marius	Agent de recherche	INRS-Eau
Parent, André	Graphiste	INRS-Eau

TABLE DES MATIERES	PAGE
LISTE DES TABLEAUX .....	iii
LISTE DES FIGURES .....	iv
CONCLUSIONS ET RECOMMANDATIONS .....	1
INTRODUCTION .....	4
1. Description des mesures .....	4
1.1 Mesure optique .....	4
1.2 Mesure physique .....	5
1.3 Données .....	5
2. Approche utilisée .....	6
3. Resultats .....	9
3.1 Description des séries de mesures .....	9
3.2 Analyse globale .....	18
3.3 Corrélation entre les séries de mesure .....	27
3.4 Complément à l'étude des relations entre les séries de mesure .....	32
3.4.1 Test des observations pairées .....	32
3.4.2 Linéarité des relations entre A et B et entre A et C .....	33
Références bibliographiques .....	38
Annexe A: données .....	39
Annexe B: méthode d'analyse factorielle des correspondances .....	46
Annexe C: représentations graphiques .....	57

LISTE DES TABLEAUX	PAGE
3.1 Description de la distribution des valeurs obtenues pour les séries A, B et C -----	13
3.2 Caractéristiques statistiques des trois séries de mesures---	15
3.3 Séparation des données en 5 classes: seuils et effec- tif par classe -----	20
3.4 Coordonnées et contributions relatives sur les premiers axes factoriels des points représentatifs des classes de valeurs pour chacune des séries de mesure -----	21
3.5 Tableaux de contingence obtenus en croisant 2 à 2, les séries A, B et C.-----	25
3.6 Corrélations entre les séries A, B et C pour différents intervalles de A. -----	29
3.7 Nombre, moyenne et écart-type des valeurs de A regroupées en 19 classes équiprobables et moyennes et écart-types des valeurs de B et C correspondantes. -----	35

## LISTE DES FIGURES

PAGE

3.1	Histogramme montrant la distribution des valeurs obtenues pour la série A .....	10
3.2	Histogramme montrant la distribution des valeurs obtenues pour la série B .....	11
3.3	Histogramme montrant la distribution des valeurs obtenues pour la série C .....	12
3.4	Comparaison entre les méthodes A, B et C: Représentation des distances euclidiennes calculées dans l'espace facto- riel (5 dimensions) entre les points représentant les 5 classes des trois séries de mesures .....	23
3.5	Représentation graphique montrant la linéarité et la dispersion de la relation de la série B avec la série A .....	36
3.6	Représentation graphique montrant la linéarité et la dispersion de la relation de la série C avec la série A .....	37

## CONCLUSIONS ET RECOMMANDATIONS

L'étude mathématique des données de 900 prises d'échantillon lors d'une campagne de contrôle d'empoussièrage s'est voulu d'être globale. En effet, volontairement on a mis de côté plusieurs facteurs pouvant influencer les résultats, entre autres:

- a) la diversité des postes de contrôle ,
- b) le temps de la mesure ,
- c) la représentativité de la mesure ,
- d) la précision des mesures .

On peut dire que la seule hypothèse de départ était la suivante: Si la composition du minerai est demeurée la même durant toute la campagne de contrôle, on devrait trouver une liaison entre les deux mesures différentes faites au même poste et au même moment.

Cette liaison ou relation pourrait par exemple refléter la variation dans la concentration de la fibre par rapport à la localisation du poste de contrôle.

L'étude mathématiques effectuée ici, est divisée en trois parties principales:

- 1) analyse des données ,
- 2) application de l'analyse factorielle des correspondances à l'ensemble des données ,
- 3) calculs des corrélations entre les données relatives aux différentes méthodes de mesure .

De ces analyses, on peut en tirer les conclusions suivantes:

- a) Pour les 900 postes étudiés, près de 80% des données obtenues par la méthode optique (A) sont inférieures à 6 fibres/cm<sup>3</sup>,

- b) Il existe une très grande dispersion dans les valeurs obtenues pour chacune des mesures, et pour chaque série ces valeurs sont distribuées en une courbe de forte asymétrie positive. La dispersion peut sans doute être expliquée en partie par la disparité des postes de contrôle;
- c) L'asymétrie positive et quelque peu différente entre les courbes relatives aux 3 séries de données, peut être expliquée en partie par la précision différente des méthodes et probablement par la nature du matériel analysé;
- d) Il existe une relation entre la mesure optique et la mesure physique (RDM) surtout évidente pour des valeurs supérieures à 6 fibres/cm<sup>3</sup>;
- e) La relation entre les deux méthodes ne peut être prédite avec précision en raison de la grande dispersion des valeurs, dispersion en grande partie due à la disparité des postes de mesure;
- f) L'étude mathématique démontre très nettement qu'il n'y a pas de différence significative entre les mesures du RDM effectuées au début (B) et à la fin (C) d'une expérience.

Plusieurs recommandations peuvent être avancées à partir de nos résultats.

- 1) Il est primordial de faire une étude sur la représentativité des deux mesures. Pour cela un essai préliminaire peut être effectué au microscope électronique à balayage. Il s'agirait de comparer morphologiquement et chimiquement la composition des deux échantillonnages (RDM et optique). Suite à cet essai préliminaire, on pourrait proposer une étude systématique de plusieurs échantillons.



- 2) A notre avis, pour mieux définir ou établir une relation entre les deux méthodes de mesure, en un premier temps, il serait intéressant d'appliquer notre méthode statistique sur quelques zones bien définies de contrôle et non sur l'ensemble des postes.
- 3) Le choix de groupes de postes devrait être fait selon un critère d'homogénéité; ainsi dans un même groupe on pourrait retrouver les postes de comportement semblable et on éliminerait ainsi la grande dispersion entre les valeurs observées. En pratique, l'A.F.C. serait un outil précieux pour déterminer les groupes possédant une certaine homogénéité. La seule condition d'utilisation de l'A.F.C. est de connaître de manière précise la localisation des sites d'échantillonnage dans l'usine.
- 4) Finalement, même si les premiers résultats ne sont pas très prometteurs, nous croyons que quelques essais additionnels doivent être faits avant de conclure définitivement qu'on ne peut prédire, avec plus de précision, le nombre de fibres/cm<sup>3</sup> à partir de mesures de RDM.

## INTRODUCTION

La Société Asbestos Corporation nous a confié l'analyse des résultats obtenus lors d'une campagne de mesure d'empoussiérage dans l'air à différents endroits de leurs usines et de leurs mines.

Le but de ce travail est en particulier d'examiner, en appliquant différentes méthodes statistiques de traitement de données par ordinateur, s'il existe une relation entre les mesures dites optiques (comptage de fibres au microscope) et celles dites physiques (mesure de l'absorption des rayons  $\beta$  par une masse de poussière). Il avait été entendu que les méthodes de traitement statistiques de données seraient appliquées aux données provenant d'une campagne de contrôle d'empoussiérage, c'est-à-dire, sur 900 mesures effectuées à divers endroits des usines et des mines pendant la période du 13 février au 21 avril 1978.

### 1. DESCRIPTION DES MESURES

#### 1.1 Mesure optique

Une quantité d'air prédéterminée ( $180000\text{cm}^3$ ) plus ou moins chargée de poussière est pompée sur un filtre millipore de section efficace de  $855\text{mm}^2$ . Normalement la prise d'échantillon dure 90 minutes. Une partie du filtre est traitée pour étude subséquente au microscope optique. A un grandissement de 400 fois, l'observateur compte le nombre de fibres de dimensions supérieures à 5 micromètres sur des plages de  $0.003\text{mm}^2$ . Toute autre particule est ignorée. Tenant compte des facteurs tels que la dimension du filtre, la quantité d'air pompée, la grandeur des plages et le nombre de fibres observées, on détermine mathématiquement le nombre de fibres par  $\text{cm}^3$  d'air.

## 1.2 Mesure physique

L'instrument utilisé consiste en un cyclone précollecteur qui retient la fraction non respirable de la poussière totale dans l'air. Essentiellement le collecteur retient toute particule dont la sphère de diamètre équivalent et densité unitaire serait supérieur à 10 micromètres. Les autres particules sont projetées sur un disque et retenues par une couche de "Vaseline". La poussière ainsi récupérée absorbe plus ou moins, selon la quantité (masse), les rayons  $\beta$  émis par une source au carbone 14. La détection du rayonnement non absorbé se fait par l'entremise d'un compteur Geiger.

Les résultats des mesures sont transformés automatiquement en mg de poussière totale respirable par mètre cube d'air pompé. La mesure dure en général 4 minutes et est effectuée au début et à la fin de la prise d'échantillon de 90 minutes pour la mesure optique.

## 1.3 Données

Les données à analyser consistent en trois séries de mesures: la première série notée A correspond aux mesures en fibres/cm<sup>3</sup> obtenues au moyen de la mesure optique tandis que les deux autres séries correspondent aux mesures en mg/m<sup>3</sup> obtenues avec la mesure physique; la série notée B correspond aux mesures sur des échantillons d'air prélevés au cours des 4 premières minutes de l'expérience tandis que la série notée C est obtenue sur des échantillons prélevés pendant les 4 dernières minutes de l'expérience, c'est-à-dire 82 minutes plus tard. Chacune des mesures provient d'échantillons d'air prélevés à différents endroits des usines et des mines au cours d'une campagne d'échantillonnage. Lorsqu'une des trois mesures était manquante, on a éliminé les deux autres mesures. Au total, on dispose de trois séries de 900 valeurs. Ces données ont été entrées sur cartes informatiques dans l'ordre dans lequel elles ont été fournies et apparaissent dans l'annexe A.

## 2. APPROCHE UTILISEE

L'approche qui a été utilisée dans la comparaison des trois séries de mesures se divise en trois parties:

- Une description de la distribution des données est faite en présentant des histogrammes de fréquence de distribution.
- Une comparaison globale des trois séries de mesures est faite au moyen de l'analyse factorielle des correspondances.
- Une étude des interrelations entre chacune des séries de mesures prises deux à deux est faite à partir des coefficients de corrélation pour différentes gammes de variations.

Pour compléter et finaliser l'analyse, on a effectué en plus un test de "Student" pour comparer les moyennes des valeurs de B et C, et une vérification de la linéarité des relations de la série A avec les séries B et C.

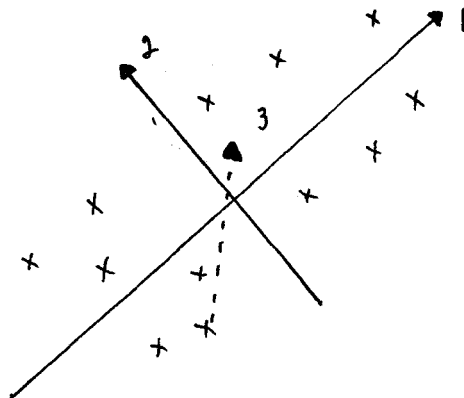
Puisqu'une partie essentielle de l'étude est effectuée au moyen de l'analyse factorielle des correspondances (A.F.C.), et que cette technique d'analyse est encore peu connue en Amérique du Nord, on présente ici les caractéristiques essentielles de cette méthode. L'annexe B donne une description plus détaillée des aspects théoriques de la méthode.

### *Analyse factorielle des correspondances (A.F.C.)*

Dans l'analyse factorielle des correspondances, on considère un tableau de  $n$  lignes et  $p$  colonnes représentant par exemple les valeurs de  $p$  variables en  $n$  stations. La description de ce tableau peut se faire en représentant par un nuage de  $n$  points dans un espace à  $p$  dimensions, l'ensemble des vecteurs (lignes) correspondant aux valeurs des  $p$  variables pour chacune des  $n$  stations. On peut également représenter par un nuage de  $p$  points, dans un espace comprenant autant de dimensions que de stations, l'ensemble des vecteurs (colonnes cor-

respondant aux valeurs obtenues aux  $n$  stations pour chacun des  $p$  paramètres. Dans le cas qui nous intéresse, on pourrait par exemple d'une part, représenter les 900 mesures par un nuage de 900 points dans un espace à trois dimensions (variables), chacune des dimensions correspondant aux séries A, B et C, d'autre part, représenter par trois points les séries A, B et C dans un espace à 900 dimensions. L'A.F.C. permet de représenter les nuages de points dans un système d'axes plus adéquat, de sorte qu'il est possible de réduire le nombre de dimensions tout en étant capable de bien représenter le tableau initial de valeurs.

On fait subir au tableau de départ une transformation. (cf. annexe B) afin de rendre le tableau homogène. On recherche ensuite les axes principaux du nuage des  $n$  points dans l'espace à  $p$  dimensions. L'axe principal 1 est l'axe d'allongement maximal des  $n$  points (voir schéma ci-dessous), c'est-à-dire celui qui explique le maximum de variabilité (variance).



Après avoir calculé ce premier axe, on en calcule un second qui est perpendiculaire au premier et qui explique encore le maximum de variabilité résiduelle (excluant celle expliquée par le premier axe). Puis, on en calcule un troisième qui est perpendiculaire aux deux autres et qui explique encore le maximum de variance résiduelle. On extrait ainsi les axes principaux par ordre décroissant de variance expliquée et on s'arrête au moment où toute la variation du nuage de points est expliquée.

Les axes étant extraits par ordre d'importance, il est possible de négliger les derniers axes extraits et d'assimiler le nuage initial à sa projection dans un espace à  $q$  dimensions  $q < p$ . Par exemple, à partir d'un nuage de 900 points dans un espace à 15 dimensions, on peut décider de retenir les 5 premiers axes factoriels si on considère que l'essentiel des variations de chacune des variables est expliqué. Ainsi, la représentation d'un nuage de points dans cet espace de dimension réduite permet d'éliminer des fluctuations aléatoires tout en conservant l'essentiel de la variation.

La transformation préalable effectuée sur le tableau de départ implique que deux vecteurs colonnes (ou lignes) de ce tableau qui ont des composantes proportionnelles sont représentés dans le nuage (donc en projection dans le plan des axes principaux) par deux points confondus. Cette particularité de l'A.F.C. permet donc de mettre en évidence des similitudes de comportement ou des types de variations identiques, puisque les points représentatifs des paramètres (ou des observations) seront confondus ou voisins dans la représentation graphique et auront donc des coordonnées identiques ou du même ordre de grandeur.

En pratique, pour permettre une analyse plus fine, il peut être intéressant de transformer les données initiales en données binaires. On divise l'intervalle d'un paramètre  $j$  en  $k$  classes d'effectifs approximativement égaux. Une observation appartenant seulement

à une des classes est codée 1 pour cette classe et 0 pour les autres classes. On obtient ainsi un tableau de valeurs binaires composé de  $n$  lignes et  $(p.k)$  colonnes (si chaque paramètre est séparé en  $k$  classes). Dans le cas qui nous intéresse, si on divise l'intervalle de variation des séries A, B et C en 5 classes d'effectif semblable, on obtient ainsi en transformant le tableau initial, un tableau de valeurs binaires composé de 15 colonnes et 900 lignes. L'analyse de ce tableau de valeurs binaires, au moyen de l'A.F.C., permet une analyse plus stable et un niveau d'interprétation plus fin. On peut ainsi, par l'étude des proximités entre les points représentant les mêmes classes, mettre en évidence des types de variation identique sur différents intervalles de variation du paramètre.

### 3. RESULTATS

Comme nous l'avons indiqué précédemment (section 2), l'analyse comprend 3 parties qui sont présentées en trois sections séparées. La complémentarité des résultats partiels de ces trois parties de l'étude permettra de tirer des conclusions plus complètes sur les relations existant entre les trois séries.

#### 3.1 Description des séries de mesures

L'étude de relations entre des séries des valeur nécessite une connaissance de l'allure de la distribution de ces valeurs. A cette fin, on a séparé l'intervalle de variation de chacune des séries en 14 petits intervalles égaux (classes). Pour la série A, la largeur des intervalles de chacune des classes correspond à 1 fibre /cm<sup>3</sup> tandis que pour les séries B et C, la largeur des intervalles est égale à 0.1 mg/m<sup>3</sup>. 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 3.1. Les figures 3.1, 3.2 et 3.3 montrent les histogrammes résultant de ces comptages pour les séries A, B et C respectivement. En ce qui concerne la dernière classe de chaque histogramme, on a regroupé les valeurs supérieures à 14 fibres/cm<sup>3</sup> pour la série A et les valeurs

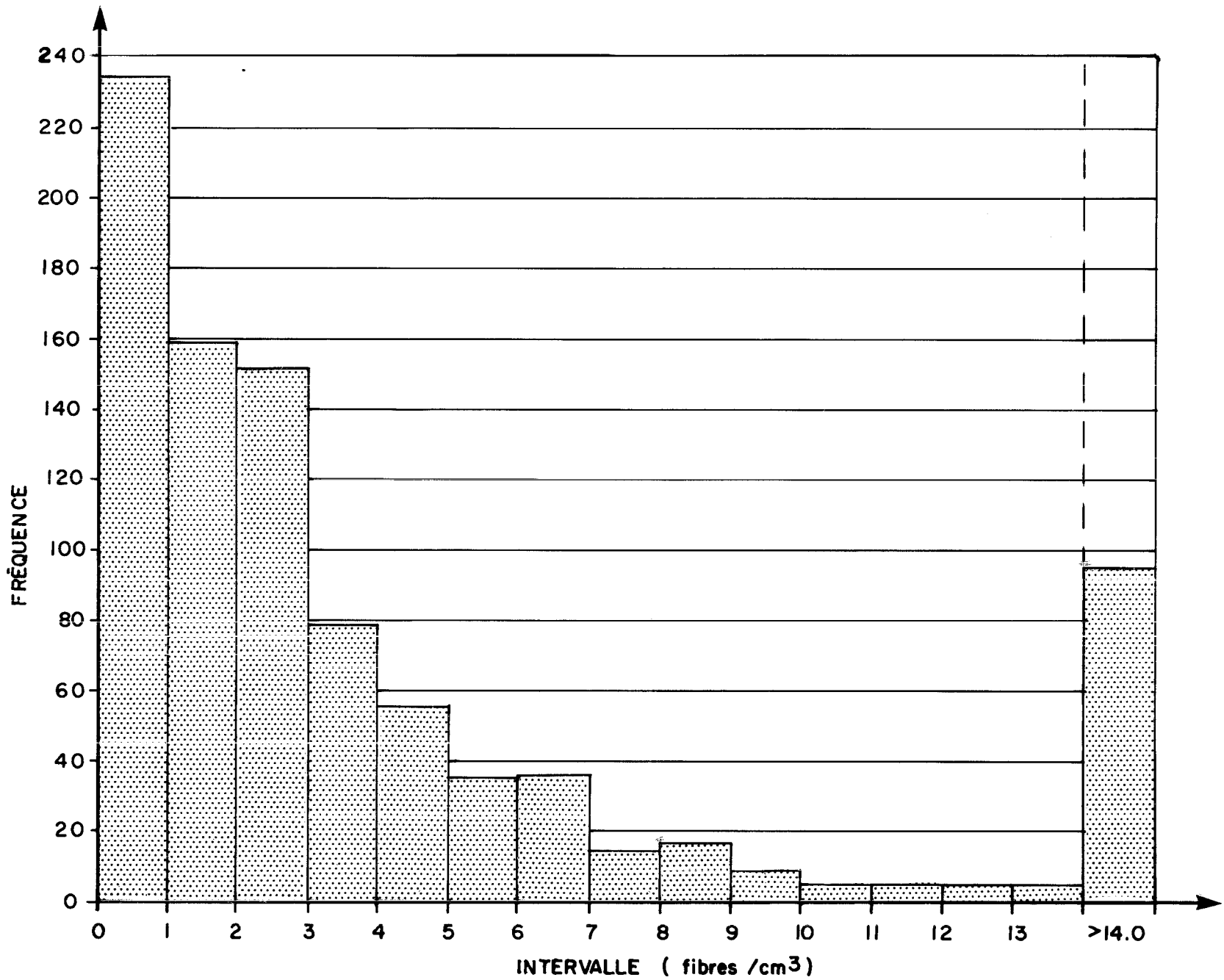


Figure 3.1. Histogramme montrant la distribution des valeurs obtenues pour la série A.



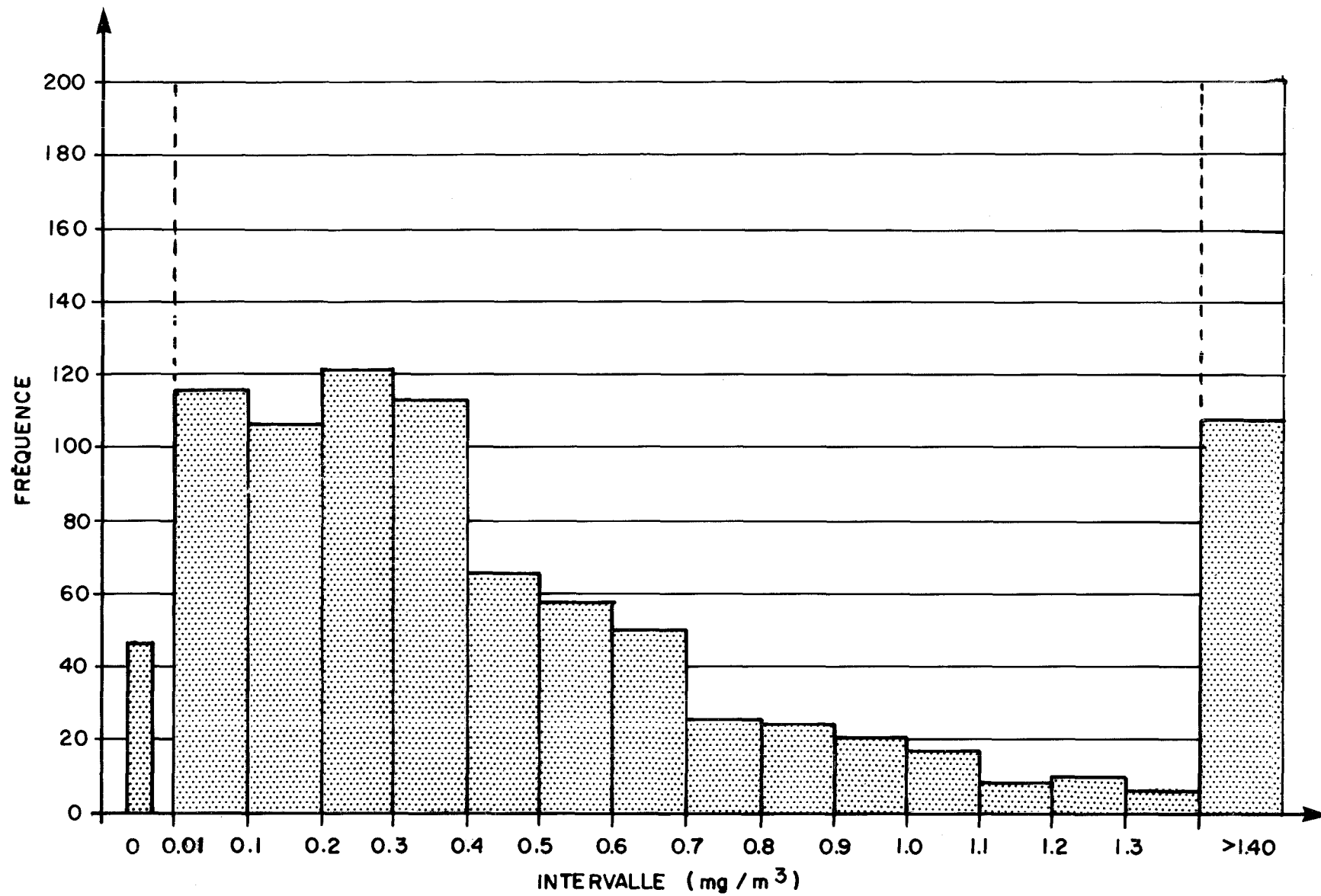


Figure 3.2. Histogramme montrant la distribution des valeurs obtenues pour la série B.

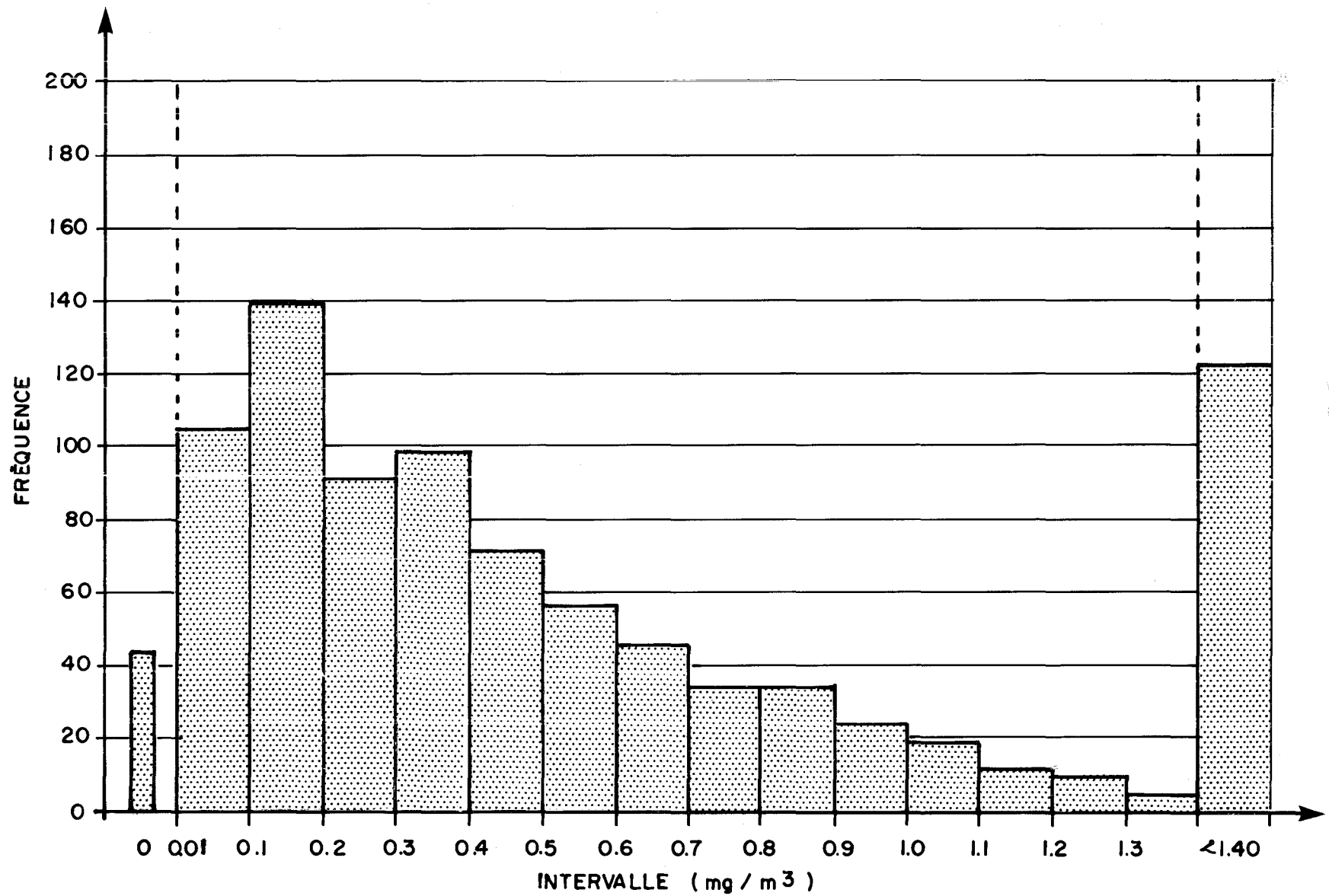


Figure 3.3. Histogramme montrant la distribution des valeurs obtenues pour la série C.

TABLEAU 3.1: Description de la distribution des valeurs obtenues pour les séries A, B et C.

classe	intervalle (fibre/cm <sup>3</sup> )	SERIE A			intervalle (mg/m <sup>3</sup> )	SERIE B			SERIE C		
		frequence	frequence relative (%)	fréquence cumulative (%)		fréquence	fréquence relative (%)	fréquence cumulative (%)	fréquence	fréquence relative (%)	fréquence cumulative (%)
1	- 0.0	0	0.0	0.0	- 0.00	47	5.2	5.2	43	4.8	4.8
2	0.0 - 1.0	234	26.0	26.0	.00 - .10	115	12.8	18.0	105	11.7	16.4
3	1.0 - 2.0	159	17.7	43.7	.10 - .20	106	11.8	29.8	138	15.3	31.8
4	2.0 - 3.0	151	16.8	60.4	.20 - .30	121	13.4	43.2	91	10.1	41.9
5	3.0 - 4.0	79	8.8	69.2	.30 - .40	113	12.6	55.8	98	10.9	52.8
6	4.0 - 5.0	54	6.0	75.2	.40 - .50	66	7.3	63.1	70	7.8	60.6
7	5.0 - 6.0	35	3.9	79.1	.50 - .60	57	6.3	69.4	55	6.1	66.7
8	6.0 - 7.0	35	3.9	83.0	.60 - .70	50	5.6	75.0	45	5.0	71.7
9	7.0 - 8.0	13	1.4	84.4	.70 - .80	27	3.0	78.0	32	3.6	75.2
10	8.0 - 9.0	16	1.8	86.2	.80 - .90	25	2.8	80.8	32	3.6	78.8
11	9.0 - 10.0	8	.9	87.1	.90 - 1.00	21	2.3	83.1	22	2.4	81.2
12	10.0 - 11.0	5	.6	87.7	1.00 - 1.10	19	2.1	85.2	19	2.1	83.3
13	11.0 - 12.0	6	.7	88.3	1.10 - 1.20	8	.9	86.1	12	1.3	84.7
14	12.0 - 13.0	5	.6	88.9	1.20 - 1.30	10	1.1	87.2	9	1.0	85.7
15	13.0 - 14.0	4	.4	89.3	1.30 - 1.40	6	.7	87.9	6	.7	86.3
16	> 14.0	96	10.7	100.0	1.40	109	12.1	100.0	123	13.7	100.0

supérieures à  $1.4 \text{ mg/m}^3$  pour les séries B et C. D'autre part, on a considéré dans la première classe, les valeurs nulles pour chacune des séries.

Pour les trois séries, on observe une distribution avec une forte asymétrie positive, c'est-à-dire avec une longue queue vers la droite. En effet, si on n'avait pas regroupé dans une même classe les valeurs élevées, on aurait dû augmenter le nombre d'intervalles d'un facteur 100 pour représenter la distribution de chacune des séries.

La distribution de la série A a une allure différente de celle obtenue pour les séries B et C:

- un nombre relativement élevé (26%) de valeurs de A sont comprises entre 0 et 1 fibre/cm<sup>3</sup>. De plus, un nombre moins élevé, mais important se retrouve dans les deux classes suivantes (1-2, 2-3 fibres/cm<sup>3</sup>) de telle sorte qu'on retrouve 60% des valeurs comprises dans les 3 classes;
- les valeurs des séries B et C sont étalées plus uniformément sur plusieurs classes et spécialement sur les classes 2 à 5 correspondant aux intervalles 0.0 à 0.4 mg/m<sup>3</sup>. En effet, on retrouve environ 50% des valeurs dans ces 4 classes.

Les principales caractéristiques statistiques des trois séries de mesures apparaissent au tableau 3.2. Pour chacune des séries, on a calculé la moyenne ( $m$ ), l'écart-type ( $s$ ), le minimum, le maximum, le coefficient de variation ( $s/m$ ) et le coefficient d'asymétrie ( $c_s$ ). Pour la moyenne, l'écart-type et le coefficient d'asymétrie, on a utilisé les formules suivantes:

TABLEAU 3.2: Caractéristiques statistiques des trois séries de mesures.

caractéristiques	séries		
	A	B	C
moyenne (m)	15.7	1.20	1.24
écart-type (s)	78.4	4.94	4.84
minimum	0.02	0.00	0.00
maximum	1251.5	99.4	99.4
coefficient de variation (s/m)	5.0	4.1	3.9
coefficient d'asymétrie ( $c_s$ )	10.5	13.0	12.7

Note: La moyenne, l'écart-type, le minimum et le maximum sont donnés en fibres/cm<sup>3</sup> pour la série A et en mg/m<sup>3</sup> pour les séries B et C.

$$m = \frac{\sum_{i=1}^{900} X_i}{900}$$

$$s = \frac{\sum_{i=1}^{900} (X_i - m)^2}{899}$$

$$c_s = \frac{\sum_{i=1}^{900} (X_i - m)^3}{900 s^3}$$

On note que l'écart-type et conséquemment le coefficient de variation sont très élevés pour les 3 séries de mesures indiquant la grande dispersion des valeurs. Les valeurs élevées ont un grand poids dans le calcul de l'écart-type parce qu'intervient dans le calcul le carré de la différence entre cette valeur et la valeur moyenne qui est beaucoup plus faible.

Lorsque les valeurs sont distribuées normalement (symétriquement), le coefficient d'asymétrie est nul. Il est négatif si on trouve une grande densité de fortes valeurs et quelques faibles valeurs (longue queue vers la gauche). Lorsqu'il est positif, c'est que la densité est plus forte dans les faibles valeurs avec une longue queue vers la droite. Les fortes valeurs de coefficient d'asymétrie observées au tableau 3.2 confirment les observations constatées à partir des histogrammes de distribution des séries de mesure (figures 3.1 à 3.3).

#### *Interprétation*

L'allure des histogrammes présentés dans les figures 3.1, 3.2 et 3.3 dépend jusqu'à un certain point du choix des intervalles utilisés.

délimitant les classes. Le choix des intervalles résulte habituellement d'un compromis entre une résolution sommaire et une trop grande finesse. Puisque le choix des intervalles est fait de façon un peu arbitraire, on ne peut tirer de conclusions fermes sur l'allure quelque peu différente des distributions. En effet, si on avait choisi des intervalles plus petits pour A (intervalle =  $0.5 \text{ fibre/cm}^3$  par exemple) et si on s'était limité aux valeurs de A inférieures à  $6.0 \text{ fibres/cm}^3$ , on aurait probablement observé pour A une distribution un peu plus ressemblante à celle obtenue pour les séries B et C. En effet, les coefficients d'asymétrie calculés indépendamment des classes, étant à peu près égaux pour chacune des séries (tableau 3.2), montrent qu'il n'existe pas effectivement de très grandes différences dans la forme des distributions. Par contre, même si les intervalles des classes étaient modifiées, les classes de fréquences maximum se retrouveraient probablement dans la même région indiquant qu'il existe effectivement une certaine différence dans la forme des distributions.

Le fait que l'on retrouve dans les histogrammes 3.2 et 3.3 des distributions dont la forme paraît différente de celle de la figure 3.1, peut s'expliquer en tenant compte des facteurs suivants:

- a) pour les deux échantillonnages, on ne mesure pas la même chose du point de vue de la dimension des particules. Dans le cas de la mesure physique (RDM), le cyclone normalement intercepte les particules de densité unitaire dont le diamètre est supérieur à  $10\mu$  (équivalent à environ  $4\mu$  pour des phyllosilicates). Pour les mesures optiques, ne sont comptées que les particules fibreuses dont les dimensions sont  $> 5\mu$  et dont le rapport longueur sur diamètre est au moins de 3:1. En réalité, on ne mesure donc pas exactement la même chose.
- b) la précision des mesures des deux méthodes varie selon la concentration. En effet, moins il y a de fibres, plus le comptage

optique devrait être précis. Le contraire existe pour l'appareil RDM dont le bruit de fonds est sûrement non négligeable.

Etant donné la disparité existant entre les 900 postes d'échantillonnage, il est normal d'observer une grande dispersion dans les valeurs. On retrouve une forte asymétrie pour chacune des séries de mesures. En ce qui concerne les mesures optiques, il est réconfortant d'observer que la majorité (60%) des valeurs mesurées sont inférieures à 3 fibres/cm<sup>3</sup>, c'est-à-dire en deçà des normes. On peut ainsi en conclure qu'il existe dans l'ensemble des stations d'échantillonnage, une bonne qualité de contrôle de l'empoussiérage.

En résumé, on peut retenir de cette partie de l'étude:

- il existe une grande dispersion dans les valeurs obtenues pour chacune des mesures et ces valeurs sont distribuées avec une forte asymétrie positive;
- les histogrammes de distribution montrent que la majorité des mesures ont des valeurs faibles; 60% des mesures optiques ont des valeurs inférieures à 3 fibres/cm<sup>3</sup>.

### 3.2 Analyse globale

Les résultats de l'analyse faite au moyen de l'A.F.C. sont décrits dans cette section. Cette analyse qui permet de considérer globalement les trois séries vise à faire ressortir les interrelations entre chacune des séries de mesure pour différents intervalles de variation. Il est en effet essentiel, avant d'examiner les corrélations existantes, d'avoir une idée générale de l'éventuelle similitude de comportement des 3 séries.



Dans une première étape, les données relatives aux trois séries de mesure 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 3.3. Dans ce découpage de classes, on s'arrange pour que les seuils de chaque classe aient un sens physique, c'est-à-dire en évitant de placer dans des classes différentes des valeurs égales ou à peu près égales.

Le tableau de départ (900 lignes, 3 colonnes) est transformé en un tableau de valeurs binaires (900 lignes, 15 colonnes). En effet, chacune des valeurs appartenant à une des classes est codée 1 pour cette classe et 0 pour les autres classes. Par exemple, la 4e valeur de la série A égale à 2.20 fibres/cm<sup>3</sup> (annexe A) est comprise dans la classe 3, car cette valeur est supérieure à 1.90 et inférieure à 3.10 fibres/cm<sup>3</sup>. On a donc pour les 5 colonnes relatives à cette mesure 0, 0, 1, 0, 0. La première colonne correspondant à la classe 1 (faibles valeurs) de la série A est représentée par le point A1 et ainsi de suite.

La méthode d'analyse factorielle des correspondances a été appliquée au moyen d'un programme informatique implanté à l'INRS-Eau (Bobée et al. 1978). Ce programme, conçu pour analyser des tableaux de dimension réduite, a dû être modifié pour permettre l'analyse du tableau assez important de cette étude (900 lignes, 15 colonnes).

L'analyse a permis de déterminer les axes principaux du nuage des 900 points dans un espace à 15 dimensions; les 5 premiers axes principaux expliquent 73.3% de la variabilité totale du nuage de points. Au tableau 3.4, on donne les coordonnées (facteurs) des points représentant les 5 classes de chacune des séries (points variables) sur les 5 premiers axes principaux. Les contributions relatives (cf. annexe B) des points variables sur chacun des axes sont représentées par des astérisques suivant leur niveau dans ce tableau. La contribution relative cumulée après 5 axes, qui exprime la qualité de la représentation

Tableau 3.3: Séparation des données en 5 classes: seuils et effectif par classe.

méthode	Classe	borne inférieure	borne supérieure	effectif
A	1	.02	.70	177
	2	.70	1.90	175
	3	1.90	3.10	192
	4	3.10	6.40	178
	5	6.40	1251.5	178
B	1	.00	.13	175
	2	.13	.28	176
	3	.28	.45	184
	4	.45	.88	184
	5	.88	99.40	181
C	1	.00	.13	175
	2	.13	.29	183
	3	.29	.49	180
	4	.49	.95	181
	5	.95	99.40	181

Tableau 3.4: Coordonnées et contributions relatives sur les premiers axes factoriels des points représentatifs des classes de valeurs pour chacune des séries de mesure.

point variable	coordonnées sur les axes factoriels					Contribution relative cumulée
	1	2	3	4	5	
A1	1.32 ***	-1.00 *	-.55	.26	.21	.77
B1	1.30 ***	-1.02 **	-.62	.27	-.11	.77
C1	1.33 ***	-1.03 **	-.58	.21	-.06	.78
A2	.57	.33	.95 *	-.79 *	-.98 **	.70
B2	.62	.22	1.00 *	-1.00 *	.41	.63
C2	.60	.30	1.10 **	-.79 *	.30	.60
A3	.04	.96 *	.27	.72 *	1.34 ***	.90
B3	.07	.88 *	.57	1.14 **	-.45	.67
C3	-.02	.89 *	.38	1.19 **	-.59	.67
A4	-.54	.67 *	-1.00 *	-.39	-.86 *	.65
B4	-.46	.92 *	-1.20 **	-.46	.17	.70
C4	-.43	.89 *	-1.18 **	-.59	.34	.71
A5	-1.37 ***	-1.03 **	.33	.12	.17	.76
B5	-1.46 ***	-1.06 **	.27	.02	.01	.83
C5	-1.44 ***	-1.08 **	.25	.01	.00	.84

Contribution relative (CR) indiquée par:

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

Note: la contribution relative indique la qualité de la représentation sur les axes factoriels, ou en d'autres termes, la variance expliquée par chacun des axes factoriels.

se situe entre 0.60 pour la classe 2 de la méthode C (point C2) et 0.90 pour la classe 3 de la méthode A (point A3). Ces valeurs de contribution indiquent que chacun des points est suffisamment bien représenté dans un espace à 5 dimensions. L'analyse des correspondances permet donc de réduire le nuage des 900 points dans l'espace à 15 dimensions, à une représentation dans un nouvel espace à 5 dimensions avec peu de perte d'informations.

Une des propriétés de l'A.F.C. est de faire ressortir les paramètres qui ont des variations semblables. En effet, les paramètres qui ont une variation semblable sont représentés par des points confondus ou voisins dans la représentation. L'examen des coordonnées indiquées sur le tableau 3.4 montre que les points représentant la même classe pour chacune des séries sont relativement voisins spécialement sur les 3 premiers axes principaux; par exemple, les points A2, B2 et C2 ont des coordonnées presque identiques sur les 3 premiers axes principaux. On peut déjà en conclure que les 3 séries de mesures A, B et C se comportent globalement de la même manière et qu'il sera intéressant d'aller plus loin dans l'étude des relations entre chacune des séries.

A partir des coordonnées sur les 5 premiers axes factoriels, (tableau 3.4), on a calculé les distances euclidiennes (sommées 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 entre A1, B1 et C1, puis entre A2, B2 et C2 etc.). Les résultats de ces calculs, apparaissant à la figure 3.4, mettent en évidence les trois 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 distances B2-C2, B3-C3, B4-C4, B5-C5, sont beaucoup plus faibles que les distances A2-B2, A2-C2, A3-B3...

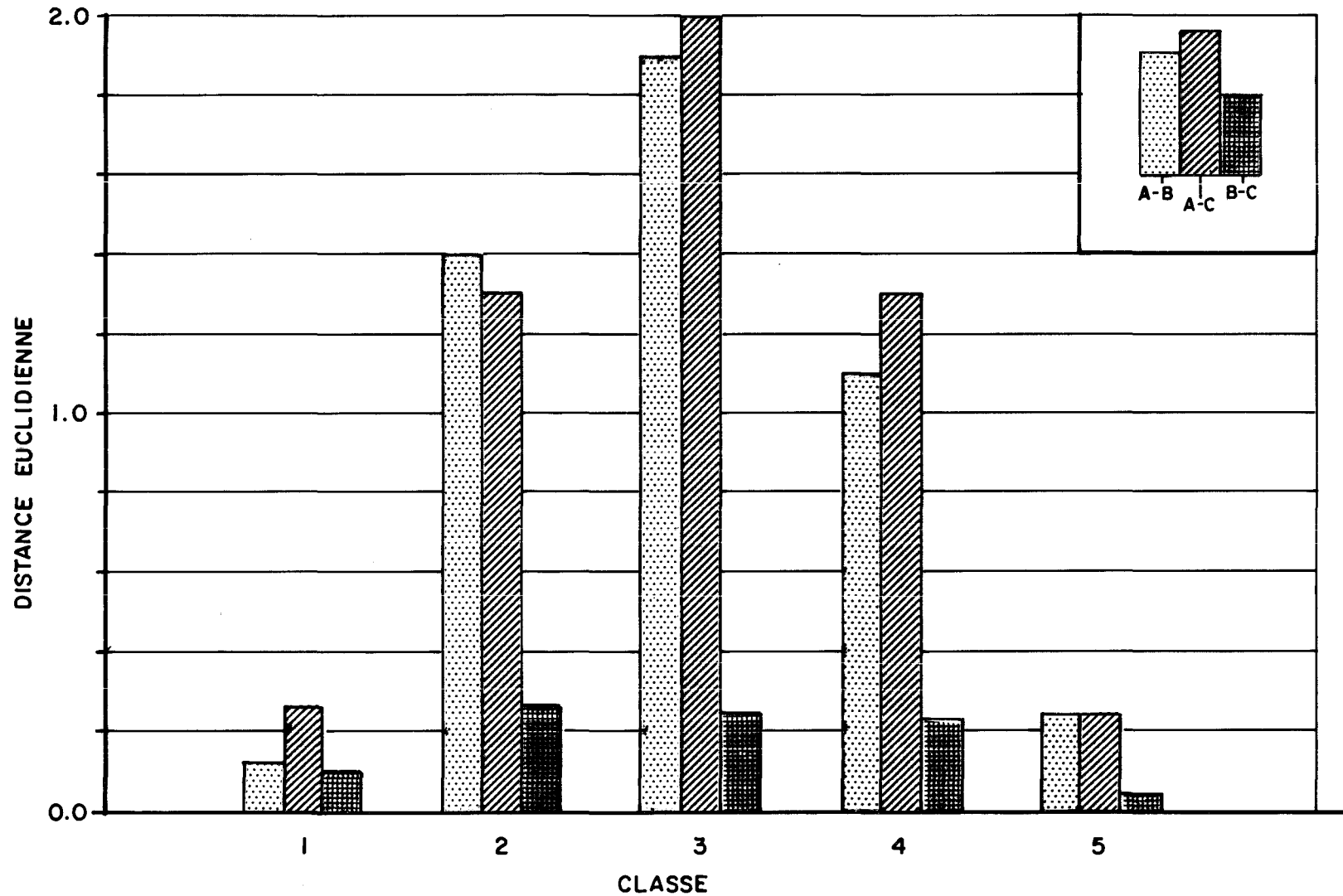


Figure 3.4 . Comparaison entre les méthodes A, B et C : Représentation des distances euclidiennes calculées dans l'espace factoriel (5 dimensions) entre les points représentant les 5 classes des trois séries de mesures .

- Pour les très faibles valeurs (classe 1), il y a autant de différence entre les mesures de B et C qu'entre A et B et entre A et C. En effet, la distance  $B1 - C1$  est comparable aux distances  $A1 - B1$  et  $A1 - C1$ . Les distances étant faibles, on ne peut noter de différence significative entre les 3 séries pour la classe 1.
- Pour les classes intermédiaires (2,3 et 4), les relations entre A et B et entre A et C sont moins fortes qu'avec les classes extrêmes (1 et 5). En effet, les distances  $A2-B2$ ,  $A2-C2$ ,  $A3-B3$ ... sont beaucoup plus grandes que les distances  $A1 - B1$ ,  $A1 - C1$ ,  $A5 - B5$  et  $A5 - C5$ .

La figure 3.4 montre donc qu'il existe 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.

La figure 3.4 traduit mathématiquement l'information contenue dans les tableaux de contingence présentés au tableau 3.5 comme nous le verrons plus loin. Ces tableaux de contingence donnent la répartition des 900 échantillons selon les 5 classes de chacune des séries prises 2 à 2. Par exemple, on a 126 échantillons dont les mesures de A et de B sont dans la classe 1, 33 échantillons dont la mesure de A est dans la classe 2 et la mesure de B est dans la classe 1, et ainsi de suite. Plus le nombre d'éléments sur la diagonale (ex.:  $A1-B1$ ,  $A2-B2$ ,  $A3-B3$  etc.) est important, plus la relation entre les deux séries est forte. A la limite, avec une relation très forte entre deux séries, la somme des éléments sur la diagonale serait égale au nombre total d'échantillons.

La présence d'un plus grand nombre de valeurs sur la diagonale se traduit par une plus faible distance entre les points représentant les classes des séries A, B et C dans l'espace factoriel (figure 3.4). C'est pourquoi, en comparant B avec C, on obtient une distance euclidienne plus faible qu'entre A et B et entre A et C, car les éléments

Tableau 3.5: Tableaux de contingence obtenus en croisant 2 à 2 les séries A, B et C.

	B1	B2	B3	B4	B5	Total
A1	126	37	11	3	0	177
A2	33	76	44	17	5	175
A3	7	46	78	54	7	192
A4	5	11	38	89	35	178
A5	4	6	13	21	134	178
Total	175	176	184	184	181	900

	C1	C2	C3	C4	C5	Total
A1	127	37	7	3	3	177
A2	34	78	43	18	2	175
A3	7	49	75	55	6	192
A4	3	13	40	86	36	178
A5	4	6	15	19	134	178
Total	175	183	180	181	181	900

	C1	C2	C3	C4	C5	Total
B1	128	34	5	5	3	175
B2	40	88	32	10	6	176
B3	5	53	90	31	5	184
B4	2	4	44	115	19	184
B5	0	4	9	20	148	181
Total	175	183	180	181	181	900

Note: Les seuils des classes pour les 3 séries sont indiqués dans le tableau 3.3.

sur la diagonale sont un peu plus abondants spécialement pour les classes 2 à 5. A cet effet, on présente dans le tableau suivant les éléments diagonaux des tableaux dans le cas des classes 2 à 5.

classe méthode	2	3	4	5
B - C	88	90	115	148
A - B	76	78	89	134
A - C	78	75	86	134

Pour la classe 1, il n'existe pas de différence significative entre les 3 séries. Le nombre d'éléments B1 - C1 (128) est presque identique à celui pour A1 - B1 et A1 - C1.

#### *Interprétation*

La méthode d'analyse factorielle des correspondances permet de comparer globalement les trois séries de mesure. Elle montre en particulier:

- il existe une forte relation entre B et C, spécialement pour les valeurs de B plus grandes que  $0.13 \text{ mg/m}^3$  (classe 2 à 5): la durée du temps entre deux échantillonnages n'affecte donc pas sensiblement la mesure.
- pour les valeurs très faibles de B (classe 1), la relation entre B et C est affaiblie par l'augmentation de l'imprécision de la mesure ;



- il existe une certaine relation entre A et B et entre A et C.  
Cette relation apparait plus forte pour les valeurs extrêmes.

Cette analyse a donc permis de détecter les relations possibles entre chacune des séries de mesures pour différents intervalles de variation. Elle montre l'intérêt de poursuivre l'analyse en examinant plus en détail les relations entre chacune des séries prises 2 à 2 (cf. section 3.3).

Rappelons que cette analyse a été effectuée sur un tableau de valeurs codées. Le codage a l'avantage d'effectuer une certaine homogénéité en éliminant l'effet d'échelle des valeurs brutes. Ainsi, une valeur de A dans la classe 5 est codée 1 peu importe si la valeur brute est égale à 7 ou 1250. C'est pourquoi, cette analyse ne fait pas ressortir la dispersion qui peut exister et masquer la relation. On examinera plus en détail cette dispersion des valeurs mesurées dans la section 3.4.

### 3.3 Corrélation entre les séries de mesure

L'étude des relations entre deux séries de mesures peut être faite en calculant le coefficient de corrélation  $r$ :

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad \text{on a: } -1 \leq r \leq 1$$

dans laquelle  $\bar{X}$  et  $\bar{Y}$  sont la moyenne de chacune des séries. Le coefficient de corrélation permet de mesurer la dépendance linéaire entre deux séries lorsqu'elle existe. En effet, il peut y avoir une dépendance

non linéaire entre X et Y et le calcul du coefficient de corrélation ne montrera pas cette dépendance. C'est pourquoi, il est nécessaire d'examiner graphiquement X en fonction de Y afin de vérifier si le coefficient de corrélation a un sens physique.

Le coefficient de corrélation est égal à 1 ou -1 lorsqu'il y a entre chacune des séries une dépendance linéaire complète, et est égal à zéro lorsqu'il n'y a aucune dépendance linéaire.

Rappelons également que les séries X et Y doivent suivre une certaine distribution normale. La présence d'une grande asymétrie dans la distribution des valeurs peut affaiblir ou renforcer la valeur du coefficient de corrélation et lui enlever en grande partie, son sens physique. Dans certains cas, la présence de quelques valeurs extrêmes peut jouer un rôle trop prépondérant. Par exemple, une valeur de A égale à 1250 fibres/cm<sup>3</sup> par rapport à une moyenne de A égale à 16 fibres/cm<sup>3</sup>, joue un rôle trop important dans le calcul, comme nous le verrons plus loin.

On a d'abord calculé les coefficients de corrélation entre chacune des séries pour les 900 échantillons (tableau 3.6). On obtient des coefficients de corrélation élevés dans chacun des cas (0.82 entre A et B et entre A et C; 0.98 entre B et C). En éliminant les deux valeurs de A supérieures à 1000 fibres/cm<sup>3</sup>, on obtient un coefficient de corrélation un peu plus élevé entre A et B et entre A et C (0.87 par rapport à 0.82).

Au niveau global, c'est-à-dire en considérant presque tout l'intervalle de variation de A ( $A < 615$  fibres/cm<sup>3</sup>), on observe que les coefficients de corrélation sont élevés, indiquant la présence d'une forte relation entre chacune des séries; la relation entre B et C est plus forte qu'entre A et B et entre A et C. Il ne faut pas oublier cependant que l'interprétation de ces valeurs de coefficients doit tenir compte de l'influence des fortes valeurs qui faussent en partie les

Tableau 3.6: Corrélations entre les séries A, B et C pour différents intervalles de A.

paramètres correlés	coefficient de corrélacion	nombre de valeurs	intervalles de A
A - B A - C B - C	.82 .82 .98	900 900 900	globale globale globale
A - B A - C B - C	.87 .87 .96	898 898 898	A < 615 A < 615 A < 615
A - B A - C B - C	.34 .11 .29	220 220 220	A < 1.0 A < 1.0 A < 1.0
A - B A - C B - C	.32 .28 .65	303 303 303	$1.0 \leq A < 3.0$ $1.0 \leq A < 3.0$ $1.0 \leq A < 3.0$
A - B A - C B - C	.28 .25 .58	187 187 187	$3.0 \leq A < 6.0$ $3.0 \leq A < 6.0$ $3.0 \leq A < 6.0$
A - B A - C B - C	.88 .86 .94	188 188 188	$6.0 \leq A < 615$ $6.0 \leq A < 615$ $6.0 \leq A < 615$
A - B A - C B - C	.62 .56 .66	710 710 710	A ≤ 6.0 A ≤ 6.0 A ≤ 6.0

calculs. C'est pourquoi, on a procédé à des calculs de coefficients de corrélation en considérant différents intervalles de variation de A:

intervalle	1	$A < 1.0$ fibres/cm <sup>3</sup>
intervalle	2	$1.0 \leq A < 3.0$ fibres/cm <sup>3</sup>
intervalle	3	$3.0 \leq A < 6.0$ fibres/cm <sup>3</sup>
intervalle	4	$6.0 \leq A < 615$ fibres/cm <sup>3</sup>
intervalle	5	$A < 6.0$ fibres/cm <sup>3</sup>

Les résultats de ces calculs de coefficient de corrélation sont donnés dans le tableau 3.6:

- Pour chacun des trois premiers intervalles de variation de A, la relation entre A et B et entre A et C est faible (r se situe autour de 0.2).
- Pour les valeurs de A supérieures à 6 fibres/cm<sup>3</sup>, la relation entre A et B et entre A et C est très bonne (r = 0.9)
- La relation entre B et C varie selon l'intervalle des valeurs de A correspondantes. Elle est faible (r = 0.3) pour des valeurs de A inférieures à 1 fibre/cm<sup>3</sup>; elle est bonne (r = 0.6) pour les deux autres intervalles de A ( $1.10 \leq A < 3.0$  et  $3.0 \leq A < 6.0$ ); elle est excellente (r = 0.9) pour les fortes valeurs de A.
- La relation entre A, B et C est bonne (r = 0.6) pour l'ensemble des valeurs de A inférieures à 6 fibres/cm<sup>3</sup>, ce qui correspond à 79% des mesures.

On présente dans l'annexe C les graphiques où sont portées les valeurs d'une série par rapport à une autre dans le premier cas, pour les valeurs de A inférieures à 6 fibres/cm<sup>3</sup> et dans le deuxième cas, pour les valeurs de A supérieures à 6 fibres/cm<sup>3</sup>. Les points superposés indiqués par un chiffre variant de 1 à 9 donnant le nombre de points superposés (le chiffre 9 comprend les nombres de points superposés égaux ou supérieurs à 9). On voit à partir de ces graphiques qu'il est difficile de visualiser la relation entre A et B et entre A et C pour les valeurs de A inférieures à 6 fibres/cm<sup>3</sup>. La relation apparait plus forte entre B et C.

### *Interprétation*

En étudiant les relations entre chacune des séries, il faut tenir compte d'un ensemble de facteurs tels que: 1) la précision des méthodes de mesure, 2) la représentativité de l'échantillonnage et 3) les fluctuations rencontrées dans les mesures. C'est pourquoi, il est préférable d'aborder l'étude des relations entre les séries A, B et C en considérant différents intervalles de mesure plutôt que de façon globale. On a donc considéré séparément quatre intervalles de variation de A. Les seuils de ces intervalles ont été choisis de façon un peu arbitraire.

On peut interpréter les résultats présentés au tableau 3.6 de la façon suivante:

- Pour les faibles valeurs de A ( $A < 1.0$ ), la relation entre chacune des séries est pratiquement inexistante. L'explication plausible de ce manque de relation proviendrait du manque de précision du RDM pour les faibles valeurs (valeurs moyennes de l'ordre de 0.13 mg/m<sup>3</sup>). Si B et C ne sont pas corrélés entre eux, on ne peut s'attendre à trouver une relation entre A et B et entre A et C. Une autre explication possible de ce manque de relation serait la non-représentativité de l'échantillonnage pour des faibles concentrations dans l'air.

- Pour des valeurs de A plus élevées ( $A < 6.0$ ), la relation entre B et C est assez bonne tandis qu'entre A et B et entre A et C, elle est faible. L'augmentation de la relation entre B et C peut s'expliquer par l'augmentation de la précision du RDM pour des valeurs moyennes de l'ordre de 0.35 et 0.62 mg/m<sup>3</sup>
- Pour les fortes valeurs ( $A > 6.0$ ), la relation entre chacune des séries est très forte. Cependant, l'interprétation de ces coefficients de corrélation élevés doit être faite avec prudence à cause de l'influence des fortes valeurs qui influencent en grande partie les calculs.

En résumé, l'étude des coefficients de corrélation confirme les observations notées lors de l'analyse globale:

- la relation entre B et C est excellente sauf pour les faibles valeurs.
- la relation entre A et B et entre A et C est faible, sauf pour les valeurs élevées.

### 3.4 Complément à l'étude des relations entre les séries

#### 3.4.1 Test des observations pairées

Dans cette section, on effectue, au moyen d'un test, une comparaison de la moyenne obtenue pour B (1.20 mg/m<sup>3</sup>) et pour C (1.24 mg/m<sup>3</sup>). On a observé dans les sections 3.2 et 3.3 la présence d'une forte relation entre B et C. L'application de ce test d'égalité des moyennes, tout en permettant de confirmer ces observations notées aux sections précédentes, comporte l'intérêt de pouvoir conclure qu'une seule mesure de B ou C est représentative d'un intervalle de temps de 90 minutes.

Puisque cette section n'est pas aussi essentielle à l'étude, on ne décrira pas ici le test des observations pairées; on trouvera dans Freund (1960) une excellente description de cette technique.

Dans ce test, on considère uniquement la différence entre la mesure pour B et la mesure pour C. Ce test permet de vérifier si la moyenne des différences (-0.035) est significativement différente de zéro. L'écart-type des différences est égale à 0.99. Pour effectuer ce test, on calcule une valeur de t:

$$t = \frac{-0.035}{0.99/\sqrt{900}} = -1.05$$

Puisque la variable t suit une loi de Student à (n - 1) degrés de liberté (n = 900 étant le nombre d'observations), on peut comparer cette valeur de t avec une valeur théorique pour une probabilité au dépassement de 5% (t = 2.58). Puisque la valeur de t calculée est inférieure en valeur absolue à la valeur théorique, on dit que la moyenne des différences observées (-0.035) n'est pas significativement différente de zéro.

On peut donc conclure qu'une seule mesure au RDM est suffisante en pratique pour une durée d'échantillonnage de 90 minutes.

#### 3.4.2 Linéarité et dispersion des relations entre A et B et entre A et C.

L'étude des relations entre chacune des séries a montré l'existence d'une certaine relation entre A et B et entre A et C. Il est important de déterminer la nature de ces relations et d'essayer de les

caractériser. C'est pourquoi, on a procédé à une vérification de la linéarité des relations observées. On a séparé l'intervalle de variation des 900 valeurs de A en 19 classes comprenant chacune entre 40 et 60 valeurs. On a calculé ensuite la moyenne et l'écart-type des valeurs de A pour chacune des classes, ainsi que la moyenne et l'écart-type des valeurs de B et C correspondantes. Ces résultats sont donnés au tableau 3.7. On a porté en graphique sur les figures 3.5 et 3.6 les données relatives aux 17 premières classes. On remarque les points suivants:

- la relation est linéaire entre A et B et entre A et C.
- il existe une dispersion très grande des mesures de B et C pour chaque classe de A. L'écart-type est, dans bien des cas, plus grand que la moyenne.

En procédant à une division des 900 mesures en un certain nombre de groupes, on arrive à démontrer que la relation qui existe entre la méthode optique et la méthode physique est linéaire. Cependant, la grande dispersion rencontrée dans les mesures rend cette relation moins intéressante. En effet, il apparaît presque impossible de prédire avec précision les valeurs individuelles de B et de C à partir des valeurs de A. Par exemple, si on examine la figure 3.6, on voit qu'à partir d'une mesure optique de 2 fibres/cm<sup>3</sup> on retrouve des mesures au RDM variant entre 0.2 et 0.6 mg/m<sup>3</sup>. Inversement, on pourrait dire qu'une mesure au RDM de 0.3 mg/m<sup>3</sup> peut se traduire en une valeur optique variant de 0.5 à 10 fibre/cm<sup>3</sup>. Il est par conséquent, presque impossible de prédire avec précision la mesure optique à partir des mesures au RDM. Cette conclusion ne s'applique, bien entendu, que dans le contexte où l'on a étudié 900 postes de contrôle dont la disparité est très grande.



Tableau 3.7: Nombre, moyenne et écart-type des valeurs de A groupées en 19 classes équiprobables et moyennes et écart-types des valeurs de B et C correspondantes.

Classe	intervalle	nom- bre	A		B		C	
			moyenne	écart- type	moyenne	écart- type	moyenne	écart- type
1	$A \leq .15$	60	.083	.031	.058	.093	.050	.062
2	$.15 < A \leq .22$	41	.200	.002	.094	.127	.151	.354
3	$.22 < A \leq .45$	45	.344	.050	.135	.143	.234	.675
4	$.45 < A \leq .75$	45	.602	.078	.148	.141	.160	.097
5	$.75 < A \leq 1.05$	43	.907	.077	.238	.283	.198	.187
6	$1.05 < A \leq 1.35$	44	1.22	.079	.308	.262	.297	.182
7	$1.35 < A \leq 1.65$	44	1.50	.075	.271	.175	.305	.207
8	$1.65 < A \leq 1.95$	52	1.82	.081	.293	.164	.307	.152
9	$1.95 < A \leq 2.25$	43	2.09	.088	.395	.196	.427	.229
10	$2.25 < A \leq 2.55$	52	2.41	.075	.398	.215	.386	.219
11	$2.55 < A \leq 2.95$	46	2.73	.105	.457	.230	.452	.244
12	$2.95 < A \leq 3.25$	49	3.07	.081	.494	.223	.622	.353
13	$3.25 < A \leq 3.85$	48	3.52	.187	.619	.398	.622	.312
14	$3.85 < A \leq 4.55$	47	4.20	.194	.533	.236	.663	.494
15	$4.55 < A \leq 5.75$	48	5.15	.342	.825	.560	.930	.598
16	$5.75 < A \leq 7.75$	48	6.63	.500	.921	.485	1.03	.567
17	$7.75 < A \leq 13.8$	49	10.1	1.81	1.35	1.15	1.28	1.07
18	$13.8 \leq A \leq 43.0$	47	25.4	8.79	1.99	1.44	2.03	1.35
19	$43.0 < A \leq 1252.$	49	218.3	265.0	12.8	17.4	12.9	16.8

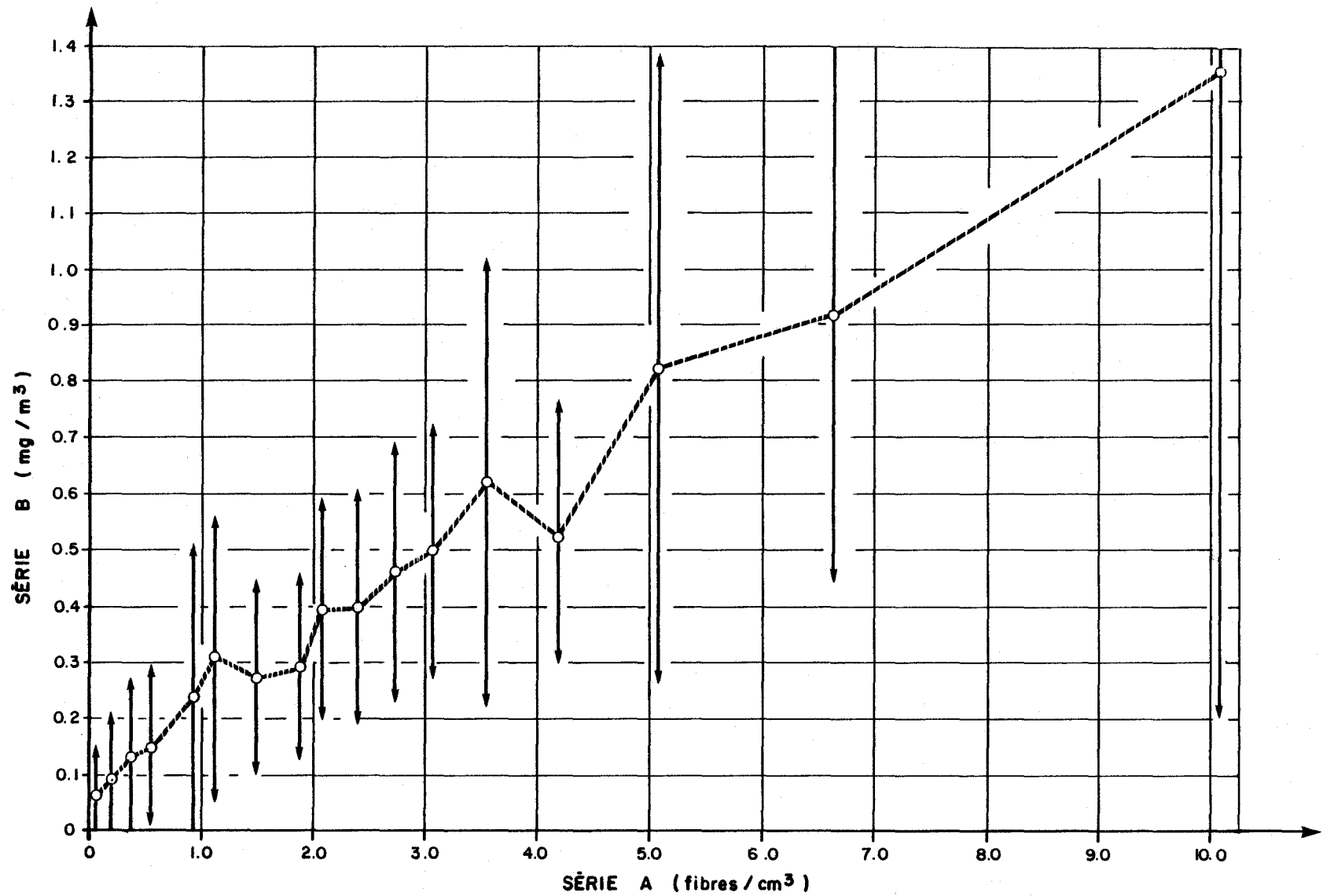


Figure 3.5 . Représentation graphique montrant la linéarité et la dispersion de la relation de la série B avec la série A .

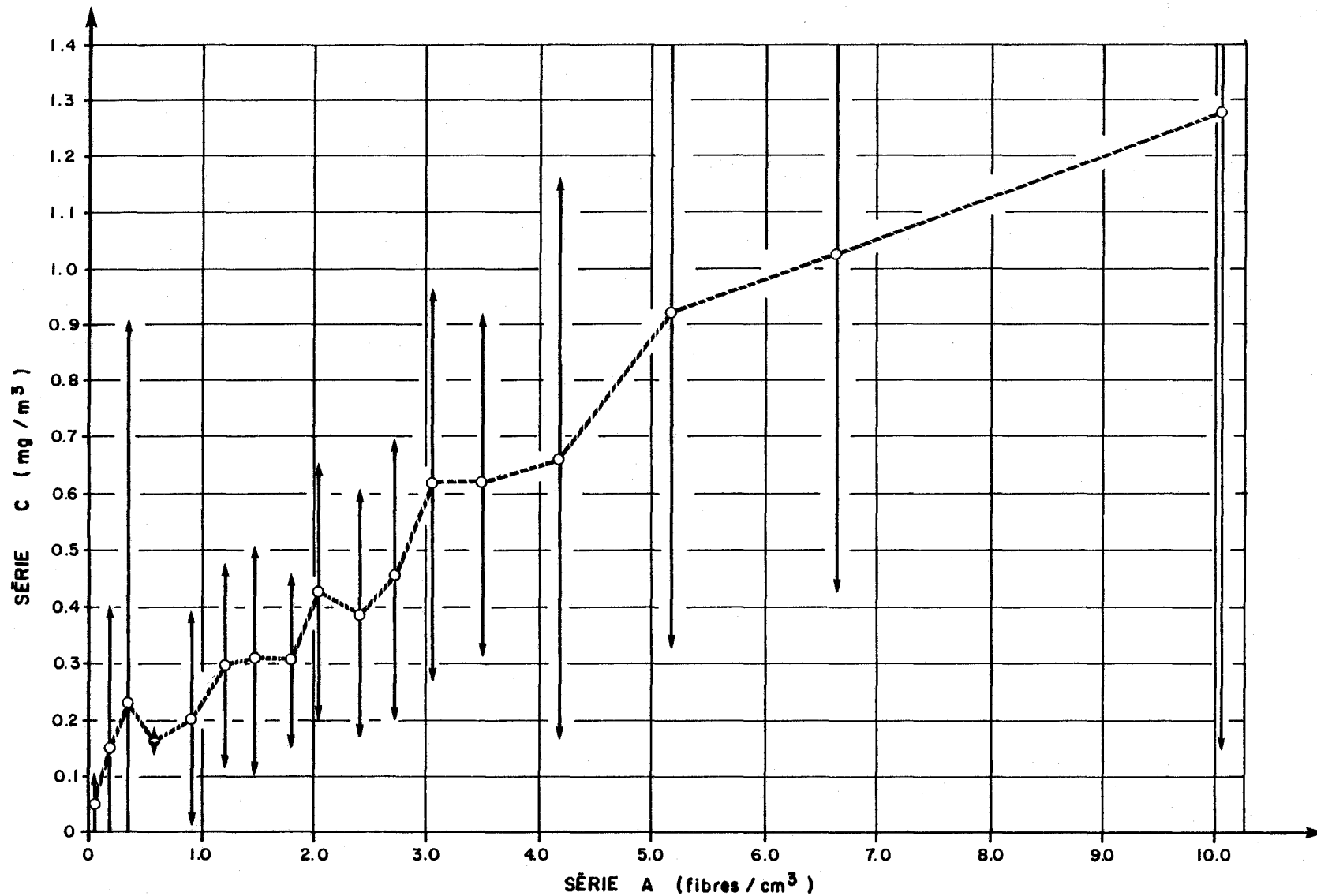


Figure 3.6 . Representation graphique montrant la linéarité et la dispersion de la relation de la série C avec la série A .

## Références bibliographiques

- BOBEE, B., BOUCHER, P., CLUIS, D., LACHANCE, M. et TESSIER A. (1978).  
Méthodologie d'évaluation de la qualité de l'eau pour différents usages. Appendice: programme d'analyse factorielle des correspondances. INRS-Eau, rapport scientifique No 89, 69 p.
- FREUND, J.E. (1960).  
Modern elementary statistics. Prentice-Hall, Inc., Englewoods Cliffs, N.J., U.S.A., second edition, 413 p.

ANNEXE A

DONNEES

NO	A	B	C	NO	A	B	C	NO	A	B	C
1	5.20	.86	.86	51	1.10	.31	.27	101	139.90	11.80	11.80
2	1.70	.20	.34	52	1.40	.32	.27	102	145.30	17.00	17.00
3	2.40	.18	.16	53	2.00	.32	.32	103	69.90	6.30	6.30
4	2.20	.17	.12	54	2.20	.31	.30	104	14.70	1.95	1.95
5	.50	0	.10	55	2.30	.29	.20	105	9.50	1.04	.41
6	.60	.18	.15	56	3.20	.28	.26	106	6.70	.96	.96
7	.90	.17	.06	57	1.70	.27	.36	107	8.00	1.02	.97
8	.80	.14	.15	58	1.90	.22	.27	108	7.70	1.29	1.20
9	1.20	.30	.16	59	1.80	.23	.32	109	8.80	1.43	1.23
10	4.10	.36	.34	60	2.30	.47	.33	110	.70	.21	.21
11	4.30	.59	.66	61	1.30	.26	.33	111	5.70	1.04	.62
12	2.00	.40	.40	62	2.20	.32	.45	112	7.50	1.54	.85
13	3.90	.46	.59	63	2.20	.47	.65	113	4.60	.84	.54
14	4.20	.33	.38	64	4.90	.18	.15	114	4.50	.88	.73
15	5.10	1.29	1.57	65	1.50	.21	.22	115	4.80	.97	.43
16	8.70	1.77	3.53	66	2.10	.38	.42	116	2.50	.56	.47
17	4.10	.53	.35	67	2.00	.47	.66	117	3.90	.78	.45
18	5.10	.49	.39	68	2.80	.41	.68	118	33.80	.96	.99
19	3.20	.45	.52	69	1.60	.24	.24	119	3.90	1.15	2.15
20	4.30	.64	.88	70	2.60	.82	.87	120	6.80	.53	1.57
21	6.10	1.02	1.60	71	1.50	.56	.53	121	7.00	.54	.87
22	2.20	.23	.19	72	1.60	.39	.54	122	8.60	.51	.62
23	2.70	.28	.16	73	.70	.34	.32	123	3.00	.71	.85
24	1.70	.29	.21	74	1.70	.37	.36	124	4.40	1.24	1.66
25	1.00	0	.16	75	4.30	.66	1.14	125	1.70	.43	.21
26	5.30	.49	.55	76	3.00	.29	.40	126	1.30	.27	.11
27	5.70	.81	.87	77	.80	.32	.24	127	2.50	.54	.23
28	7.50	.89	1.21	78	1.40	.19	.20	128	2.30	.44	.38
29	3.10	.43	.58	79	1.80	.31	.34	129	4.60	.59	.76
30	4.80	.54	.90	80	1.50	.25	.30	130	1.30	.36	.41
31	5.50	.79	2.00	81	3.70	.35	.35	131	4.80	1.39	.76
32	5.30	.55	1.06	82	3.00	.35	.26	132	40.00	2.82	2.63
33	4.30	.50	.51	83	1.90	.44	.25	133	2.50	.36	.22
34	2.70	.29	.36	84	1.70	0	0	134	2.40	.36	.18
35	1.90	.32	.48	85	2.40	0	.15	135	2.50	.29	.30
36	1.90	.31	.40	86	2.70	0	0	136	2.80	.34	.36
37	3.70	.52	.63	87	.02	.03	.04	137	2.20	.29	.27
38	6.50	.58	.57	88	.08	.02	.02	138	2.20	.59	.74
39	5.20	.58	.62	89	.12	0	0	139	2.80	.61	.72
40	6.90	.91	1.01	90	.04	.09	.09	140	3.10	.67	.66
41	3.90	.35	.33	91	1.20	.08	.10	141	3.00	.40	.36
42	4.50	.63	.41	92	.03	0	.02	142	3.00	.33	.39
43	3.00	.67	.59	93	25.20	1.92	1.92	143	4.40	.38	.36
44	3.30	.36	.32	94	32.50	2.03	2.03	144	4.30	.50	.37
45	3.50	.50	.52	95	40.70	1.63	1.63	145	2.90	.98	.48
46	4.20	.63	.51	96	51.80	2.04	2.04	146	5.40	.82	.66
47	1.50	.18	.12	97	20.80	1.47	1.47	147	23.40	.73	1.83
48	1.40	.10	.16	98	60.60	.20	0	148	4.40	.56	.29
49	2.20	.50	.32	99	60.60	2.94	2.94	149	12.60	.34	.64
50	2.50	.34	.36	100	173.00	14.95	14.95	150	44.90	.83	.83

NO	A	B	C	NO	A	B	C	NO	A	B	C
151	2.40	.28	.28	201	17.60	2.74	2.80	251	.10	.03	.04
152	81.40	2.98	2.98	202	16.50	2.55	2.21	252	.03	.03	.02
153	71.30	3.35	3.51	203	4.70	1.43	1.09	253	.20	.13	.07
154	1251.50	15.40	15.40	204	4.10	.55	.52	254	.10	.19	.12
155	251.70	20.00	20.00	205	2.20	.35	.32	255	.30	0	.01
156	465.20	19.40	19.40	206	3.30	.43	.48	256	.06	.09	.01
157	.10	.23	.13	207	2.40	.41	.38	257	.60	0	.14
158	.20	.14	.13	208	2.30	.76	.79	258	.40	.09	.05
159	.20	0	0	209	1.00	.29	.21	259	.05	.05	.02
160	.10	.04	.01	210	2.30	.43	.90	260	3.40	.77	.82
161	.02	.02	.02	211	1.80	.35	.51	261	3.30	.70	.62
162	.08	0	0	212	4.50	.48	.74	262	2.80	.79	.61
163	.10	0	0	213	2.90	.52	.71	263	3.30	.67	.55
164	.10	.05	.07	214	7.60	1.30	1.57	264	3.10	.41	.41
165	1.00	.09	.20	215	6.70	1.48	1.65	265	2.00	.43	.46
166	.80	.11	0	216	2.50	1.08	.80	266	1.80	.30	.19
167	.80	.14	0	217	5.40	1.56	1.41	267	2.50	.30	.20
168	1.20	.09	.14	218	5.30	1.75	1.59	268	3.10	.46	.86
169	.10	0	.19	219	53.60	4.16	4.16	269	1.50	.06	.19
170	.20	.27	.81	220	37.80	2.63	2.63	270	1.20	.35	.41
171	.20	.19	1.71	221	403.80	15.50	15.50	271	2.10	.35	.39
172	.10	.15	.17	222	368.30	18.60	18.60	272	2.80	.51	.38
173	.10	.13	.10	223	500.00	39.90	39.90	273	2.60	.32	.33
174	.30	.20	.23	224	49.20	2.84	2.84	274	2.20	.31	.36
175	.20	.23	.11	225	.70	.30	.10	275	1.10	.27	.54
176	.20	.49	.07	226	134.90	5.50	5.50	276	3.20	.33	.23
177	.20	.42	.17	227	4.00	.58	.42	277	2.00	.33	.25
178	.30	.42	.13	228	4.90	.42	.31	278	2.10	.42	.29
179	.20	.29	.10	229	.20	.04	0	279	2.90	.28	.47
180	.20	.17	.11	230	.50	.13	.15	280	3.10	.41	.45
181	.30	.04	.12	231	.60	.01	.18	281	2.00	.33	.66
182	.20	.02	.02	232	.70	.13	.30	282	1.80	.28	.43
183	.20	.03	.01	233	.10	.12	0	283	3.00	.44	.59
184	.10	.07	.10	234	.10	.07	.09	284	2.20	.37	.37
185	.10	.02	.16	235	.10	.01	.06	285	4.20	.29	.27
186	.30	.08	.37	236	.10	.08	.07	286	3.10	.37	.30
187	.20	0	0	237	.30	.04	.03	287	1.50	.22	.10
188	8.80	2.53	1.58	238	.50	.38	.34	288	1.80	.14	.04
189	3.90	.55	.60	239	.30	.28	.24	289	4.10	.51	.91
190	3.30	.42	.85	240	.30	.27	.26	290	2.00	.48	.61
191	3.80	1.84	1.45	241	.30	.19	.13	291	3.60	.41	.41
192	3.90	.33	.33	242	.50	.06	.32	292	3.60	.38	.43
193	6.50	.47	1.24	243	.40	.19	0	293	13.10	1.46	.64
194	6.20	.92	1.34	244	1.30	.15	.09	294	5.60	.60	1.76
195	9.90	1.09	1.43	245	.10	.02	.08	295	6.10	.62	1.64
196	.40	.30	.11	246	.10	.10	.06	296	3.00	.17	.80
197	.90	.13	.28	247	.10	.03	.36	297	4.00	.22	1.97
198	2.10	.72	.79	248	.03	.08	.02	298	6.80	1.09	.90
199	6.60	2.39	2.46	249	.10	0	.04	299	5.30	1.36	1.74
200	10.90	3.46	3.78	250	.03	.08	0	300	6.50	1.52	2.09

NO	A	B	C	NO	A	B	C	NO	A	B	C
301	8.40	2.21	1.73	351	5.70	.14	.14	401	1.70	.18	.17
302	10.60	1.98	1.80	352	5.70	.19	.19	402	1.50	.13	.12
303	12.20	2.87	2.78	353	1.10	.03	.13	403	2.00	.18	.16
304	9.00	1.24	1.78	354	.90	.08	.18	404	1.50	.11	.19
305	7.60	1.65	1.49	355	.50	0	.02	405	6.60	.52	.36
306	5.70	1.52	2.18	356	.20	.01	.03	406	1.60	.73	.43
307	1.50	.18	.29	357	.10	0	.01	407	3.70	.65	.54
308	2.50	.18	.22	358	.20	.03	.02	408	3.00	.58	.43
309	1.40	.19	.22	359	.02	0	0	409	3.30	.29	.37
310	2.60	.74	1.35	360	.20	.01	.01	410	3.00	.25	.31
311	3.60	1.19	.85	361	.04	0	.01	411	3.80	.28	.33
312	2.60	.35	.32	362	.60	.01	.02	412	3.70	.48	.59
313	7.60	.89	1.12	363	1.50	0	.01	413	3.00	.40	.42
314	4.00	.70	.48	364	.20	.01	0	414	6.10	.56	.64
315	7.10	1.38	1.08	365	.90	0	.02	415	6.30	.54	.46
316	2.10	.91	.86	366	2.00	.25	.25	416	2.40	.19	.37
317	9.20	1.77	2.17	367	31.30	.95	.95	417	2.60	.26	.21
318	6.10	.57	.67	368	18.70	.38	2.42	418	1.80	.27	.22
319	3.30	.31	.51	369	492.70	28.60	28.60	419	8.00	.50	.78
320	3.20	.57	.82	370	102.10	9.30	9.30	420	2.20	.07	.15
321	1.40	.28	.03	371	1.60	.22	.91	421	1.30	.16	.14
322	.80	.23	.16	372	1.50	.72	.54	422	6.10	.80	.79
323	.10	.17	.10	373	.60	.12	.01	423	5.10	.57	.44
324	3.80	.75	.96	374	1.10	.13	.13	424	2.70	.48	.51
325	51.70	2.27	2.71	375	.30	.08	.08	425	4.10	.64	.56
326	83.20	2.82	3.42	376	13.70	1.01	1.01	426	7.00	.87	.79
327	4.50	.84	.92	377	9.30	1.17	1.17	427	5.30	.62	.77
328	1.70	.35	.40	378	1183.80	99.40	99.40	428	4.80	.66	.69
329	.20	.02	.07	379	.60	0	.05	429	6.80	.98	.92
330	1.60	.12	.28	380	2.00	.30	.30	430	3.70	.67	.69
331	1.60	.18	.25	381	1.70	.19	.17	431	5.00	.87	1.14
332	1.60	.70	.54	382	16.80	3.26	3.26	432	10.20	1.04	1.13
333	2.40	.57	.95	383	5.00	1.00	1.00	433	3.00	.65	.61
334	2.00	.36	.15	384	.90	.26	.26	434	4.20	.61	.56
335	3.60	.62	.37	385	.60	.15	.15	435	2.70	.43	.32
336	2.80	.49	.39	386	.80	.22	.22	436	5.60	.88	.91
337	1.80	.11	.60	387	1.90	.22	.22	437	2.40	.26	.25
338	1.70	.22	.45	388	3.30	.29	.29	438	2.40	.22	.38
339	.60	.04	0	389	17.00	1.01	1.01	439	4.20	.66	.60
340	1.50	.26	.24	390	.60	.31	.31	440	1.50	.22	.21
341	2.50	.32	.10	391	7.00	.99	.99	441	5.20	.69	.61
342	2.70	.44	.70	392	34.60	2.48	2.48	442	4.60	.62	.68
343	1.70	.17	.14	393	42.80	2.29	2.29	443	11.20	.48	.42
344	3.60	.39	.74	394	15.50	1.29	1.29	444	6.10	.63	.64
345	1.90	.21	.40	395	36.00	2.58	2.58	445	3.90	.70	.80
346	1.10	.36	.35	396	17.80	2.12	2.12	446	4.30	.70	.75
347	1.90	.37	.34	397	.40	.04	.08	447	3.80	.44	.43
348	.80	.13	.03	398	1.90	.23	.19	448	3.80	.55	.38
349	1.60	.13	.36	399	5.00	0	0	449	2.50	.35	.36
350	1.10	.09	.36	400	3.20	.63	.65	450	2.40	.29	.37



NO	A	B	C	NO	A	B	C	NO	A	B	C
451	2.80	.31	.32	501	1.30	.45	.23	551	.20	.04	.01
452	5.90	.75	.72	502	1.60	.25	.30	552	.10	.02	.01
453	6.40	.37	.47	503	1.90	.27	.29	553	.40	.04	0
454	1.90	.13	.12	504	2.80	.23	.28	554	.30	.16	.19
455	1.20	.24	.11	505	2.40	.25	.11	555	.20	.02	.02
456	2.80	.18	.16	506	2.50	.26	.15	556	.30	.09	.13
457	1.90	.22	.15	507	1.90	.23	.15	557	.40	0	.04
458	1.90	.16	.24	508	.80	.18	.18	558	.10	.11	.04
459	2.80	.26	.33	509	1.20	.31	.27	559	.10	.06	.10
460	2.80	.22	.24	510	1.50	.34	.29	560	.10	.04	.04
461	3.00	.34	.36	511	.50	.03	.05	561	.30	.03	.10
462	2.20	.32	.33	512	.50	.08	.13	562	2.30	.32	.17
463	2.60	.28	.23	513	1.00	.07	.08	563	.90	.09	.13
464	2.50	.31	.43	514	20.60	7.34	7.34	564	2.50	.13	.31
465	3.00	.41	.50	515	8.70	1.93	1.93	565	2.20	.15	.50
466	4.10	.41	.50	516	13.70	3.86	3.86	566	9.00	.38	.90
467	4.30	.44	.35	517	2.60	.58	.58	567	4.00	.14	.42
468	2.70	.26	.15	518	6.00	.79	.79	568	2.80	.67	.52
469	2.80	.29	.14	519	6.30	1.10	1.10	569	4.10	.33	.49
470	1.40	.31	.33	520	8.20	1.46	1.46	570	4.50	.18	.46
471	1.30	.22	.23	521	13.00	4.55	4.55	571	6.50	.29	.13
472	2.60	.24	.23	522	8.70	1.20	1.20	572	3.10	.62	.82
473	2.40	.31	.13	523	18.50	2.41	2.41	573	4.20	.46	.65
474	2.90	.26	.19	524	12.50	1.51	1.51	574	3.60	.57	.63
475	2.70	.20	.33	525	14.40	2.04	2.04	575	2.30	.16	.30
476	1.80	.25	.19	526	8.20	1.04	1.04	576	4.20	.05	.23
477	2.00	.33	.19	527	16.30	1.87	1.87	577	7.80	.81	.36
478	1.80	.24	.22	528	1.00	.13	.13	578	8.30	.33	.19
479	1.80	.23	.19	529	38.40	1.35	1.35	579	8.00	.59	.68
480	1.40	.14	.11	530	1.40	.33	.85	580	3.40	.09	.17
481	2.00	.11	.13	531	1.60	.16	.16	581	3.40	.28	.69
482	1.90	.24	.20	532	.90	.13	.13	582	5.90	.40	.28
483	1.50	.24	.21	533	29.70	1.85	1.85	583	5.40	.70	.69
484	1.80	.30	.24	534	15.80	1.07	1.07	584	4.30	.48	.54
485	2.90	.41	.44	535	64.50	3.60	3.60	585	6.50	.34	.29
486	3.20	.51	.44	536	349.50	12.50	12.50	586	.07	.12	.04
487	1.40	.55	.44	537	127.60	10.40	10.40	587	3.90	.34	.51
488	2.00	.33	.48	538	612.40	16.40	16.40	588	3.70	.23	.53
489	2.50	.63	.58	539	37.90	1.90	1.90	589	1.50	.36	.74
490	2.90	.27	.45	540	52.70	1.74	1.74	590	5.40	.93	.93
491	3.00	.37	.54	541	28.00	1.25	1.25	591	18.80	2.54	1.32
492	2.80	.27	.34	542	18.30	1.21	1.21	592	46.80	5.27	6.07
493	3.40	.31	.42	543	6.90	1.02	1.02	593	366.10	14.00	14.00
494	2.50	.26	.25	544	.30	.23	.16	594	11.10	1.12	.74
495	3.30	.61	.59	545	.20	.02	.04	595	7.90	.74	.36
496	8.80	0	0	546	.20	.36	.85	596	49.40	5.59	5.59
497	3.10	.64	1.40	547	.30	.79	.93	597	24.10	2.41	2.41
498	2.70	.20	.27	548	.60	.17	.19	598	41.70	4.54	4.54
499	2.90	.38	.32	549	.80	1.64	1.08	599	1.30	.34	.34
500	1.90	.42	.26	550	1.20	1.48	.97	600	50.60	1.32	1.32

NO	A	B	C	NO	A	B	C	NO	A	B	C
601	.30	.05	.03	651	49.30	6.26	.34	701	.90	.17	.11
602	.20	.03	.01	652	.50	.08	.17	702	.90	.18	.11
603	.30	.07	.04	653	.90	1.09	.29	703	.90	.27	.18
604	.40	.13	.01	654	1.20	.04	.61	704	.90	.16	.18
605	.50	.07	.17	655	1.80	1.14	.37	705	11.80	2.83	1.05
606	.30	.22	.20	656	.60	.25	.11	706	.40	.31	.37
607	.10	.09	.09	657	17.90	2.04	2.04	707	.60	.15	.05
608	.20	.03	.04	658	6.50	.40	1.99	708	.70	.75	.31
609	.40	.14	.01	659	21.70	.37	.87	709	.40	.10	.09
610	.70	.13	.13	660	.30	.17	.19	710	.60	.25	.36
611	.70	.29	.15	661	11.90	1.90	.37	711	.40	.18	.22
612	.10	.02	0	662	11.50	.16	3.30	712	11.10	1.76	.85
613	.20	.03	0	663	8.40	5.24	.14	713	18.10	.40	1.30
614	1.00	.08	.15	664	13.00	1.95	1.95	714	5.90	.54	.89
615	.30	.03	0	665	20.10	.33	1.08	715	8.10	.50	.31
616	1.30	.18	.13	666	44.00	9.20	1.70	716	1.90	.04	.37
617	.10	.09	.03	667	69.00	.21	6.26	717	2.70	.99	.81
618	.10	.10	.07	668	44.60	0	0	718	5.50	1.03	.97
619	.20	.07	0	669	1.40	.11	.56	719	3.10	.74	.81
620	.10	.01	0	670	3.70	1.53	1.28	720	2.00	.74	.54
621	.02	0	0	671	.40	.10	.28	721	3.30	.92	.62
622	.20	.08	.12	672	.80	.33	.22	722	3.80	1.96	1.75
623	.20	0	.12	673	1.30	.07	.63	723	1.80	.53	.48
624	.40	.01	.05	674	9.90	0	.08	724	2.50	.30	.34
625	.20	0	.03	675	4.70	1.01	1.76	725	1.90	.42	.75
626	.10	0	.02	676	6.90	.59	2.58	726	1.10	.23	.27
627	.40	0	.19	677	.40	.11	4.54	727	.70	.25	.26
628	.03	0	.02	678	11.00	.06	2.34	728	.90	.32	.16
629	.02	0	.01	679	5.00	.17	.94	729	.70	.29	.25
630	.10	0	.05	680	4.30	.36	.45	730	.50	.31	.27
631	.10	0	0	681	9.90	.40	.42	731	.40	.31	.26
632	.10	.03	.04	682	2.60	.82	.47	732	1.70	.11	0
633	.40	.03	.01	683	2.30	.13	.09	733	3.10	.29	.17
634	.40	0	.08	684	5.30	3.45	1.81	734	5.10	.25	.03
635	1.00	.54	.56	685	4.30	.50	2.53	735	5.10	.86	.91
636	.60	.04	.18	686	1.00	.13	.19	736	6.70	.52	.55
637	.60	.08	.13	687	5.20	.98	.31	737	2.40	.54	.42
638	2.30	.21	.43	688	.60	.03	.21	738	3.60	.61	.89
639	1.30	.52	.54	689	6.40	2.12	.35	739	1.70	.56	.46
640	1.30	.20	.37	690	1.30	1.11	.14	740	9.90	.55	.50
641	.50	.10	.14	691	6.00	1.08	1.18	741	6.60	.94	.19
642	.40	.09	.09	692	5.10	1.08	2.81	742	4.50	.78	.20
643	2.70	.67	.45	693	1.30	.56	.30	743	19.70	1.84	1.20
644	7.00	.26	.31	694	1.30	.38	.38	744	3.00	.09	.09
645	1.40	.13	.14	695	1.20	.36	.35	745	1.60	.32	.17
646	23.40	1.29	.28	696	2.10	.65	.69	746	1.20	.24	.20
647	4.70	.03	1.66	697	2.40	.85	.66	747	1.80	.17	.15
648	21.60	.13	.96	698	3.20	.08	.63	748	3.40	.25	.45
649	25.30	.77	.45	699	1.20	.53	.32	749	13.50	.38	.42
650	30.10	.69	.56	700	1.40	.44	.34	750	8.40	.24	.19

NO	A	B	C	NO	A	B	C	NO	A	B	C
751	2.10	.36	.23	801	1.00	0	.18	851	.10	.04	.05
752	.70	0	.09	802	1.10	.04	.03	852	.20	0	0
753	1.60	.03	0	803	1.50	.21	.28	853	1.20	.31	.17
754	.20	.01	.10	804	.80	.33	.44	854	2.40	.83	.74
755	.90	.13	.02	805	1.00	.39	.48	855	1.00	.34	.48
756	3.30	.29	.17	806	1.00	.02	.14	856	.20	.06	.06
757	.70	.01	.11	807	2.00	.53	.69	857	.90	.11	.11
758	.60	.19	.02	808	2.40	.51	.43	858	4.10	.90	.76
759	.70	.09	.03	809	2.40	.45	.60	859	7.10	2.05	1.16
760	.20	0	0	810	2.60	.27	.63	860	2.60	.24	.18
761	1.50	.23	.12	811	2.70	.74	1.02	861	3.00	.90	.74
762	4.30	.68	.90	812	2.10	.88	1.01	862	2.30	.51	.40
763	2.10	.81	.81	813	1.90	.43	.53	863	1.00	.22	.12
764	505.10	68.20	57.90	814	3.00	1.11	1.00	864	1.10	.30	.22
765	71.60	4.54	7.10	815	1.30	.25	.33	865	1.20	.50	.20
766	51.50	1.48	17.50	816	.90	.33	.26	866	3.20	.56	.49
767	431.40	27.00	29.50	817	2.50	.69	.73	867	4.60	.43	.66
768	4.50	.23	.09	818	3.10	.97	1.11	868	2.00	.35	.26
769	20.10	1.64	1.43	819	3.60	.91	1.07	869	3.40	.66	.47
770	.70	.02	.12	820	2.70	.80	.70	870	1.90	.23	.29
771	.30	.10	.09	821	2.80	.67	.77	871	3.40	.38	.54
772	.20	.03	.09	822	2.10	.44	.46	872	2.40	.49	.33
773	.20	.27	1.22	823	.30	0	0	873	2.40	.28	.30
774	.30	.16	.12	824	.19	.01	0	874	.90	.17	0
775	.40	0	.17	825	.21	0	0	875	3.00	.31	1.04
776	1.30	.07	.14	826	.12	.04	0	876	3.00	.70	.91
777	.20	0	0	827	.11	.01	.02	877	3.50	1.50	.72
778	.20	.09	.01	828	.04	0	0	878	5.60	.94	1.08
779	.30	.17	.12	829	1.50	.53	.50	879	1.20	.36	.27
780	.50	.18	.11	830	1.90	.31	.21	880	2.70	.55	.54
781	1.90	.25	.46	831	1.20	.15	.27	881	3.60	.80	.53
782	1.50	.35	.44	832	1.40	.19	.20	882	2.40	.74	.64
783	1.80	.35	.41	833	3.50	.40	.28	883	2.60	.87	.37
784	1.30	.23	.61	834	100.30	16.87	6.87	884	3.30	.53	.54
785	2.60	.61	.58	835	15.60	.90	2.41	885	2.50	.39	.54
786	2.30	.68	.57	836	9.40	.91	1.12	886	3.80	.63	.48
787	10.40	.64	1.49	837	2.60	.48	.45	887	2.40	.43	.51
788	3.20	.65	1.99	838	1.00	.25	.13	888	6.10	.93	.84
789	3.70	.68	.65	839	198.70	23.70	24.10	889	3.30	.48	.82
790	3.30	.85	1.07	840	2.40	.04	0	890	3.00	.71	.71
791	1.80	.47	.56	841	145.80	8.59	5.85	891	1.10	.16	.31
792	.70	.15	.17	842	102.00	3.62	3.62	892	2.00	.26	.72
793	4.70	.65	.71	843	2.00	.13	.13	893	2.60	.40	.59
794	.30	.04	0	844	49.20	1.32	4.11	894	2.60	.43	.65
795	1.90	.33	.44	845	34.30	1.99	1.99	895	2.70	.51	.32
796	6.70	1.30	1.74	846	.10	.06	.11	896	140.20	12.60	22.17
797	2.50	.51	.57	847	.10	.64	0	897	157.90	14.90	14.60
798	.50	.13	.16	848	.10	0	.02	898	36.10	5.80	2.93
799	1.30	.47	.45	849	.20	.20	.04	899	524.20	14.22	12.27
800	1.90	.35	.35	850	.10	.02	0	900	32.60	5.63	7.00

ANNEXE B

METHODE D'ANALYSE  
FACTORIELLE DES  
CORRESPONDANCES

## ANNEXE B:

La description de grands tableaux de données nécessite l'utilisation de méthodes très générales. La méthode d'analyse factorielle des correspondances décrite ci-dessus, permet la description de ces grands tableaux de données, faisant ressortir suivant la constitution du tableau initial, les similitudes de comportement du groupes de stations ou de dates d'échantillonnage et les identités de variations des paramètres considérés.

Analyse factorielle des correspondances (A.F.C.)

La méthode d'analyse factorielle des correspondances développée par Benzécri (1973) est un outil mathématique utilisé de plus en plus pour l'analyse de grands tableaux de données. Dans l'analyse, on considère un tableau de nombres positifs comprenant n lignes et p colonnes,  $x_{ij}$  étant un élément du tableau. Ce tableau peut, par exemple correspondre à un ensemble d'observations de différents paramètres chimiques,  $x_{ij}$  représentant alors l'observation i du paramètre j. La description d'un tel tableau en montrant des similitudes de paramètres ou d'observations ne peut se faire sans la définition d'une distance entre observations ou paramètres permettant d'observer ces similitudes. Puisque les paramètres mesurés sont souvent d'ordres de grandeur différents, on est amené à faire subir une transformation au tableau des données afin de résoudre le problème d'échelle et obtenir une stabilité dans l'analyse. C'est pourquoi, dans l'A.F.C. on calcule:

- la somme de chacune des lignes

$$x_j = \sum_{i=1}^n x_{ij}$$

- la somme de chacune des colonnes  $x_i = \sum_{j=1}^p x_{ij}$

- la somme de tous les éléments du tableau  $x = \sum_{i=1}^n x_i = \sum_{j=1}^p x_j$

Chacun des éléments du tableau est alors transformé:

$$p_{ij} = \frac{x_{ij}}{x}$$

La somme des lignes  $x_i$  et la somme des colonnes  $x_j$  sont transformées et l'on obtient

$$p_i = \frac{x_i}{x} = \frac{x_j}{x}$$

Dans certains ouvrages (Lebart et Fenelon, 1973 par exemple) on différencie les caractéristiques  $p_i$  et  $x_i$  relatives aux lignes  $i$  en les notant ( $p_i$  et  $x_i$ ) des caractéristiques relatives aux colonnes  $j$  en les notant ( $p_j$  et  $x_j$ ). Ici, puisqu'il ne peut y avoir de confusion, on conserve  $p_i$ ,  $x_i$ ,  $p_j$  et  $x_j$ .

On a le choix, pour raisonner, de se placer dans l'espace des paramètres ou dans celui des observations. On se place dans l'espace des paramètres à  $p$  dimensions ( $R^p$ ) dans lequel se retrouvent les  $n$  points observations.

$$\sum_{j=1}^p \frac{p_{ij}}{p_i} = 1 \quad (i = 1, 2, \dots, n)$$

Les distances entre deux points observations  $i$  et  $i'$  sont notées:

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

Si pour une coordonnée (par exemple la troisième), les valeurs absolues des chiffres sont grandes, leur différence peut être grande. Dans la somme des carrés des différences pour obtenir la distance  $d^2$ , le terme correspondant  $(P_{i3} / P_i - P_{i'3} / P_{i'})^2$  sera prépondérant et jouera alors un rôle excessif dans l'analyse.

Pour atténuer ces disparités, on utilise une expression pondérée pour obtenir  $d^2$ :

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

$d^2(i, i')$  peut aussi prendre la forme:

$$d^2(i, i') = \sum_{j=1}^p \left( \frac{P_{ij}}{P_i \sqrt{P_j}} - \frac{P_{i'j}}{P_{i'} \sqrt{P_j}} \right)^2$$

Cette dernière relation définit la distance dite du  $\chi^2$  entre deux points observations  $i$  et  $i'$ . Elle est équivalente à une distance euclidienne entre deux points observations, les coordonnées du point  $i$  étant

$$P_{ij} / P_i \sqrt{P_j} \quad (j=1, 2, \dots, p)$$

La recherche des axes principaux d'un nuage de points consiste à trouver un ensemble de droites orthogonales deux à deux telles que les projections des distances entre les couples de points sur ces droites sont maximums. On peut montrer (Lebart et Fenelon, 1973) que dans l'A.F.C., la recherche des axes principaux du nuage des  $n$  points  $i$  conduit à diagonaliser une matrice  $S=R'R$  après avoir défini  $r_{ij}$ , terme général de la matrice  $R$ , de la façon

suivante:

$$r_{ij} = \frac{P_{ij} - P_i P_j}{\sqrt{P_i P_j}}$$

Ce qui donne:

$$S_{jk} = \sum_{i=1}^n \frac{P_{ij} P_{ik}}{P_i \sqrt{P_j P_k}} \quad \left( \begin{matrix} j \\ k \end{matrix} \right) = 1, 2, \dots, p$$

Si on pose  $u_\alpha$ , le vecteur unitaire d'ordre  $\alpha$  de la matrice  $S$  avec comme composantes  $u_{j\alpha}$  ( $j=1, 2, \dots, p$ ), la projection du point  $i$  sur l'axe factoriel  $\alpha$  est  $F_\alpha(i)$  telle que

$$F_\alpha(i) = \sum_{j=1}^p u_{j\alpha} \frac{P_{ij}}{P_i \sqrt{P_j}}$$

ou encore, en considérant les données initiales  $P_{ij}/P_i$

$$F_\alpha(i) = \sum_{j=1}^p \phi_{j\alpha} \frac{P_{ij}}{P_i}$$

avec  $\phi_{j\alpha} = u_{j\alpha} / \sqrt{P_j}$



Si on avait choisi de se placer de la même manière dans l'espace des observations à  $n$  dimensions ( $\mathbb{R}^n$ ) dans lequel se retrouvent les  $p$  points-paramètres, on aurait été amené à définir les distances entre deux points paramètres  $j$  et  $j'$

$$d^2(j, j') = \sum_{i=1}^n \frac{1}{P_i} \left( \frac{P_{ij}}{P_j} - \frac{P_{ij'}}{P_{j'}} \right)^2 \quad (2)$$

La recherche des axes principaux consiste alors à diagonaliser une matrice  $T = R R'$  qui a les mêmes valeurs propres  $\lambda_\alpha$  que  $S$ ; le vecteur propre unitaire d'ordre  $\alpha$  de  $T$  est  $V_\alpha$  avec comme composantes  $V_{i\alpha}$  ( $i=1, 2, \dots, n$ ). La projection du point  $j$  sur l'axe  $\alpha$  est alors  $G_\alpha(j)$ :

$$G_\alpha(j) = \sum_{i=1}^n \psi_{i\alpha} \frac{P_{ij}}{P_j}$$

$$\text{avec } \psi_{i\alpha} = \frac{V_{i\alpha}}{\sqrt{P_i}}$$

On peut montrer qu'il existe une relation entre les facteurs  $\phi_\alpha$  et  $\psi_\alpha$ :

$$\psi_{i\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j=1}^p \frac{P_{ij}}{P_i} \quad \phi_{j\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} F_\alpha(i)$$

$$\phi_{j\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^n \frac{P_{ij}}{P_j} \quad \psi_{i\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} G_\alpha(i)$$

Ces deux relations montrent les liens existant entre les représentations dans les espaces  $R^p$  et  $R^n$ . On voit qu'il suffit de connaître une seule représentation pour déduire l'autre. En pratique, on diagonalise la plus petite des matrices  $S$  et  $T$ . Ces relations justifient une représentation simultanée des points  $i$  du nuage dans  $R^p$  et des points  $j$  du nuage dans  $R^n$ . En effet, les  $\psi_i$  (ou les  $F_\alpha(i)$ ) sont les centres de gravité des  $\phi_j$  (ou des  $G_\alpha(j)$ ) avec un poids  $P_{ij}/P_i$ . Ainsi, la représentation simultanée montrant une proximité entre un point paramètre et un groupe d'observations montre que ce paramètre est caractéristique du comportement de ce groupe d'observations. Cette particularité de l'A.F.C. permet un niveau d'interprétation intéressant. De plus, il est généralement possible, tenant compte de la position des points paramètres et des points observations par rapport aux axes, de donner un sens physique au groupe d'observations.

La transformation du tableau initial ( $x_{ij}$  devenant  $P_{ij}/P_i$  ou  $P_{ij}/P_j$ ) implique que deux vecteurs lignes ou colonnes de ce tableau, avec des composantes proportionnelles, sont représentées dans le nuage (donc, dans la projection dans le plan des axes principaux) par deux points confondus. L'A.F.C. permet donc de mettre en évidence des similitudes de variations pour les paramètres et de comportement pour les observations en considérant les points dont les coordonnées sont identiques ou du même ordre de grandeur.

#### *contributions*

L'inertie totale du nuage des points  $i$  dans  $R^p$  est définie:

$$\text{inertie totale} = \sum_{i=1}^n P_i \rho^2(i) = \sum_{\alpha} \lambda_{\alpha}$$

dans laquelle  $\rho^2(i)$ , la distance d'un point  $i$  au centre de gravité du nuage est donnée par:

$$\rho^2(i) = \sum_{j=1}^p \left( \frac{P_{ij}}{P_i} - P_j \right)^2 / P_j = \sum_{\alpha} F_{\alpha}^2(i)$$

et dans laquelle  $\lambda_{\alpha}$ , la valeur propre associée à l'axe  $\alpha$  est l'inertie expliquée par l'axe  $\alpha$ .

$$\lambda_{\alpha} = \sum_{i=1}^n F_{\alpha}^2(i) P_i$$

La contribution absolue du point  $i$  à l'axe  $\alpha$  est ainsi  $F_{\alpha}^2(i) P_i$ . De façon parallèle, on a :

$$\lambda_{\alpha} = \sum_{j=1}^p G_{\alpha}^2(j) P_j$$

$G_{\alpha}^2(j) P_j$  représente la contribution absolue du point  $j$  à l'axe  $\alpha$ . En pratique, ces contributions absolues sont exprimées en pourcentages et indiquent le rôle joué par chacun des points du nuage dans l'inertie expliquée par un axe.

La contribution relative du point  $i$  à l'axe  $\alpha$  est donnée par :

$$C_{\alpha}(i) = \frac{F_{\alpha}^2(i)}{\rho^2(i)} \quad \text{avec} \quad \sum_{\alpha} C_{\alpha}(i) = 1$$

De la même façon, on définit la contribution relative du point  $j$  à l'axe  $\alpha$

$$C_{\alpha}(j) = \frac{G_{\alpha}^2(j)}{\rho^2(j)} \quad \text{avec} \quad \sum_{\alpha} C_{\alpha}(j) = 1$$

Ces contributions sont utiles à l'interprétation : par exemple,

- la contribution absolue montre quels paramètres sont responsables d'un facteur  $\alpha$
- la contribution relative indique quel axe factoriel représente le mieux un paramètre.

#### *Reconstitution des données*

Il est intéressant parfois d'examiner comment les données sont reconstituées quand, dans l'analyse, on retient un nombre réduit d'axes factoriels. On obtient

$$P'_{ij} = P_i P_j \left[ 1 + \sum_{\alpha} \left( \frac{1}{\sqrt{\lambda_{\alpha}}} F_{\alpha}(i) G_{\alpha}(j) \right) \right]$$

En pratique on ne cherche pas à reconstituer intégralement les données initiales. On peut éliminer d'un tableau de données ce qu'on peut appeler un bruit de fond attribuable à l'imprécision des mesures ou aux fluctuations d'échantillonnage.

#### *Points supplémentaires.*

On peut être intéressé à connaître la position sur les axes factoriels de points représentatifs d'un paramètre ou d'une observation qui pour une raison ou une autre n'ont pas été inclus dans l'analyse. Les coordonnées de ces

points, de poids nul dans l'analyse, peuvent être calculées de la façon suivante: pour une observation supplémentaire,

$$F'_\alpha(i) = \sum_{j=1}^p \frac{p'_{ij}}{p'_i} G'_\alpha(j)$$

$$\text{avec } p'_{ij} = \frac{x'_{ij}}{x} \quad \text{et } p'_i = \frac{\sum_{j=1}^p x'_{ij}}{x}$$

et pour une colonne supplémentaire,

$$G'_\alpha(j) = \sum_{i=1}^n \frac{p'_{ij}}{p'_j} F'_\alpha(i)$$

$$\text{avec } p'_{ij} = \frac{x'_{ij}}{x} \quad \text{et } p'_j = \frac{\sum_{i=1}^n x'_{ij}}{x}$$

### *Codage des données*

En pratique, il peut être intéressant (Benzecri, 1973) de transformer les données initiales en données binaires. L'intervalle d'un paramètre  $j$  peut être divisé en  $k$  classes d'effectif approximativement égal. Une observation appartenant seulement à une des classes est codée 1 pour cette classe et 0 pour les autres classes. On obtient ainsi un tableau de valeurs binaires composé de  $n$  lignes et  $(p \cdot k)$  colonnes (si chaque paramètre est séparé en  $k$  classes). L'analyse de ce tableau de valeurs codées permet une analyse plus stable et un autre niveau d'interprétation. Par exemple, on peut

décrire dans le plan des axes principaux les liaisons entre les paramètres en utilisant la trajectoire de chacun des paramètres comme guide à l'interprétation.

*Equivalence distributionnelle*

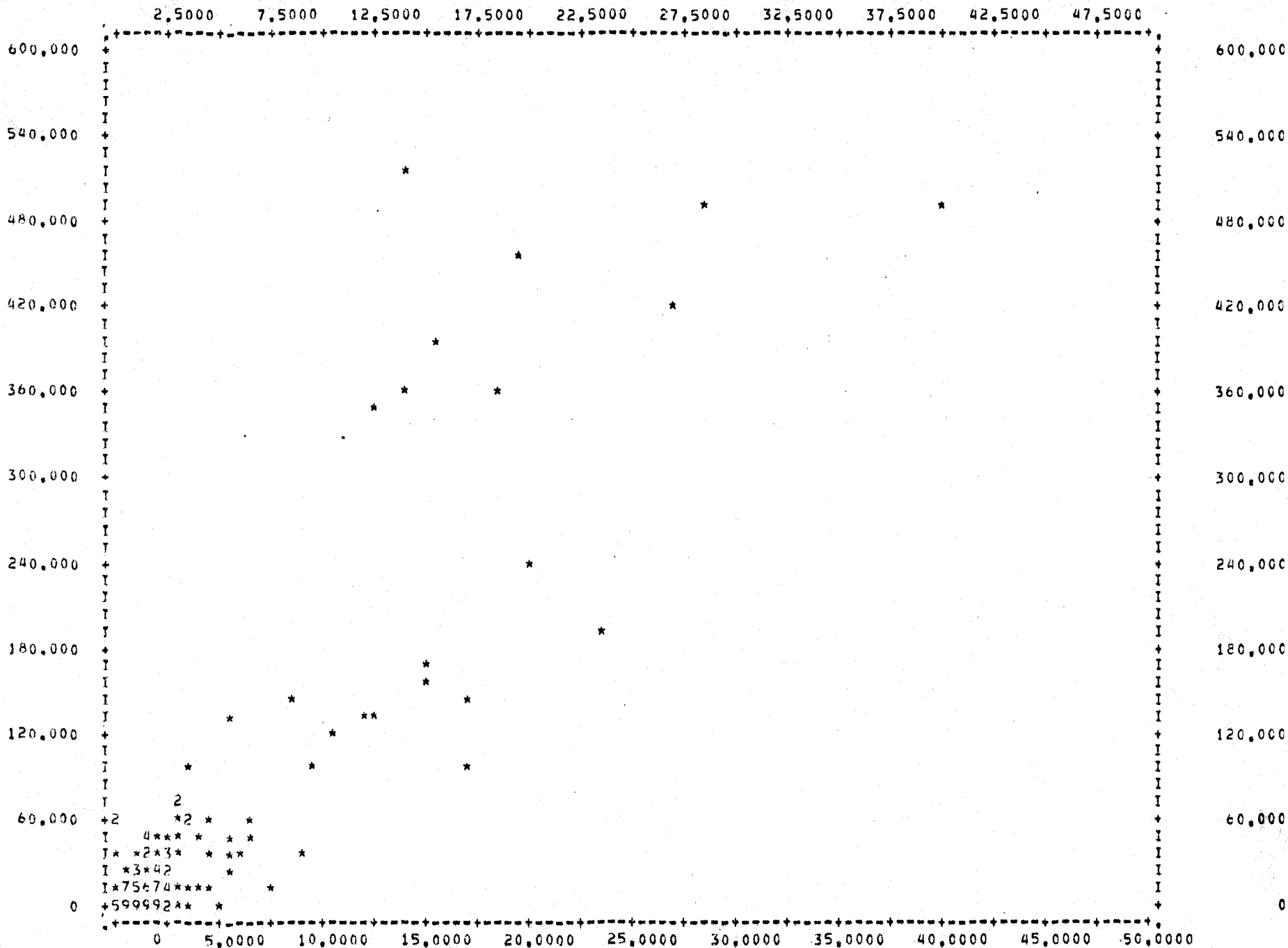
Les distances définies au dessus (relation 1 et 2 ) ont un avantage important. En effet, on peut montrer (Benzecri, 1973) que deux points confondus et possédant chacun un poids donné peuvent être remplacés par un seul point affecté d'un poids qui est la somme du poids des deux points sans modifier les résultats de l'analyse. Par conséquent, l'analyse n'est pas modifiée en appliquant l'A.F.C. sur un tableau dans lequel on a regroupé les observations identiques, c'est-à-dire sur un tableau de comptage.

ANNEXE C

REPRESENTATIONS  
GRAPHIQUES

SCATTERGRAM OF (DOWN) A

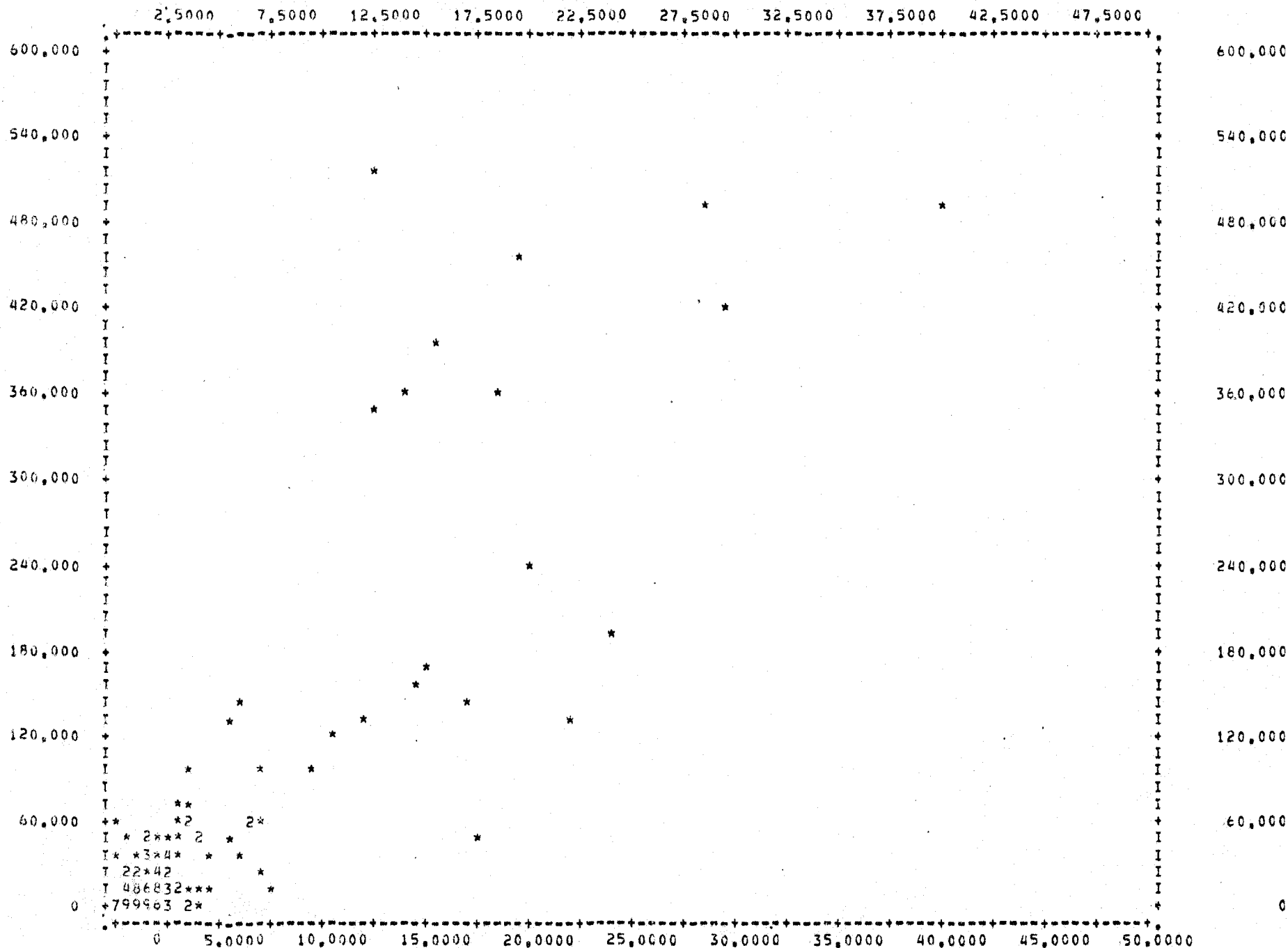
(ACROSS) B





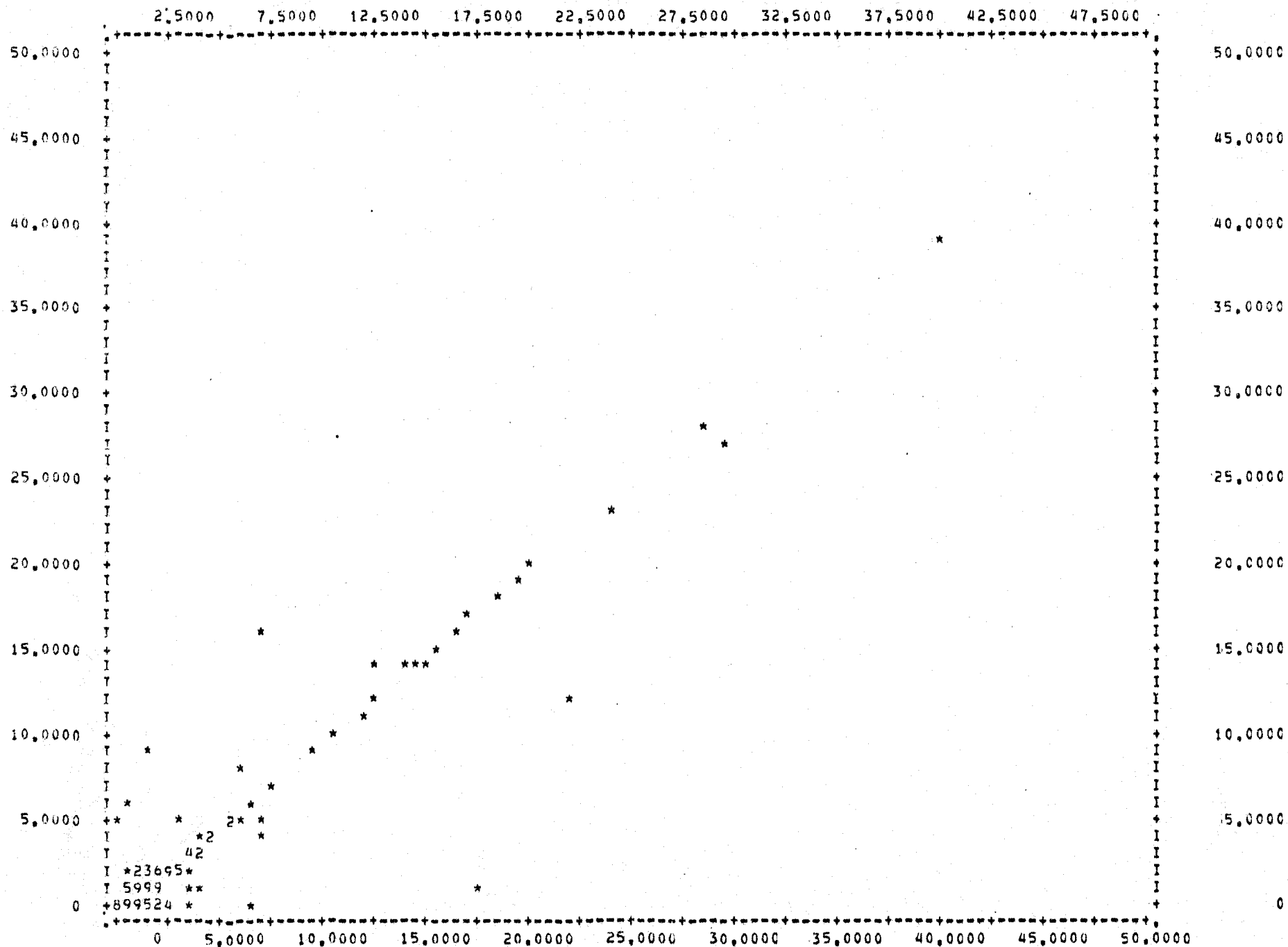
SCATTERGRAM OF (DOWN) A

(ACROSS) C



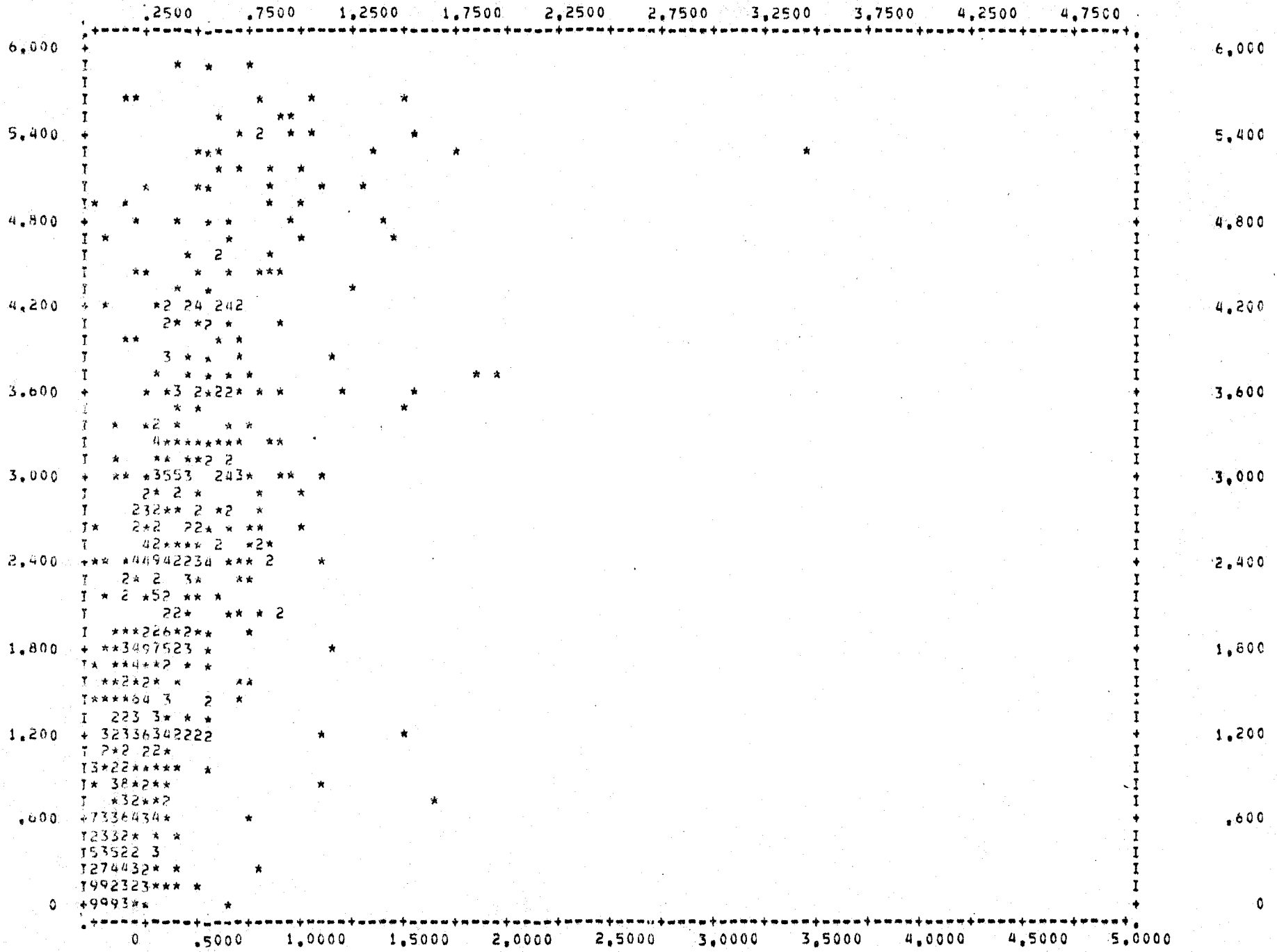
SCATTERGRAM OF (DOWN) R

(ACROSS) C



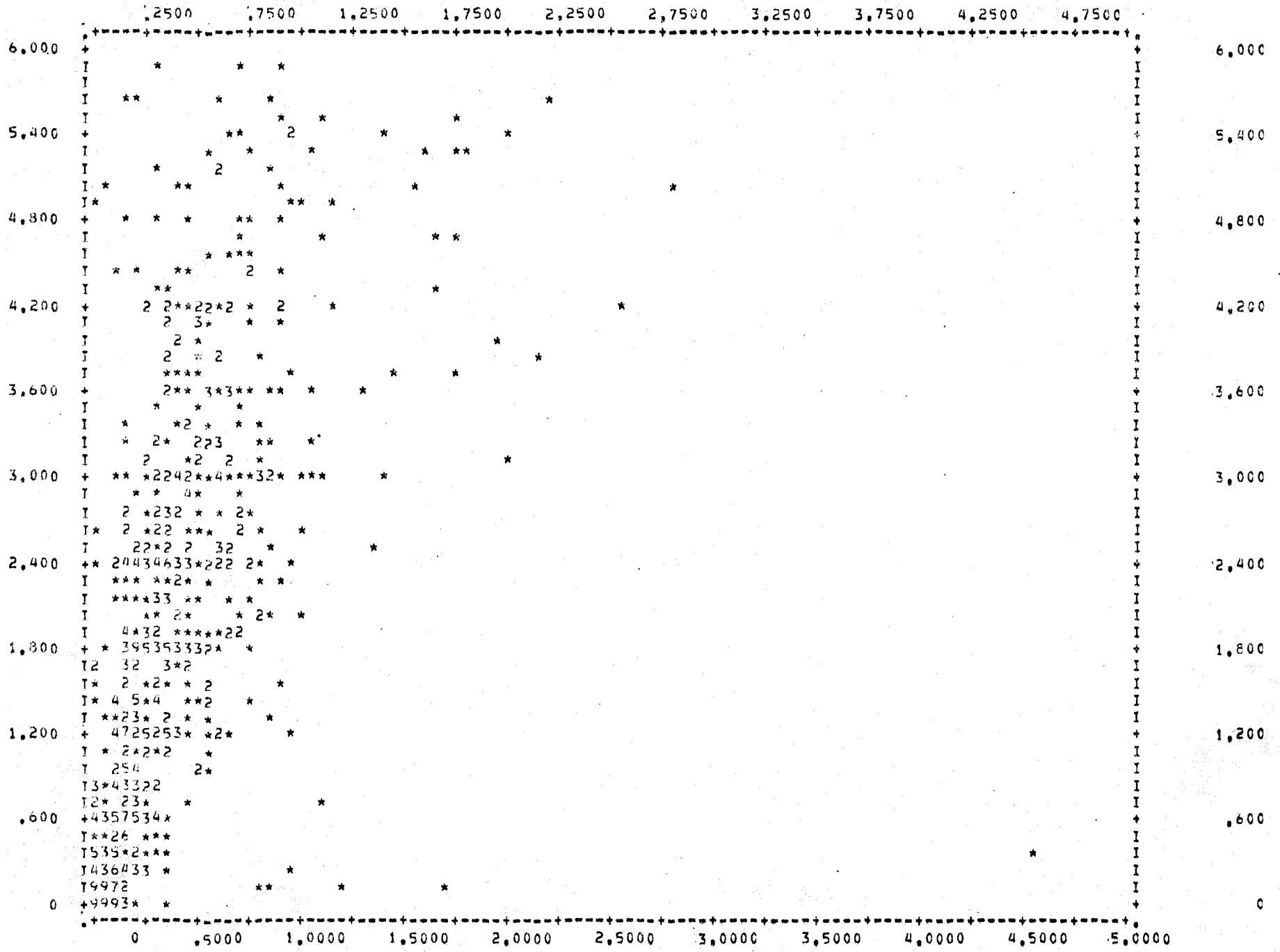
SCATTERGRAM OF (DOWN) A

(ACROSS) B



SCATTERGRAM OF (DOWN) A

(ACROSS) C



SCATTERGRAM OF (DOWN) B

(ACROSS) C

