

APPLICATION DE LA GÉOSTATISTIQUE
AUX PROBLÈMES DE
L'ESTIMATION DU CHARBON

THÈSE

présentée à

l'École Nationale Supérieure des Mines de Paris

par

Margaret ARMSTRONG

pour obtenir le titre

de

Docteur-Ingénieur

en Sciences et Techniques Minières - Option Géostatistique

Soutenue le 18 Juillet 1980

devant le jury composé de :

M. G. MATHERON	Président
M. A. MARECHAL	Rapporteur
M. P. BLAZY	Examineur
M. J.-P. BROCHIER	Examineur
M. Y. CHAMPETIER	Examineur
M. A. JOURNEL	Examineur

APPLICATION DE LA GEOSTATISTIQUE AUX PROBLEMES

=====

DE L'ESTIMATION DU CHARBON

=====

R E M E R C I E M E N T S

=====

Je voudrais d'abord remercier tout particulièrement MM MATHERON et MARECHAL qui m'ont permis de travailler au Centre de Géostatistique et qui n'ont cessé de m'aider et de me conseiller pendant plus de deux ans et demi. Je dois à leur accueil et celui de l'ensemble des chercheurs du Centre d'avoir passé un séjour aussi agréable à Fontainebleau. Je leur sais gré de la patience dont ils ont fait preuve à l'égard de mon mauvais français.

Je dois aussi des remerciements sincères à M. WHITMORE de l'Université de Queensland qui m'a fourni les données de lavabilité utilisées dans cette étude et qui m'a encouragée pendant toute la durée de ce projet. J'ai particulièrement apprécié ses conseils sur les aspects techniques du lavage du charbon ainsi que ceux de M. DELL de l'Université de Leeds et de MM. SMITH et ABBOTT du National Coal Board.

Mes remerciements vont également à Utah Development Co pour m'avoir aimablement autorisée à utiliser les données de la mine de Blackwater sans lesquelles le projet n'aurait jamais pu être mené à bien.

La mise en oeuvre de nouveaux estimateurs de variogramme n'aurait pas été possible sans l'aide et l'assistance de M. DELFINER.

Je voudrais aussi remercier MM. BLAZY, CHAMPETIER, BROCHIER et JOURNEL pour avoir bien voulu faire partie du jury.

Enfin, je suis très reconnaissant à M. CHAUVET pour ses conseils et pour les améliorations qu'il a eu la tâche ingrate d'apporter à la traduction de cette thèse.

TABLE DES MATIERES

	Page
CHAPITRE 1 : INTRODUCTION	
1-1 : Remarques préliminaires.....	3
1-2 : La géologie de la région.....	5
1-3 : Les variables étudiées.....	9
1-4 : Les données fournies.....	10
CHAPITRE 2 : LE CHARBON BRUT	
2-1 : Remarques préliminaires.....	15
2-2 : L'analyse structurale.....	17
2-3 : Calcul de la maille optimale d'échantillonnage : un exemple de l'utilisation du variogramme.....	29
2-4 : Développement d'un variogramme plus robuste.....	43
CHAPITRE 3 : ESTIMATION DES CARACTERISTIQUES DU CHARBON LAVE	
3-1 : Deux approches à ce problème d'estimation.....	57
3-2 : Description des données de lavabilité.....	58
CHAPITRE 4 : PREMIERE APPROCHE	
4-1 : Ajustement des courbes du facteur de récupération et de la teneur en cendre.....	63
4-2 : Estimation du facteur de récupération et de la teneur en cendre.....	63
4-3 : Variogrammes.....	64
4-4 : Statistiques sur les variogrammes.....	66
4-5 : Modèles ajustés des variogrammes.....	73
CHAPITRE 5 : DEUXIEME APPROCHE	
5-1 : Remarques préliminaires.....	85
5-2 : Modèle paramétrique.....	86
5-3 : Tests des modèles.....	89
5-4 : Analyse statistique sur les valeurs des paramètres..	99
5-5 : Le problème fondamental.....	105
5-6 : Tentative d'étude de la distribution de μ	111
5-7 : Procédure à suivre.....	117

CHAPITRE 6 : CONCLUSION.....	123
ANNEXE : ANALYSE STATISTIQUE DES PARAMETRES DU MODELE DE LAVABILITE	
A-1 : Introduction.....	127
A-2 : Statistiques.....	129
A-3 : Conclusion de la première partie.....	136
A-4 : Estimation de la valeur de ρ_a pour une valeur de ρ_c fixée.....	136
A-5 : Valeurs estimées de μ , σ et A_{\min}	139
A-6 : Contrôle de la qualité de l'ajustement.....	140
A-7 : Conclusions.....	144
BIBLIOGRAPHIE.....	147

CHAPITRE 1

INTRODUCTION

CHAPITRE I

INTRODUCTION

1-1 Remarques préliminaires

Au cours des dix dernières années, l'industrie minière a pris de plus en plus conscience du besoin de prédictions très précises de la qualité et de la quantité des réserves de minerai. Aussi s'est-elle intéressée aux méthodes géostatistiques développées initialement par D.G. KRIGE et G. MATHÉRON. Pourtant, malgré l'importance accrue du charbon comme source d'énergie, très peu d'études ont été faites sur l'application de ces techniques aux problèmes spécifiques à l'estimation des réserves de charbon. La principale raison de cette lacune est que, pour la plupart des gisements de charbon, les réserves peuvent être estimées tout simplement par les méthodes habituelles, les difficultés venant des problèmes de corrélation des couches et de localisation des failles. Néanmoins, à cause de l'augmentation des coûts de développement minier et de l'importance des questions de contrôle de l'environnement, les sociétés charbonnières s'intéressent de plus en plus à l'utilisation de la géostatistique pour le calcul des réserves et la prédiction de certains paramètres de qualité tels que la teneur en cendre, la teneur en soufre et l'indice de gonflement.

Le but de ce projet de recherche est d'étudier quelques applications possibles de la géostatistique à ces problèmes d'estimation et de prédiction. Comme la compagnie qui a fourni les données de cette étude tire la plupart de son revenu de la vente de charbon à cokéfier et non de charbon vapeur, l'étude est centrée sur les variables relatives au charbon à cokéfier.

La plus grande partie du charbon à cokéfier exploité en Australie est vendue à l'industrie sidérurgique japonaise qui impose des limites très strictes à la qualité du charbon, notamment en ce qui concerne la teneur en cendre, celle en soufre ainsi que l'indice de gonflement (qui donne une idée de la qualité du charbon à cokéfier). Comme l'infraction à ces limites entraîne de lourdes pénalités financières, il faut absolument fournir du charbon d'une qualité acceptable. Afin de maintenir le niveau de qualité du

charbon, il faut en concasser la plus grande partie et la laver pour éliminer le stérile, matière minérale qui se trouve dans les bandes de schiste entre les couches de charbon. Dans la mine étudiée, un quart seulement du charbon est assez propre pour pouvoir être expédié au Japon sans lavage préalable.

Le lavage et le séchage étant chers, on voudrait pouvoir prévoir, pendant que le charbon est encore in situ, quels blocs sont au dessus des limites prescrites, et, pour ceux qui doivent être lavés, prévoir le facteur de récupération et la teneur en cendre (ou en soufre) s'il est lavé à densité donnée. L'étude envisagée comporte donc deux étapes :

1. Estimation de la qualité des blocs de charbon in situ.
2. Prédiction pour le charbon qui doit être lavé, du facteur de récupération et de la teneur en cendre.

La première partie de cette thèse (Chapitre II) est relative au problème de l'estimation des valeurs des paramètres du charbon brut. Cette étude fait appel aux seules techniques classiques de la géostatistique; ce que laissaient déjà prévoir les études précédemment effectuées sur la teneur en soufre (AGTERBERG et CHUNG (1973), SABOURIN (1975)). Cependant, on a rencontré quelques difficultés lors du calcul des variogrammes de la teneur en soufre, ce qui a nécessité le développement d'un estimateur plus robuste du variogramme. Mis à part cette nouveauté, ce chapitre ne fait que développer l'application de techniques déjà bien connues à un nouveau domaine.

Le reste de cette thèse traite le problème de la prédiction du facteur de récupération et de la teneur en cendre du charbon lavé. Deux façons d'aborder cette question sont présentées au chapitre III. Le chapitre IV donne une description détaillée de la première de ces deux approches, qui consiste à estimer d'abord le facteur de récupération et la teneur en cendre, puis à kriger ces valeurs sur le domaine étudié. L'alternative à cette approche est de développer un modèle paramétrique exprimant le facteur de récupération et la teneur en cendre en fonction de la densité de lavage, puis de kriger les paramètres de ce modèle sur le domaine étudié et finalement d'évaluer le facteur de récupération et la teneur en cendre à densité donnée en injectant les valeurs krigées des paramètres dans le modèle. Cette méthode présente l'avantage de permettre l'évaluation du facteur de récupération et de la teneur en cendre à une nouvelle densité de lavage sans répéter tous les calculs: il suffit de rentrer la nouvelle valeur de la densité

dans le modèle. Le bilan de ces techniques est présenté dans le chapitre final (Chapitre VI). Une annexe contient les détails d'une étude statistique des valeurs des paramètres.

Avant d'aborder les questions géostatistiques proprement dites, nous présenterons une brève description de la géologie du gisement de charbon du Bowen Basin (§ 1-2), et nous donnerons quelques indications sur les problèmes techniques liés à ces couches de charbon et aux variables étudiées (§ 1-3). Enfin, au paragraphe 1-4, nous décrirons les données fournies par la compagnie minière.

1-2 La géologie de la région.

La description suivante est extraite d'un article écrit par un des géologues de la compagnie (HOURIGAN, 1977).

" Le gisement de charbon du Bowen Basin couvre approximativement une superficie de 65 000 km² et s'étend sur une distance de près de 500 km, de la région de Theodore et Cracow au Sud jusqu'à Collinsville au Nord, et sur une largeur de 130 km, de 100 km de la côte à l'extrémité Est jusqu'à Blair Athol à l'Ouest. (Voir Figure 1).

Le Bowen Basin est une structure d'origine sédimentaire qui s'est développée au début du Permien. L'histoire détaillée du dépôt et de la structure de ce bassin est très compliquée, mais le bassin peut être divisé grossièrement en deux parties : une profonde dépression à l'Est et un plateau à l'Ouest. Pendant la première moitié du Permien, une séquence volcanique recouverte par des sédiments marins s'est déposée dans la dépression orientale, pendant que se déposait sur le plateau occidental une mince couche de sédiments marins et que se développaient des gisements de charbon à la périphérie du bassin. Pendant la deuxième moitié du Permien, des gisements de charbon plus étendus se sont formés à travers tout le bassin à mesure que la mer se retirait de la région. La majeure partie du bassin est recouverte d'argiles, de sables et de graviers qui datent de l'époque tertiaire. C'est pourquoi les couches de charbon affleurent rarement.

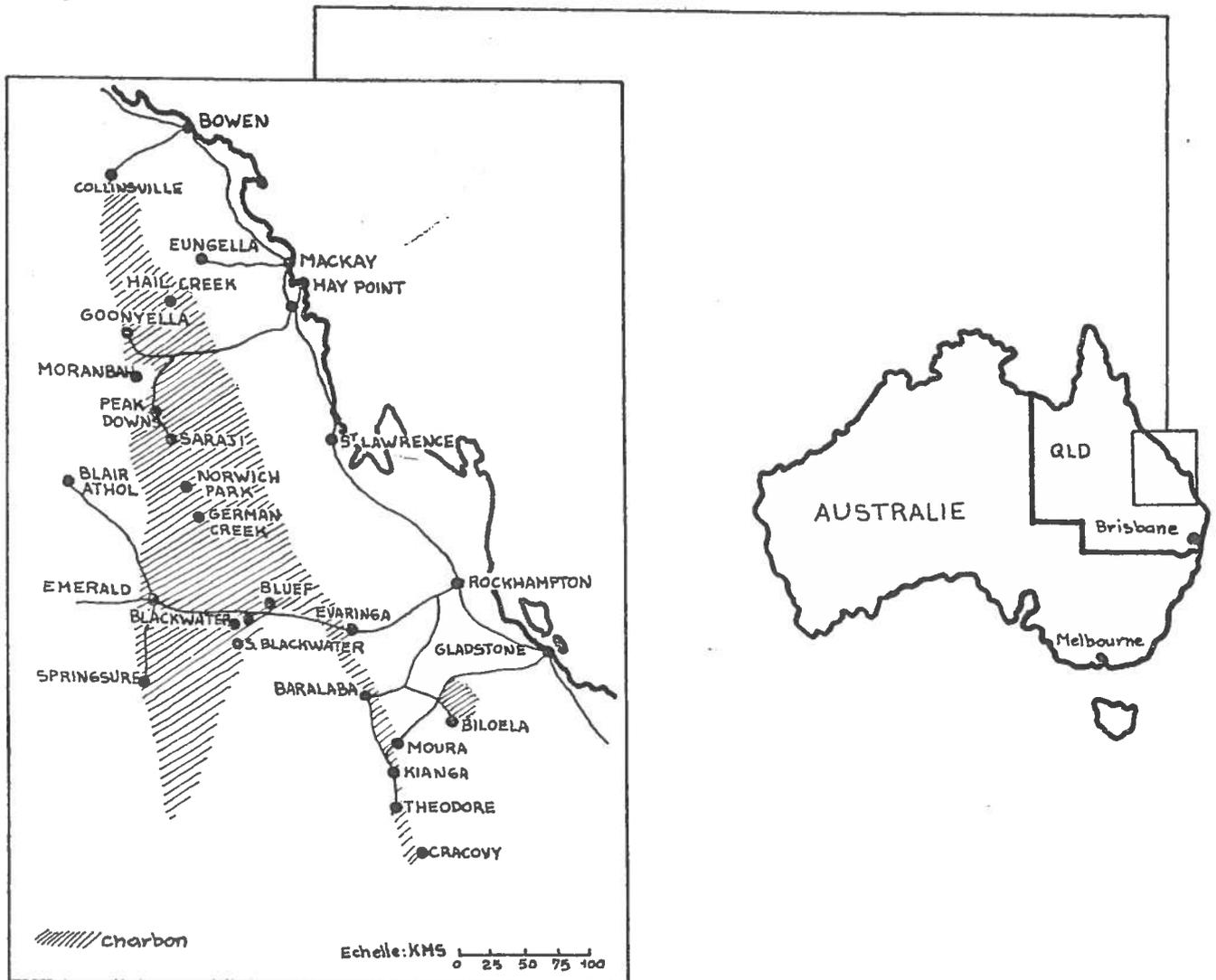


Figure 1 : Le gisement de charbon dans le Bowen Basin.

Après le dépôt des sédiments triassiques, la séquence Permien-Triassique du Bowen Basin a été pliée par de fortes forces tectoniques qui agissaient dans la direction Nord-Est. En conséquence de cette activité, les sédiments épais de la zone de dépression proche du plateau occidental ont été

fortement pliés et faillés, et par endroits poussés vers le plateau, ce qui a formé un axe structural appelé "Dawson Tectonic Zone". A l'Ouest de cette zone tectonique, les sédiments plus minces du plateau ont été pliés faiblement et les couches de charbon ont un pendage de 2 à 8 degrés."

Comme les mines de la compagnie se situent à l'Ouest de cette zone tectonique, les couches exploitées sont pliées et inclinées mais pas fortement faillées. Cependant, elles se composent de quelques sous-couches qui se séparent et se rejoignent ici et là, comme on peut le voir dans le schéma des couches dans la région de la mine étudiée. (Figure extraite de l'article de STAINES (1972)).

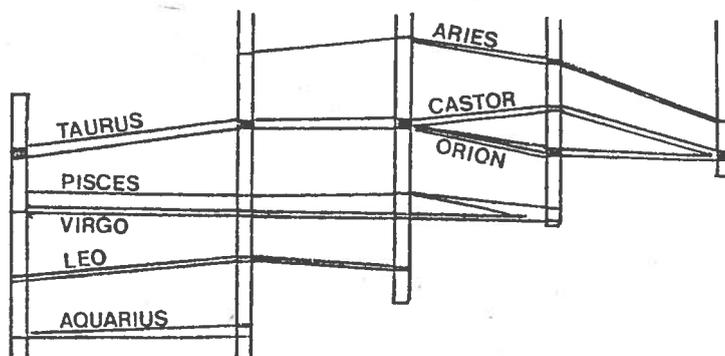
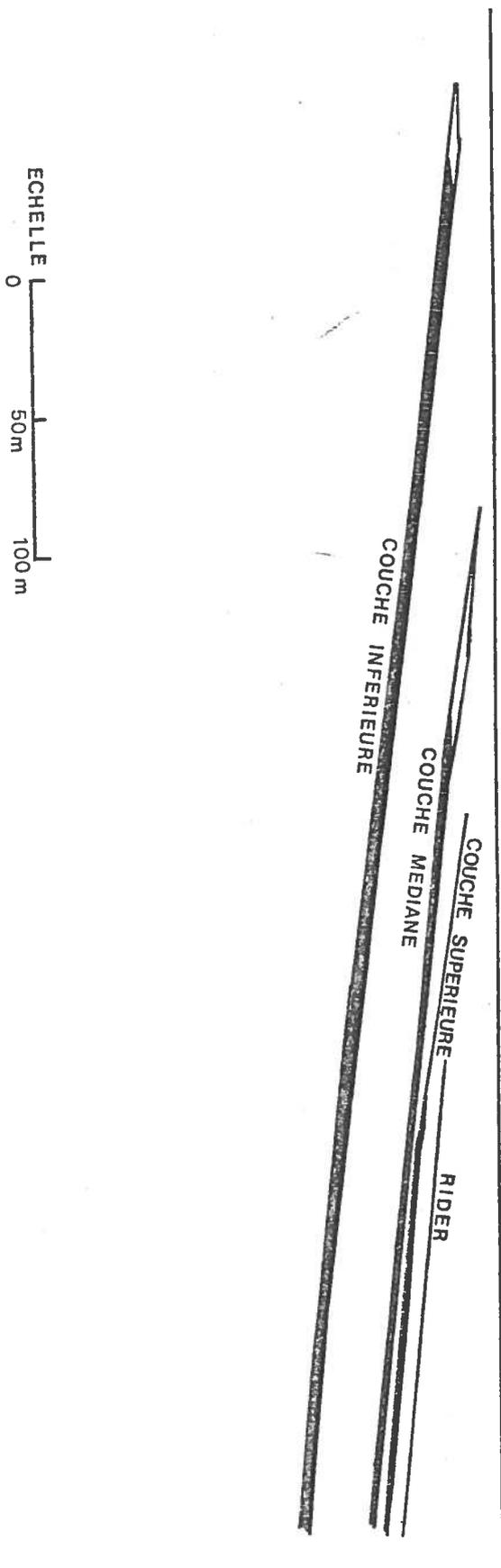


Figure 2 : Exemple de section verticale dans la direction E-W du gisement de Blackwater (d'après STAINES (1972)).

A proximité de la mine étudiée, il y a trois couches qui seront désignées par : "couche inférieure", "couche médiane" et "couche supérieure". Au-dessus de la couche supérieure se trouve une autre couche très mince qui s'appelle le "rider" de la couche supérieure. La couche inférieure et la couche médiane sont composées de deux sous-couches ayant des caractéristiques tout à fait différentes. L'existence de ces deux sous-couches complique le travail du géostatisticien, parce que, en toute rigueur, il faudrait faire tous les calculs séparément pour les deux sous-couches. Une deuxième complication se présente lorsque la couche de charbon s'approche de la surface, parce que le charbon, étant alors oxydé, ne peut pas être utilisé pour faire du coke. Une illustration des trois couches et des zones d'oxydation est présentée à la figure 3.

FIGURE 3: Les trois couches



1-3 Les variables étudiées.

Les variables étudiées sont relatives au charbon à cokéfier et non au charbon vapeur. Ce sont : la profondeur de la couche, la puissance de la couche, la teneur en cendre, la teneur en soufre et l'indice F.S.I. (Free Swelling Index). La dernière de ces variables demande quelques mots d'explication. Le "Free Swelling Index" (ou "Crucible swelling number") donne une indication de la qualité du charbon à cokéfier. Pour l'évaluer, un gramme de charbon pulvérisé est maintenu à 820° C dans un four jusqu'à ce que la composante volatile ait été éliminée. L'indice de gonflement qui varie de 0 à 9 par demi-unité est déterminé par comparaison de la forme du champignon résultant aux profils standard (Figure 4). Pour les détails sur la procédure, voir British Standard BS1016 (1977).

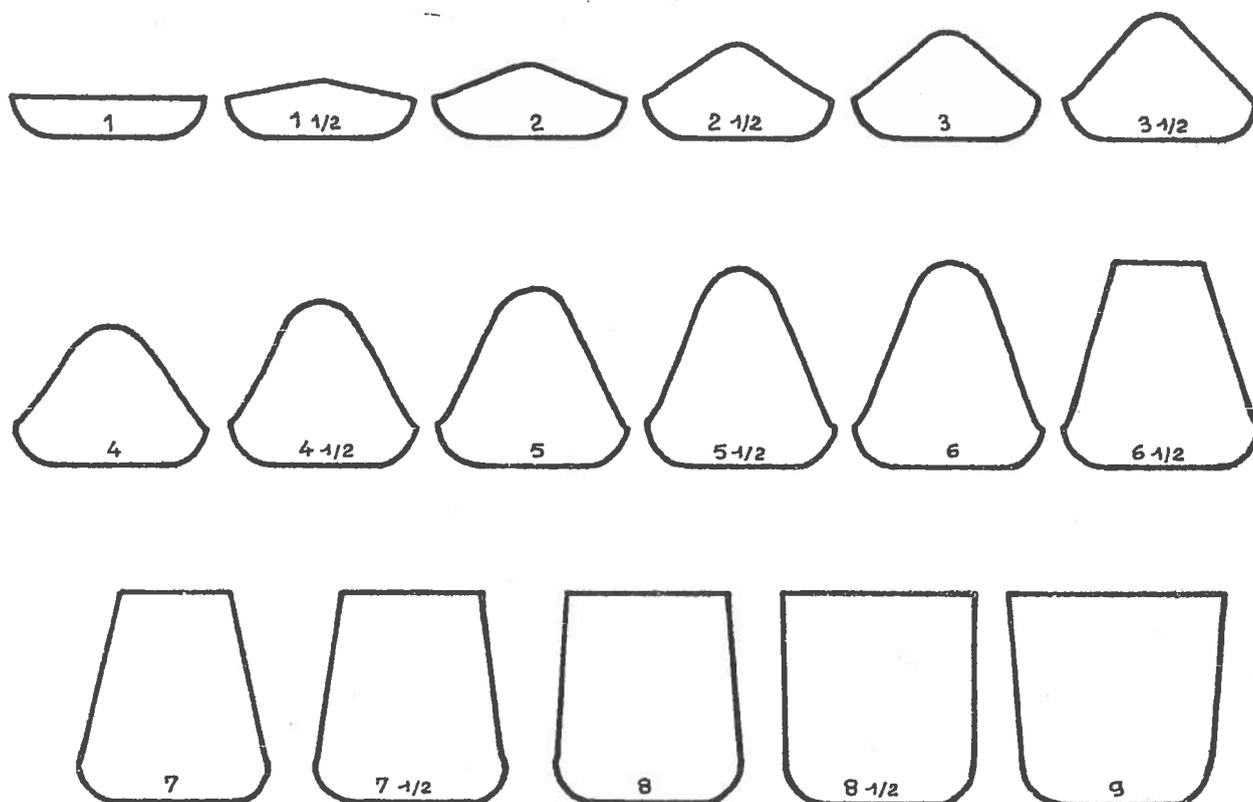


Figure 4 : Profils standard pour la détermination du F.S.I.

Au contraire des quatre autres variables, cet indice (F.S.I.) n'est pas additif; c'est-à-dire que si x grammes de charbon d'un indice de 3 sont mélangés avec y grammes de charbon d'un indice de 7, le résultat n'a

pas nécessairement un indice de $(3x + 7y)/(x + y)$. BENNETT et TAYLOR (1970) ont démontré que pour des charbons australiens, cet indice est lié à la classification et à la composition pétrographique du charbon. Les abaques accompagnant leur article permettent l'estimation de cet indice pour des mélanges de charbons où la classification et la composition pétrographique sont connues. Mais cette information n'est pas disponible pour les analyses standard des sondages. Faute d'autre procédure possible, il a fallu supposer l'additivité de cet indice. Dans une communication personnelle M. ISOKANGAS, responsable de la section technique de la compagnie, a remarqué que bien qu'elle ne soit pas exacte, l'approximation donne des résultats qui sont corrects à $\pm 1/2$ près pour le charbon de la mine.

1-4 Les données fournies.

La compagnie a fourni des données sous forme d'une bande magnétique contenant le numéro de chaque sondage, ses coordonnées et les résultats des analyses de tous les sondages de 2" de diamètre provenant de sa concession de Blackwater. Les résultats des analyses comportent le nom de la couche rencontrée, sa profondeur et le type de charbon, ainsi que des analyses du charbon brut, celles de quelques tranches densimétriques et de temps à autre de la fraction très fine. Dans la plupart des cas, le charbon a été analysé pour sa teneur en cendre, sa teneur en soufre et son F.S.I. Les densités auxquelles étaient faites les analyses varient considérablement d'un sondage à l'autre. En plus de la bande magnétique, la compagnie a fourni une carte de la concession indiquant l'implantation des sondages et quelques sections verticales représentant le profil des couches.

Quelques difficultés s'étant présentées lors de l'estimation de l'effet de pépite du variogramme de soufre, la mine a communiqué les résultats d'une série de 18 sondages disposés en 3 lignes de 6 sondages écartés de 10 m l'un de l'autre. Dans un premier temps, le diamètre de ces sondages a été donné égal à 4", mais il est apparu par la suite qu'il serait en fait de 2". Cette incertitude n'affecte pas la méthode exposée, mais dans le cas d'une étude réelle, il serait important de connaître la valeur exacte de ce diamètre : l'ajustement de l'effet de pépite est très sensible à l'effet de support.

Vers la fin de l'étude, la compagnie a également fourni des données de lavabilité très détaillées en provenance de 21 sondages de 8" de diamètre, qui ont été faits dans la moitié Nord de la concession. Dans le cas de 7 de

ces sondages, les deux sous-couches ont été testées séparément ainsi que la couche entière. Dans trois autres cas, les données venaient d'un mélange de deux sondages très proches: N^{OS} 902, 903; 904, 905; 906, 907 . Cela donne 32 séquences de données au total.

Nombre de sondages	Numéro des sondages	Echantillon en provenance de la
11	901-949	Couche entière
7	1155-1161	Couche entière
"	"	Sous-couche inférieure
"	"	Sous-couche supérieure

TABLEAU 1 : Résumé des sondages de 8".

Une description détaillée de ce jeu de données est présenté à la section 3-2, ainsi qu'une description de la procédure d'analyse de lavabilité.

Le Professeur WHITMORE de l'Université de Queensland m'a permis d'utiliser des données de lavabilité de 28 échantillons de charbon en provenance des mines de l'Est de l'Australie. Ces données ont servi à tester les deux modèles mathématiques proposés au Chapitre V.

CHAPITRE II

LE CHARBON BRUT

CHAPITRE II

LE CHARBON BRUT

2-1 - Remarques préliminaires

Les problèmes de l'estimation de l'épaisseur de la couche, la teneur en cendres, etc..., du charbon brut sont peu différents de ceux qu'on rencontre en géostatistique classique. Puisque ces méthodes sont maintenant bien connues, il nous suffit de décrire brièvement leur application dans ce cas et de montrer où elles sont différentes des techniques habituelles. Avant de commenter l'analyse structurale, qui constitue toujours la première étape d'une étude géostatistique, il faut faire quelques remarques préliminaires au sujet des variables étudiées.

Dans l'introduction, on a noté l'existence d'une différence entre les deux sous-couches des couches inférieure et médiane. (On n'étudie pas la couche supérieure parce qu'elle n'est pas exploitée). Dans les deux cas, la sous-couche supérieure n'est pas aussi bonne à cokéfier, et dans certaines régions du gisement, elle n'est utilisée que pour fournir du charbon vapeur. La différence entre les deux sous-couches est déjà évidente au niveau des valeurs moyennes des variables (Tableau 2).

	Teneur en cendre	Teneur en soufre	F.S.I.
Sous-couche supérieure	11,9 %	0,41 %	2,5
Sous-couche inférieure	8,7 %	0,39 %	8,0

TABEAU 2 : Valeurs moyennes des variables des deux sous-couches pour la couche médiane.

Il aurait été préférable d'étudier les sous-couches séparément, mais cette option est exclue du fait de l'organisation du fichier des données. Dans le cas

de quelques sondages, les analyses n'ont été effectuées que pour la couche entière (au lieu de sous-couche par sous-couche); dans d'autres endroits, il y a plus de deux analyses en provenance de la même couche et il est impossible de distinguer de quelle sous-couche provient l'analyse. Etant donné l'impossibilité de traiter les sous-couches séparément, on a préféré combiner toutes les analyses d'une même couche. Par conséquent les variables étudiées sont la puissance totale de la couche, la valeur moyenne de F.S.I., les teneurs moyennes en cendre et en soufre. L'inconvénient d'étudier la couche entière au lieu des sous-couches est que la variabilité due à l'hétérogénéité des deux sous-couches masque la variabilité spatiale et a pour conséquence d'augmenter l'effet de pépité et de rendre plus difficile l'analyse structurale. Mais malgré tous ces inconvénients, c'est la seule solution possible. A long terme, il serait préférable de remplacer le système actuel par un autre plus complet qui indiquera de quelle sous-couche provient l'analyse.

On rencontre un problème semblable en bordure du gisement où la couche se trouve très proche de la surface et où par conséquent le charbon est oxydé (voir Figure 5).

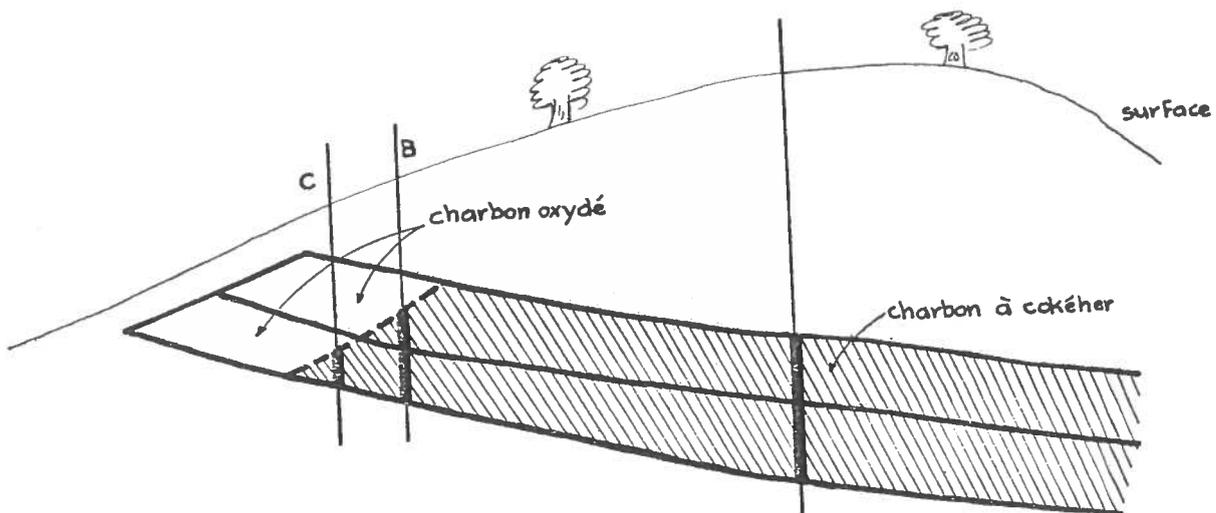


Figure 5 : Section verticale du gisement près de la surface.

Contrairement aux échantillons prélevés aux points B et C qui contiennent du charbon oxydé ainsi que du charbon à cokéfier, le charbon au point A n'est pas oxydé du tout. On peut alors se demander comment définir la puissance de la couche aux points B et C : par la puissance totale de la couche y compris le charbon oxydé, ou seulement par la puissance du charbon utilisable ? Si on considère la puissance totale comme variable à étudier, la variable reste stationnaire, mais il faut trouver un moyen de corriger l'estimateur de la puissance moyenne en prenant en compte le pourcentage de charbon non-utilisable. Si, au contraire, on n'étudie que la fraction non-oxydée, on n'a plus le problème de correction de l'estimateur de la puissance moyenne, mais la variable n'est plus stationnaire. Le même problème de non-stationnarité se présente pour définir les autres variables même si on se restreint au cas du charbon à cokéfier, à cause de la présence de deux sous-couches. Par exemple, la teneur moyenne en cendre au point B est due à la teneur en cendre d'une partie de la sous-couche supérieure ainsi que celle de toute la sous-couche inférieure. Au contraire, au point C, elle est due entièrement à la sous-couche inférieure. Il est évident que ces deux teneurs sont différentes de celle en A.

Après avoir considéré les deux possibilités, on a décidé d'étudier seulement le charbon de qualité à cokéfier et de ne pas tenir compte du reste. La raison de cette décision est que, en fin de compte, la société ne vend que le charbon non-oxydé. Ce choix entraîne le problème de non-stationnarité en bordure du gisement, mais il a l'avantage de donner une certaine unité aux définitions des variables. Le problème des définitions de variables résolu, on peut poursuivre l'analyse structurale.

2-2 - L'analyse structurale.

Un aspect important de l'analyse structurale est de déterminer si la régionalisation des variables est stationnaire et isotrope. Si des dérives ou des anisotropies existent, il est probable qu'elles suivront la direction ou la pente de la couche. Par conséquent, il est intéressant d'étudier la régionalisation des variables dans une section du gisement où la couche est plane, ce qui permet d'affecter un angle précis à la direction et à la pente de la couche. C'est pour cette raison que la moitié sud du gisement a été choisie pour l'analyse initiale des statistiques et de la structure. Cette partie du gisement présente un autre avantage : les sondages

sont assez proches les uns des autres.

2-2-1 L'analyse statistique

Pour commencer l'étude, nous avons analysé les données relatives au charbon brut, pour mettre en évidence leurs éventuelles particularités. On a calculé les histogrammes des cinq variables (profondeur de la couche, puissance, teneurs en cendre et en soufre, F.S.I.). Les histogrammes sont unimodaux, sauf celui de la puissance de la couche médiane qui présente un pic vers les fortes valeurs (Figure 6). A cause de l'étalement des fortes valeurs, ces variables ne peuvent pas être gaussiennes, ce qui est confirmé par un test de χ^2 . Les moyennes et les variances des variables ont été calculées. Le tableau des moyennes et variances n'est pas présenté pour des raisons de discrétion. Les moyennes des puissances et des teneurs en cendre sont différentes d'une couche à l'autre mais celle des F.S.I. et des teneurs en soufre sont très semblables pour les deux couches. Cette similitude est un résultat inattendu de l'étude, surtout si l'on considère l'écart entre les dates de dépôt des deux couches. Les matrices de corrélation ont également été calculées, mais il n'y a pas de fortes corrélations entre les variables.

2-2-2 Les variogrammes

(i) Pour mettre en évidence une éventuelle anisotropie, les variogrammes ont été calculés dans les quatre directions N-S, O-E, NE-SO et NO-SE. La profondeur de la couche exceptée, les variables sont isotropes jusqu'à une distance de 2 km.

La géologie de la région indique que la couche s'enfonce vers l'est. Cette dérive de la profondeur de la couche est la cause d'une forte augmentation à grandes distances du variogramme O-E. On peut s'attendre à ce que la profondeur de la couche soit presque continue et qu'elle ne présente pas beaucoup de fluctuation à courtes distances. Ces deux caractéristiques expliquent le comportement du variogramme près de l'origine : l'absence d'un effet de pépite indique peu de fluctuations à courtes distances et la forme parabolique est liée à la continuité de la variable. Dans des cas de variables très continues et avec une forte dérive, il vaut mieux utiliser un programme automatique de krigeage (comme BLUEPACK) qui estime la dérive et le variogramme simultanément. Cette variable ne nécessite alors plus d'étude détaillée, et il

FIGURE 6 : Histogrammes de la couche mediane

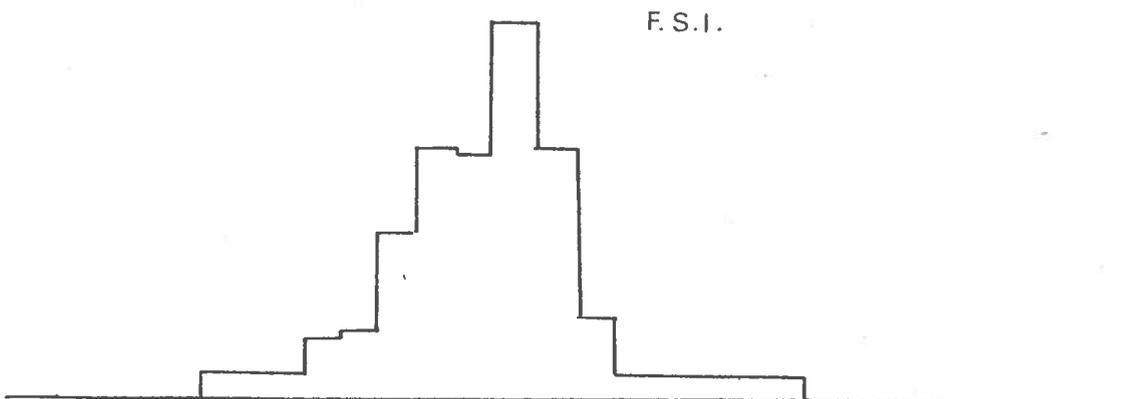
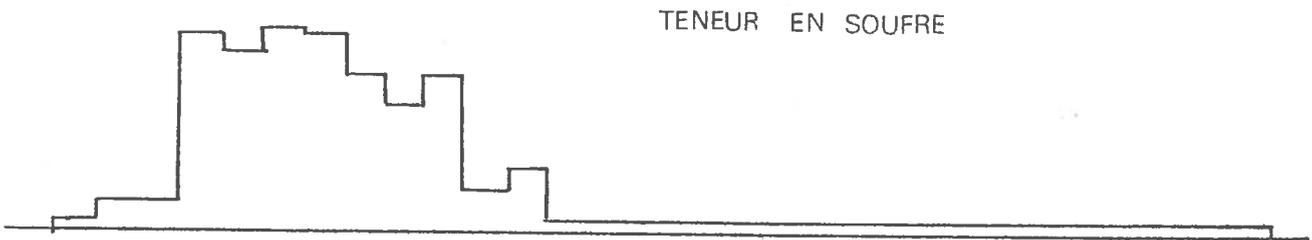
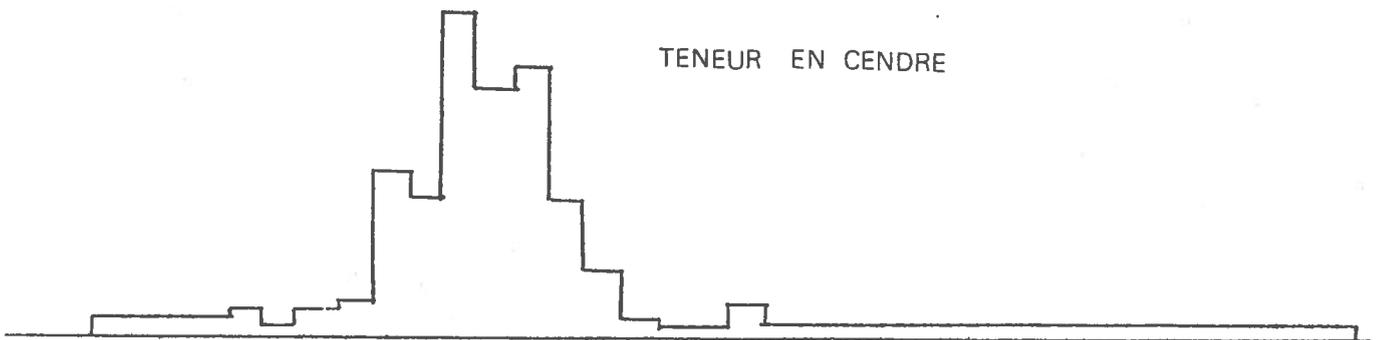
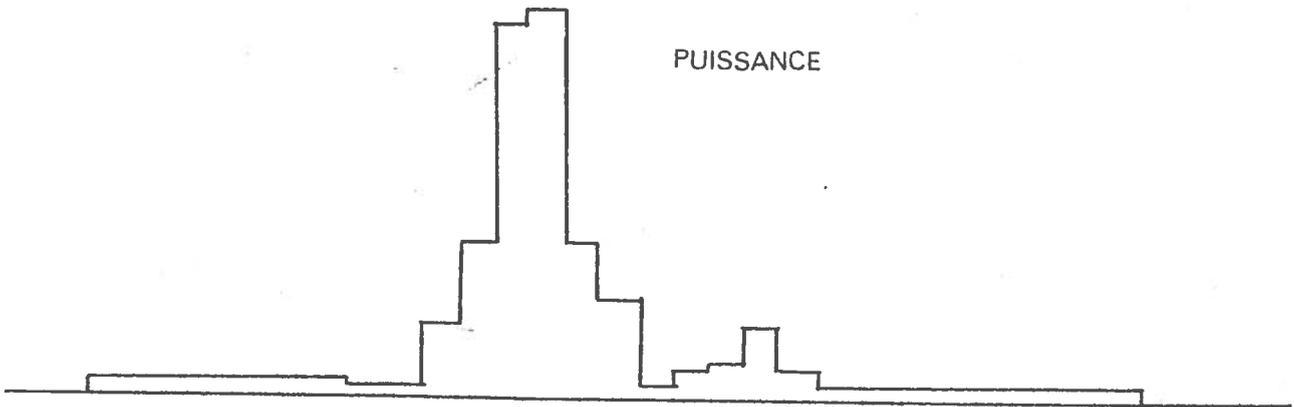
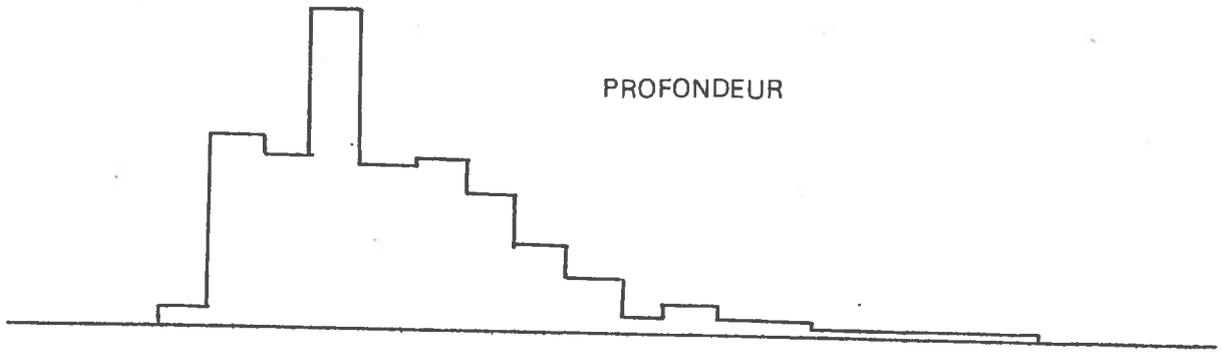
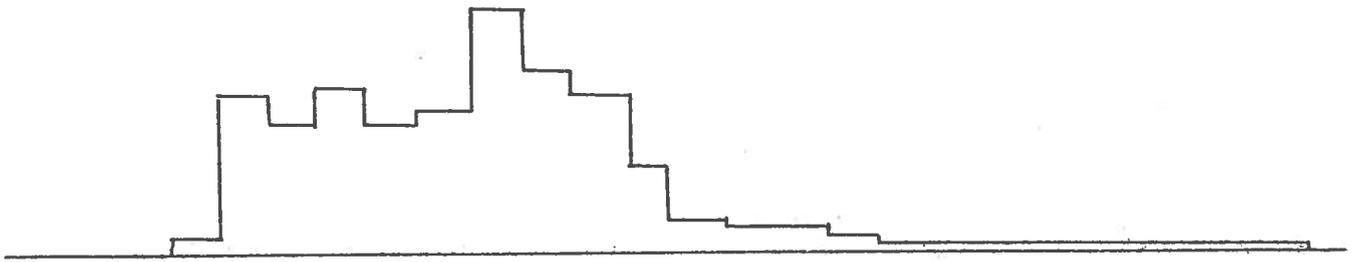
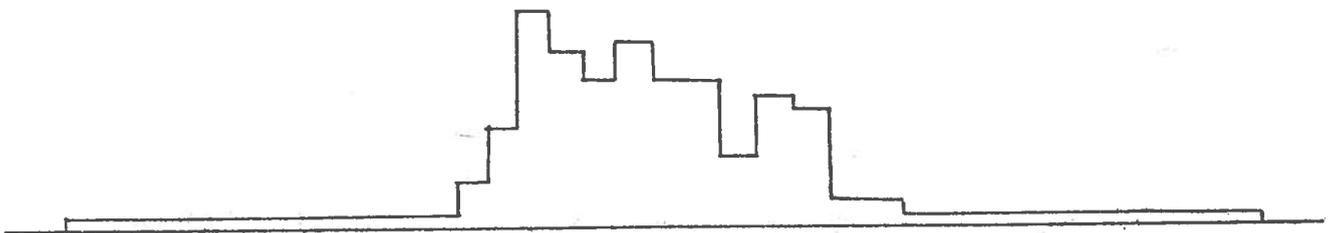


FIGURE 6 Histogrammes de la couche inferieure

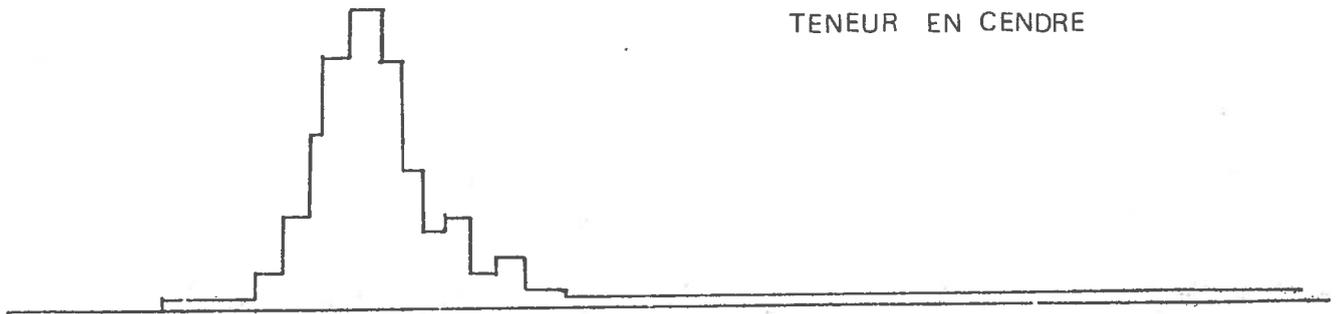
PROFONDEUR



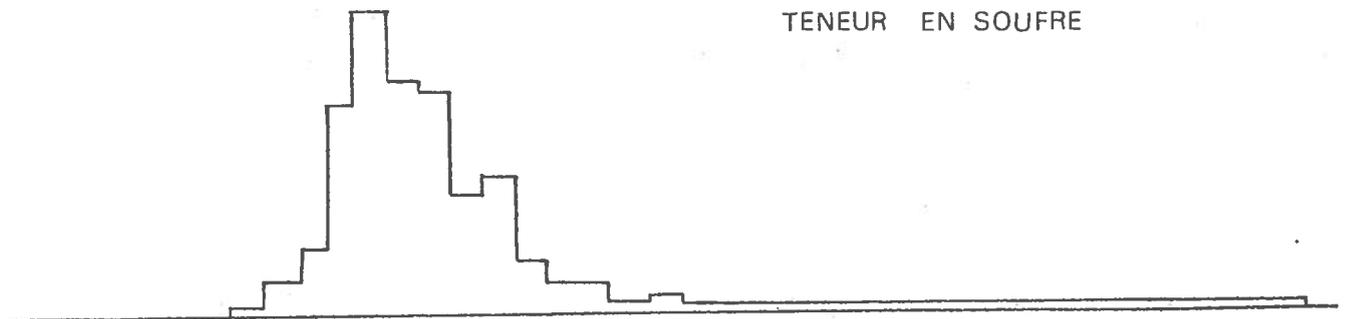
PUISSANCE



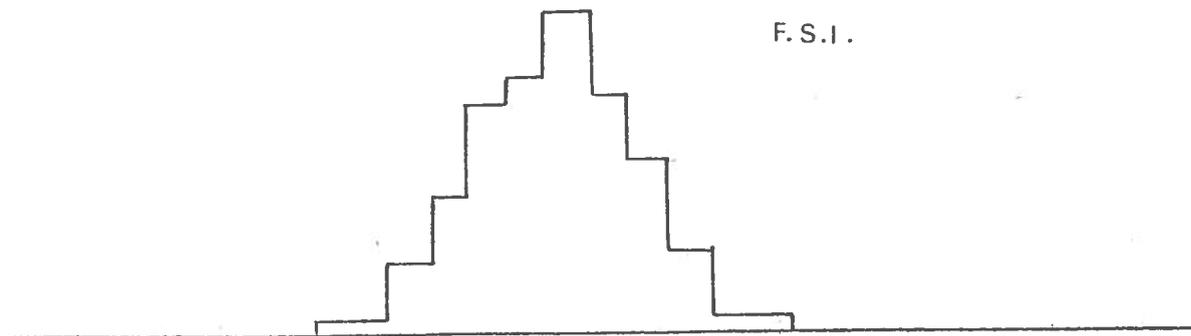
TENEUR EN CENDRE



TENEUR EN SOUFRE



F.S.I.



n'y sera plus fait référence dans la suite du texte.

Puisque les quatre autres variables sont isotropes et sans dérive jusqu'à une distance de 2 km, les quatre variogrammes directionnels peuvent être regroupés. Malheureusement, même après le regroupement, les variogrammes ne sont pas bien définis à l'origine.

(ii) Le comportement des variogrammes à l'origine.

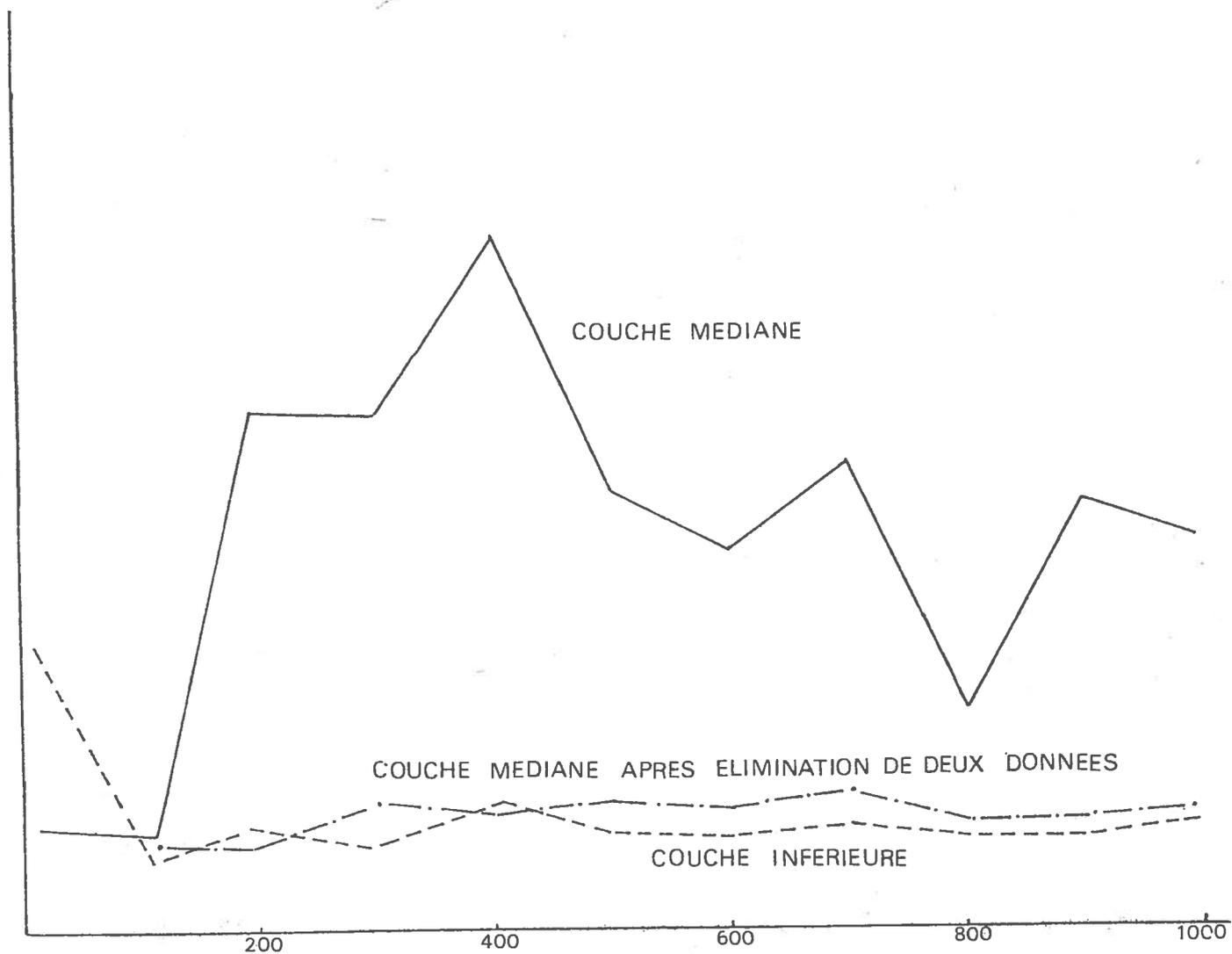
Le nombre de points d'échantillonnage n'est pas suffisant pour donner une indication claire du comportement des variogrammes à l'origine. Compte tenu de l'importance des variogrammes dans la qualité des estimations, il faut donc chercher à déterminer plus précisément leur comportement.

En raison de l'absence d'isotropie et de dérive, on peut calculer les variogrammes sur le gisement entier. Sauf pour la teneur en soufre, les variogrammes des deux couches se ressemblent beaucoup. Ils sont presque identiques pour la teneur en cendre et la puissance de la couche; ils ont la même forme mais des effets de pépite différents pour le F.S.I. Au contraire, le variogramme de la teneur en soufre de la couche médiane a un ordre de grandeur plus important que celui de la couche inférieure. La différence est due à la présence de deux données aberrantes parmi les 207. Après élimination de ces deux valeurs, le variogramme relatif à la couche médiane se retrouve au niveau de celui de la couche inférieure. Les variogrammes initiaux des deux couches ainsi que le variogramme obtenu en éliminant les deux valeurs très élevées sont présentés dans la figure 7.

En pareil cas, il est important de savoir si les données statistiquement "aberrantes" sont dues effectivement à une erreur de transcription, ou si elles sont exactes. Pour élucider cette question, toutes les teneurs en soufre de plus de 1,5 fois la moyenne de la couche ont été localisées. Ces valeurs étant sensiblement alignées dans une petite partie du gisement, il semblerait qu'elles soient exactes. La société minière l'a confirmé, mais elle n'a pas pu suggérer une explication physique de ce phénomène.

Dans ce cas, le variogramme sous-jacent a été obtenu après élimination de deux données. Mais souvent la forme du variogramme est affectée par l'existence d'une queue de valeurs extrêmes et il est difficile de déterminer à quel point il faut tronquer l'histogramme pour obtenir un variogramme représentatif. En pareil cas, il est utile d'avoir un deuxième estimateur du variogramme. Quelques possibilités sont présentées au paragraphe 2.4.

FIGURE 7: Variogrammes de la teneur en soufre



D'autre part, le calcul des variogrammes sur l'ensemble de la concession n'a pas permis de déterminer les structures à petite échelle. Pour ce faire, il faut des données espacées de quelques mètres l'une de l'autre; une campagne de sondages supplémentaires a été entreprise par la société pour permettre ce calcul. Au début, on a cru que ces sondages avaient un diamètre de 4", mais plus tard, alors que tous les calculs étaient déjà faits, il est apparu que les carottes avaient, en fait, un diamètre de 2". Cette différence est très importante parce que la taille du support a une très grande influence sur la variabilité des teneurs en cendre et en soufre, et aussi du F.S.I. Comme dans le cas des carottes de 4", les teneurs sont calculées sur des volumes quatre fois plus grands que pour les carottes de 2", l'effet de pépité sera quatre fois plus petit. En ce qui concerne le F.S.I., qui n'est pas additif, on ne peut pas prévoir l'effet précis de l'augmentation du support sur l'effet de pépité. Cependant, on suppose que l'effet de pépité sera plus grand que pour les carottes de 2". En fait, on gardera l'hypothèse de linéarité faite au paragraphe 1-3. Quant à l'effet de pépité du variogramme de la puissance, il ne sera pas changé par le changement du support.

En estimant l'effet de pépité de cette façon, on fait l'hypothèse implicite que toute la variabilité à une échelle inférieure à 10 m est due à des microstructures de portée inférieure à 2". S'il y avait une structure à une échelle plus petite, l'extrapolation à l'origine du variogramme à un pas de 10 m conduirait à sous-estimer l'effet de pépité du modèle ponctuel.

Comme tous les calculs ont été déjà faits en supposant que le diamètre était de 4", on conservera cette hypothèse par la suite. Mais s'il apparaissait que le diamètre réel soit de 2", il faudrait réajuster toutes les valeurs estimées des effets de pépité.

(iii) L'ajustement des modèles.

En raison de la ressemblance des variogrammes des deux couches, les moyennes pondérées par le nombre de couples ont été calculées et les modèles ont été ajustés à ces variogrammes. Ces variogrammes moyens fluctuent moins, ce qui rend l'ajustement plus facile.

Nous avons commencé par ajuster un modèle au variogramme de la teneur en cendre, parce qu'il est mieux structuré. Le modèle choisi est la somme de deux variogrammes sphériques de portées 200 m et 1 800 m, et de paliers, 1 et 2,3 respectivement, avec un effet de pépité de 1,2. Puisqu'on peut supposer que les autres variables ont été influencées par les mêmes facteurs

FIGURE 8a : Variogramme ajusté de la teneur en cendre

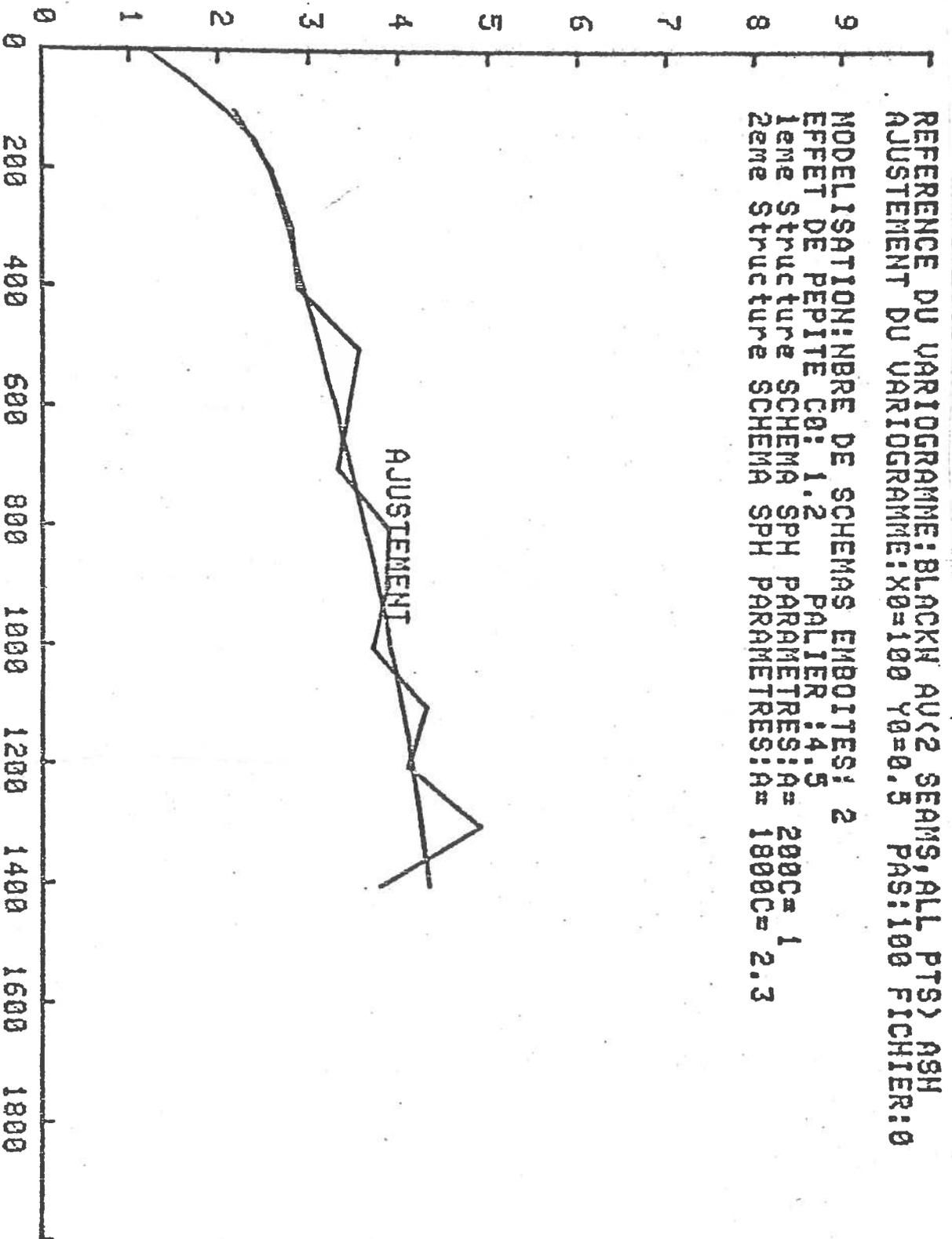


FIGURE 8b: Variogramme ajusté de la puissance

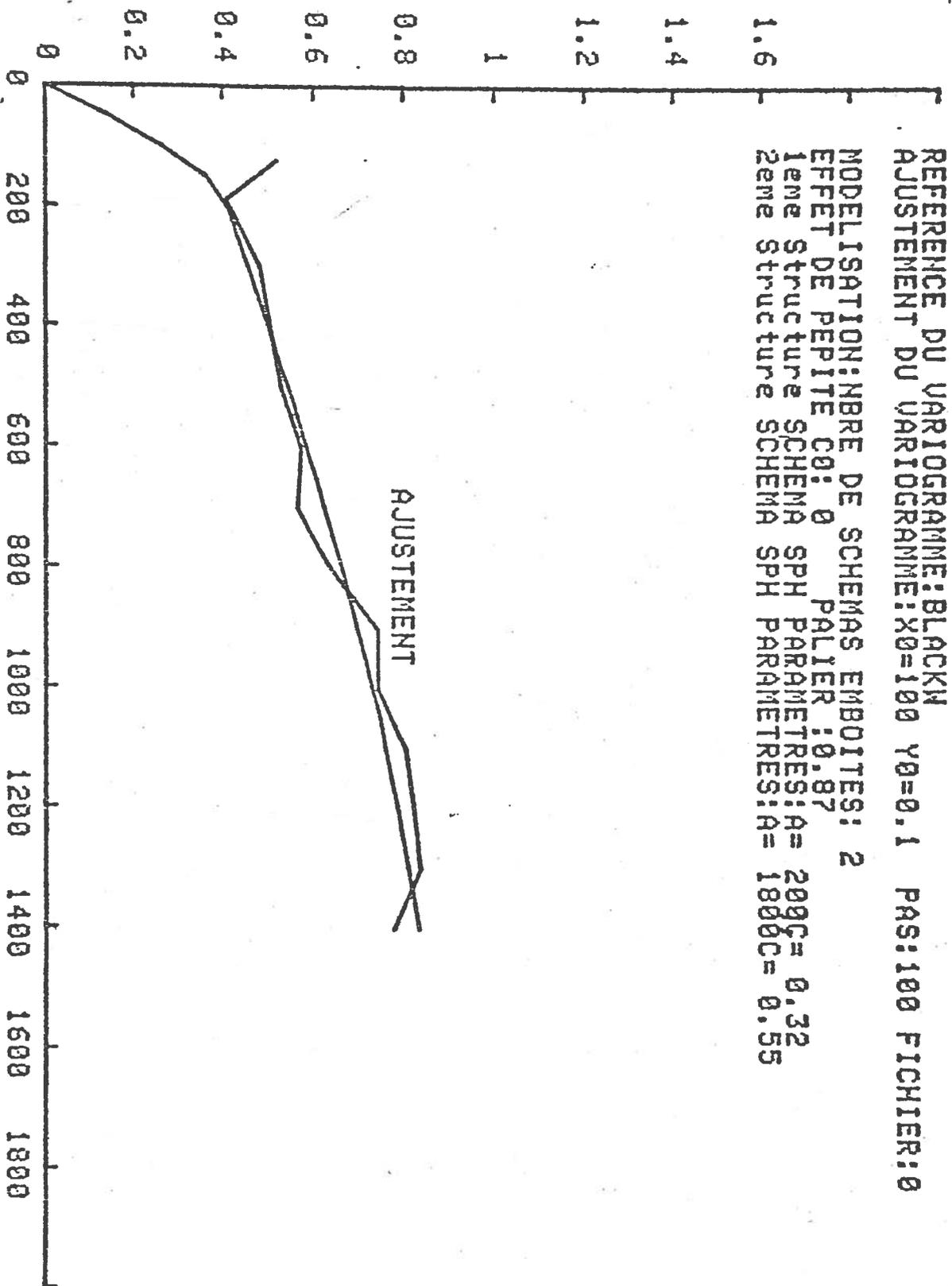


FIGURE 8c : Variogramme ajusté de la teneur en soufre

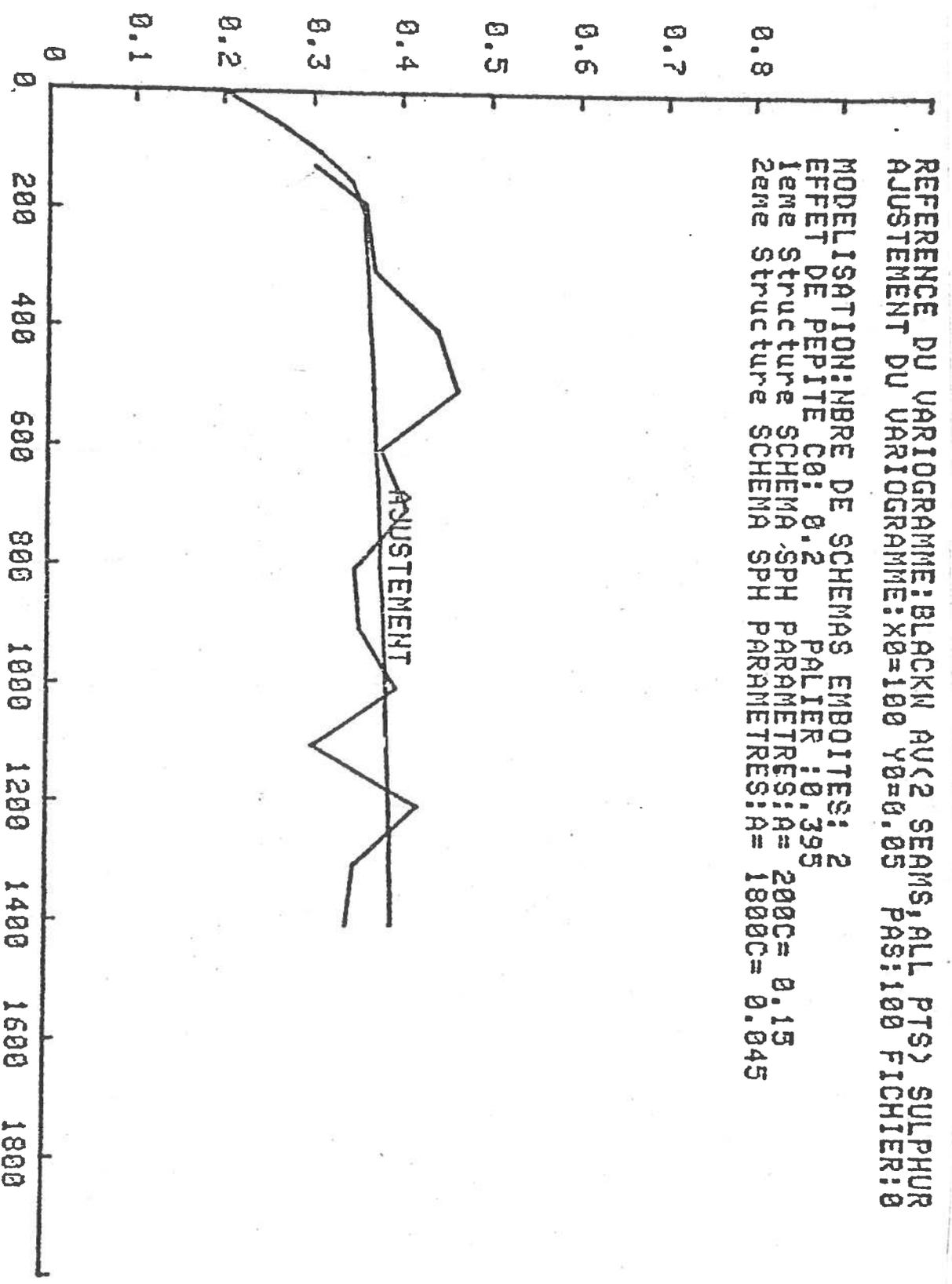
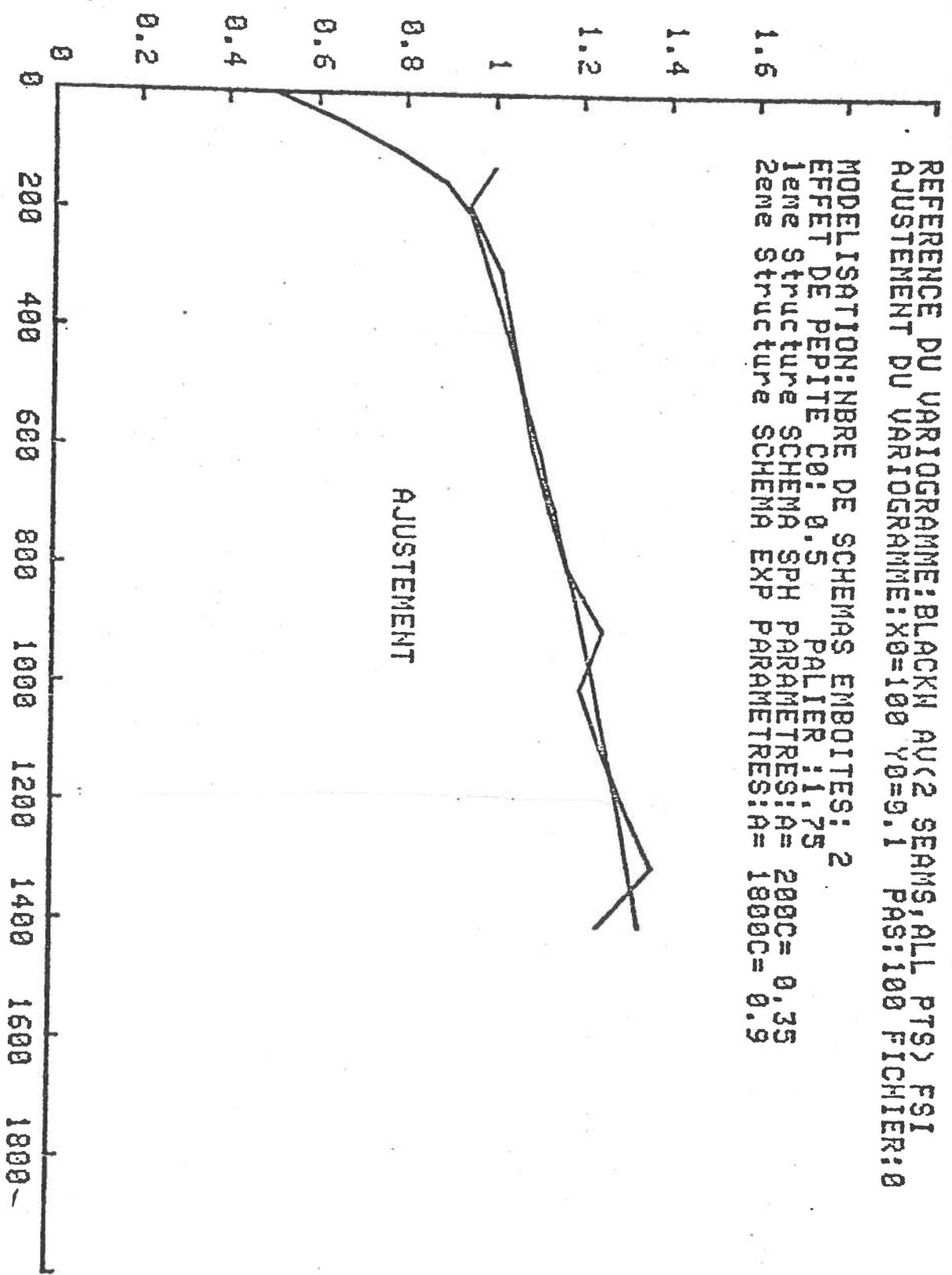


FIGURE 8d : Variogramme ajusté du F.S.I.



que la teneur en cendre, il est raisonnable de supposer que les portées de tous les variogrammes devront être les mêmes, bien que les paliers et les effets de pépité soient changés. Les variogrammes moyens et les modèles ajustés sont présentés aux figures 8 a, b, c, d.

Les variogrammes de la puissance et des teneurs en cendre et en soufre sont les mêmes pour les deux couches, mais ceux du F.S.I. ont des effets de pépité différents, qu'il faut particulariser.

Ils doivent être cohérents avec les variogrammes expérimentaux et avoir une moyenne pondérée de 0,5 (d'après l'approximation de linéarité du F.S.I.). Après quelques essais, on a trouvé que des valeurs de 0,64 pour la couche médiane et de 0,43 pour la couche inférieure remplissent ces conditions. Les modèles ajustés sont donc les suivants :

Couche médiane :

	Effet de pépité	1ère portée	1er palier	2ème portée	2ème palier
Puissance	0	200	0,32	1 800	0,55
Teneur en cendre	1,2	200	1,0	1 800	2,3
Teneur en soufre	0,2	200	0,15	1 800	0,45
F.S.I.	0,64	200	0,35	1 800	0,9

Couche inférieure :

	Effet de pépité	1ère portée	1er palier	2ème portée	2ème palier
Puissance	0	200	0,32	1 800	0,55
Teneur en cendre	1,2	200	1,0	1 800	2,3
Teneur en soufre	0,2	200	0,15	1 800	0,45
F.S.I.	0,43	200	0,35	1 800	0,9

TABLEAU 3 : Coefficients des modèles sphériques.

2-3 Calcul de la maille optimale d'échantillonnage : un exemple de l'utilisation du variogramme.

2-3-1 Définition du problème.

Les modèles des variogrammes peuvent être utilisés dans des applications diverses : estimation des ressources totales, krigeage de blocs, calcul de la maille optimale d'échantillonnage, etc... Dans notre cas, la société minière qui avait fourni les données souhaitait déterminer la maille optimale. Par conséquent, cette étude est présentée ici comme exemple de l'utilisation des modèles de variogramme. De même que l'analyse structurale et l'ajustement des variogrammes, cette application suit la méthodologie classique de la géostatistique linéaire.

Il faut d'abord définir le critère d'optimalité. La société a constaté que, pour ses estimations des réserves, il ne fallait pas que l'erreur dans l'estimation des réserves totales dépasse 10 %. La société doit aussi remettre une estimation des réserves au gouvernement de l'état de Queensland qui demande que "le tonnage et la teneur soient inscrits dans une fourchette de ± 20 % des vraies valeurs". De plus, le ministère des mines impose une distance maximale de 1/2 mile (800 m) entre les sondages. Il est permis de penser que le ministère accepterait un écartement plus grand si on pouvait montrer que les estimateurs du tonnage et de la teneur sont assez précis.

On peut imaginer que, pour un modèle de variogramme donné, la maille optimale est déterminée; mais cela n'est pas vrai. En fait, la maille optimale dépend directement du nombre total de sondages et seulement indirectement de la distances entre les sondages. Par conséquent, la maille optimale pour un domaine ne le serait pas pour un autre. Elle fournirait des estimateurs n'ayant pas la précision requise sur un domaine plus petit, et sur un domaine plus grand elle entraînerait plus de sondages qu'il n'en faut.

Un autre facteur important pour déterminer la maille optimale est le niveau de connaissance de la minéralisation. Il est évident que si on connaît a priori les limites du gisement, il faut une stratégie différente de celle que l'on utiliserait si, en plus, il fallait déterminer des limites pendant la campagne de sondages. Les questions posées par la compagnie étant de nature très générale, nous abordons le problème en étudiant trois cas à

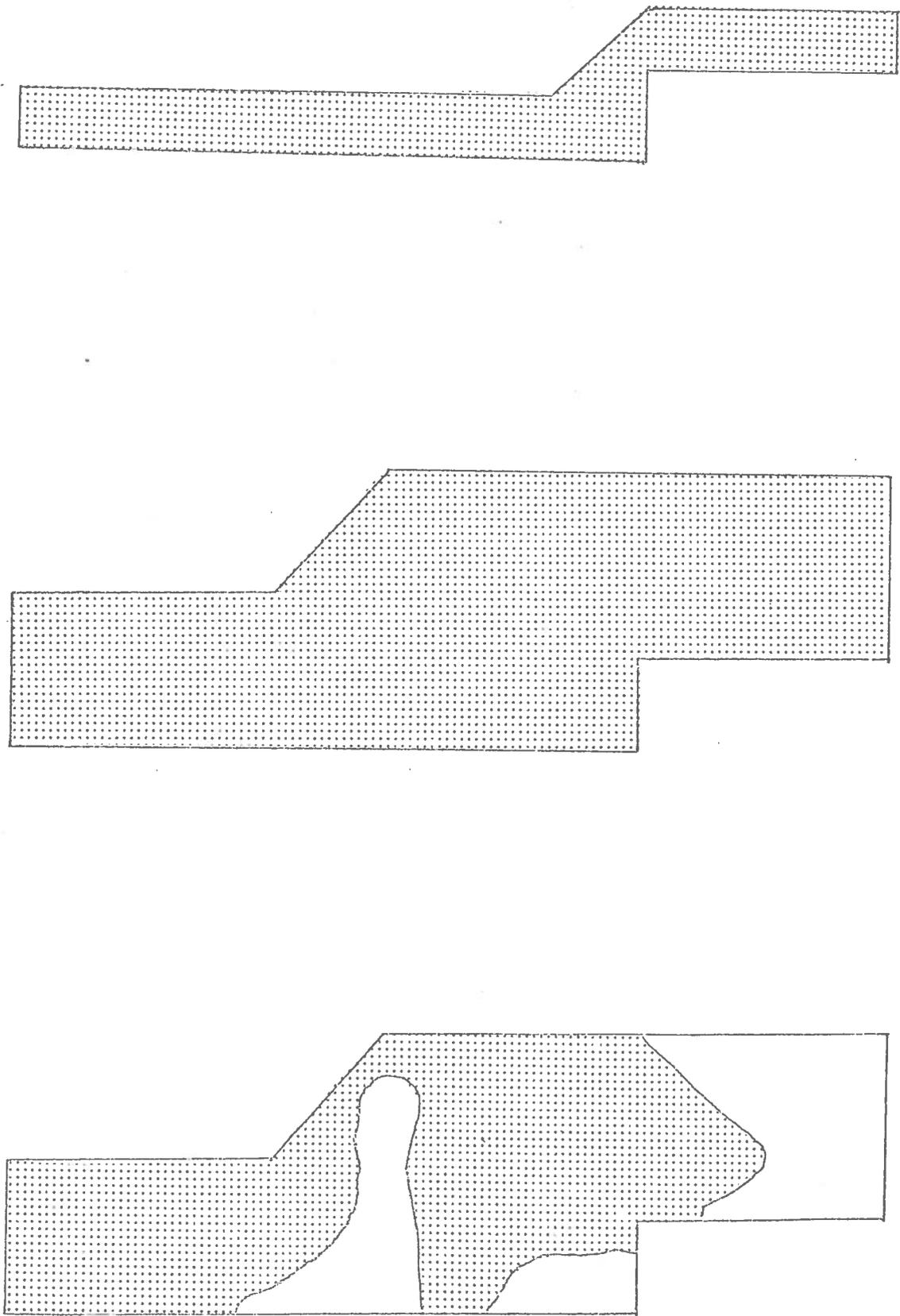
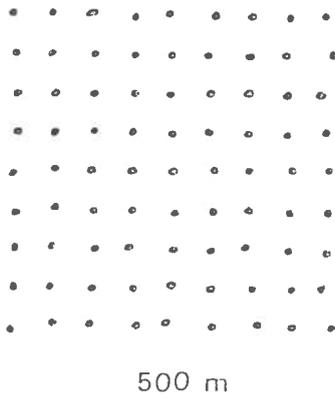
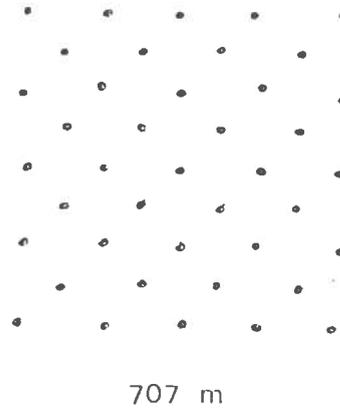
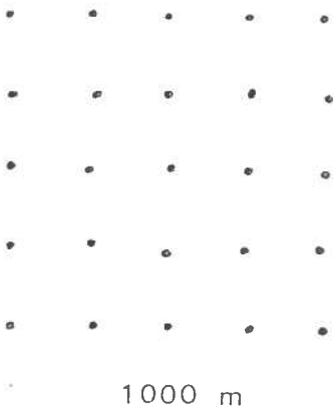
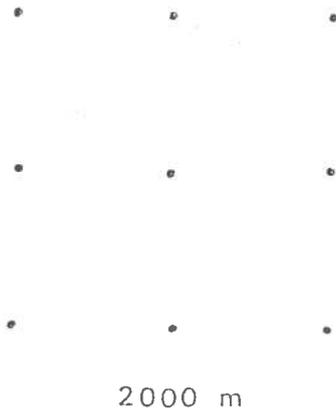


FIGURE 9 : Les trois régions indiquant les limites de la minéralisation

FIGURE 10: Les cinq mailles considérées



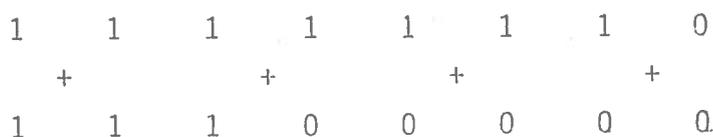
à titre de démonstration. Ce sont :

- (i) un petit domaine que l'on suppose minéralisé partout,
- (ii) un domaine plus grand supposé minéralisé partout, et
- (iii) le même grand domaine, en supposant que les limites de la minéralisation sont à déterminer.

Ces domaines sont illustrés aux figures 9 a, b, c. Dans chacun des cas, les variances des réserves totales ont été estimées pour chacune des quatre variables (la puissance, les teneurs en cendre et en soufre et le F.S.I.). En regardant les variances estimées pour les mailles différentes, on peut apprécier l'influence de la distance entre les sondages sur la précision des estimateurs des réserves, au moins pour ces trois domaines. Une comparaison entre les résultats du premier et du deuxième cas donne une idée de l'effet de la superficie totale. Finalement, une comparaison entre le deuxième cas et le troisième montre de combien augmente la variance quand les limites de la minéralisation ne sont pas encore connues.

Puisque la maille maximale approuvée par le gouvernement est à l'heure actuelle de 800 m, on a étudié une gamme de mailles variant de 2 000 m jusqu'à 500 m. Si on divise la maille en deux à chaque étape de la campagne de sondages, le nombre de sondages quadruple. Mais si on remplit la maille en quinconce à chaque étape, on diminue la maille d'un facteur de $\sqrt{2}$ et on double le nombre de sondages. Comme cette approche semble plus intéressante et plus réaliste, c'est elle qui a été choisie. Les mailles étudiées sont donc de 2 000m, 1 414m, 1 000m, 707m et 500m (voir figure 10).

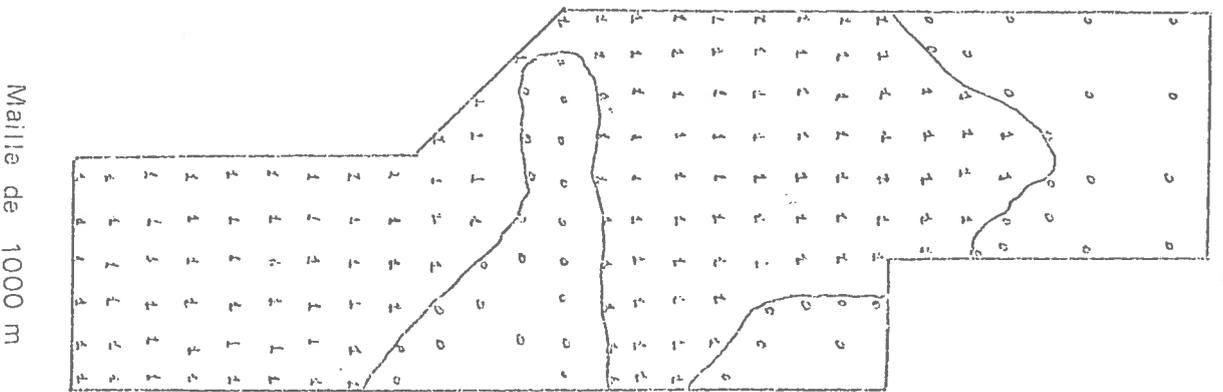
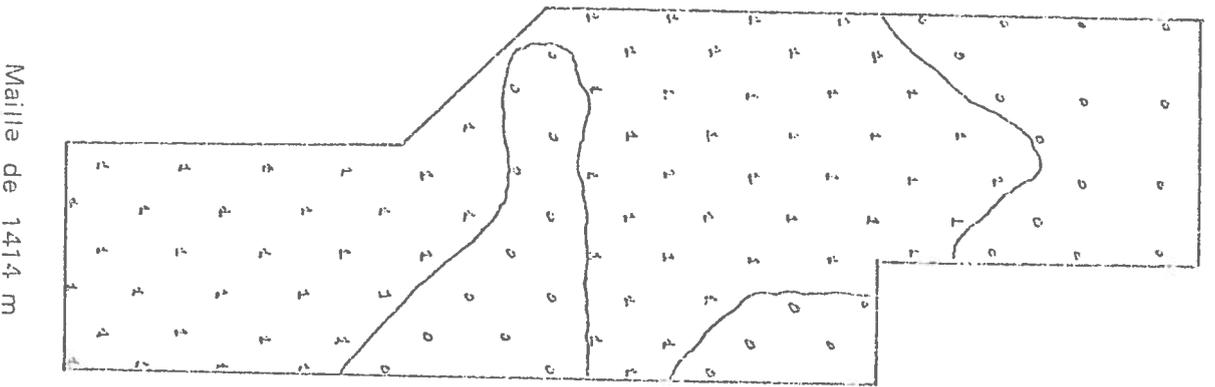
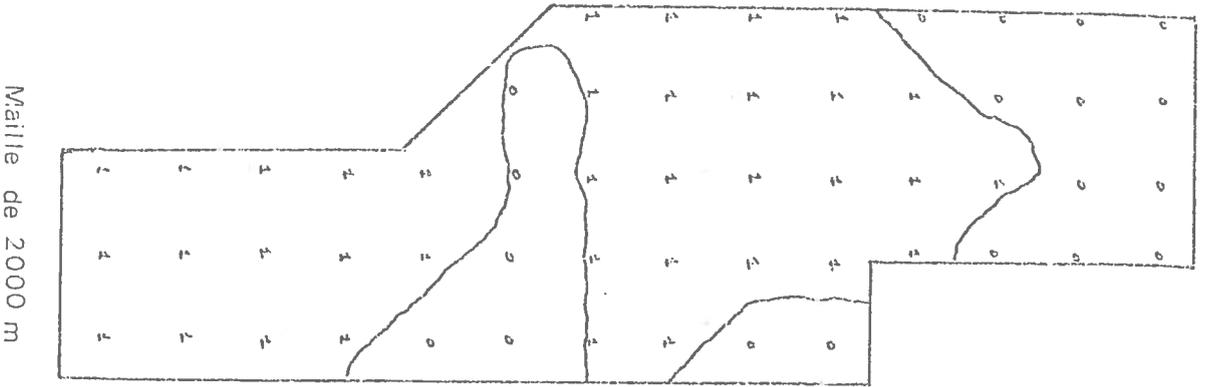
Quand les limites de la minéralisation ne sont pas connues a priori, il vaut mieux commencer par une grande maille et remplir seulement les sections qui sont minéralisées ou qui peuvent l'être. C'est-à-dire qu'on met un sondage supplémentaire si au moins un des quatre sondages voisins a donné un résultat positif à l'étape précédente. Le dessin suivant peut servir à illustrer cette idée.



1 : résultat positif à l'étape précédente
0 : résultat négatif à l'étape précédente
+ : un sondage supplémentaire

Figure 11 : La stratégie de la campagne de sondage.

FIGURE 12 : Dispositions des sondages



Cette stratégie d'exploration fournit des informations sur la présence ou l'absence de minerai, et, dans les régions minéralisées, sur les teneurs. Elle permet également de définir la frontière entre le minerai et le stérile sans gaspillage de sondages.

L'emplacement des sondages à une maille de 2 000 m et leurs résultats (positifs ou négatifs) sont présentés à la figure 12a. De la même manière, les figures 12b et c montrent les résultats des deuxième et troisième étapes de la campagne (c'est-à-dire les mailles de 1 414 m et de 1 000 m).

2-3-2 Deux méthodes pour estimer les réserves totales.

Deux méthodes de calcul des réserves totales ont été étudiées. Ce sont :

1. Les polygones d'influence : la valeur du sondage central est attribuée au bloc.
2. Le krigeage des blocs : les réserves sont estimées pour chaque bloc par krigeage. A titre d'exemple, on considère un krigeage en utilisant le sondage central du bloc et ceux des 8 blocs voisins. Les réserves sont obtenues par sommation des estimateurs krigés.

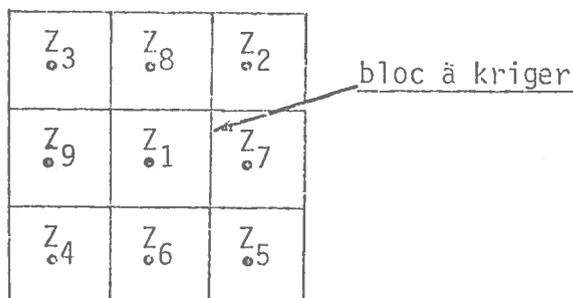


Figure 13 : Le système de krigeage.

Le désavantage de la deuxième méthode est qu'elle demande plus d'effort et de temps de calcul tout en donnant un résultat peu différent de celui fourni par la première. On peut le démontrer en considérant un bloc qui n'est pas au bord du domaine à estimer (c'est-à-dire un bloc avec l'ensemble de ses 8 blocs voisins).

Dans le cas présent, la régionalisation est isotrope (c'est-à-dire que les variogrammes sont les mêmes dans toutes les directions). Dans ces conditions, un même facteur de pondération s'applique à Z_2 , Z_3 , Z_4 et Z_5 ,

ainsi qu'à Z_6 , Z_7 , Z_8 et Z_9 . L'estimateur krigé Z_K du bloc est donc la somme pondérée de ces 9 valeurs :

$$Z_K = \lambda_1 Z_1 + \lambda_2 (Z_2 + Z_3 + Z_4 + Z_5) + \lambda_3 (Z_6 + Z_7 + Z_8 + Z_9) \quad (1)$$

où

$$\lambda_1 + 4 \lambda_2 + 4 \lambda_3 = 1 \quad (2)$$

Les facteurs de pondération (λ_1 , λ_2 , λ_3) résultent des équations de krigeage. Puisque l'estimateur krigé d'un bloc est la somme pondérée de la valeur du sondage dans ce bloc et de ceux dans ses 8 voisins, la valeur du sondage dans un bloc apparaît dans les estimateurs de ces 8 voisins. La contribution de la valeur Z_1 aux estimateurs de ses voisins est donc :

	Contribution de la valeur Z_1
Bloc central	$\lambda_1 Z_1$
Bloc autour de Z_2	$\lambda_2 Z_1$
" " " Z_3	$\lambda_2 Z_1$
" " " Z_4	$\lambda_2 Z_1$
" " " Z_5	$\lambda_2 Z_1$
" " " Z_6	$\lambda_3 Z_1$
" " " Z_7	$\lambda_3 Z_1$
" " " Z_8	$\lambda_3 Z_1$
" " " Z_9	$\lambda_3 Z_1$
Contribution totale	$(\lambda_1 + 4 \lambda_2 + 4 \lambda_3) Z_1 = 1 \cdot Z_1$

TABLEAU 4 : Contribution de la valeur Z_1 aux estimateurs krigés.

Donc on voit que, pour les blocs internes, chaque valeur aura un poids de 1 dans la somme des réserves totales. Par conséquent, la complication introduite par le krigeage ne se justifie pas si on ne cherche

que l'estimateur des réserves totales. Bien entendu, les valeurs krigées des blocs sont un estimateur beaucoup plus précis que celles obtenues par la méthode des polygones d'influence.

2-3-3 La variance de l'estimateur des réserves totales.

La variance d'estimation σ_{EST}^2 d'un bloc est la variance de la différence entre la valeur vraie (mais inconnue) du bloc et celle de son estimateur.

$$\begin{aligned}\sigma_{EST}^2 &= E (Z_B - Z_B^*)^2 \\ &= 2 \bar{\gamma} (B,C) - \bar{\gamma} (B,B) - \bar{\gamma} (C,C)\end{aligned}\quad (3)$$

où B est le bloc,

C est le sondage central,

$\bar{\gamma} (B,C)$ est la valeur moyenne du variogramme entre un point arbitraire dans le bloc et le sondage central,

$\bar{\gamma} (B,B)$ est la valeur moyenne du variogramme entre deux points arbitraires du bloc,

$\bar{\gamma} (C,C)$ est la valeur moyenne du variogramme entre le sondage central et lui-même, c'est-à-dire 0.

La variance d'estimation peut être calculée par intégration numérique mais cela n'est pas nécessaire pour les modèles sphériques parce qu'il existe des abaques. (Voir Abaque n° 7, P. 131 de "Mining Geostatistics" par A. Journel et Ch. Huijbregts). Les variances d'estimation correspondant à toutes les variables et à toutes les mailles ont été calculées en utilisant ces abaques (Tableau 5).

Variance d'estimation d'un bloc				
Maille	Puissance	Teneur en cendre	Teneur en soufre	F.S.I.
2 000 m	0,576	2,072	0,170	0,77
1 414 m	0,492	1,724	0,162	0,634
1 000 m	0,427	1,456	0,135	0,53
707 m	0,385	1,288	0,150	0,465
500 m	0,343	1,133	0,137	0,380

TABLEAU 5: Variances d'estimation

Ayant calculé les variances d'estimation d'un bloc, il faut les sommer pour obtenir la variance des réserves totales. Si les estimateurs des réserves des blocs sont indépendants, on peut tout simplement additionner les variances; sinon, il faut tenir compte des covariances entre les blocs. Dans cette étude, on a calculé ces covariances: au pire, la vraie variance est sous-estimée d'environ 14 %, ce qui correspond à une erreur de 7 % sur l'écart-type. Il ne semble donc pas que le calcul supplémentaire des covariances se justifie dans la pratique. La variance d'estimation a donc été évaluée sans en tenir compte; c'est-à-dire par sommation des variances d'estimation de tous les blocs. Après avoir divisé par le nombre de blocs, on obtient la variance de la valeur moyenne de chaque variable (Tableau 6).

Petit domaine :

Maille	Puissance	Teneur en cendre	Teneur en soufre	F.S.I.
2 000 m	0,0457	0,1625	0,0133	0,0604
1 414 m	0,0205	0,0718	0,0068	0,0264
1 000 m	0,0093	0,0318	0,0034	0,0116
707 m	0,0036	0,0122	0,00142	0,0044
500 m	0,0016	0,0053	0,00064	0,0018

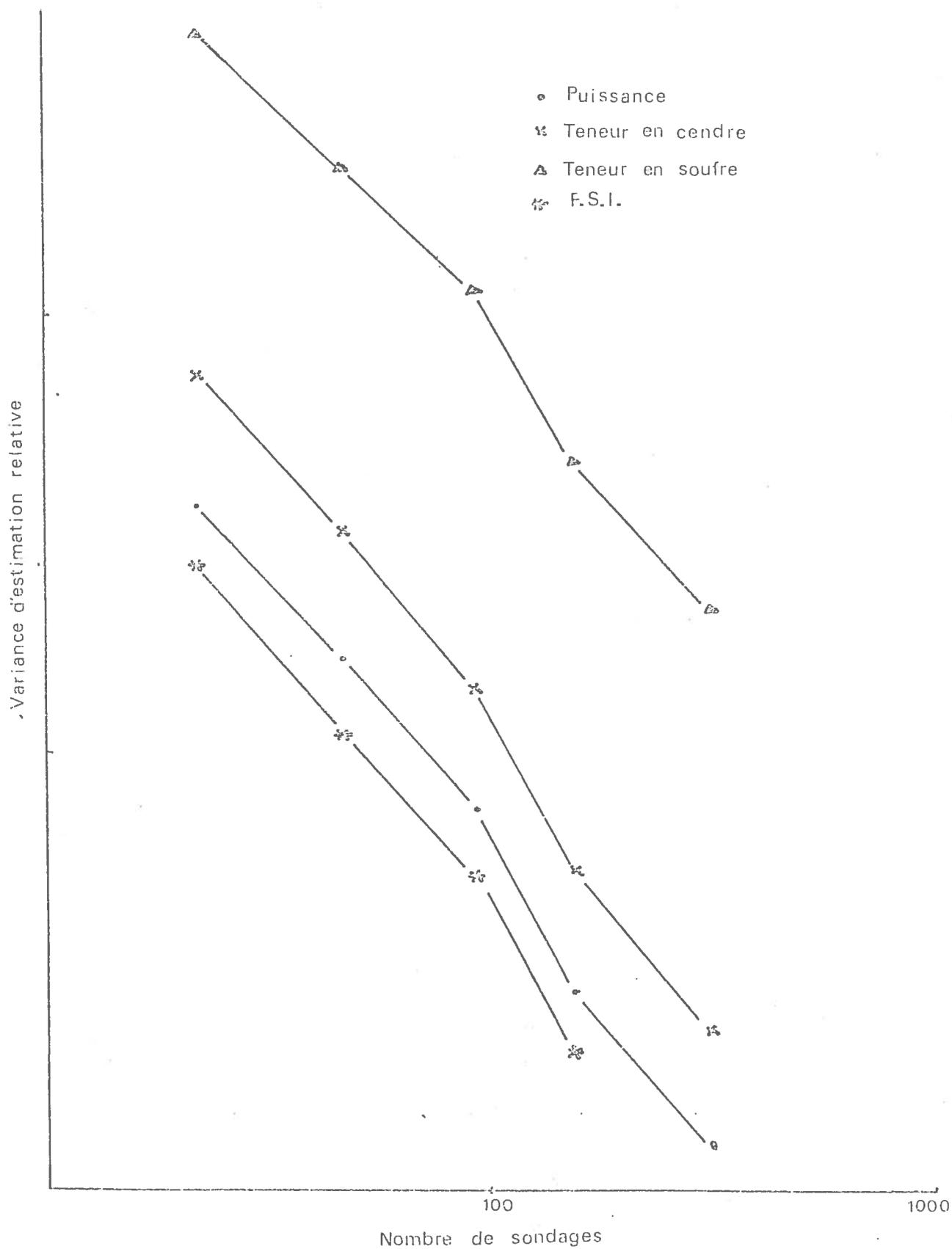
Grand domaine (supposé entièrement minéralisé) :

Maille	Puissance	Teneur en cendre	Teneur en soufre	F.S.I.
2 000 m	0,0142	0,0512	0,00419	0,0196
1 414 m	0,0055	0,0193	0,00182	0,0071
1 000 m	0,0024	0,0082	0,00087	0,00298
707 m	0,00106	0,00356	0,00041	0,00128
500 m	0,00047	0,00156	0,00019	0,00052

TABLEAU 6 : Variances des moyennes des variables.

La relation sous-jacente entre la variance et le nombre total de sondages devient évidente quand elle est représentée sur papier bi-logarithmique (Figure 14). Pour faciliter la comparaison entre les 4 variables, on a tracé les variances relatives (c'est-à-dire les variances divisées par

FIGURE 14 Relation entre la variance et le nombre de sondages



les carrés des moyennes). Cela confirme le commentaire fait précédemment que la précision d'estimation dépend du nombre total de sondages.

2-3-4 Variance d'estimation quand les limites de la minéralisation ne sont pas connues a priori.

Dans la section précédente, on a supposé que la couche de charbon existe partout dans le domaine à estimer. Dans ce cas, on fait une première erreur en attribuant au bloc la valeur du sondage central. Quand il faut estimer de plus les limites de la minéralisation, on introduit une deuxième source d'erreur liée au manque de précision sur les limites. Cette erreur, qui s'appelle l'erreur géométrique, peut être estimée par la méthodologie de l'approche transitive de G. MATHIERON. Une description de cette approche se trouve pp. 428-443 de "Mining Geostatistics" par A. JOURNEL et Ch. HUIJBREGTS. Les résultats sont les suivants :

Pour les problèmes à deux dimensions d'estimation de surface à partir d'une maille régulière rectangulaire (a_1, a_2), on démontre la formule au premier ordre suivante qui donne la variance relative d'estimation de surface :

$$\frac{\sigma_s^2}{s^2} = \frac{1}{n^2} \left[\frac{1}{6} N_2 + 0,06 \frac{(N_1)^2}{N_2} \right] + \dots \quad (4)$$

où $n > 10$ est le nombre de sondages positifs. La surface estimée est donc $n a_1 a_2$. Les variables $2 N_1, 2 N_2$ ($N_1 < N_2$) sont les nombres d'éléments parallèles aux 2 côtés de la maille (a_1, a_2) que constituent le périmètre et la réunion des n zones d'influence positives. Le périmètre des lacunes éventuelles de minéralisation doit intervenir dans le décompte des valeurs $2 N_1$ et $2 N_2$.

Dans le cas actuel, un deuxième problème se présente parce que le domaine minéralisé est borné par endroits par les limites de la concession qui sont connues sans erreur. On voudrait en tenir compte mais on ne peut pas étendre l'approche transitive de G. Matheron à ce cas. Par conséquent, il faut proposer une correction ad hoc de cette formule. On peut admettre que si la limite de la concession constitue p % du périmètre du domaine minéralisé, l'erreur géométrique doit être réduite dans la même proportion. Ce facteur de correction a été appliqué à l'erreur géométrique. Après avoir

ajouté l'erreur géométrique à l'erreur à domaine connu, on retrouve les variances suivantes :

Maille	Puissance	Teneur en cendre	Teneur en soufre	F.S.I.
2 000 m	0,0255	0,08779	0,005859	0,0352
1 414 m	0,00888	0,03069	0,002682	0,0116
1 000 m	0,00364	0,01238	0,001296	0,00454
707 m	0,00170	0,00569	0,000653	0,00208
500 m	0,00103	0,003416	0,000407	0,00115

TABLERAU 7 : Variances d'estimation des moyennes des variables quand les limites de la minéralisation ne sont pas connues a priori.

Les variances présentent le même comportement linéaire qu'on a vu dans la section 2-3-3.

2-3-5 La maille optimale.

Au commencement de l'étude, le but était de trouver la maille optimale, c'est-à-dire celle pour laquelle la variance d'estimation atteint le niveau de précision demandé. Rappelons que pour le compte-rendu des réserves au gouvernement, le double écart-type relatif maximal permis est de 20 %, tandis que pour le rapport interne, il est de 10 %. On peut exprimer ces niveaux de précision dans les termes statistiques suivants :

" A un niveau de confiance de 95 %, la différence entre les valeurs estimées et les vraies valeurs des réserves totales et des teneurs moyennes ne doit pas dépasser 10 % ou 20 % selon le cas."

Le choix du niveau de confiance est, bien entendu, arbitraire, mais 95 % est généralement accepté comme satisfaisant. Soit m la valeur estimée de la moyenne d'une variable et soit μ sa valeur vraie (mais inconnue). On veut que

$$\Pr (|m - \mu| < \frac{1}{10} m) = 0,95 \quad \text{pour la société} \quad (5)$$

$$\Pr (|m - \mu| < \frac{2}{10} m) = 0,95 \quad \text{pour le gouvernement} \quad (6)$$

PETIT DOMAINE

Maille	Puissance	Teneur en cendre	Teneur en soufre	F.S.I.
2 000 m	3,656	7,287	44,33	2,753
1 414 m	1,640	3,220	22,66	1,203
1 000 m	0,744	1,426	11,33	0,529
707 m	0,288	0,547	4,73	0,201
500 m	0,128	0,238	2,13	0,082

GRAND DOMAINE

Maille	Puissance	Teneur en cendre	Teneur en soufre	F.S.I.
2 000 m	1,136	2,296	13,97	0,893
1 414 m	0,440	0,865	6,07	0,324
1 000 m	0,192	0,368	3,90	0,136
707 m	0,085	0,150	1,84	0,058
500 m	0,038	0,070	0,85	0,024

GRAND DOMAINE (les limites de la minéralisation n'étant pas connues a priori)

Maille	Puissance	Teneur en cendre	Teneur en soufre	F.S.I.
2 000 m	2,040	3,937	19,53	1,604
1 414 m	0,710	1,376	8,94	0,529
1 000 m	0,291	0,555	4,32	0,207
707 m	0,136	0,255	2,18	0,095
500 m	0,082	0,153	1,36	0,052

TABEAU 8 : Variances relatives (multipliées par 1 000).

Pour résoudre ces équations, il faut connaître la distribution des erreurs. Dans le cas de la plupart des minerais, les variables elles-mêmes ne sont pas gaussiennes, mais la distribution des erreurs est, en général, presque symétrique. Pour faciliter les calculs, on supposera que la distribution des erreurs est gaussienne bien que ce ne soit pas rigoureusement juste. On sait alors que l'intervalle de confiance correspondant à un niveau de confiance de 95 % est large de 2 écarts-type de part et d'autre de la moyenne. Par conséquent, les équations (5) et (6) deviennent :

$$\frac{1}{10} m = 2 \sigma \quad (7)$$

$$\frac{2}{10} m = 2 \sigma \quad (8)$$

où σ^2 est la variance d'estimation. C'est-à-dire qu'on demande une variance relative d'estimation (σ^2/m^2) de 0,0025 pour la société et de 0,01 pour le gouvernement.

Les variances relatives (multipliées par un facteur de 1 000) sont présentées dans le tableau 8. Après la multiplication par 1 000, les niveaux critiques deviennent 10 et 2,5. D'après ce tableau, on voit que pour le petit domaine une maille de 2 000 m donnerait la précision demandée par le gouvernement pour la puissance, la teneur en cendre et le F.S.I., mais qu'il faudrait une maille de moins de 1 000 m pour la teneur en soufre. Pour la société elle-même, une maille de 1 000 m suffira pour trois des variables, mais il faudrait une maille de 500 m pour la teneur en soufre.

Les conclusions sont semblables pour le grand domaine. Une maille de 2 000 m donne le niveau de précision voulu sauf pour la teneur en soufre pour laquelle il faudrait une maille de 1 414 m pour satisfaire le gouvernement et de 707 m pour la société. Dans le cas où les limites de la minéralisation sont comme sur la figure 9c; bien qu'elles ne soient pas connues a priori, les mailles optimales seraient les mêmes que précédemment sauf pour la teneur en cendre. Dans ce dernier cas, on aurait besoin d'une maille de 1 414 m pour atteindre la limite de précision voulue par la société.

Comme la précision de l'estimation dépend de deux facteurs principaux: nombre total de sondages et superficie du domaine à estimer, on ne peut pas

donner de règle générale pour la maille optimale. On a vu qu'elle varie énormément d'une variable à l'autre. Les résultats de cette étude suggèrent que le niveau de précision voulu par le ministère des mines peut être atteint en utilisant une maille plus large que celle stipulée à l'heure actuelle.

2-4 Développement d'un variogramme plus robuste.

N.B. : Ce travail a été fait en collaboration avec P. DELFINER.

Au cours du calcul des variogrammes, on a rencontré quelques difficultés avec le variogramme de la teneur en soufre de la couche médiane. Avant l'élimination des deux valeurs très élevées, le variogramme fluctuait beaucoup et était quatre ou cinq fois plus grand que celui de l'autre couche; mais après élimination de ces deux points, il est retombé au même niveau que l'autre. Dans ce cas, la différence entre les variogrammes des deux couches nous suggère que le variogramme n'est pas un bon estimateur de la variabilité régionale. Mais, comme dans la plupart des cas on ne dispose pas d'autre variogramme comme élément de comparaison, on a donc besoin d'un deuxième estimateur du variogramme pour tester l'estimateur classique.

2-4-1 Variogramme quantile.

L'estimateur classique du variogramme est la moyenne des carrés des écarts entre des paires de valeurs distantes de h l'une de l'autre. Comme la médiane est un estimateur de la tendance centrale beaucoup plus robuste que la moyenne, il est intéressant de considérer un estimateur obtenu en joignant les médianes des carrés dans chaque classe de distances au lieu des moyennes. Pour ce faire, il faut former l'histogramme des carrés des écarts pour les couples de valeurs séparées d'une distance h . Une représentation schématique de cette idée est présentée Figure 15. La forme des histogrammes est purement illustrative :

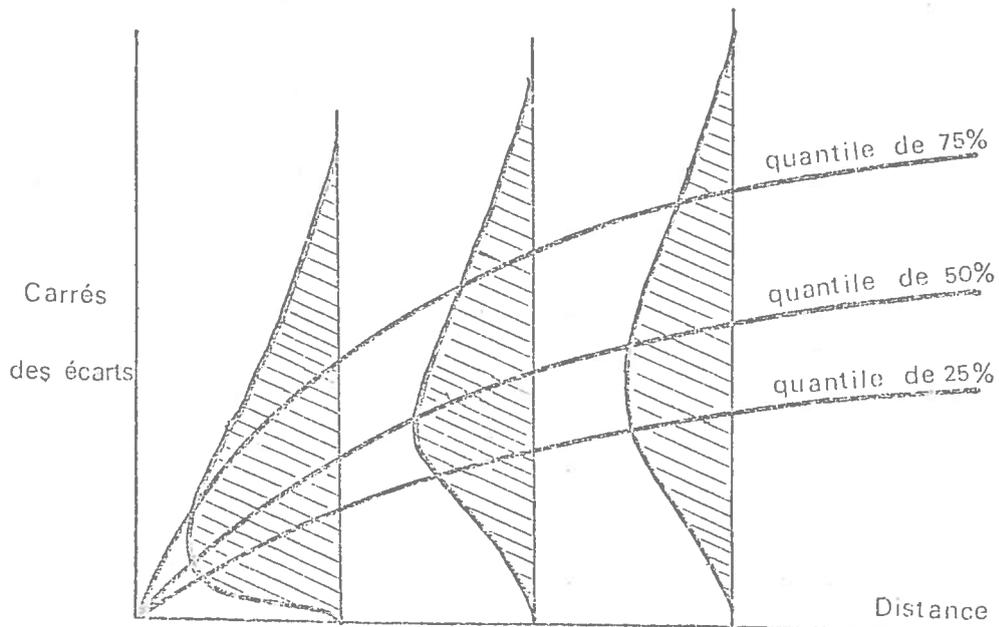


Figure 15 : Variogrammes quantiles.

Comme le variogramme "médian" est formé en joignant les médianes des histogrammes pour chaque classe de distances, on le définit par :

$$2 \gamma_{50}(h) = Q_{50} [Z_{X_{i+h}} - Z_{X_i}]^2 \quad (9)$$

où $Q_{50} [\cdot]$ est le quantile de 50 % des écarts. Le facteur 2 reste devant le γ_{50} pour raisonner en termes d'écart au lieu d'écart divisés par $\sqrt{2}$. De la même façon, on peut définir d'autres variogrammes quantiles. Par exemple, le variogramme des quantiles de 25 % et de 75 % définis par :

$$2 \gamma_{25}(h) = Q_{25} [Z_{X_{i+h}} - Z_{X_i}]^2 \quad (10)$$

$$2 \gamma_{75}(h) = Q_{75} [Z_{X_{i+h}} - Z_{X_i}]^2 \quad (11)$$

donnent une idée de la dispersion autour de la médiane.

Ces trois nouveaux variogrammes ont été calculés pour les données de teneur en soufre de la couche médiane et ils sont présentés à la figure 16 avec le variogramme classique. Les variogrammes quantiles donnent une identification nette du comportement linéaire du variogramme sous-jacent obtenu en enlevant les valeurs extrêmes. En ce sens, ils fournissent un estimateur beaucoup plus robuste de la forme du variogramme sous-jacent, bien qu'il ne donnent pas d'estimation immédiate de sa valeur. Les variogrammes quantiles

FIGURE 16 : Variogrammes de la teneur en soufre

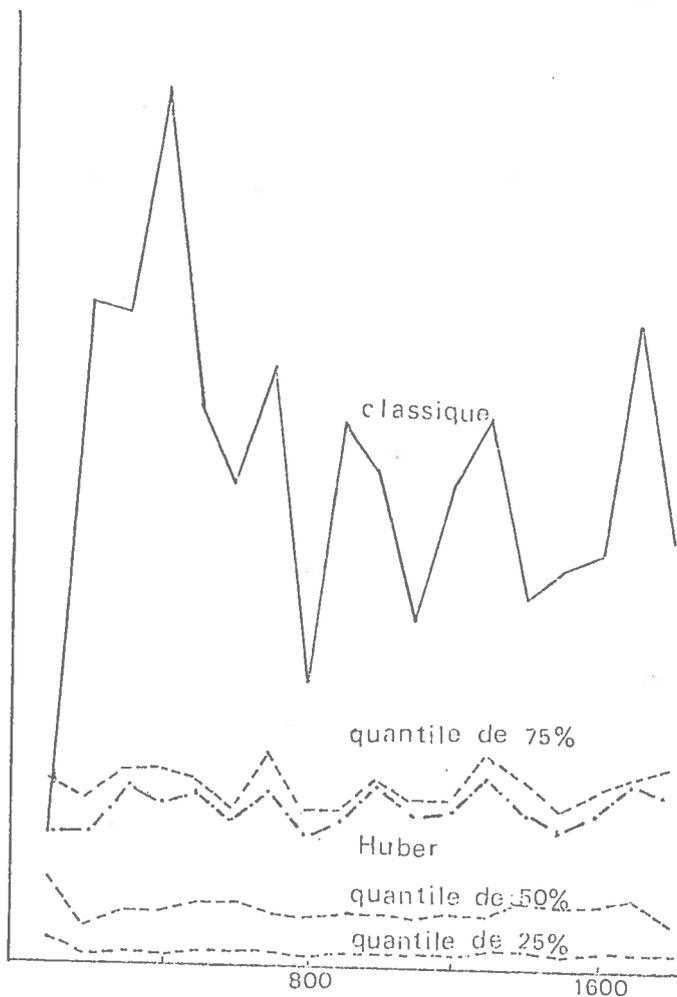
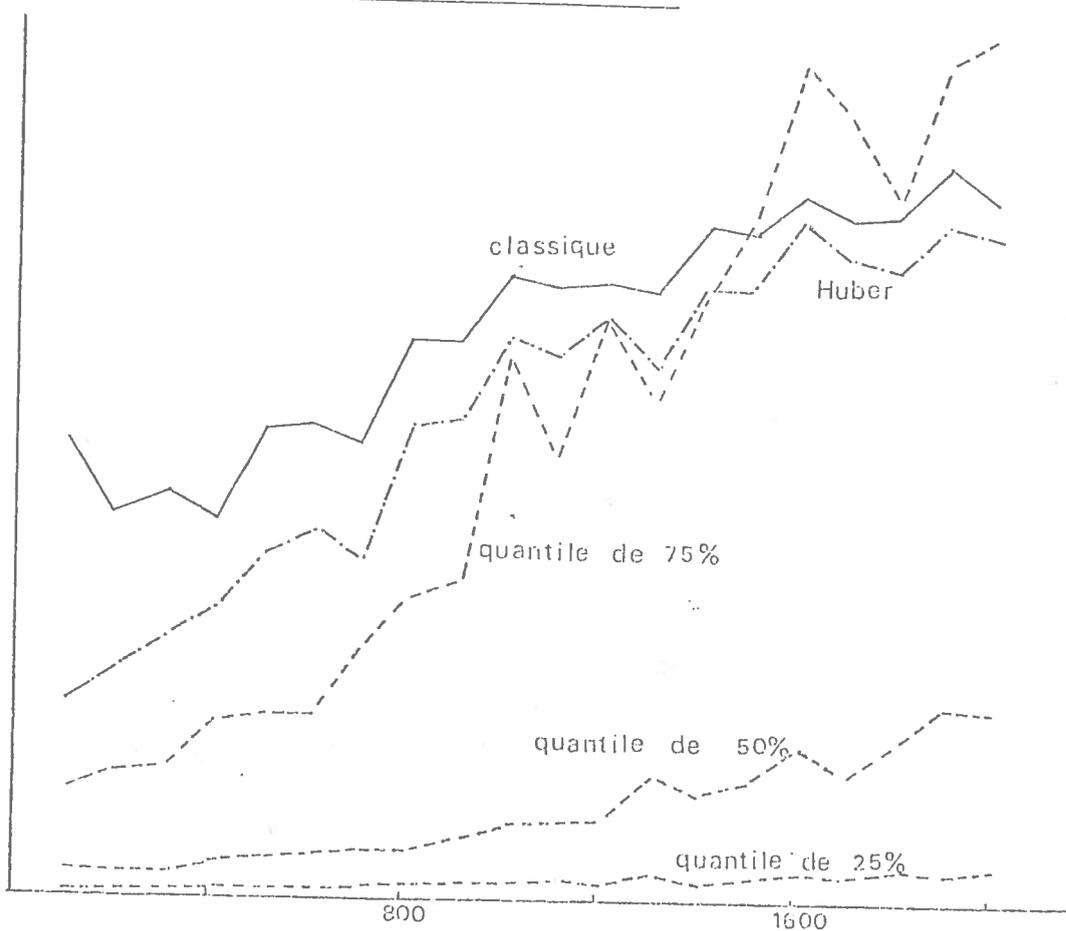


FIGURE 17 : Variogrammes de la puissance



ont été aussi calculés pour les trois autres variables (figures 17, 18 et 19). Comme dans le cas de la teneur en soufre, ils ont la même allure que le variogramme classique. Le variogramme des quantiles de 25 % et de 75 % révèlent la forte asymétrie de la distribution des carrés des écarts (ce qui est normal quand on étudie des carrés). Ce qui est étonnant est que le variogramme classique soit au dessus du niveau du variogramme du quantile de 75 % dans le cas de la puissance et de la teneur en cendre.

Le modèle de référence le plus simple par lequel on puisse juger le comportement des variogrammes des quantiles est le modèle gaussien. Si la variable Z_x a une distribution multivariante gaussienne de moyenne constante, les différences entre les points séparés d'une distance h sont gaussiennes de moyenne 0 et de variance $2 \gamma(h)$:

$$(Z_{x+h} - Z_x) \sim N(0, 2 \gamma(h)) \quad (12)$$

et les carrés des différences ont une distribution $2 \gamma(h) \times \chi_1^2$: (ou χ_1^2 est un χ^2 à un degré de liberté)

$$[Z_{x+h} - Z_x]^2 \sim 2 \gamma(h) \times \chi_1^2 \quad (13)$$

Une conséquence fort intéressante de (13) est que les histogrammes correspondant aux différentes valeurs de h diffèrent seulement d'un facteur d'échelle qui est précisément $2 \gamma(h)$. Autrement dit, tous les variogrammes quantiles ont la même forme, étant des transformées affines du variogramme classique. En fait, par sa définition, le variogramme du quantile de L % satisfait à :

$$\Pr \{ [Z_{x+h} - Z_x] < 2 \gamma_L(h) \} = L \quad (14)$$

D'après (13), cette relation est équivalente à :

$$\Pr \{ \chi_1^2 < \gamma_L(h) / \gamma(h) \} = L$$

c'est-à-dire :

$$\gamma_L(h) = \gamma(h) \times Q_L(\chi_1^2) \quad (15)$$

où $Q_L(\chi_1^2)$ est le quantile de $L\%$ de la distribution de χ_1^2 . En particulier, $Q_L(\chi_1^2) = 1$ quand $L = 58 \%$ de sorte que le variogramme classique coïncide

FIGURE 18 : Variogrammes du F.S.I.

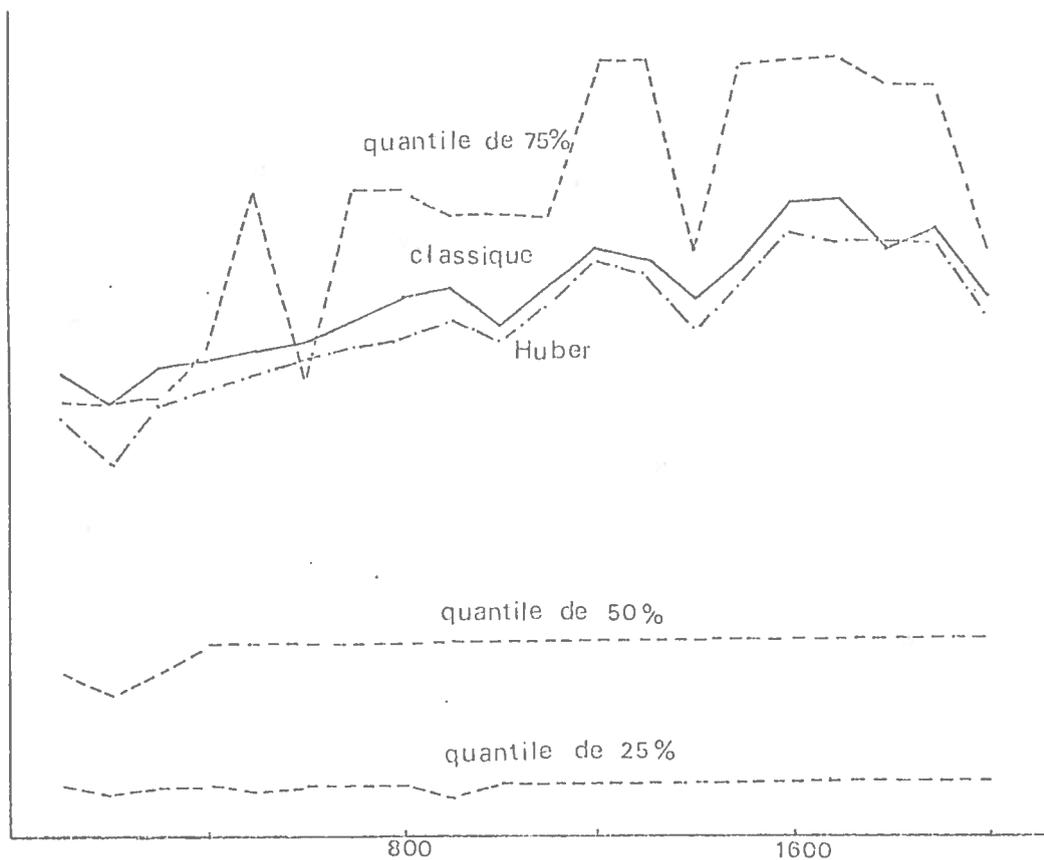
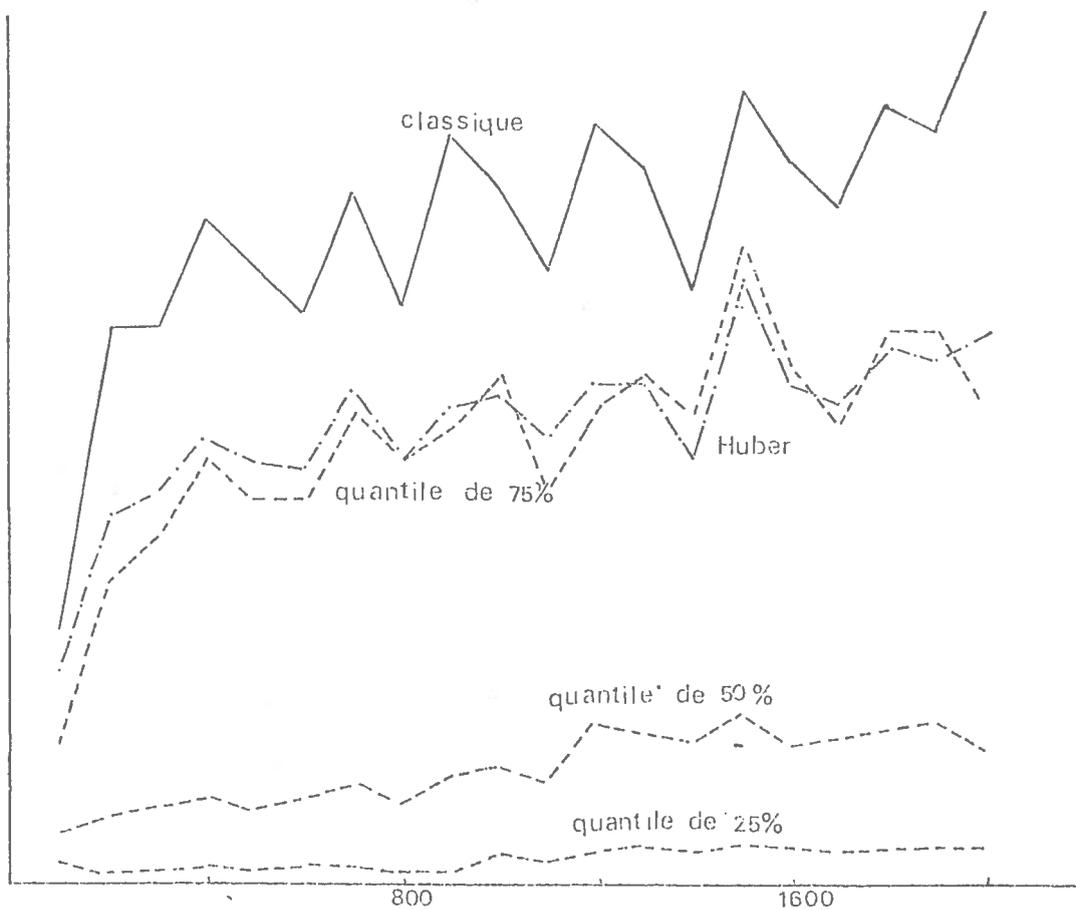


FIGURE 19 : Variogrammes de la teneur en cendre



avec le variogramme du quantile de 68 %. Il existe d'autres quantiles intéressants, tels :

$$\begin{aligned} Q_{25}(\chi_1^2) &= 0,10 \\ Q_{50}(\chi_1^2) &= 0,45 \\ Q_{75}(\chi_1^2) &= 1,32 \end{aligned} \quad (16)$$

(voir Handbook of Mathematical Functions, par ABRAMOWITZ et STEGUN, p. 984).

Dans le cas du modèle gaussien, on peut s'attendre à ce que le variogramme classique soit 10 fois plus grand que γ_{25} , 2 fois plus grand que γ_{50} et 25 % plus petit que γ_{75} . Les quantiles eux-mêmes doivent satisfaire approximativement à :

$$\gamma_{50} = 4,5 \gamma_{25} \quad \text{et} \quad \gamma_{75} = 3 \gamma_{50} \quad (17)$$

Les variogrammes du F.S.I., ainsi que ceux de la teneur en soufre (après élimination des deux valeurs extrêmes) satisfont ces relations assez bien. Mais elles sont violées dans le cas de la puissance et celui de la teneur en cendre dont les histogrammes des carrés sont évidemment plus asymétriques que le χ_1^2 . Non seulement le variogramme classique est beaucoup trop élevé, mais le variogramme du quantile de 75 % l'est également. Ce phénomène est dû à l'asymétrie de la variable Z_x , comme on le voit par exemple sur l'histogramme des teneurs en cendre (figure 20a). Les deux valeurs très élevées exceptées, la distribution des teneurs en soufre est plus symétrique (figure 20b).

Un ensemble de variogrammes quantiles constitue un bilan statistique de la distribution des carrés des différences correspondant aux différentes valeurs de h . Il est évidemment plus instructif que le variogramme classique considéré isolément. Cependant, pour le krigeage, il nous faut un estimateur du variogramme classique. Peut-on le déduire des variogrammes quantiles ? Cela sera certainement le cas si l'on peut accepter un modèle de la distribution des carrés des différences. Mais le seul modèle utilisable est celui dérivé des hypothèses gaussiennes. Cependant, on a déjà vu qu'elle n'est pas acceptable pour la puissance et la teneur en cendre. D'autant plus, dans le cas où cette hypothèse est réaliste, le variogramme classique est un bon estimateur après l'élimination des valeurs très élevées.

FIGURE 20a : Histogramme de la teneur en cendre

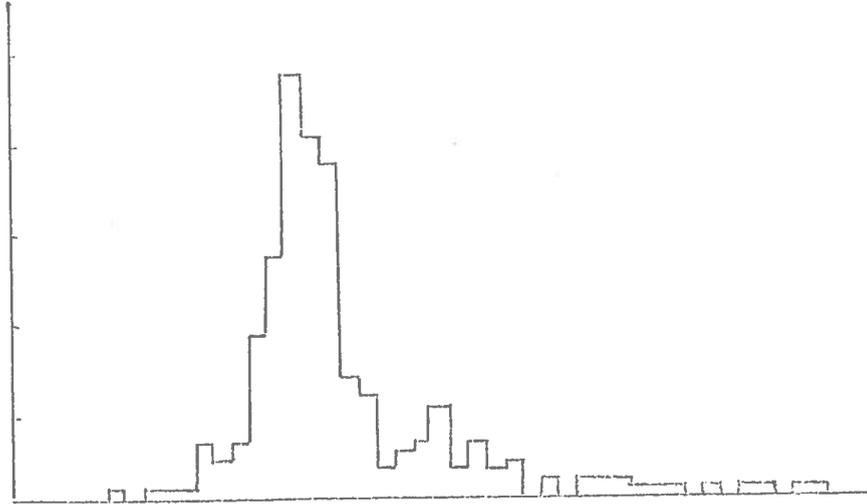
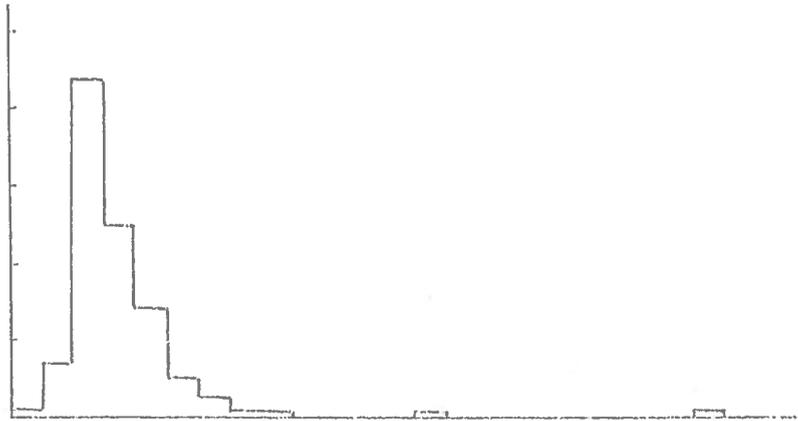


FIGURE 20b Histogramme de la teneur en soufre



En résumé : les variogrammes quantiles peuvent nous avertir de la présence des valeurs extrêmes ou de l'existence d'une distribution fortement asymétrique, mais ils ne donnent pas un estimateur des valeurs du variogramme. Cependant, comme on a besoin d'une estimation du variogramme pour le krigeage, il faut chercher un autre estimateur.

2-4-2 Variogramme hubérisé.

Lors de la recherche d'un estimateur plus robuste que le variogramme classique, les difficultés sont doubles :

1. La plupart des estimateurs robustes sont des estimateurs de la tendance centrale d'une distribution symétrique; mais la distribution des carrés est fortement asymétrique.
2. Même si l'on peut définir un tel estimateur, ce que l'on cherche pour les besoins du krigeage est l'espérance mathématique :

$$2 \gamma(h) = E (Z_{x+h} - Z_x)^2 \quad (18)$$

Une autre approche à cette question est de chercher un estimateur d'échelle (et non de tendance centrale). Pour ce faire, on considère la population symétrique formée en associant $-[Z_{x+h} - Z_x]$ à chaque $[Z_{x+h} - Z_x]$. Cet artifice double le nombre de couples, mais ne cause aucun changement au variogramme qui ne dépend que des carrés. Désignons la variable symétrique par :

$$\text{Sgn} \cdot (Z_{x+h} - Z_x)$$

où Sgn est indépendant de $Z_{x+h} - Z_x$ et prend la valeur 1 ou -1 avec la probabilité 1/2. Le variogramme prend alors la forme :

$$2 \gamma(h) = \text{var} [\text{Sgn} \cdot (Z_{x+h} - Z_x)] \quad (19)$$

Pour l'estimer, on se sert de l'estimateur robuste proposé par HUBER (1964). Pour h fixé, soit

$$Y_i = \text{Sgn} \cdot [Z_{x_i+h} - Z_{x_i}] \quad (20)$$

et

$$S^2 = 2 \gamma(h) \quad (21)$$

Selon la méthode de HUBER, S est défini implicitement par l'équation :

$$\frac{1}{n} \sum_{i=1}^n \psi_c^2 \left(\frac{Y_i}{S} \right) = \beta_c \tag{22}$$

où $\psi_c(x)$ est la fonction définie par :

$$\psi_c = \begin{cases} -c & x < -c \\ x & -c \leq x \leq c \\ c & x > c \end{cases} \tag{23}$$

et

$$\beta_c = E_{\phi} \left[\psi_c^2 (Y/S) \right] \tag{24}$$

est calculé sous l'hypothèse que Y/S a une distribution gaussienne, N(0,1) dont la fonction de répartition est ϕ

On montre que cet estimateur (22) a des propriétés d'optimalité pour la classe des distributions de Y obtenues par une transformation d'échelle des distributeurs de la forme :

$$F = (1 - \epsilon) \Phi + \epsilon H \tag{25}$$

où ϵ est un pourcentage et H est une distribution de contamination inconnue dont l'ensemble du poids est réparti hors de l'intervalle (-c,c). (Les valeurs de ϵ et c sont liées). Le paramètre d'échelle S^2 peut être interprété comme un estimateur de l'échelle de la composante gaussienne de F.

Dans la pratique, comme $\psi_c(\cdot)$ est une fonction impaire, le signe n'a pas d'importance et on peut travailler directement sur les différences $Z_{x+h} - Z_x$.

Il faut alors résoudre l'équation (22) itérativement, mais l'algorithme est simple grâce à la forme particulière de la fonction ψ utilisée. Si S_m^2 est l'estimateur obtenu après m itérations, le suivant est donné par :

$$S_{m+1}^2 = \sum_j [Z_{x_j+h} - Z_{x_j}]^2 / (n \beta_c - n_2 c^2) \tag{26}$$

où

$$I = \{i : |Z_{X_i+h} - Z_{X_i}|^2 \leq c^2 S_{10}^2\}$$

et

$$n_2 = \text{le nombre de } i \text{ non-inclus dans } I.$$

Les résultats de HUBER (1964), dont le présent estimateur n'est qu'un cas particulier, montrent que cet algorithme converge vers une solution unique. Mais on peut aussi le montrer directement en multipliant les deux membres de (22) par S^2 et en substituant t à S^2 . On a donc l'équation $h(t) = t$. En particulier, on voit que la convergence est assurée pour toute valeur initiale non-nulle. La solution sera zéro si la proportion des différences carrées égales à zéro dépasse $1 - \beta_c/c^2$. L'existence d'une solution unique et la simplicité de l'algorithme sont deux avantages de cette fonction ψ par rapport à d'autres choix de ψ , notamment des fonctions qui ne sont pas monotones.

Cette méthode a été appliquée au calcul des variogrammes "hubérisés" correspondant aux valeurs de $c = 2,0$ et $c = 2,5$ pour les 4 variables de la couche médiane. Le choix de la valeur de c est arbitraire : dans ce cas, on a volontairement pris une valeur élevée pour changer le variogramme classique le moins possible dans le cas gaussien. Les valeurs de la constante sont :

$$\begin{aligned} c = 2 \quad \beta_c &= 0,9310 \\ c = 2,5 \quad \beta_c &= 0,9775 \end{aligned} \tag{27}$$

Pour d'autres valeurs de c

$$\beta_c = \Pr (X_1^2 < c^2) + 2 c^2 (-c) \tag{28}$$

La valeur du variogramme classique est utilisée comme valeur initiale de S^2 . Les variogrammes hubérisés correspondant à $c = 2,5$ sont présentés sur les figures 16, 17, 18 et 19. Quand on prend $c = 2,0$, les variogrammes hubérisés ont la même allure, mais ils sont plus bas.

Le variogramme hubérisé de la teneur en soufre est semblable au variogramme du quantile de 75 %, ainsi qu'au variogramme obtenu après élimination des deux valeurs très élevées. Les variogrammes hubérisés de trois autres variables ont même allure que les variogrammes classiques et que les

variogrammes quantiles. Tandis que les variogrammes hubérisés de la puissance et du F.S.I. sont à peine plus bas que le variogramme classique, celui de la teneur en cendre est beaucoup plus bas. Ce résultat était inattendu et doit être expliqué. La distribution des teneurs en cendre a une longue queue de valeurs élevées. Contrairement aux teneurs en soufre où les valeurs élevées sont évidentes, les valeurs élevées de teneur en cendre sont plus nombreuses et en même temps moins extrêmes. Comme pour les teneurs en soufre, les valeurs élevées ont été tracées pour voir si elles sont groupées. Elles manifestent la même linéarité que les teneurs en soufre, sauf que le groupe est déplacé de 1 000 m au N-0 de la poche de soufre. Il serait très intéressant de trouver une explication géologique à ce phénomène.

Comme il faut résoudre itérativement l'équation (22) du variogramme hubérisé, il est important de savoir à quelle vitesse converge l'algorithme. Pour la valeur $c = 2,5$, la valeur de S a été obtenue à une précision de 1 % après deux ou trois itérations pour toutes les classes de distances et pour toutes les variables. Des résultats semblables ont été obtenus pour $c = 2,0$ sauf que dans quelques cas, il faut une itération supplémentaire.

2-4-3 Discussion et conclusion.

Au cours de cette étude sur le charbon brut, deux nouveaux estimateurs du variogramme ont été proposés. Les variogrammes quantiles permettent de déterminer la forme du variogramme sous-jacent, mais ils ne donnent pas un estimateur unique du variogramme. Un deuxième désavantage est que les données doivent être ordonnées, ce qui implique de garder en mémoire tous les carrés des différences (ou au moins tous ceux d'une classe de distances). Pour un grand nombre de données, cela peut être prohibitif.

Au contraire des variogrammes quantiles, le variogramme hubérisé est un estimateur correct du variogramme. Dans les cas étudiés, il a fourni de bons résultats et il a l'avantage d'avoir de bonnes propriétés statistiques qui appartiennent à l'estimateur d'échelle de HUBER. Bien que l'algorithme utilisé dans ce cas demande le stockage de tous les carrés, cela peut être évité en décidant de s'arrêter après un nombre d'itérations fixé, et en relisant les données chaque fois au lieu de les garder en mémoire jusqu'à la convergence de l'algorithme. Du point de vue du temps de calcul,

Le variogramme hubérisé est équivalent à trois ou quatre calculs du variogramme classique, et en plus cette procédure ne demande que quelques modifications simples au programme de calcul du variogramme classique.

Un autre aspect de cette étude est qu'elle prouve que le manque de robustesse du variogramme classique peut être lié à diverses causes, parmi lesquelles la présence de valeurs extrêmes ou d'une queue de valeurs. Le problème de quelques valeurs extrêmes est facilement résolu; on peut les éliminer des données ou prendre le variogramme hubérisé. Mais le problème d'une distribution fortement asymétrique est plus délicat. Il n'est pas facile de déterminer à quel niveau il faut tronquer la distribution. Le choix du point de troncature est très subjectif. Par exemple, dans le cas de la figure 18, il n'y a pas de critère objectif qui permette de savoir si le variogramme hubérisé avec $c = 2,5$ est un meilleur estimateur de la variabilité sous-jacente des teneurs en cendre que celui avec $c = 2,0$. En augmentant la valeur de c , on obtiendrait une série de variogrammes hubérisés qui se rapprochent du variogramme classique.

Le but de cette section est d'attirer l'attention sur le manque de robustesse du variogramme classique et de suggérer deux nouveaux estimateurs pouvant servir de contrôle du variogramme classique. Nous n'avons pas l'intention de rejeter l'estimateur classique. Au contraire, il sera toujours calculé d'abord, pour voir son allure et ensuite pour servir de valeur initiale dans l'algorithme de HUBER. Nous préférons considérer le variogramme robuste comme une "sonnette d'alarme". Une différence prononcée entre celui-ci et le variogramme classique soulignerait la nécessité de bien examiner les données afin d'en déceler les causes.

CHAPITRE III

ESTIMATION DES CARACTERISTIQUES DU CHARBON LAVE

CHAPITRE III

ESTIMATION DES CARACTERISTIQUES DU CHARBON LAVE

3-1 Deux approches à ce problème d'estimation.

Dans les sections suivantes, deux approches différentes au problème de l'estimation du facteur de récupération et de la teneur en cendre du charbon après lavage seront présentées et leurs mérites seront examinés. Dans un premier temps, on ne considère pas la teneur en soufre ni le F.S.I.

L'estimation du facteur de récupération et de la teneur en cendre se divise en trois étapes : l'ajustement aux données de lavabilité des courbes de récupération et de teneur en cendre, l'estimation du facteur de récupération et de la teneur en cendre à la densité voulue, et la phase d'interpolation que l'on fera par krigeage. A priori, il semblerait préférable de réaliser ces trois étapes dans cet ordre; c'est-à-dire d'ajuster aux données de lavabilité venant de chaque sondage ces deux courbes (soit à la main, soit en utilisant une autre méthode plus sophistiquée comme par exemple les fonctions spline), puis d'évaluer pour la densité voulue le facteur de récupération et la teneur en cendre de chaque sondage et enfin de kriger ces valeurs sur le domaine étudié.

Pourtant, cette approche présente le désavantage que toute la procédure doit être recommencée s'il faut refaire ultérieurement l'estimation à une autre densité.

Une deuxième manière d'aborder la question est de développer un modèle paramétrique qui exprime le facteur de récupération et la teneur en cendre en fonction de la densité de séparation. Les paramètres peuvent être krigés sur le domaine étudié, et les estimations du facteur de récupération et de la teneur en cendre peuvent alors être réalisées à partir de ces paramètres krigés. La différence essentielle entre ces deux approches est que l'ordre des deuxième et troisième étapes est inversé. On peut le résumer par le tableau suivant (Figure 21).

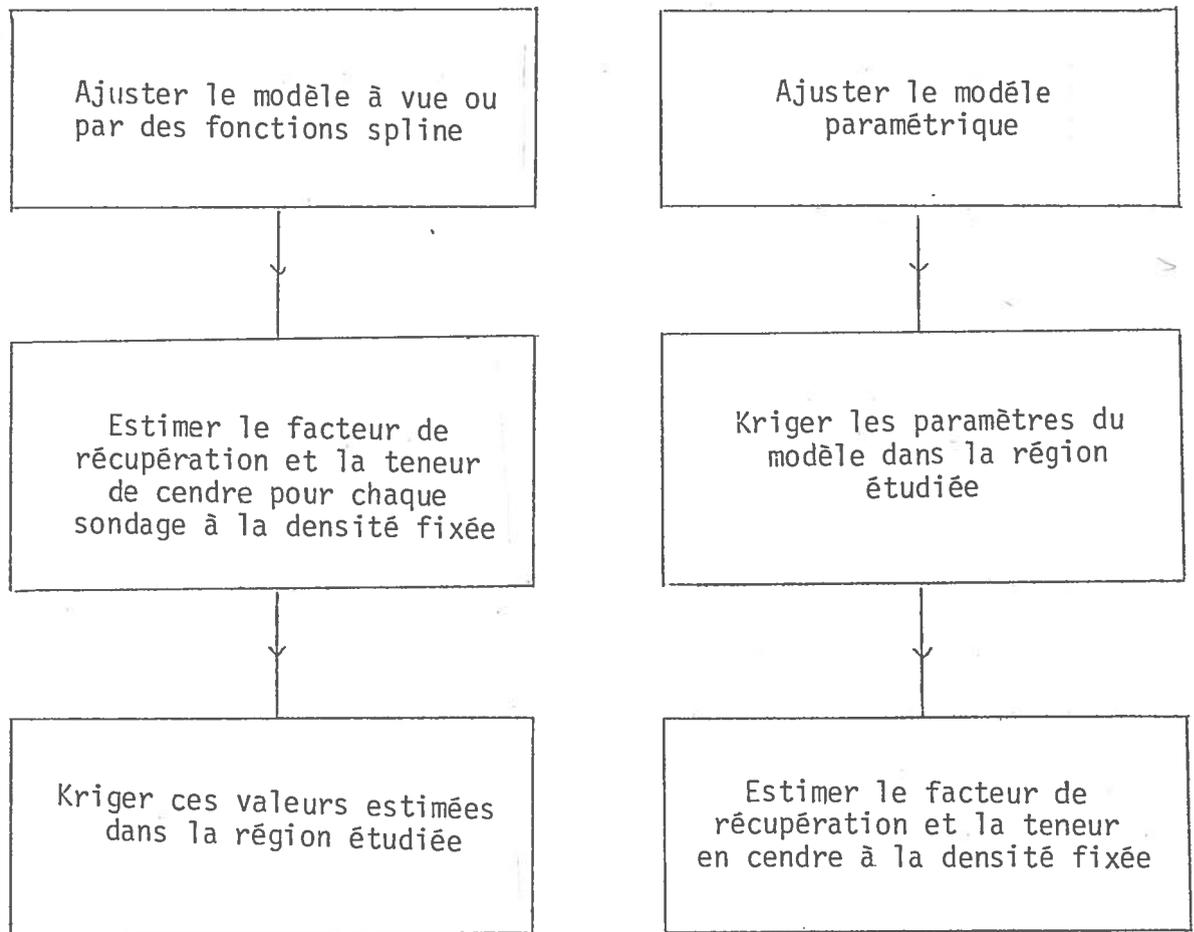


Figure 21 : Schémas des deux approches du problème.

Un avantage de la deuxième approche est que, pour estimer le facteur de récupération et la teneur en cendre à une nouvelle densité, il suffit de répéter la troisième étape au lieu de toute la procédure. Mais il faut pour cela développer un modèle paramétrique qui puisse exprimer le facteur de récupération et la teneur en cendre en fonction de la densité de lavage, ce qui n'est pas évident.

Avant de considérer ces deux approches en détail, il faut décrire la nature des données de lavabilité.

2-2 Description des données de lavabilité.

Les tests de lavabilité sont effectués pour une série de valeurs de densité (décroissantes). Après avoir été concassé et tamisé, le charbon est mis dans un liquide de densité connue et la matière flottante est

récupérée et pesée. La teneur en cendre de la partie flottante est déterminée ensuite en chauffant sous des conditions standard jusqu'à ce que cesse la dissipation des produits volatiles. Ainsi, les données de lavabilité donnent pour chaque classe de densité le pourcentage de produits flottants et leur teneur en cendre.

Le nombre de classes de densité varie de 2 à 5 pour des carottes standard de 2" de diamètre, et de 7 à 14 pour des carottes de 8" de diamètre. Un exemple typique de données de lavabilité est présenté dans le tableau 9.

Classe de densité	% en poids	% en cendre
Floats 1.300	29,47	4.2
1.300 - 1.350	35.22	7.4
1.350 - 1.375	10.91	10.0
1.375 - 1.400	8.17	11.5
1.400 - 1.425	5.48	13.5
1.425 - 1.450	2.65	16.6
1.450 - 1.500	2.95	20.2
1.500 - 1.55	1.53	24.4
1.55 - 1.60	0.52	27.9
Sinks at 1.60	3.10	55.1

TABLEAU 9 : Exemple typique de données de lavabilité en provenance d'une carotte de 8" de diamètre.

De temps en temps, les analyses sont effectuées jusqu'à une densité de 1,8 ou même 2,0. Au contraire, les analyses des carottes de 2" de diamètre ne sont effectuées que pour quelques densités. Il est regrettable que la société qui a fourni les données ait changé les valeurs des densités durant les dix années où les analyses ont été faites. Ni le nombre de classes de densité, ni les valeurs des densités sélectionnées ne sont restés constants, ce qui gênera l'estimation des deux paramètres ρ_c et ρ_a vers la fin de l'étude.

Le tableau 10 présente la gamme de densités utilisées dans les analyses de données de lavabilité.

		Densités utilisées							
Fréquence		1.4	1.35	1.35	1.375	1.375	1.35	1.3	1.45
		1.5	1.6	1.45		1.45	1.4	1.4	1.6
		1.6		1.6		1.6	1.45	1.5	
		1.8					1.6	1.6	
Couche médiane		20	-	59	49	32	139	5.	-
Couche inférieure		21	31	90	159	103	106	22	76

TABLEAU 10 : Densités utilisées dans les analyses de lavabilité.

Au début de cette étude, on espérait que certaines densités seraient communes à toutes les analyses et que par suite, il aurait été possible de kriger le facteur de récupération et la teneur en cendre pour ces densités sur le domaine d'étude, et donc d'estimer ces variables pour la valeur intéressante de la densité. Ce n'est malheureusement pas possible.

CHAPITRE IV

PREMIERE APPROCHE

CHAPITRE IV

PREMIERE APPROCHE

4-1 Ajustement des courbes du facteur de récupération et de la teneur en cendre.

La première approche se compose successivement de l'ajustement des courbes du facteur de récupération et de la teneur en cendre (soit à la main, soit par une méthode mathématique plus sophistiquée), de l'estimation à la densité voulue de ces deux quantités, puis du krigeage sur le domaine étudié. Comme les données fournies par la compagnie minière se composent de plusieurs centaines de sondages de 2" de diamètre, il n'est pas question de faire l'ajustement des courbes à la main; il faut le faire par ordinateur. Comme dans ses propres calculs la compagnie minière utilise des fonctions spline cubiques pour interpoler entre des densités connues, cette méthode a été également adoptée dans cette étude. Contrairement aux courbes de régression, les fonctions spline sont des interpolateurs exacts aux points d'échantillonnage.

4-2 Estimation du facteur de récupération et de la teneur en cendre.

On a utilisé des fonctions spline cubiques pour estimer le facteur de récupération et les unités cumulées de cendres (c'est-à-dire $\sum p_i a_i$, où p_i est le pourcentage en poids de la $i^{\text{ème}}$ tranche densimétrique flottant et a_i est sa teneur en cendre) et ceci pour 4 valeurs de densité (1,375; 1,425; 1,475; 1,55). En général, les valeurs estimées sont satisfaisantes, sauf en ce qui concerne les valeurs estimées à une densité de 1,375 lorsque toutes les analyses de lavabilité pour le sondage étaient effectuées à des densités supérieures à 1,375. (Figure 22). Un autre problème est que, vers les densités fortes (1,475 et 1,55), les valeurs estimées des facteurs de récupération ont tendance à dépasser 100 %. A priori, cela apparaît comme un défaut sérieux de la méthode, mais en fait, ce n'est pas très gênant parce que personne ne s'intéresse au lavage du charbon à des densités où presque tout le charbon flotte.

Il faut remarquer que les calculs n'ont été effectués que dans les cas où il y avait au moins 4 classes de densités. Il n'aurait pas été réaliste de les estimer à partir de deux ou trois classes seulement. En regardant le tableau 10, on voit que plus de la moitié des échantillons de la couche

inférieure est éliminée pour cette raison. Il est regrettable que le pourcentage soit si élevé. On pourrait pallier cette difficulté en faisant des hypothèses sur la forme exacte des courbes du facteur de récupération et de la teneur en cendre en fonction de la densité, puis les ajuster en leur imposant de passer par 1 ou 2 points disponibles. Mais, en l'absence de modèles généraux admis par ces courbes, on a préféré ne pas en tenir compte dans l'étude plutôt que d'introduire une deuxième source d'erreur.

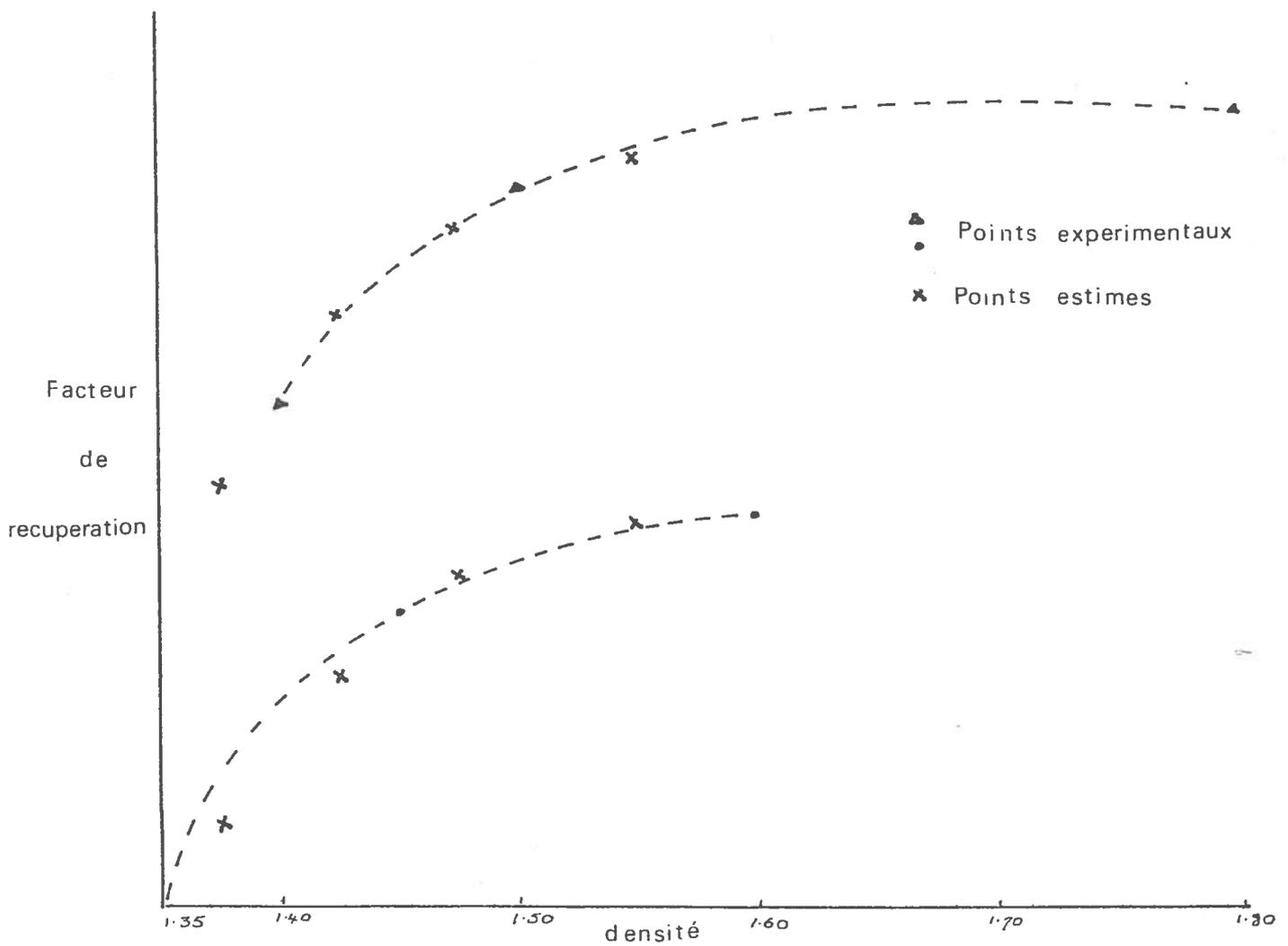
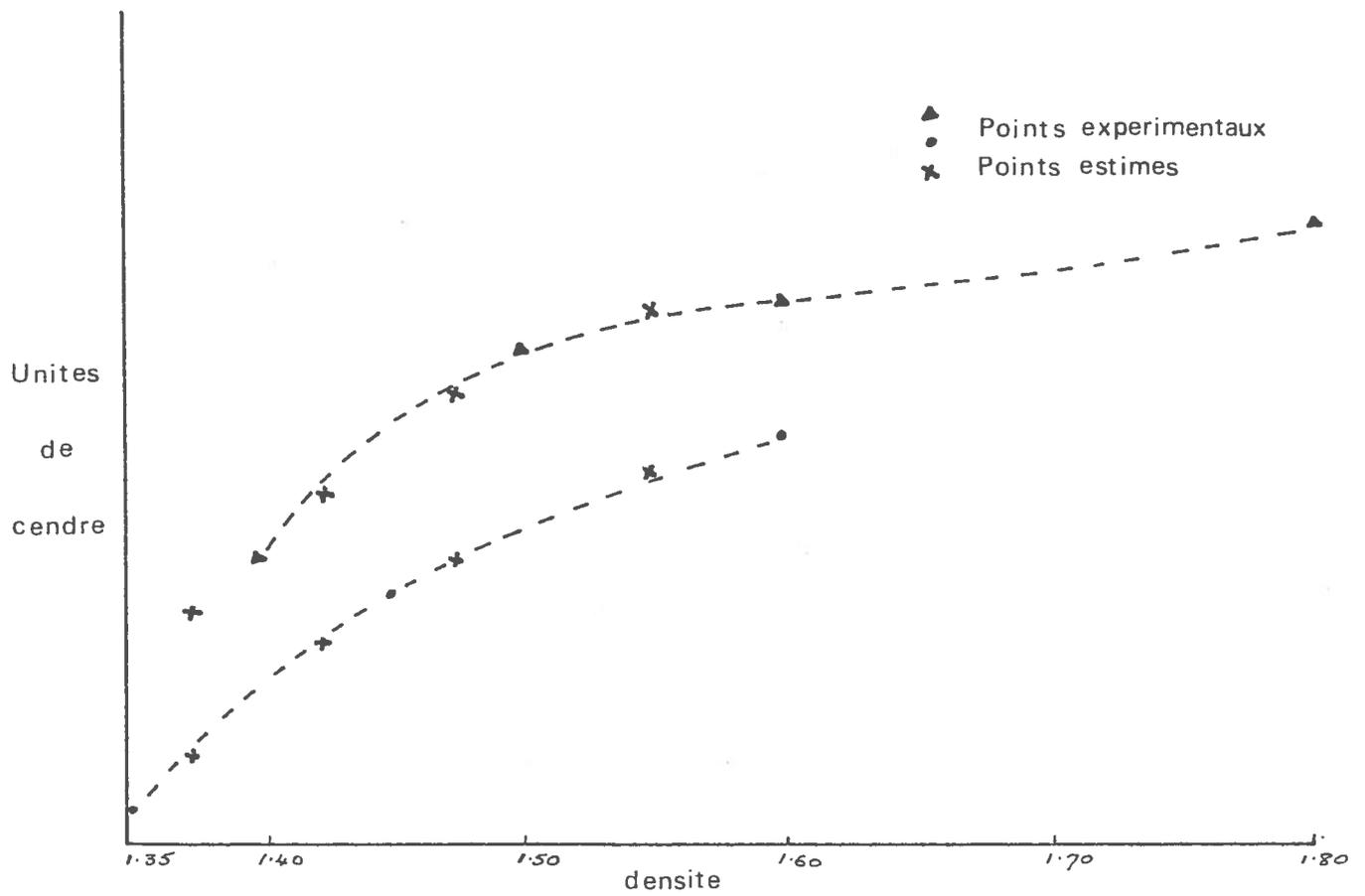
4-3 Variogrammes.

Avant de commencer l'analyse structurale, il faut décider des variables à étudier. Vaut-il mieux traiter l'accumulation des facteurs de récupération (c'est-à-dire le facteur de récupération multiplié par la puissance de la couche) ou le facteur de récupération ? Comme les variables d'une étude géostatistique doivent être additives, il convient a priori d'étudier les accumulations.

En ce qui concerne la teneur en cendre, le problème devient même plus compliqué. Le poids du produit dans chaque classe de densité étant très variable, les teneurs en cendre sont d'abord converties en unités de cendre (c'est-à-dire la teneur en cendre multipliée par le pourcentage du produit dans la tranche). Ainsi, en ce qui concerne la cendre, on pourrait étudier comme variable soit les teneurs en cendre elles-mêmes, soit les unités de cendre, soit encore les accumulations des unités. En raison de leur additivité, on a d'abord traité les accumulations des unités.

La première étape de toute étude géostatistique consiste à faire quelques statistiques classiques sur les variables (moyennes, variances, corrélations, etc...) avant de procéder au calcul des variogrammes, mais pour des raisons pédagogiques, il est plus intéressant dans ce cas de regarder d'abord les variogrammes. Ils ont été calculés pour les 9 variables (la puissance de la couche, les accumulations des unités de cendre et des facteurs de récupération correspondant aux 4 densités fixées). On remarque tout de suite la ressemblance entre ces variogrammes (Figure 23); ce qui traduit la prédominance de la variabilité de la puissance sur les variabilités du facteur de récupération et des unités de cendre. Ceci veut dire que la variabilité de ces deux dernières variables est beaucoup moins forte que celle de la puissance et, par suite, il vaut mieux les traiter séparément. Mais auparavant, il faut noter un fort effet proportionnel sur

FIGURE 22 : Valeurs estimees a partir des fonctions splines



les quatre variogrammes d'accumulation d'unités de cendre. Après division par le carré de la moyenne ou par la variance, les variogrammes sont presque superposables (Figure 24).

On a ensuite calculé les variogrammes des unités de cendre et des facteurs de récupération et on remarque comme précédemment un effet proportionnel (Figure 25) et aussi une forte ressemblance entre l'ensemble des huit variogrammes. A la figure 26, on montre les deux variogrammes correspondant à la densité 1,375. On peut donc conclure que la variabilité des facteurs de récupération domine celle des teneurs en cendre, et par conséquent on passe à l'étude des teneurs en cendre au lieu des unités de cendre.

Comme on pouvait le prévoir, les quatre variogrammes des teneurs en cendre se ressemblent beaucoup (Figure 27). On remarque aussi la linéarité de ces variogrammes jusqu'à une distance de 2 000 m, ce qui est un peu surprenant parce que le variogramme correspondant au charbon brut avait atteint un palier dès 1 500 m. Pour étudier cette différence de comportement, on a calculé les quatre variogrammes et jusqu'à 4 000 m, ils ont tous démontré une croissance linéaire (Figure 28). Après avoir été recalculé, le variogramme de la teneur en cendre du charbon brut montre une augmentation rapide après 2 500 m qui est probablement liée à la présence d'une dérive à grande distance (Figure 29). L'hypothèse de stationnarité qui est valide pour des distances de moins de 2 000 m ne tient plus au delà, mais, comme en pratique on n'a besoin que de stationnarité locale, cela n'est pas très gênant.

4-4 Statistiques sur les variables.

Comme le calcul des variogrammes montre que les variables à étudier sont la puissance de la couche, le facteur de récupération et les teneurs en cendre correspondant aux quatre densités, on calcule les statistiques habituelles sur ces variables. Les moyennes, les variances et les erreurs relatives sont présentées au Tableau 11 tandis que les coefficients de corrélations sont au Tableau 12.

L'erreur relative sur la puissance est beaucoup plus grande que celles de toutes les autres variables (sauf le facteur de récupération à une densité de 1,375). Cela explique pourquoi les variogrammes ressemblent à celui de la puissance. Comme prévu, les moyennes des teneurs en cendre et celles des facteurs de récupération augmentent avec la densité. La décroissance de la

FIGURE 23 : Variogrammes des accumulations

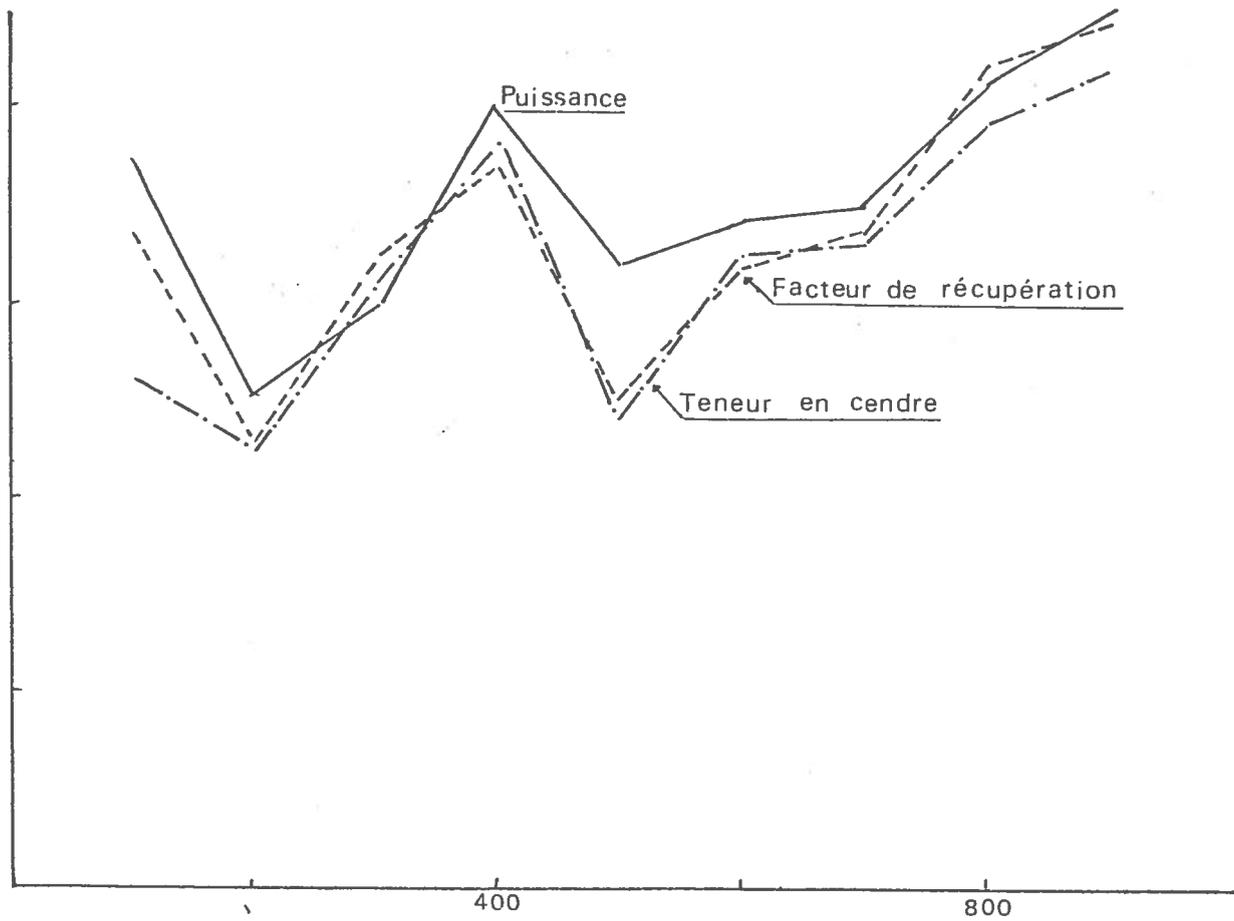


FIGURE 24 : Variogrammes des accumulations de la teneur en cendre

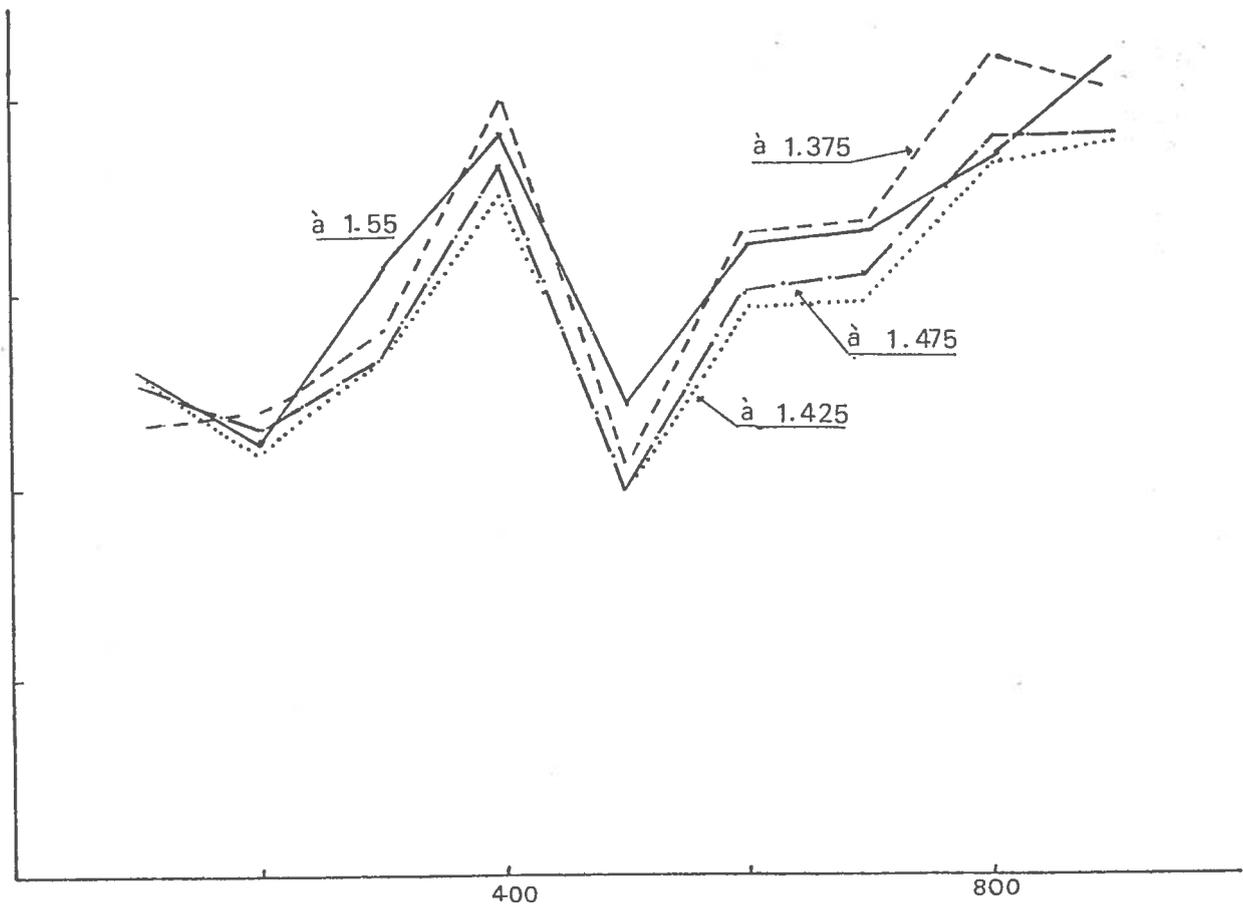


FIGURE 25 : Variogrammes des unités de cendre

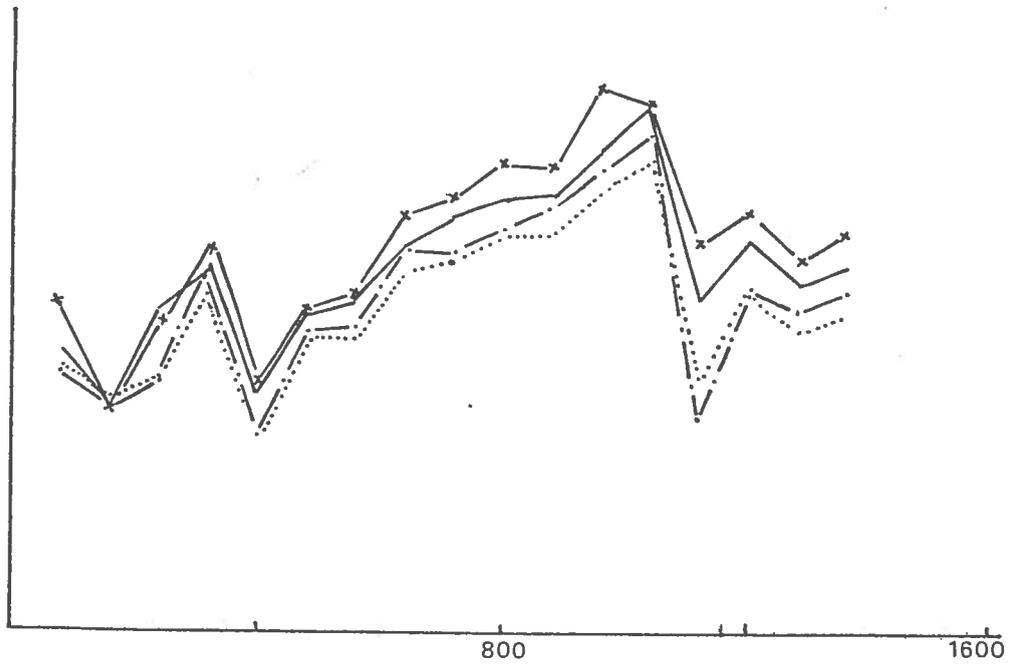


FIGURE 26 : Variogrammes

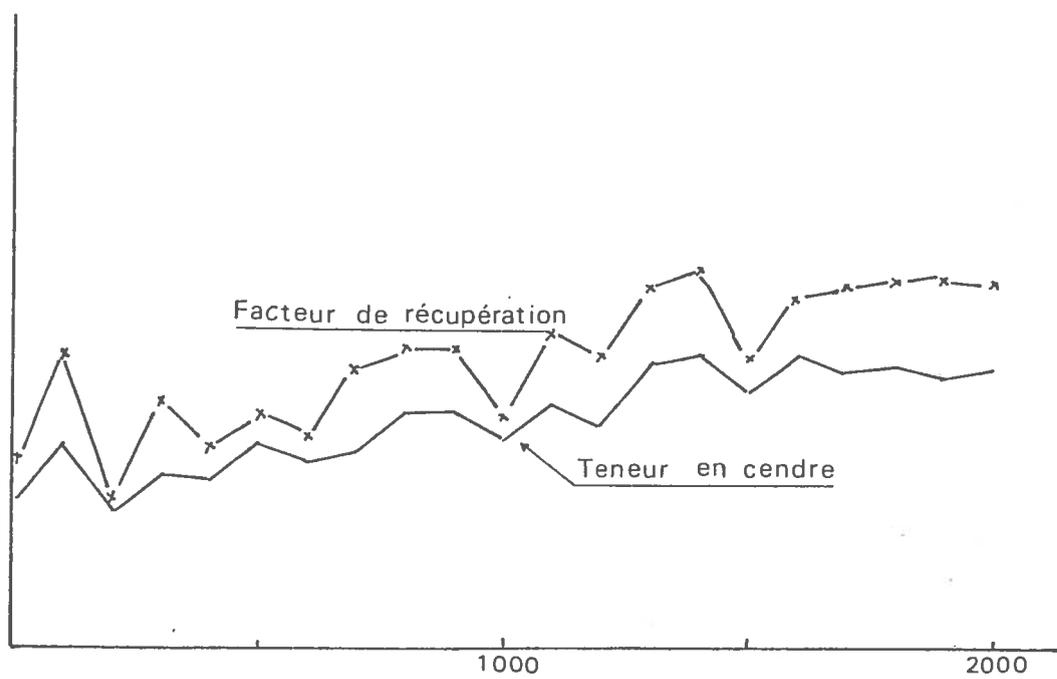


FIGURE 27 : Variogrammes de la teneur en cendre

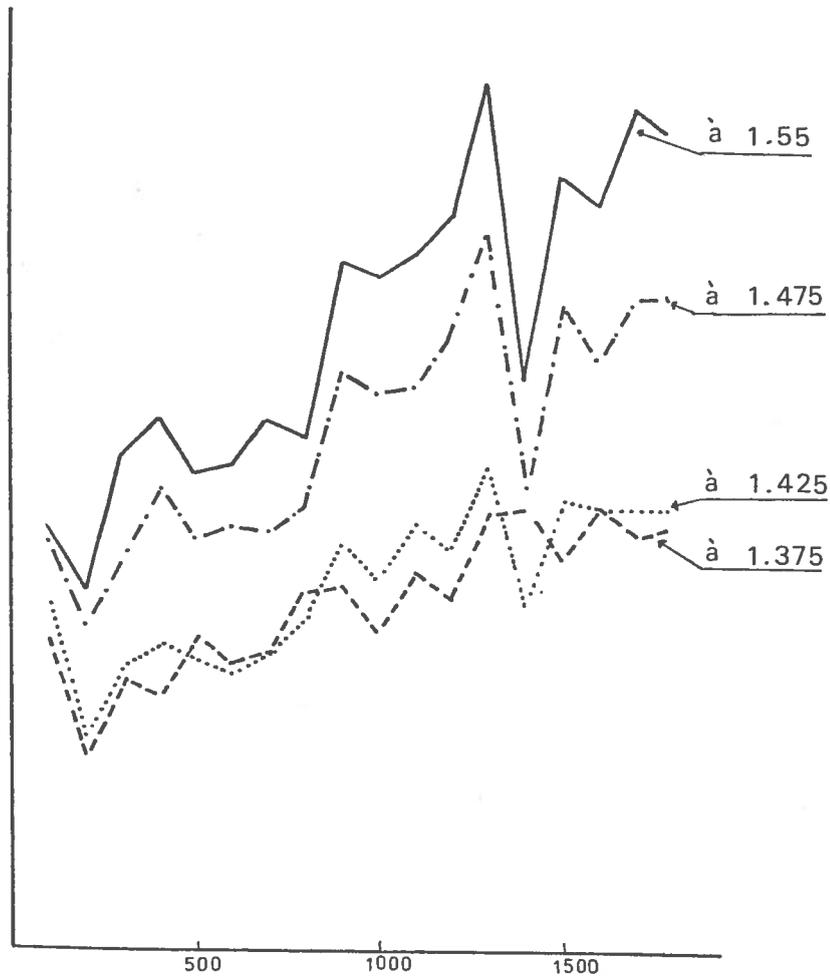
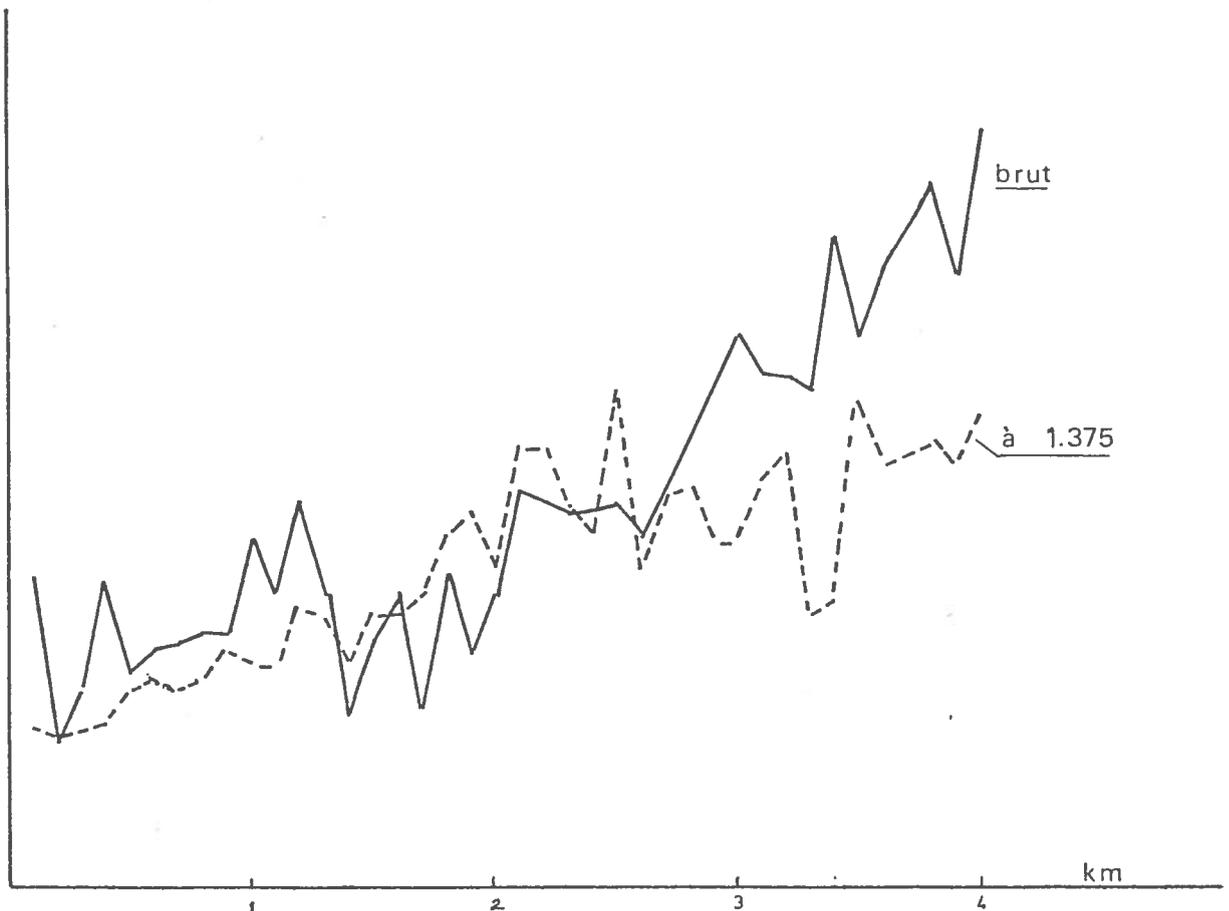


FIGURE 28 : Variogrammes de la teneur en cendre jusqu'à 4 km



variance est également prévisible. En regardant les histogrammes (Figure 29), on voit que plus la densité croît, plus les facteurs de récupération s'approchent de 100 % et par suite la variance décroît.

On remarque que les coefficients de corrélation sont très faibles; aucun ne dépasse 0,5. Les corrélations entre les teneurs en cendre et les facteurs de récupération sont positives tandis que celles entre les teneurs en cendre et la puissance sont toutes négatives.

A première vue, cela peut sembler curieux parce que les facteurs de récupération et les teneurs en cendre augmentent tous avec la densité. Mais réflexion faite, on voit que les facteurs de récupération élevés sont associés, en général, aux faibles teneurs en cendre. Pour le voir, considérons deux échantillons de charbon de facteurs de récupération différents. Soit $f_1(x)$ et $f_2(x)$ les distributions des densités des particules de charbon dans ces deux échantillons (voir figure 30).

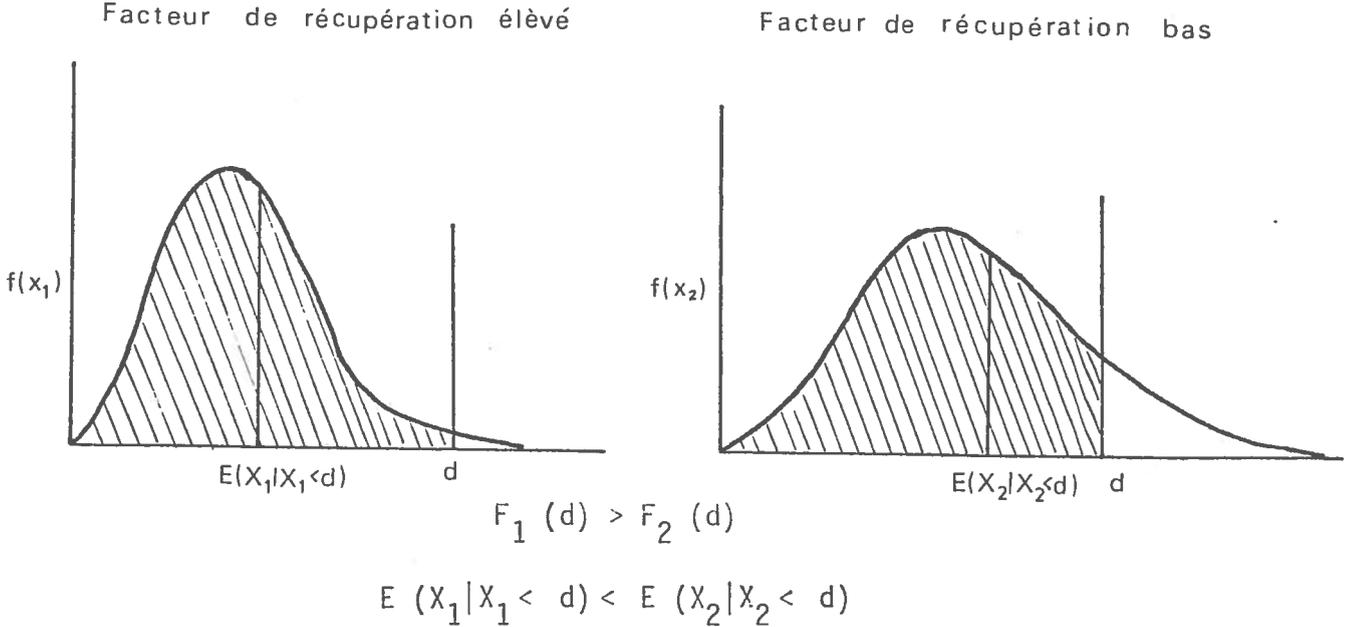


Figure 30 : Relation teneur en cendre vs facteur de récupération.

Les distributions ont des queues très longues de valeurs élevées. Comme la courbe exprimant la teneur en cendre en fonction de la densité est presque linéaire pour des densités comprises entre 1,3 et 1,6, on peut poser $a(x) = k x$ et ainsi trouver que la teneur en cendre moyenne jusqu'à une densité d est donnée par :

FIGURE 29 : Histogrammes du facteur de récupération

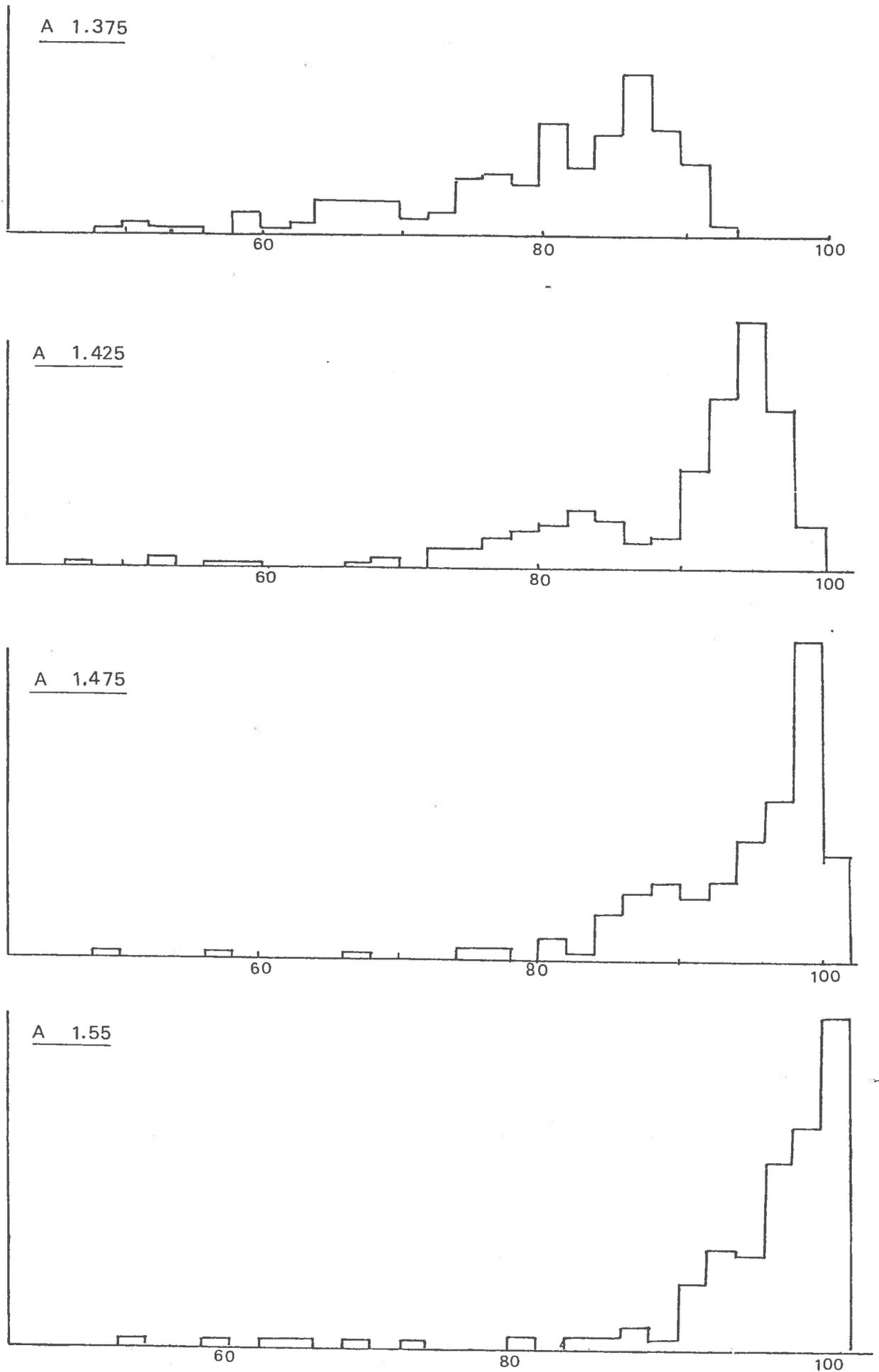
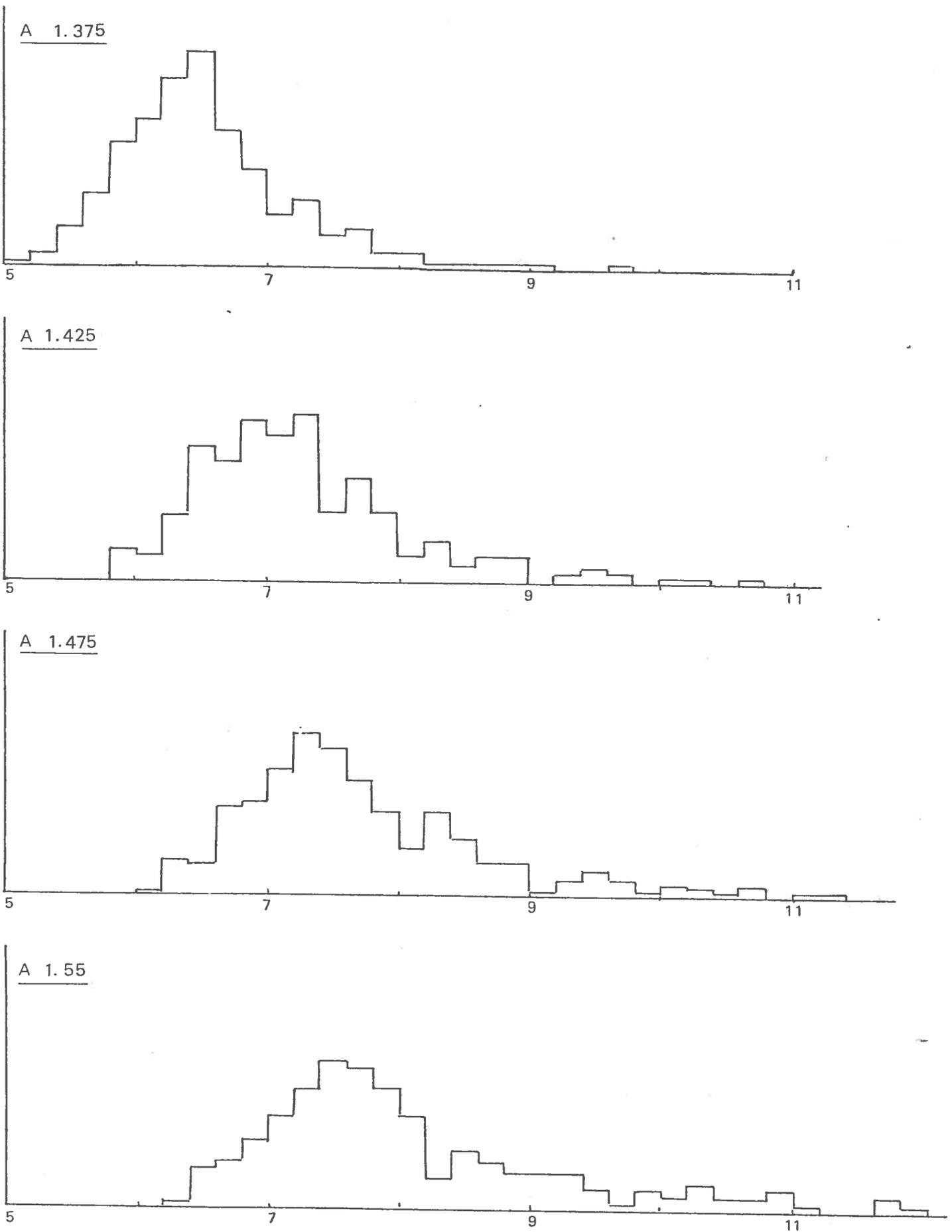


FIGURE 29 : Histogrammes de la teneur en cendre



$$\begin{aligned}
 & \frac{\text{Poids en cendre}}{\text{Poids récupéré}} \\
 & \frac{\int_0^d a(x) f(x) dx}{\int_0^d f(x) dx} \\
 & = k \frac{\int_0^d x f(x) dx}{\int_0^d f(x) dx} \\
 & = \frac{k E(X|X < d)}{F(d)} \tag{29}
 \end{aligned}$$

Par conséquent, il est évident que pour des densités supérieures à la médiane de la distribution, il y aura une corrélation entre la teneur en cendre dans les produits flottants et le facteur de récupération.

Après avoir calculé les statistiques, il est intéressant de tracer les données pour voir s'il existe une dérive. La Figure 31 indique la disposition de tous les échantillons en provenance de la couche inférieure dont les analyses de lavabilité contiennent plus de 2 classes de densité; ainsi que leurs teneurs en cendre estimées pour une densité de lavage de 1,375. Les teneurs estimées sont plus élevées dans la moitié Nord. Bien que cette dérive soit assez nette, elle ne se manifeste qu'à partir de 5 ou 10 km, distance beaucoup plus grande que celles qu'on utilise dans le krigeage. Pour les distances plus petites, on peut considérer la variable comme stationnaire.

4-5 Modèles ajustés des variogrammes.

Le calcul de variogrammes effectué à la section 4-3 a démontré que les variables à étudier sont la puissance, les facteurs de récupération et les teneurs en cendre. Les corrélations entre ces trois groupes de variables étant très faibles, on peut les considérer comme indépendantes.

Le modèle du variogramme de la puissance de la couche qui a déjà été choisi en traitant le charbon brut est la somme de deux modèles sphériques, d'effet de pépité nul.

$$\begin{aligned}
 \gamma_T(h) &= C_1 \left(\frac{1}{2} \frac{h}{a_1} - \frac{3}{2} \frac{h^3}{a_1^3} \right) + C_2 \left(\frac{1}{2} \frac{h}{a_2} - \frac{3}{2} \frac{h^3}{a_2^3} \right) & h \leq a_1 \\
 &= C_1 + C_2 \left(\frac{1}{2} \frac{h}{a_2} - \frac{3h^3}{2a_2^3} \right) & a_1 < h \leq a_2 \\
 &= C_1 + C_2 & h > a_2
 \end{aligned}
 \tag{30}$$

où

$$\begin{aligned}
 C_1 &= \text{1er palier} = 0,32 \\
 C_2 &= \text{2ème palier} = 0,55 \\
 a_1 &= \text{1ère portée} = 200 \text{ m} \\
 a_2 &= \text{2ème portée} = 1\ 800 \text{ m}
 \end{aligned}$$

En raison de l'effet proportionnel sur les quatre variogrammes des facteurs de récupération et sur les quatre variogrammes des teneurs en cendre, il vaut mieux établir pour chaque groupe de variables la forme sous-jacente du variogramme en divisant les variogrammes expérimentaux par des facteurs d'échelle appropriés. On a essayé comme facteur d'échelle le carré de la moyenne de la variable, puis la variance. Le premier donne de bons résultats avec les teneurs en cendre mais il ne convient pas aux facteurs de récupération. La variance a fourni de meilleurs résultats (Tableaux 13 et 14). Même en utilisant la variance comme facteur d'échelle, les variogrammes ont des niveaux légèrement différents; on remarque encore des différences systématiques entre les colonnes des tableaux. La dernière colonne de chaque tableau contient les moyennes des autres colonnes. Les modèles des variogrammes à un facteur d'échelle près ont été ajustés aux valeurs de ces deux dernières colonnes. Comme celui de la puissance, ils sont les sommes des deux modèles sphériques, plus un effet de pépité dont les valeurs sont indiquées au tableau 15.

	1ère portée	1er palier	2ème portée	2ème palier	Effet de pépité
Teneur en cendre	200	0,02	2 600	0,43	0,15
Facteur de récupération	200	0,0005	1 800	0,005	0,003

TABLEAU 15 : Modèles ajustés aux variogrammes ÷ variances.

Les variogrammes moyens et les modèles ajustés ont été tracés pour visualiser l'ajustement (Figures 32 et 33). Il faut maintenant reconstituer les modèles ajustés en multipliant les coefficients du tableau 15 par les facteurs d'échelle. Mais ce faisant, il faut corriger les inégalités systématiques notées plus haut. Par exemple, en regardant le tableau 13, on remarque que les colonnes 1 et 2 sont notablement plus grandes que la moyenne (colonne 5). Si on calcule le modèle du variogramme à une densité de 1,375 en multipliant simplement par la variance sans tenir compte de cette différence systématique, on risque de se trouver avec un modèle du variogramme bien plus bas que le variogramme expérimental. Une méthode très simple pour résoudre cette difficulté consiste à estimer le facteur de correction comme le rapport de la valeur moyenne du variogramme expérimental à celle du variogramme moyen, ce qui donne les facteurs suivants :

	à 1,375	à 1,425	à 1,475	à 1,525
Teneur en cendre	1,10	1,05	0,97	0,88
Facteur de récupération	0,92	0,97	1,01	1,14

TABLEAU 16 : Facteurs de correction des modèles ajustés.

Après avoir multiplié par la variance et par le facteur de correction, on obtient les modèles suivants :

FIGURE 31 : Variogramme ajusté

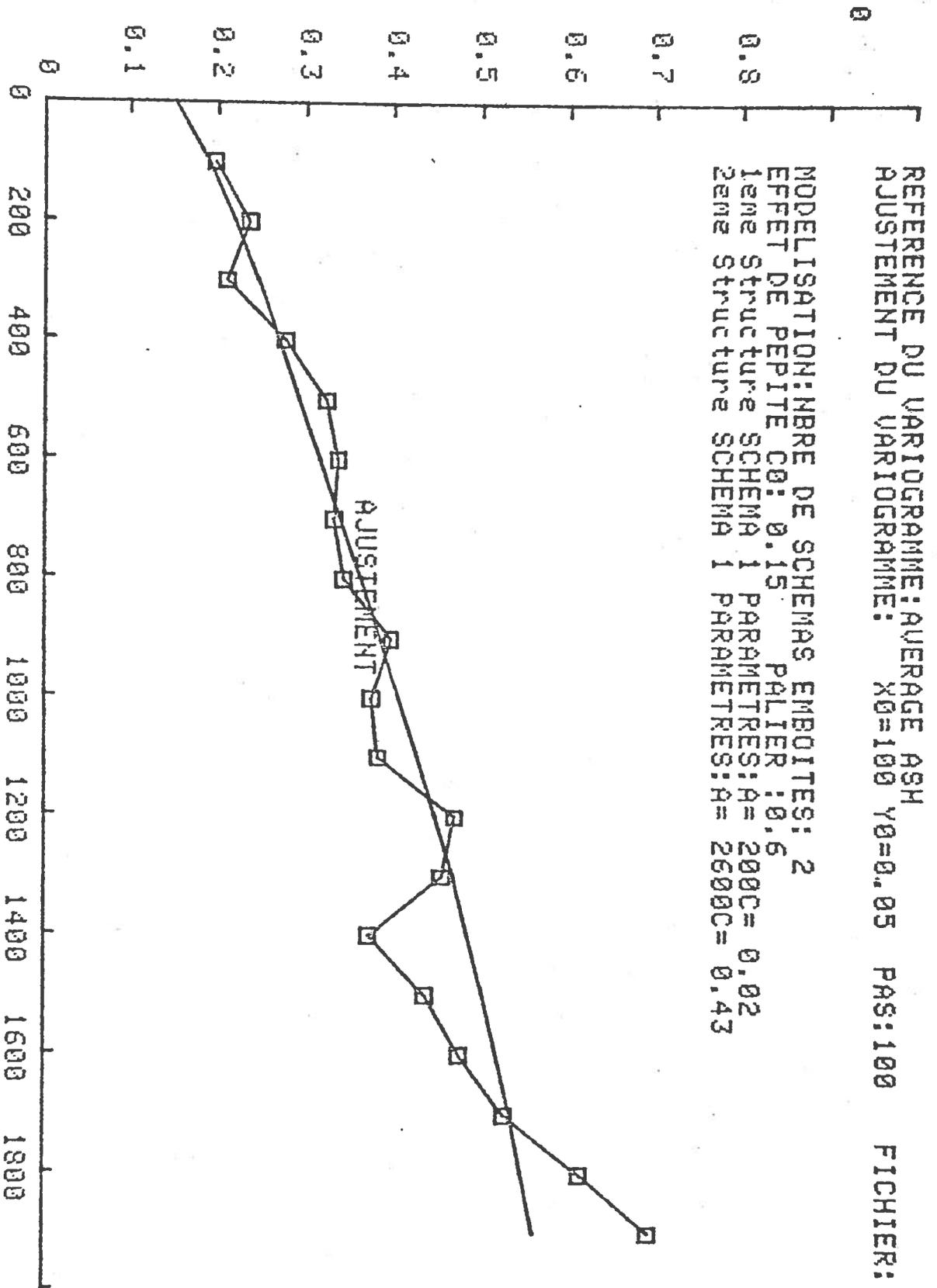


FIGURE 33 : Variogramme ajusté

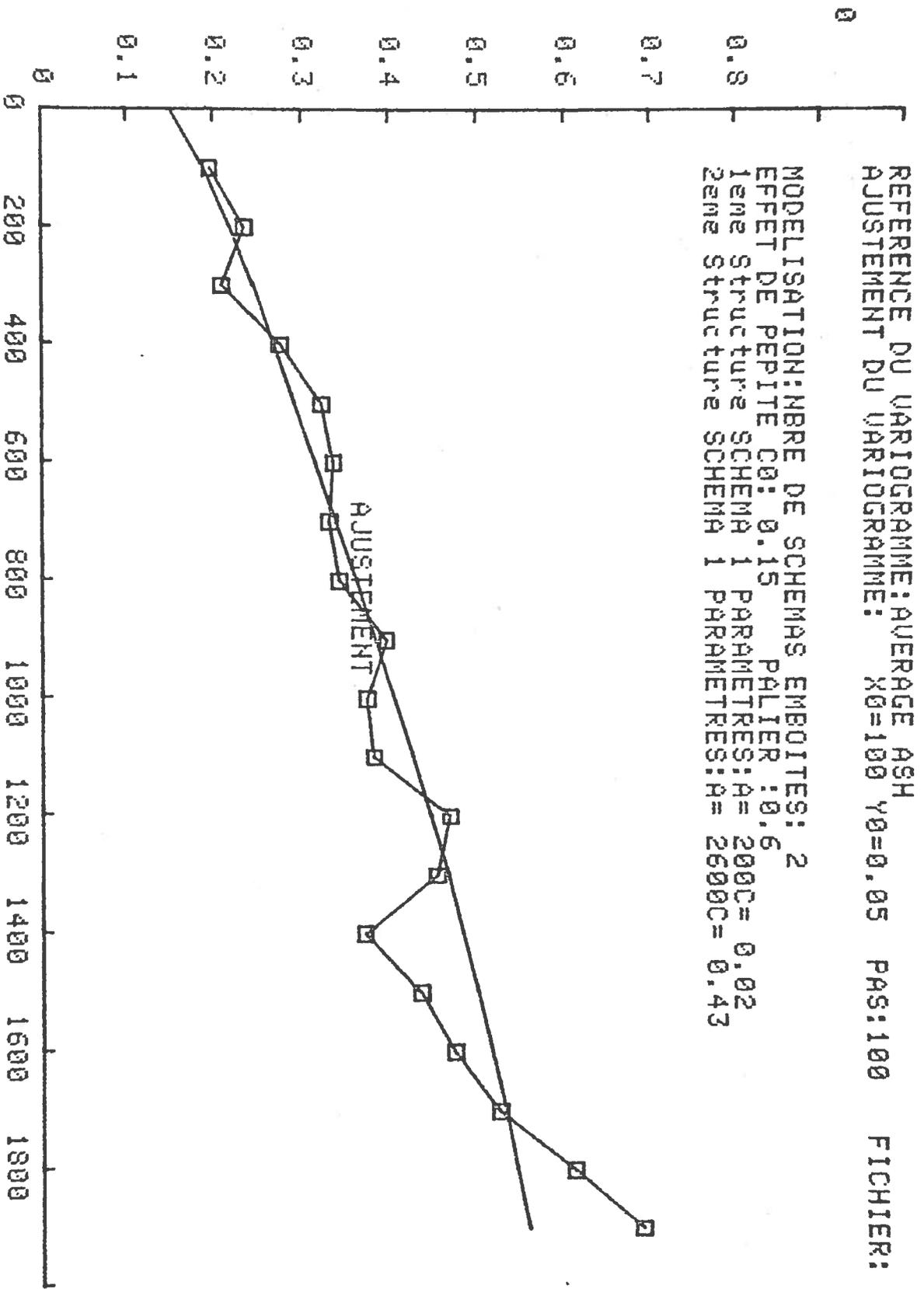
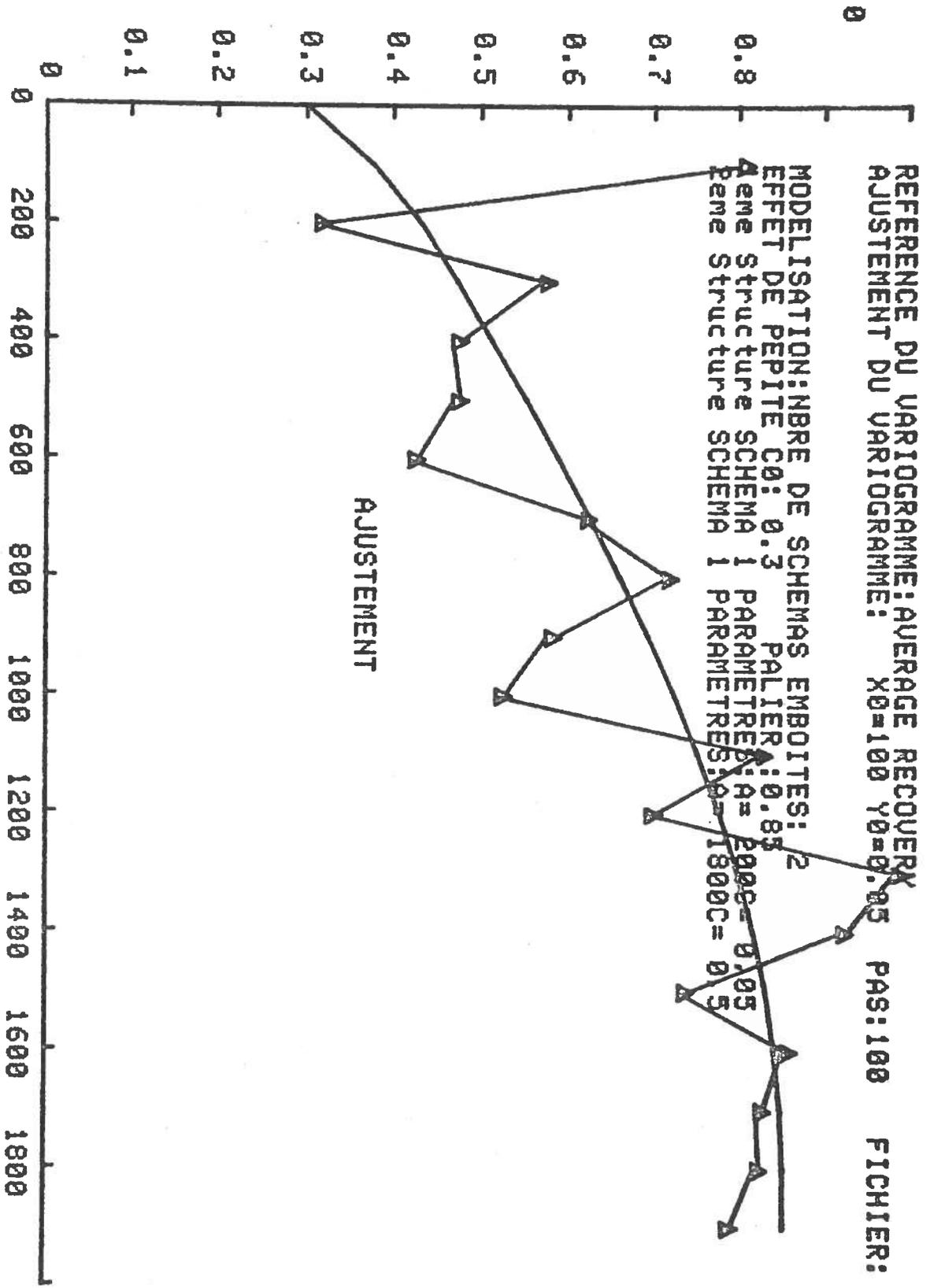


FIGURE 33 : Variogramme ajusté



Teneur en cendre

	1,375	1,425	1,475	1,525
1ère portée	200	200	200	200
1er palier	0,0116	0,0165	0,0193	0,0224
2ème portée	2 600	2 600	2 600	2 600
2ème palier	0,2490	0,3548	0,4158	0,4820
Effet de pépité	0,0869	0,1238	0,1415	0,1682

Facteur de récupération

	1,375	1,425	1,475	1,525
1ère portée	200	200	200	200
1er palier	0,0440	0,0330	0,0292	0,0293
2ème portée	1 800	1 800	1 800	1 800
2ème palier	0,4403	0,3300	0,2917	0,2929
Effet de pépité	0,2642	0,1980	0,1751	0,1757

TABLEAU 17 : Valeurs des paramètres des modèles ajustés.

Au premier coup d'oeil, il semble étrange d'utiliser une méthode aussi compliquée pour ajuster des variogrammes. Pourquoi ne pas ajuster les modèles directement aux variogrammes expérimentaux ? Cela éviterait tous les problèmes de facteurs d'échelle. Les raisons sont doubles. Premièrement, la forte ressemblance des variogrammes expérimentaux suggère qu'ils ont tous le même type de variabilité sous-jacente à un facteur d'échelle près. Faire une moyenne des variogrammes expérimentaux élimine une partie de la variabilité purement aléatoire, ce qui facilite l'ajustement des modèles. La deuxième raison d'utiliser la même forme de modèle est plus subtile et plus profonde : c'est que cela conduit au même système de

pondérateurs pour kriger les variables pour les 4 densités. Cela a l'avantage d'économiser du temps de calcul, parce qu'il n'y a qu'une matrice à inverser au lieu de quatre. Mais, ce qui est plus important, cela garde quelques unes des propriétés des données originales. Comme une augmentation de la densité de lavage entraîne une augmentation de la teneur en cendre et du facteur de récupération, il est raisonnable d'attendre la même relation au niveau des valeurs krigées, ce que l'on ne peut garantir, à moins d'utiliser le même système de pondérateurs de krigeage pour chacune des quatre densités.

Après avoir ajusté les modèles de variogrammes, on peut kriger les valeurs des blocs, ou estimer les valeurs moyennes sur le gisement entier, ou encore déterminer la maille optimale. Une fois l'ajustement fait, les problèmes peuvent être résolus en utilisant les techniques de géostatistique linéaire. Comme celles-ci sont déjà bien connues, il n'est pas nécessaire de les redécrire en détail ici.

CHAPITRE V

DEUXIEME APPROCHE

CHAPITRE V

DEUXIEME APPROCHE

5-1 Remarques préliminaires

On peut aborder le problème d'estimation d'une autre façon. Si on avait un modèle paramétrique exprimant le facteur de récupération et la teneur en cendre en fonction de la densité de lavage, on pourrait kriger les paramètres de ce modèle dans le domaine étudié, puis évaluer le facteur de récupération et la teneur en cendre à partir des valeurs krigées. La différence essentielle entre cette approche et la première est l'inversion de la deuxième et de la troisième étape. Cette approche présente l'avantage qu'on peut calculer le facteur de récupération et la teneur en cendre pour une nouvelle densité de lavage sans répéter toute la procédure; il suffit d'injecter la nouvelle densité et les paramètres krigés dans le modèle.

Avant l'arrivée des données, on espérait que les analyses de lavabilité auraient été effectuées pour les mêmes classes de densité partout dans la concession, mais cela n'était pas le cas. S'il en avait été ainsi, on aurait pu envisager le krigeage du facteur de récupération ainsi que des unités de cendre aux densités communes, puis l'interpolation de ces quantités à une autre densité. Dans ce cas, le facteur de récupération et la teneur en cendre apparaissent comme des paramètres du modèle. Ou bien, on aurait pu ajuster des fonctions spline cubiques à la courbe de récupération et à celle de la teneur en cendre, puis kriger les coefficients de ces fonctions spline, le facteur de récupération et la teneur en cendre étant évalués à partir des coefficients krigés. Mais ces deux approches, qui auraient pu être fructueuses, ne peuvent pas être testées à cause de l'hétérogénéité des densités utilisées pour les analyses de lavabilité. Faute d'autre possibilité, on a dû chercher une représentation paramétrique des deux variables (le facteur de récupération et la teneur en cendre) en fonction de la densité.

5-2 Modèle paramétrique

ARMSTRONG et WHITMORE (1979) ont proposé deux modèles pour caractériser la lavabilité du charbon. Le premier de ces deux modèles est une fonction déterministe exprimant la teneur en cendre d'une particule de charbon en

fonction de la densité, tandis que la deuxième est un modèle lognormal de la distribution en masse de la densité des particules.

5-2-1 Modèle à deux paramètres

Les particules de charbon vont des particules à faible densité et à faible teneur en cendre aux particules à forte densité et à forte teneur en cendre. On remarque que la teneur en cendre augmente avec la densité. On considère que les particules les moins denses sont du charbon "pur" ayant une densité ρ_c et une teneur en cendre A_{min} , et que les particules les plus denses sont de la cendre "pure" ayant une densité ρ_a et une teneur en cendre de A_{max} . Les particules intermédiaires peuvent être considérées comme un mélange de charbon "pur" et de cendre "pure". Pour une de ces particules (de densité x), soient W_a , W_c , V_a et V_c les masses et les volumes de cendre "pure" et de charbon "pur" dans la particule. On a donc :

$$x = \frac{W_a + W_c}{V_a + V_c} \quad (31)$$

Comme la teneur en cendre $a(x)$ est la proportion en masse de la cendre dans la particule :

$$a(x) = \frac{A_{min} W_c + A_{max} W_a}{W_a + W_c} \quad (32)$$

En utilisant les relations $W_a = \rho_a V_a$ et $W_c = \rho_c V_c$, on obtient :

$$\frac{a(x) - A_{min}}{A_{max} - a(x)} = \frac{\rho_a}{\rho_c} \frac{(x - \rho_c)}{(\rho_a - x)} \quad (33)$$

D'où :

$$a(x) = \frac{A_{max} \rho_a (x - \rho_c) + A_{min} \rho_c (\rho_a - x)}{(\rho_a - \rho_c) x} \quad (34)$$

Dans des communications personnelles, ABBOTT et SMITH du National Coal Board ainsi que BELL de l'Université de Leeds, considèrent 90 % comme une limite supérieure raisonnable de la valeur de A_{max} , puisque cela correspond à la teneur en cendre de schiste, les 10 % restant étant volatiles. Avec $A_{min} = 0$, et $A_{max} = 0,9$, l'équation (34) devient :

$$a(x) = 0,9 \frac{(1/\rho_c - 1/x)}{(1/\rho_c - 1/\rho_a)} \quad (35)$$

Le changement des valeurs de A_{\min} ou de A_{\max} ne change pas la forme de la courbe de la teneur en cendre en fonction de la densité; la courbe est seulement tronquée aux teneurs en cendre de A_{\min} et de A_{\max} :

ALPERN (1977) a remarqué que le CERCHAR avait établi une relation semblable entre la teneur en cendre des particules de charbon et sa teneur pour les gisements d'Agades (Niger) et d'Alumance.

5-2-2 Distribution lognormale de la densité des particules.

Dans la section précédente, on a proposé une relation déterministe simple entre la teneur en cendre d'une particule et sa densité; dans cette section, on adoptera une approche probabiliste. Choisissons une particule au hasard. On peut considérer sa densité comme une variable aléatoire. Les analyses de lavabilité fournissent une détermination expérimentale de la distribution des densités, parce qu'elle donnent le pourcentage en masse dans chaque classe de densité.

Les densités étant bornées par les valeurs ρ_c et ρ_a , il peut être utile de transformer cette variable (la densité) pour obtenir une grandeur sans dimension. On pourrait choisir $(X - \rho_c)/(\rho_a - \rho_c)$ qui varie entre 0 et 1, mais la transformation suggérée par AITCHISON et BROWN (1957) semble plus intéressante. Elle consiste à utiliser $(X - \rho_c)/(\rho_a - X)$. Comme cette nouvelle variable varie entre 0 et $+\infty$, son logarithme varie entre $-\infty$ et $+\infty$, ce qui suggère une distribution lognormale.

Une autre raison plus fondamentale pour choisir cette transformation est qu'elle apparaît naturellement en considérant la division aléatoire d'une masse homogène en particules hétérogènes. KOLMOGOROV (1941) a proposé ce modèle pour expliquer la distribution lognormale de la taille des particules de minerai concassé. DE WIJS (1951) et plus tard MATHERON (1960) ont utilisé ce même raisonnement pour démontrer que la teneur en or dans les gisements d'Afrique du Sud est également lognormale. Nous étendrons ce raisonnement aux teneurs en cendre. Considérons une particule de charbon (que l'on appellera "de première génération") qui se divise en deux parties inégales. Supposons qu'une de ces deux parties contienne une proportion p_1 de la cendre "pure"

de la génération précédente, ainsi qu'une proportion q_1 du charbon "pur".
(Voir Figure 34).

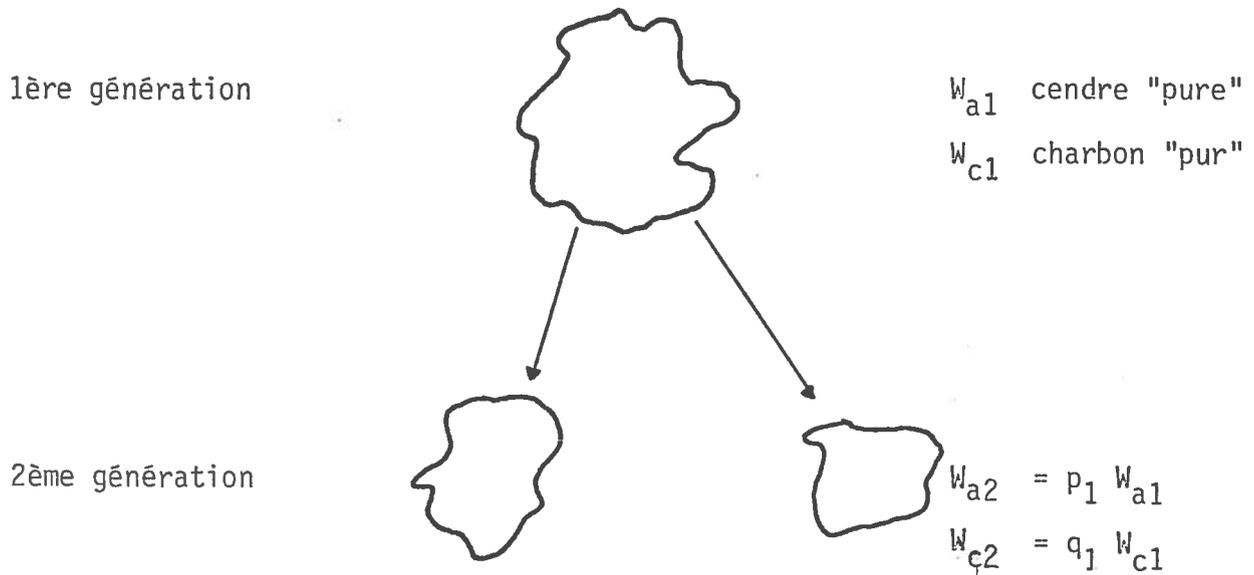


Figure 34 : Division d'une particule en deux.

Soient X_1 et X_2 les densités respectives des particules de la première et de la deuxième génération. Il est évident que :

$$\frac{X_1 - \rho_c}{\rho_a - X_1} = \frac{W_{a1}}{W_{c1}} \times \frac{\rho_a}{\rho_c} \quad (36)$$

et

$$\frac{X_2 - \rho_c}{\rho_a - X_2} = \frac{p_1}{q_1} \frac{W_{a1}}{W_{c1}} \frac{\rho_c}{\rho_a} = \frac{p_1}{q_1} \frac{(X_1 - \rho_c)}{\rho_a - X_1} \quad (37)$$

Si ces subdivisions continuent jusqu'à la $n^{\text{ième}}$ génération, on a :

$$\frac{X_n - \rho_c}{\rho_a - X_n} = \frac{p_1}{q_1} \times \frac{p_2}{q_2} \times \dots \times \frac{p_{n-1}}{q_{n-1}} \times \frac{(X_1 - \rho_c)}{(\rho_a - X_1)} \quad (38)$$

où p_i, q_i sont les proportions de cendre "pure" et de charbon "pur" transmises de la $i^{\text{ème}}$ génération à la $(1+i)^{\text{ème}}$. Supposons que les rapports p_i/q_i soient indépendants et qu'ils aient la même distribution. En raison du théorème de la limite centrale, le produit $p_1/q_1 \times \dots \times p_{n-1}/q_{n-1}$ est asymptotiquement gaussien, et par suite la variable $(X - \rho_c)/(\rho_a - X)$ a une distribution lognormale.

Un corollaire intéressant de ce résultat est que la variable $(a(X) - A_{\min})/(A_{\max} - a(X))$ est également lognormale. Par conséquent, on a pu étudier la distribution en masse des teneurs en cendre des particules au lieu de celle des densités.

5-3 Tests de ces modèles

5-3-1 Test du modèle à deux paramètres.

Pour tester ce modèle, il faut connaître les valeurs de ρ_c, ρ_a, A_{\min} et A_{\max} , mais à l'heure actuelle, elles ne sont pas disponibles. On peut s'attendre à ce qu'elles varient d'un charbon à un autre. A priori, il semble raisonnable de prendre 0 et 0,9 respectivement pour les valeurs de A_{\min} et de A_{\max} . Comme les analyses de lavabilité sont effectuées à des densités entre 1,3 et 2, on ne connaît pas la valeur de ρ_c ni celle de ρ_a . Mais on peut les estimer à partir des données détaillées de la lavabilité en ajustant par la méthode des moindres carrés la courbe de teneur en cendre en fonction de la densité et en l'extrapolant aux teneurs de 0 et de 0,9. Les estimateurs de moindre carré sont donc :

$$\rho_c = \frac{n \sum (a_i/x_i) - \sum a_i \sum (1/x_i)}{\sum (a_i/x_i) \sum (1/x_i) - \sum a_i \sum (1/x_i^2)} \quad (39)$$

$$\rho_a = \frac{\rho_c \sum (a_i/x_i)}{1 \cdot 1 \sum (a_i/x_i) - \sum (1/x_i) + \rho_c \sum (1/x_i^2)} \quad (40)$$

où a_i est la teneur en cendre d'une particule de densité x_i , et n est le nombre d'échantillons.

A ce point, un deuxième problème se présente, parce que les analyses donnent la teneur moyenne en cendre pour une classe de densité et non la teneur des particules d'une certaine densité. On peut surmonter cette difficulté en utilisant la valeur moyenne des densités de la classe comme valeur initiale de la densité qui correspond à la teneur moyenne en cendre, puis en ajustant la densité pour optimiser l'ajustement de la courbe (c'est-à-dire pour maximiser la valeur de la statistique F).

On dispose de 28 ensembles de données de lavabilité en provenance d'une sélection de charbons australiens. On a ajusté une courbe de la forme (35) à ces données. Les tests statistiques ont indiqué que l'ajustement est acceptable dans tous les cas sauf un, l'exception étant le charbon N° 17 qui ne contient que 5 classes de densité. Les figures 35(a) et (b) montrent combien les points expérimentaux sont proches des courbes ajustées dans le cas des deux charbons-type (N^{OS} 1 et 6). Les colonnes 2 et 3 du tableau 17 présentent les valeurs estimées de ρ_c et ρ_a pour l'ensemble des 28 charbons. A l'exception d'une valeur de 1,347, toutes les valeurs se trouvent entre 1,2 et 1,3, ce qui est en accord satisfaisant avec la valeur de 1,25 citée par TIEMEN (1973). Bien que les valeurs estimées de ρ_a soient plus étalées (entre 1,847 et 5,45), la plupart d'entre elles sont encore dans la gamme de valeurs (de 2,5 à 5) suggérée par TIEMEN. Le fait que ces valeurs estimées sont en accord avec les valeurs indiquées dans la littérature, nous donne encore plus confiance dans le modèle à deux paramètres. La forme de la courbe de la teneur en cendre en fonction de la densité obtenue indépendamment par ALPERN (1977) confirme notre confiance.

5-3-2 Test de la distribution lognormale

On aurait voulu tester la distribution statistiquement, mais cela n'est pas possible. Les tests statistiques habituels (par exemple, le test de χ^2 et le test de Geary-Pearson) ne sont pas applicables parce qu'ils partent de l'hypothèse que l'on a un échantillon de n dérivations X_1, \dots, X_n ce qui n'est pas le cas. Les analyses de lavabilité ne fournissent que le poids dans chaque classe de densité. Puisque les tests classiques ne peuvent pas être appliqués, il faut réfléchir au critère sur lequel on jugera la qualité de l'ajustement, mais avant de le faire, il faut mentionner un autre problème.

Dans la pratique, le charbon exploité contient un pourcentage de stérile en provenance du toit ou du mur de la mine, ou des bandes de schiste entre les couches, qui atteint un niveau de 30 % ou 40 % du total exploité.

FIGURE 35 : Teneur en cendre en fonction de la densité

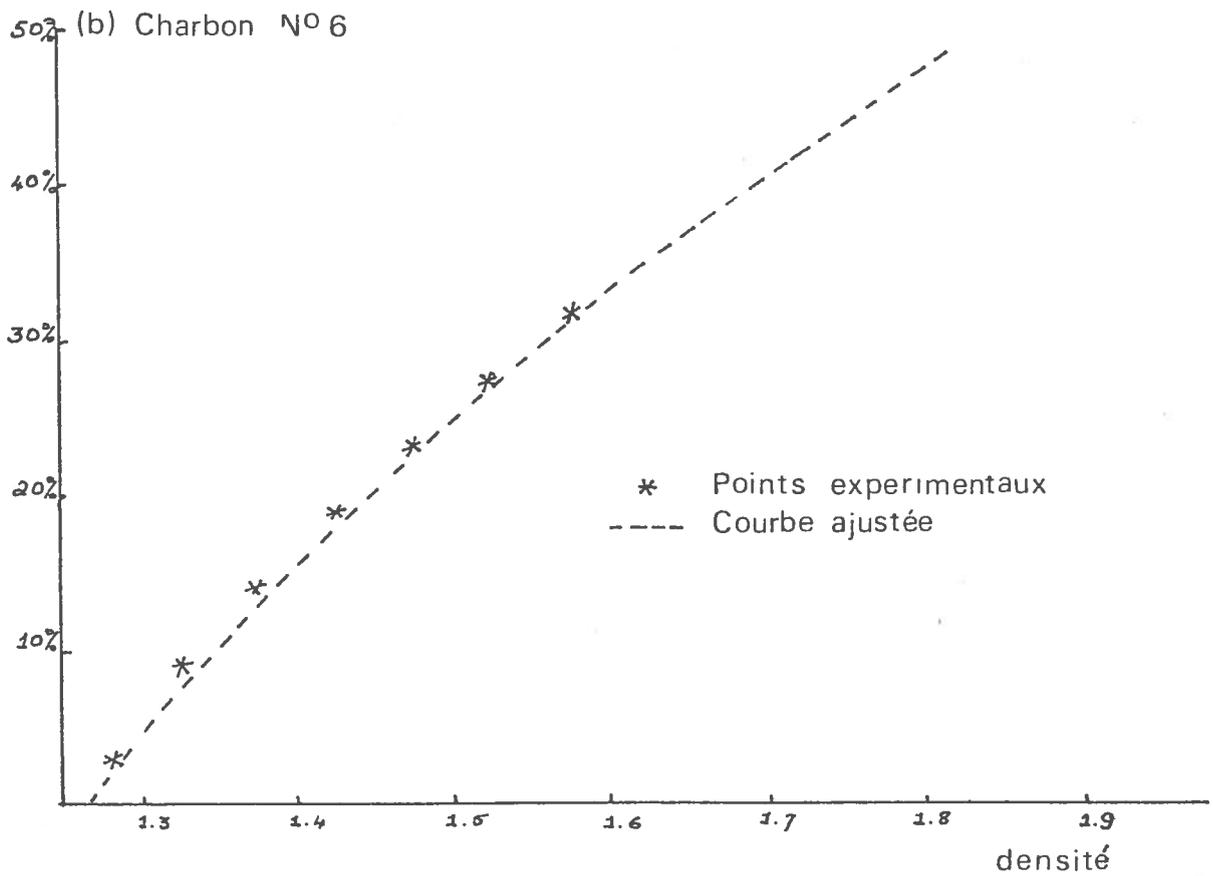
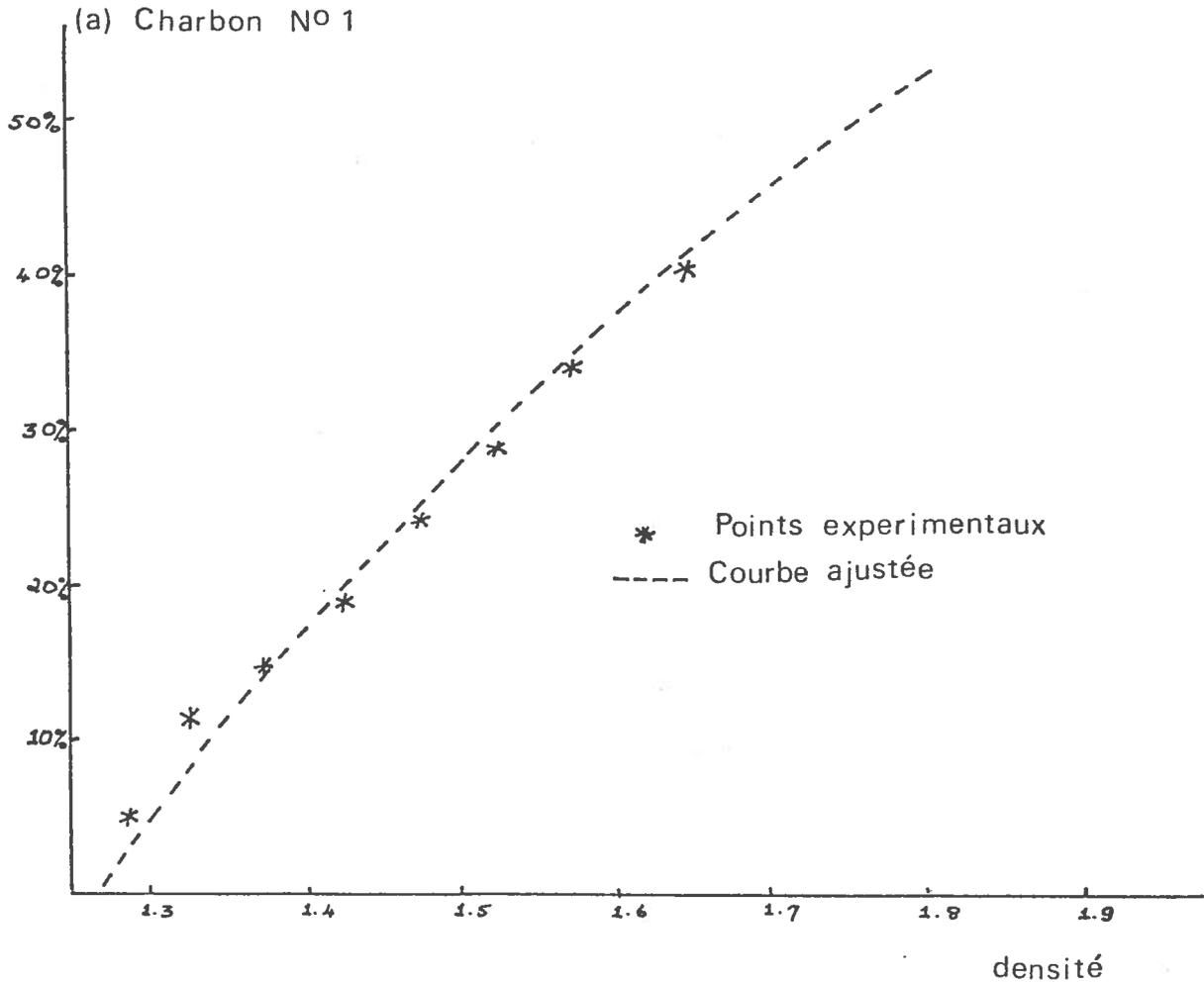


TABLEAU 17 : Valeurs des paramètres

N° du charbon	ρ_c	ρ_a	ρ_{min}	A_{min}	P	μ	σ	
1	1.250	2.022	1.288	4.7%	99.3%	-6.04	4.25	
2	1.274	2.528	1.337	8.6%	99.7%	-3.67	3.90	
3	1.256	2.304	1.282	4%	99.4%	-2.17	2.46	
3 (4")	1.252	2.270	1.255	0.5%	76.3%	-2.21	1.06	
3 (1/2"-3/4")	1.258	2.269	1.283	4%	85.1%	-2.46	2.25	
4	1.202	3.198	1.229	3.2%	99.7%	-5.61	1.94	
5	1.269	2.455	1.290	3.0%	87.6%	-3.56	1.38	*
6	1.260	2.869	1.260	0%	93.9%	-2.45	0.57	†
7	1.244	2.534	1.273	4%	88.5%	-2.60	0.99	
8	1.219	2.821	1.219	0%	85.6%	-2.46	0.74	
9	1.246	3.336	1.298	5.8%	99.5%	-6.44	3.42	
10	1.255	2.577	1.288	4.5%	99.3%	-7.97	6.59	
11 (< 12.7 mm)	1.261	2.818	1.277	2%	76.6%	-1.49	1.22	
11 (> 12.7 mm)	1.259	3.018	1.282	2.8%	99.6%	-2.08	1.99	
12	1.272	2.706	1.291	2.5%	97.7%	-2.60	0.93	†
13	1.265	2.990	1.265	0%	99.8%	-2.54	1.01	*
14	1.288	3.934	1.343	5.5%	99.7%	-4.66	2.26	
15	1.347	1.847	1.386	9.4%	99.4%	-6.54	5.15	
16	1.262	2.590	1.291	4%	87.7%	-2.80	0.85	
17	1.287	2.932	1.320	4%	90.7%	-3.59	1.04	
18	1.229	4.371	1.259	3%	99.6%	-3.18	1.41	
19	1.237	3.491	1.237	0%	90.2%	-2.43	0.78	*
20	1.254	3.714	1.292	4%	99.6%	-2.95	0.75	*
21	1.199	4.247	1.257	5.8%	88.3%	-5.86	3.07	
22	1.219	4.315	1.219	0%	100%	-2.78	1.22	
23	1.226	5.451	1.248	2.0%	90.8%	-3.44	0.98	
24	1.195	2.432	1.195	0%	88.6%	-1.49	1.08	
25	1.220	2.409	1.262	6%	91.7%	-2.15	1.47	

† Indique qu'on a jugé trop grande la différence entre le facteur de récupération et sa valeur estimée.

* Indique qu'on a jugé trop grande la différence entre la teneur en cendre et sa valeur estimée.

Comme le modèle, dans sa forme présente, ne peut pas tenir compte du niveau de salissage, il faut introduire un facteur de dilution. Supposons que l'échantillon contienne p % de charbon de la couche, le reste étant du stérile. On ajuste la distribution lognormale aux p% du charbon et non au total de 100 %.

Un exemple suffit pour montrer l'importance du paramètre p. Considérons la différence entre la qualité de l'ajustement pour le charbon N° 6, d'abord sans le paramètre p, puis quand p prend la valeur de 24% (Figure 36); l'amélioration est frappante.

Ayant introduit le paramètre p, on peut considérer le critère sur lequel on juge de la qualité de l'ajustement du modèle. Soit l_i la variable après transformation logarithmique :

$$l_i = \log \left(\frac{X_i - \rho_c}{\rho_a - X_i} \right) \quad (41)$$

Soit y_i le pourcentage de matière flottant à une densité x_i . Soit $g(u)$ la densité de la loi normale, et $G(u)$ sa fonction de répartition. Soit G^{-1} la transformation inverse de G . Le problème est d'ajuster aux données une distribution lognormale et de juger la qualité de l'ajustement. On peut le faire par la minimisation de la somme des carrés des écarts entre le pourcentage cumulé expérimental (après correction pour le stérile) et la valeur estimée à partir du modèle lognormal. C'est à dire qu'on cherche à minimiser :

$$Q = \sum_{i=1}^n \left[p G\left(\frac{l_i - \mu}{\sigma}\right) - y_i \right]^2 \quad (42)$$

où μ et σ^2 sont la moyenne et la variance de la distribution normale. Après avoir dérivé l'équation (42) par rapport aux variables à estimer (p , μ et σ), on obtient trois équations :

$$p \sum G\left(\frac{l_i - \mu}{\sigma}\right)^2 = \sum y_i G\left(\frac{l_i - \mu}{\sigma}\right) \quad (43)$$

$$p \sum g\left(\frac{l_i - \mu}{\sigma}\right) G\left(\frac{l_i - \mu}{\sigma}\right) = \sum y_i g\left(\frac{l_i - \mu}{\sigma}\right) \quad (44)$$

$$p \sum l_i g\left(\frac{l_i - \mu}{\sigma}\right) G\left(\frac{l_i - \mu}{\sigma}\right) = \sum l_i g\left(\frac{l_i - \mu}{\sigma}\right) \quad (45)$$

FIGURE 36 : Facteur de recuperation

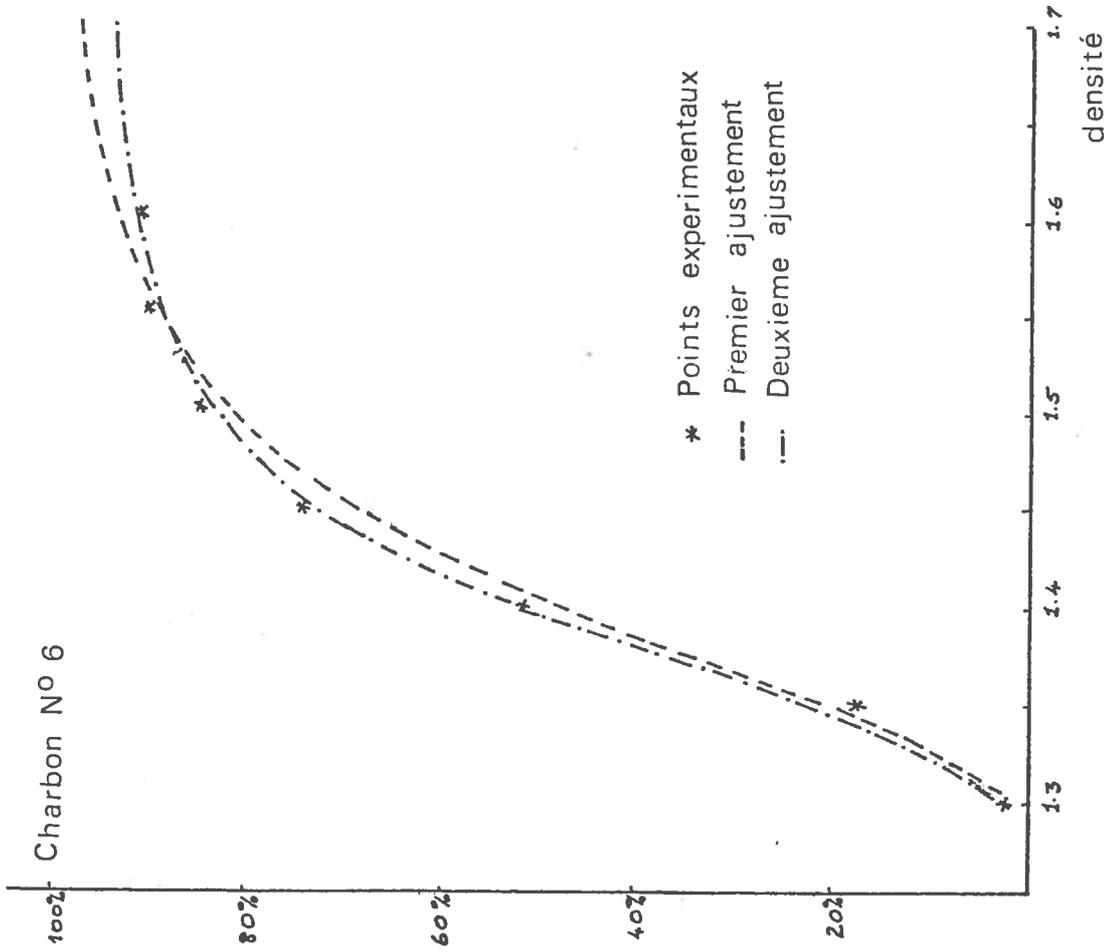
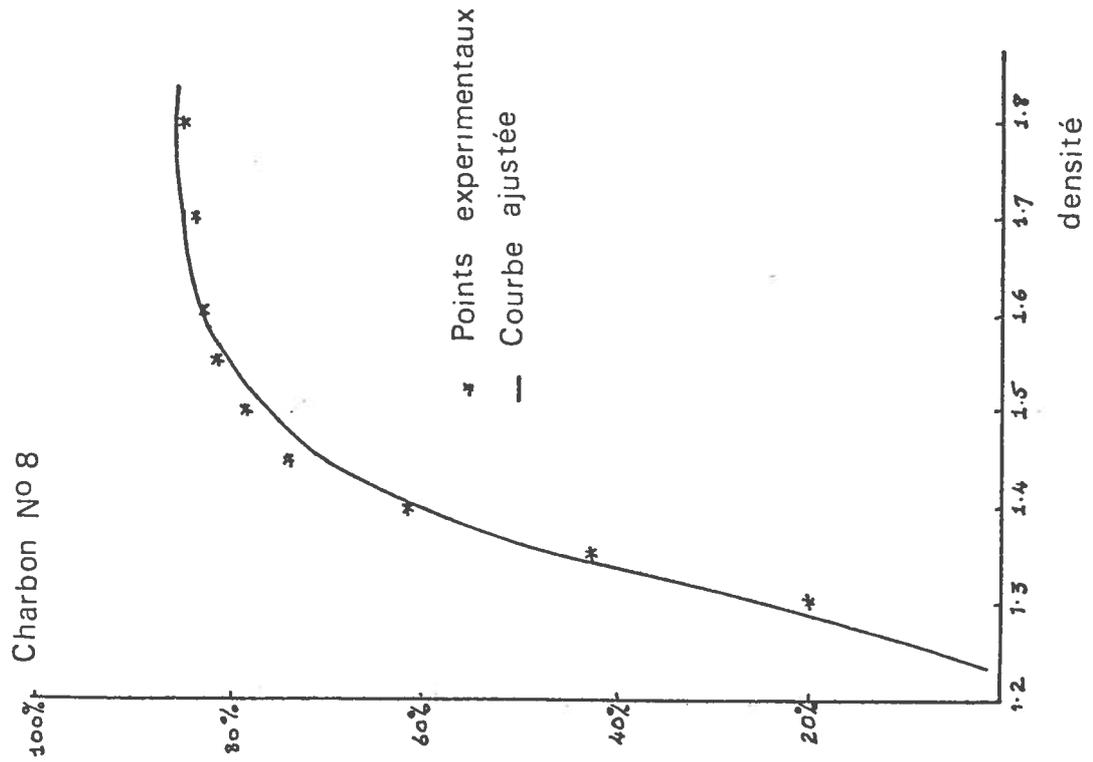


FIGURE 37 : Facteur de recuperation



Ce système ne semble pas avoir une solution explicite classique, et bien qu'on puisse le résoudre itérativement par des méthodes numériques, cela reste difficile. Cependant, il existe une autre approche à ce problème, qui est de transformer les facteurs de récupération en équivalents gaussiens, c'est-à-dire de prendre la transformation inverse de la distribution de Gauss. Par exemple, un facteur de récupération de 50 % correspond, après cette transformation, à 0 équivalent gaussien, tandis que celui de 84 % correspond à 2 équivalents gaussiens, etc... Les valeurs d'une variable gaussienne tracées en fonction de ces unités transformées forment une droite. Exprimée en ces unités transformées, la somme des carrés des écarts est :

$$Q' = \sum \left(\frac{x_i - \mu}{\sigma} - Z_i \right)^2 \quad (46)$$

où

$$Z_i = G^{-1} \left(\frac{y_i}{p} \right)$$

Après dérivation par rapport à μ , σ et p , on obtient trois équations :

$$\sum x_i^2 - \sigma \sum x_i Z_i - \mu \sum x_i = 0 \quad (47)$$

$$\sum x_i - \sigma \sum Z_i - n\mu = 0 \quad (48)$$

$$\sum (x_i - \mu - \sigma Z_i) \frac{\partial Z_i}{\partial p} = 0 \quad (49)$$

Les équations (47) et (48) sont les équations d'une régression linéaire pour ajuster une droite aux données transformées, du moins pour p fixé. L'équation (49) n'est pas aussi simple à résoudre, mais malgré cette difficulté, on peut résoudre le système d'équations en fixant la valeur de p et en minimisant Q' pour cette valeur de p , puis en faisant varier p . La valeur minimale de Q' obtenue de cette façon est évidemment le minimum absolu de Q' .

Il faut noter que l'on obtient des estimateurs de μ , σ et p différents par deux systèmes d'équations. La métrique de l'espace de Q est :

$$p G \left(\frac{x_i - \mu}{\sigma} \right) - y_i$$

tandis que celle de Q' est :

$$\frac{x_i - \mu}{\sigma} = G^{-1}\left(\frac{y_i}{p}\right)$$

Ni l'une ni l'autre de ces deux métriques n'est intrinsèquement meilleure : elles sont associées à des espaces différents. Comme les équations associées au deuxième espace sont beaucoup plus faciles à résoudre que celles du premier, on adopte la deuxième métrique. La procédure d'ajustement du modèle lognormal est donc une minimisation en deux étapes de la somme des carrés des écarts, d'abord pour une valeur fixée de p et ensuite en faisant varier p. La première étape correspond à l'ajustement d'une droite de régression aux données après qu'elles aient été tracées sur papier gauss-logarithmique. On remarque que, lors de l'utilisation du papier gauss-logarithmique, l'ajustement est souvent fait à la main, mais ici, on le fait par régression parce que cette méthode n'est pas subjective et parce qu'on peut le calculer par ordinateur, ce qui est plus rapide.

La minimisation à deux étapes a été faite par ordinateur pour les 28 ensembles de données et elle a convergé assez rapidement, en général après 6 ou 7 itérations. Dans quelques cas, on a eu des difficultés pour trouver le minimum exact parce que la concavité était faible. Mais on a toujours pu déterminer le minimum dans une fourchette de valeurs de $\pm 2\%$ et, par suite, p dans une fourchette de $\pm 0,01$. Dans tous les cas sauf deux exceptions, on a jugé bon l'ajustement du modèle lognormal aux données. Un exemple typique de l'ajustement est le charbon N° 8 (Figure 37).

5-3-3 Tests simultanés des deux modèles

Si le modèle lognormal et celui à deux paramètres sont tous les deux justes, il doit être possible d'estimer la teneur cumulée en cendre pour chacune des densités des analyses de lavabilité. Soit $f(t)$ la densité de distribution de x. Le facteur de récupération à une densité x_i est donc :

$$p \int_{p_c}^{x_i} f(t) dt$$

et la quantité de cendre incluse est

$$p \int_{\rho_c}^{x_i} a(t) f(t) dt$$

La teneur en cendre est donc le rapport entre ces deux quantités. On a estimé la teneur en cendre à chaque densité de lavage pour l'ensemble des 28 charbons australiens. Bien que les prédictions soient bonnes dans un grand nombre de cas (par exemple pour le charbon N° 3 (classe granulométrique de 4"), (Tableau 18(a)), dans d'autres cas elles étaient sous-estimées, et cela pour toutes les densités. Voir le charbon N° 1 (Tableau 18(b)).

Un examen plus précis de ces cas anormaux suggère que le mauvais ajustement est peut être dû à une hypothèse énoncée sur la valeur de A_{\min} . A la section 5-3-1, on a supposé que cette valeur était 0 %. On a donc recalculé les teneurs estimées en cendre pour des valeurs de densité de charbon "pur" correspondant à une série de valeurs croissantes de A_{\min} . (Pour ne pas les confondre, soit ρ_{MIN} la valeur de la densité du charbon pur correspondant à une valeur de A_{\min} supérieure à 0 %; le paramètre ρ_c correspondant donc à A_{\min} égal à 0%). Par cette procédure d'ajustement dont un exemple détaillé est donné par ARMSTRONG et WHITMORE (1979), on a obtenu de bons résultats dans la grande majorité des cas. Les prédictions des facteurs de récupération sont proches des valeurs expérimentales dans tous les cas sauf deux exceptions (N°s 5 et 12) tandis que celles des teneurs en cendre ne diffèrent pas plus de 1,4 % des valeurs expérimentales à toutes les densités de prédiction sauf 5 exceptions (N°s 5, 12, 13, 19, 20), une différence de 1,4 % étant jugée acceptable.

On a examiné ces 5 cas pour déterminer pourquoi le modèle a échoué. On a d'abord recalculé les courbes de teneurs en cendre en fonction de la densité, mais comme les points expérimentaux se trouvent très près des courbes ajustées, cela ne semble pas être la source de l'erreur. On s'est ensuite intéressé au modèle lognormal. On a comparé les valeurs expérimentales de $(x_i - \rho_{\text{MIN}})/(\rho_a - x_i)$ au pourcentage cumulé de charbon de densité de moins de x_i (après avoir tenu compte du facteur de dilution p). Tracés sur papier log-probabilité, les points doivent être alignés, mais dans les cas N°s 5 et 12, ceci n'est pas vérifié, ce qui suggère que les résultats dans ces deux cas sont dus à un modèle lognormal mal inadapté. Contrairement à ces deux cas, les points sont bien alignés pour les trois autres exceptions. La raison de ces mauvaises prédictions n'est donc pas encore claire.

Densité	Facteur de récupération (estimé)	Facteur de récupération (expérimental)	Teneur en cendre (estimée)	Teneur en cendre (expérimentale)
1,3	0,169	0,176	0,0491	0,0500
1,35	0,379	0,373	0,0835	0,0827
1,4	0,512	0,507	0,1089	0,1090
1,45	0,597	0,598	0,1287	0,1292
1,5	0,651	0,652	0,1436	0,1433
1,55	0,688	0,687	0,1553	0,1544
1,6	0,714	0,711	0,1645	0,1631

TABLEAU 18 (a) : Charbon N° 3 (4").

Densité	Facteur de récupération (estimé)	Facteur de récupération (expérimental)	Teneur en cendre (estimée)	Teneur en cendre (expérimentale)
1,3	0,623	0,629	0,0181	0,0490
1,35	0,731	0,718	0,0283	0,0569
1,4	0,784	0,775	0,0361	0,0633
1,45	0,814	0,812	0,0420	0,0688
1,5	0,834	0,837	0,0470	0,0739
1,55	0,848	0,846	0,0507	0,0762
1,6	0,859	0,858	0,0544	0,0799
1,7	0,874	0,870	0,0599	0,0844

TABLEAU 18 (b) : Charbon N° 1.

En conclusion, le modèle lognormal et celui à deux paramètres ont donc fourni de bonnes prédictions du facteur de récupération et de la teneur en cendre pour 23 des 28 cas étudiés, et dans trois autres cas, les prédictions du facteur de récupération (mais non de la teneur en cendre) ont été acceptables. Les prédictions sont présentées en détail au Tableau 19.

Les valeurs de tous les paramètres (ρ_c , ρ_a , ρ_{MIN} , A_{min} , p , μ et σ) sont présentées au Tableau 17. Au vu de ces valeurs, il est évident qu'il existe de fortes corrélations entre certains paramètres. On a fait une analyse statistique sur ces paramètres mais on ne la présentera pas parce que les résultats sont très semblables à ceux des sondages de 8" de diamètre, qui sont présentés à la section 5-4 et dans l'annexe.

Le but de cette étude est de développer un modèle paramétrique qui puisse être utilisé pour l'estimation du facteur de récupération et la teneur en cendre après lavage du charbon in-situ. Pour ce faire, le nombre de paramètres doit être réduit, et ceci pour deux raisons. Premièrement, les analyses de routine du charbon en provenance des carottes ne sont effectuées que pour quelques densités, et par suite on ne peut estimer que 2 ou 3 paramètres pour chaque échantillon. Deuxièmement, même si on peut estimer un nombre plus élevé de paramètres, on ne pourrait pas étudier leur régionalisation. Pour ces deux raisons, il faut limiter à 1 ou 2 le nombre des paramètres à estimer pour chaque sondage. Par conséquent, il faut étudier les relations entre les 7 paramètres pour voir comment réduire leur nombre. On pourrait étudier les valeurs obtenues pour les 28 charbons australiens, mais il serait plus fructueux d'étudier des échantillons pris dans un même gisement. Cela élimine la variabilité due aux différentes techniques d'échantillonnage ou aux différences entre les types de charbon.

5-4 Analyse statistique sur les valeurs des paramètres

On dispose d'un deuxième ensemble de données de lavabilité provenant d'une série de sondages de 8" de diamètre faits à la mine. Comme ces échantillons proviennent de la même couche et de la même mine, on peut étudier la régionalisation des valeurs des paramètres. Les résultats complets de cette étude étant développés dans l'annexe, on n'en présentera qu'un résumé dans cette section. Les conclusions principales sont :

- 1) La valeur de ρ_c est constante dans des régions assez larges.

TABLEAU 19

*** COAL NO 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

*** COAL NO 1
 RHC = 1.2500 RHA = 2.8220
 ASH MIN = .0470 MIN DENSITY = 1.2875
 AV = -6.0405 SD = 4.2505
 PROP = .9931
 DENSITY OBSERVED PREDICTED OBSERVED PREDICTED
 ASH ASH YIELD YIELD
 1.3000 .0490 .0193 .6290 .6099
 1.3500 .0569 .0553 .7180 .7435
 1.4000 .0633 .0603 .7750 .7883
 1.4500 .0688 .0644 .8120 .8128
 1.5000 .0739 .0687 .8370 .8329
 1.5500 .0762 .0718 .8460 .8447
 1.6000 .0799 .0759 .8580 .8581
 1.7000 .0844 .0818 .8700 .8747

SS FOR LOGNORMAL = .21322 SS FOR ASH = .00010
 *** COAL NO 2
 RHC = 1.2740 RHA = 2.5280
 ASH MIN = .0860 MIN DENSITY = 1.3374
 AV = -3.6661 SD = 3.8968
 PROP = .9966
 DENSITY OBSERVED PREDICTED OBSERVED PREDICTED
 ASH ASH YIELD YIELD
 1.4000 .1000 .0998 .5750 .5733
 1.4500 .1098 .1089 .6310 .6386
 1.5000 .1183 .1162 .6750 .6752
 1.5500 .1258 .1234 .7080 .7033
 1.6000 .1297 .1305 .7200 .7268

SS FOR LOGNORMAL = .00997 SS FOR ASH = .00001
 *** COAL NO 3 BLEND
 RHC = 1.2560 RHA = 2.3040
 ASH MIN = .0400 MIN DENSITY = 1.2819
 AV = -2.1709 SD = 2.4574
 PROP = .9936
 DENSITY OBSERVED PREDICTED OBSERVED PREDICTED
 ASH ASH YIELD YIELD
 1.3000 .0520 .0504 .2380 .2222
 1.3500 .0801 .0719 .4150 .4182
 1.4000 .0960 .0902 .4970 .5165
 1.4500 .1163 .1065 .5830 .5833
 1.5000 .1318 .1211 .6400 .6329
 1.5500 .1424 .1346 .6740 .6732
 1.6000 .1533 .1477 .7030 .7079

SS FOR LOGNORMAL = .02825 SS FOR ASH = .00004
 *** COAL NO 4
 RHC = 1.2020 RHA = 3.1980
 ASH MIN = .0320 MIN DENSITY = 1.2293
 AV = -5.6092 SD = 1.9414
 PROP = .9967
 DENSITY OBSERVED PREDICTED OBSERVED PREDICTED
 ASH ASH YIELD YIELD
 1.3000 .0450 .0455 .8799 .8799
 1.3500 .0521 .0498 .9390 .9275
 1.4000 .0539 .0527 .9520 .9496
 1.4500 .0564 .0550 .9670 .9626
 1.5000 .0580 .0568 .9740 .9708
 1.5500 .0586 .0582 .9760 .9762
 1.6000 .0595 .0593 .9790 .9802
 1.7000 .0605 .0611 .9820 .9857
 1.8000 .0633 .0624 .9890 .9888
 1.9000 .0637 .0634 .9900 .9910
 2.0000 .0642 .0642 .9910 .9926

SS FOR LOGNORMAL = .01023 SS FOR ASH = .00013
 *** COAL NO 5
 RHC = 1.2580 RHA = 2.2690
 ASH MIN = .0400 MIN DENSITY = 1.2834
 AV = -2.4561 SD = 2.2519
 PROP = .8510
 DENSITY OBSERVED PREDICTED OBSERVED PREDICTED
 ASH ASH YIELD YIELD
 1.3000 .0510 .0502 .2060 .2006
 1.3500 .0800 .0727 .4050 .3986
 1.4000 .0937 .0904 .4770 .4895
 1.4500 .1119 .1065 .5540 .5517
 1.5000 .1254 .1201 .6000 .5946
 1.5500 .1353 .1326 .6280 .6290
 1.6000 .1462 .1447 .6530 .6582

*** COAL NO 8 ***

*** RHC = 1.2190 RHA = 2.8210
*** ASH MIN = .0001 MIN DENSITY = 1.2191
*** AV = -2.4621 SO = .7358
*** PROP = .8557

DENSITY	OBSERVED	PREDICTED	OBSERVED	PREDICTED
	ASH	ASH	YIELD	YIELD
1.3000	.0650	.0689	.2000	.2206
1.3500	.0919	.0972	.4230	.4448
1.4000	.1142	.1183	.6170	.6039
1.4500	.1319	.1335	.7330	.7026
1.5000	.1411	.1446	.7810	.7636
1.5500	.1479	.1523	.8100	.7997
1.6000	.1518	.1578	.8240	.8224
1.7000	.1559	.1641	.8370	.8436
1.8000	.1597	.1669	.8480	.8512

SS FOR LOGNORMAL = .07322 SS FOR ASH = .00025

*** COAL NO 9 ***

*** RHC = 1.2460 RHA = 3.3360
*** ASH MIN = .0580 MIN DENSITY = 1.2984
*** AV = -6.4434 SO = 3.4213
*** PROP = .9954

DENSITY	OBSERVED	PREDICTED	OBSERVED	PREDICTED
	ASH	ASH	YIELD	YIELD
1.4000	.0690	.0690	.8460	.8421
1.4500	.0720	.0724	.8650	.8687
1.5000	.0750	.0753	.8810	.8864
1.5500	.0769	.0782	.8900	.9007
1.6000	.0819	.0806	.9130	.9105

SS FOR LOGNORMAL = .05175 SS FOR ASH = .00000

*** COAL NO 10 ***

*** RHC = 1.2550 RHA = 2.5770
*** ASH MIN = .0450 MIN DENSITY = 1.2880
*** AV = -7.9658 SO = 6.5946
*** PROP = .9927

DENSITY	OBSERVED	PREDICTED	OBSERVED	PREDICTED
	ASH	ASH	YIELD	YIELD
1.3000	.0460	.0463	.6830	.6832
1.3500	.0524	.0503	.7710	.7678
1.4000	.0541	.0539	.7840	.7964
1.4500	.0591	.0570	.8110	.8125
1.5000	.0616	.0602	.8350	.8252
1.5500	.0658	.0651	.8350	.8350
1.6000	.0665	.0656	.8370	.8421
1.8000	.0783	.0764	.8640	.8663

SS FOR LOGNORMAL = .07617 SS FOR ASH = .00002

*** COAL NO 6 ***

*** RHC = 1.2690 RHA = 2.4550
*** ASH MIN = .0300 MIN DENSITY = 1.2898
*** AV = -3.5832 SO = 1.4911
*** PROP = .8810

DENSITY	OBSERVED	PREDICTED	OBSERVED	PREDICTED
	ASH	ASH	YIELD	YIELD
1.3000	.0520	.0375	.2070	.1944
1.3500	.0758	.0594	.4930	.5937
1.4000	.0981	.0728	.7430	.7140
1.4500	.1047	.0827	.7930	.7742
1.5000	.1095	.0898	.8190	.8072
1.5500	.1084	.0954	.8410	.8288
1.6000	.1117	.1000	.8520	.8434
1.6500	.1131	.1036	.8560	.8535
1.7000	.1150	.1064	.8610	.8606
1.7500	.1159	.1087	.8630	.8658
2.0000	.1194	.1152	.8690	.8778

SS FOR LOGNORMAL = .88751 SS FOR ASH = .00253

*** COAL NO 7 ***

*** RHC = 1.2600 RHA = 2.8900
*** ASH MIN = .0001 MIN DENSITY = 1.2601
*** AV = -2.4648 SO = .5651
*** PROP = .9388

DENSITY	OBSERVED	PREDICTED	OBSERVED	PREDICTED
	ASH	ASH	YIELD	YIELD
1.3000	.0300	.0412	.0230	.0144
1.3500	.0821	.0807	.1740	.2360
1.4000	.1207	.1094	.5210	.5326
1.4500	.1411	.1291	.7440	.7317
1.5000	.1526	.1420	.8480	.8395
1.5500	.1596	.1498	.8990	.8925
1.6000	.1615	.1543	.9100	.9176

SS FOR LOGNORMAL = .03667 SS FOR ASH = .00066

*** COAL NO 7 ***

*** RHC = 1.2440 RHA = 2.5340
*** ASH MIN = .0400 MIN DENSITY = 1.2728
*** AV = -2.5957 SO = .9934
*** PROP = .8848

DENSITY	OBSERVED	PREDICTED	OBSERVED	PREDICTED
	ASH	ASH	YIELD	YIELD
1.3000	.0530	.0644	.0860	.0969
1.3500	.0955	.0960	.4330	.3934
1.4000	.1179	.1187	.6060	.5834
1.4500	.1308	.1349	.6900	.6922
1.5000	.1432	.1474	.7560	.7607
1.5500	.1535	.1567	.8040	.8035
1.6000	.1587	.1637	.8230	.8313
1.8000	.1769	.1779	.8730	.8745

SS FOR LOGNORMAL = .02959 SS FOR ASH = .00020

*** COAL NO 12 ***

*** RHC = 1.2720 RHA = 2.7060
 *** ASH MIN = .0250 MIN DENSITY = 1.2910
 AV = -2.6028 SD = .9338
 PROP = .9767

DENSITY	OBSERVED ASH	PREDICTED ASH	OBSERVED YIELD	PREDICTED YIELD
1.3000	.0410	.0339	.0060	.0044
1.3500	.0726	.0712	.1420	.2775
1.4000	.1004	.0971	.5510	.5348
1.4500	.1155	.1164	.7920	.6492
1.5000	.1246	.1309	.8440	.7994
1.5500	.1292	.1414	.9130	.8603
1.6000	.1315	.1495	.9230	.9007
1.7000	.1365	.1600	.9410	.9433
1.8000	.1408	.1657	.9530	.9617
1.9000	.1442	.1688	.9610	.9701
2.0000	.1472	.1705	.9670	.9739

SS FOR LOGNORMAL = .88899 SS FOR ASH = .00290
 *** COAL NO 13 ***

*** RHC = 1.2650 RHA = 2.9900
 *** ASH MIN = .0001 MIN DENSITY = 1.2651
 AV = -2.5381 SD = 1.0128
 PROP = .9978

DENSITY	OBSERVED ASH	PREDICTED ASH	OBSERVED YIELD	PREDICTED YIELD
1.3000	.0400	.0285	.1360	.0917
1.3500	.0623	.0587	.3300	.3363
1.4000	.0819	.0821	.5000	.5266
1.4500	.0985	.1003	.5990	.6579
1.5000	.1169	.1151	.6950	.7503
1.6000	.1514	.1368	.8620	.8623
1.8000	.1761	.1605	.9690	.9545
2.0000	.1808	.1711	.9820	.9844

SS FOR LOGNORMAL = .14334 SS FOR ASH = .00071
 *** COAL NO 14 ***

*** RHC = 1.2880 RHA = 3.9340
 *** ASH MIN = .0550 MIN DENSITY = 1.3432
 AV = -4.6628 SD = 2.2603
 PROP = .9648

DENSITY	OBSERVED ASH	PREDICTED ASH	OBSERVED YIELD	PREDICTED YIELD
1.3500	.0630	.0575	.2830	.2742
1.4000	.0763	.0686	.6060	.6254
1.4500	.0833	.0763	.7270	.7224
1.5000	.0879	.0821	.7800	.7713
1.5500	.0914	.0876	.8040	.8068
1.6000	.0947	.0918	.8200	.8295

SS FOR LOGNORMAL = .02843 SS FOR ASH = .00020

*** COAL NO 11 12.7 ***

*** RHC = 1.2610 RHA = 2.8180
 *** ASH MIN = .0200 MIN DENSITY = 1.2767
 AV = -1.4886 SD = 1.2198
 PROP = .7659

DENSITY	OBSERVED ASH	PREDICTED ASH	OBSERVED YIELD	PREDICTED YIELD
1.3000	.0500	.0412	.0120	.0103
1.3500	.0817	.0756	.0840	.0825
1.4000	.1046	.1050	.1630	.1645
1.4500	.1292	.1310	.2420	.2416
1.5000	.1514	.1540	.3110	.3114
1.5500	.1733	.1748	.3780	.3709
1.6000	.1910	.1931	.4300	.4224
1.7000	.2227	.2247	.5150	.5078
1.8000	.2469	.2508	.5710	.5734
1.9000	.2684	.2727	.6170	.6288
2.0000	.2917	.2901	.6620	.6635

SS FOR LOGNORMAL = .00712 SS FOR ASH = .00019
 *** COAL NO 11 12.7 ***

*** RHC = 1.2590 RHA = 3.0100
 *** ASH MIN = .0275 MIN DENSITY = 1.2818
 AV = -2.0804 SD = 1.9876
 PROP = .9982

DENSITY	OBSERVED ASH	PREDICTED ASH	OBSERVED YIELD	PREDICTED YIELD
1.3000	.0490	.0380	.1250	.1054
1.3500	.0651	.0602	.2700	.2831
1.4000	.0832	.0791	.3770	.3921
1.4500	.0993	.0951	.4550	.4666
1.5000	.1135	.1090	.5180	.5257
1.5500	.1276	.1230	.5720	.5731
1.6000	.1419	.1347	.6230	.6119
1.7000	.1644	.1559	.6920	.6750
1.8000	.1807	.1754	.7350	.7265
1.9000	.1943	.1922	.7660	.7675
2.0000	.2064	.2085	.7910	.8047

SS FOR LOGNORMAL = .07837 SS FOR ASH = .00038

*** COAL NO 15 ***

RHC = 1.3470 RHA = 1.8470
 *** ASH MIN = .0940 MIN DENSITY = 1.3862
 AV = -6.5379 SD = 5.1523
 PROP = .9917

DENSITY	OBSERVED	PREDICTED	OBSERVED	PREDICTED	YIELD	PREDICTED	YIELD
1.4000	.0960	.0972	.7220	.7181			
1.4500	.1070	.1054	.7990	.8133			
1.5000	.1150	.1126	.8460	.8476			
1.5500	.1219	.1194	.8710	.8692			
1.6000	.1291	.1265	.8820	.8868			

SS FOR LOGNORMAL = .07679 SS FOR ASH = .00002

*** COAL NO 16 ***

RHC = 1.2290 RHA = 4.3710
 *** ASH MIN = .0300 MIN DENSITY = 1.2592
 AV = -3.1809 SD = 1.4135
 PROP = .9959

DENSITY	OBSERVED	PREDICTED	OBSERVED	PREDICTED	YIELD	PREDICTED	YIELD
1.3500	.0650	.0695	.4100	.4075			
1.4500	.1019	.0967	.6260	.6231			
1.5500	.1199	.1160	.7240	.7361			
1.6000	.1290	.1236	.7750	.7732			

SS FOR LOGNORMAL = .00237 SS FOR ASH = .00009

*** COAL NO 17 ***

RHC = 1.2870 RHA = 2.9320
 *** ASH MIN = .0400 MIN DENSITY = 1.3199
 AV = -3.5907 SD = 1.0305
 PROP = .9073

DENSITY	OBSERVED	PREDICTED	OBSERVED	PREDICTED	YIELD	PREDICTED	YIELD
1.3500	.0580	.0599	.3250	.3260			
1.4500	.0960	.0904	.6000	.7860			
1.5500	.1098	.1025	.8590	.8695			
1.6000	.1151	.1056	.8840	.8842			

SS FOR LOGNORMAL = .01033 SS FOR ASH = .00018

*** COAL NO 19 ***

RHC = 1.2370 RHA = 3.4910
 *** ASH MIN = .0001 MIN DENSITY = 1.2371
 AV = -2.4259 SD = .7797
 PROP = .9023

DENSITY	OBSERVED	PREDICTED	OBSERVED	PREDICTED	YIELD	PREDICTED	YIELD
1.3500	.0660	.0807	.2390	.2268			
1.4000	.0929	.1050	.3850	.3904			
1.4500	.1179	.1251	.5200	.5261			
1.5000	.1370	.1409	.6280	.6273			
1.5500	.1519	.1510	.7110	.7032			
1.6000	.1589	.1642	.7470	.7567			

SS FOR LOGNORMAL = .00278 SS FOR ASH = .00046

*** COAL NO 20 ***

RHC = 1.2540 RHA = 3.7140
 *** ASH MIN = .0400 MIN DENSITY = 1.2920
 AV = -2.9537 SD = .7518
 PROP = .9958

DENSITY	OBSERVED	PREDICTED	OBSERVED	PREDICTED	YIELD	PREDICTED	YIELD
1.3500	.0820	.0807	.1990	.1558			
1.4000	.1062	.1054	.4010	.4388			
1.4500	.1272	.1233	.6020	.6455			
1.5000	.1401	.1367	.7610	.7790			
1.5500	.1490	.1462	.8660	.8583			
1.6000	.1500	.1533	.9190	.9092			

SS FOR LOGNORMAL = .03186 SS FOR ASH = .00006

*** COAL NO 18 ***

RHC = 1.2290 RHA = 4.3710
 *** ASH MIN = .0300 MIN DENSITY = 1.2592
 AV = -3.1809 SD = 1.4135
 PROP = .9959

DENSITY	OBSERVED	PREDICTED	OBSERVED	PREDICTED	YIELD	PREDICTED	YIELD
1.3500	.0650	.0695	.4100	.4075			
1.4500	.1019	.0967	.6260	.6231			
1.5500	.1199	.1160	.7240	.7361			
1.6000	.1290	.1236	.7750	.7732			

SS FOR LOGNORMAL = .00237 SS FOR ASH = .00009

*** COAL NO 21 ***

*** RHC = 1.1990 RHA = 4.2470
 *** ASH MIN = .0580 MIN DENSITY = 1.2571
 *** AV = -5.8569 SD = 3.0676
 *** PROP = .8831
 DENSITY OBSERVED PREDICTED PREDICTED
 ASH ASH YIELD YIELD
 1.3500 .0700 .0703 .6890 .6909
 1.4000 .0750 .0747 .7260 .7274
 1.4500 .0782 .0782 .7460 .7492
 1.5000 .0818 .0814 .7630 .7651
 1.5500 .0851 .0844 .7770 .7779
 1.6000 .0873 .0872 .7850 .7881
 SS FOR LOGNORMAL = .00120 SS FOR ASH = .00000

*** COAL NO 22 ***

*** RHC = 1.2190 RHA = 4.3140
 *** ASH MIN = .0001 MIN DENSITY = 1.2191
 *** AV = -2.7811 SD = 1.0232
 PROP = 1.0000
 DENSITY OBSERVED PREDICTED PREDICTED
 ASH ASH YIELD YIELD
 1.3000 .0360 .0498 .2200 .2062
 1.3500 .0601 .0715 .3600 .3669
 1.4000 .0820 .0900 .4910 .5000
 1.4500 .1021 .1049 .6100 .5987
 1.5000 .1151 .1178 .6810 .6775
 1.5500 .1261 .1288 .7360 .7391
 1.6000 .1341 .1379 .7810 .7852
 SS FOR LOGNORMAL = .00502 SS FOR ASH = .00042

*** COAL NO 23 ***

*** RHC = 1.2260 RHA = 5.4510
 *** ASH MIN = .0200 MIN DENSITY = 1.2475
 *** AV = -3.4360 SD = .9766
 PROP = .9081
 DENSITY OBSERVED PREDICTED PREDICTED
 ASH ASH YIELD YIELD
 1.3000 .0380 .0502 .1340 .1533
 1.3500 .0695 .0708 .4260 .3610
 1.4000 .0797 .0867 .5290 .5116
 1.4500 .0902 .0992 .6160 .6153
 1.5000 .1008 .1094 .6790 .6885
 1.5500 .1094 .1177 .7260 .7408
 1.6000 .1173 .1241 .7640 .7766
 1.7000 .1301 .1345 .8260 .8277
 1.8000 .1364 .1415 .8490 .8564
 1.9000 .1431 .1465 .8730 .8712
 2.0000 .1470 .1499 .8850 .8851
 SS FOR LOGNORMAL = .05110 SS FOR ASH = .00053

*** COAL NO 24 ***

*** RHC = 1.1950 RHA = 2.4230
 *** ASH MIN = .0001 MIN DENSITY = 1.1951
 *** AV = -1.4896 SD = 1.0610
 PROP = .8860
 DENSITY OBSERVED PREDICTED PREDICTED
 ASH ASH YIELD YIELD
 1.3000 .0860 .0912 .1600 .1827
 1.4000 .1517 .1513 .4420 .4045
 1.5000 .1971 .1960 .6180 .5644
 1.6000 .2210 .2305 .6970 .6772
 1.7000 .2378 .2565 .7430 .7537
 1.8000 .2545 .2766 .7830 .8076
 1.9000 .2736 .2909 .8250 .8424
 2.0000 .2929 .3009 .8670 .8646
 SS FOR LOGNORMAL = .12859 SS FOR ASH = .00132

*** COAL NO 25 ***

*** RHC = 1.2200 RHA = 2.4090
 *** ASH MIN = .0600 MIN DENSITY = 1.2615
 *** AV = -2.1506 SD = 1.4662
 PROP = .9167
 DENSITY OBSERVED PREDICTED PREDICTED
 ASH ASH YIELD YIELD
 1.3000 .0850 .0881 .1800 .1864
 1.4000 .1517 .1370 .5200 .4983
 1.5000 .1826 .1711 .6580 .6500
 1.6000 .2062 .1971 .7400 .7405
 1.7000 .2253 .2169 .7940 .7980
 1.8000 .2410 .2352 .8330 .8398
 1.9000 .2526 .2458 .8600 .8686
 2.0000 .2662 .2553 .8890 .8881
 SS FOR LOGNORMAL = .02856 SS FOR ASH = .00074

- 2) On peut prendre 1 comme valeur de p .
- 3) Les paramètres peuvent être divisés en 2 groupes indépendants :
 ρ_c, ρ_a d'une part, μ, σ, A_{\min} d'autre part.
- 4) La valeur de ρ_{MIN} est donc déterminée par les valeurs de ρ_c , de ρ_a et de A_{\min} et par l'équation (35).

On peut maintenant traiter le problème fondamental de la section, l'estimation du facteur de récupération et de la teneur en cendre de charbon lavé. Pour ce faire, on considère la question à partir des principes de base.

5-5 Le problème fondamental

Il est évident qu'il faut deux relations pour caractériser la lavabilité du charbon. La première est la relation entre la teneur en cendre $a(x)$ d'une particule de charbon et sa densité x . La deuxième peut être soit la distribution en masse des densités des particules, soit celle des teneurs en cendre. (La première relation donne une équivalence entre ces deux distributions). Soit $f(x)$ la distribution des densités des particules et soit d la densité de séparation. On voudrait donc estimer les quantités suivantes :

$$\text{Quantité récupérée} = \int_0^d f(x) dx \quad (50)$$

$$\text{Quantité de cendre récupérée} = \int_0^d a(x) f(x) dx \quad (51)$$

En écrivant ces équations, on fait implicitement l'hypothèse que la séparation est parfaite, c'est-à-dire que toutes les particules de densité de moins de d sont récupérées et aucune autre. En pratique, le pourcentage de particules affectées de façon erronée dépend du type de séparateur, de la densité de séparation et de la granulométrie des particules; mais dans cette étude, on ne tient pas compte du type de séparateur ni de la granulométrie.

Soit $E_d(x)$ la probabilité qu'une particule de densité x soit récupérée avec le "charbon" quand la densité de séparation est d . Pour une séparation

parfaite (ou une séparation effectuée en laboratoire), on a :

$$E_d(x) = \begin{cases} 1 & x \leq d \\ 0 & x > d \end{cases} \quad (52)$$

mais dans la pratique on a :

$$E_d(x) \begin{cases} > 1/2 & x < d \\ \approx 1/2 & x = d \\ < 1/2 & x > d \end{cases} \quad (53)$$

Dans les deux cas, $E_d(x)$ est une fonction monotone décroissante. La forme de cette fonction est connue par expérience pour les différents types de séparateurs. Tenant compte des incertitudes de la séparation, on a les équations suivantes :

$$\text{Quantité récupérée} = \int_0^{\infty} E_d(x) f(x) dx \quad (54)$$

$$\text{Quantité de cendre récupérée} = \int_0^{\infty} E_d(x) a(x) f(x) dx \quad (55)$$

Dans les sections 5-3-2 et 5-3-3, on a montré que la variable $(X - \rho_{\text{MIN}})/(\rho_a - X)$ a une distribution lognormale de la forme :

$$\log \left(\frac{X - \rho_{\text{MIN}}}{\rho_a - X} \right) \sim N(\mu, \sigma^2)$$

On a donc intérêt à prendre la transformée :

$$Z = \log \left(\frac{X - \rho_{\text{MIN}}}{\rho_a - X} \right) \quad (56)$$

Soit $TE(Z)$ et $TA(Z)$ les transformées des fonctions $E_d(x)$ et $a(x)$ respectivement. Les équations (54) et (55) deviennent :

$$\text{Quantité récupérée} = \int_{-\infty}^{\infty} TE(Z) g\left(\frac{Z - \mu}{\sigma}\right) dz \quad (57)$$

$$\text{Quantité de cendre récupérée} = \int_{-\infty}^{\infty} TE(Z) TA(Z) g\left(\frac{Z - \mu}{\sigma}\right) dz \quad (58)$$

où $g(\cdot)$ est la dérivée de la loi normale. Les deux équations (57) et (58) sont des intégrales de la forme :

$$I = \int_{-\infty}^{\infty} B(z) g\left(\frac{z - \mu}{\sigma}\right) dz \quad (59)$$

Notons que si l'expérience montre que la distribution lognormale ne convient pas à un charbon particulier, on peut choisir une autre distribution. Il faudrait alors une autre transformation pour normaliser la variable, ce qui changerait la forme exacte des fonctions $TE(z)$ et $TA(z)$, mais les deux quantités à estimer seraient, néanmoins, des intégrales de la forme (59).

Comme on ne voit pas, de façon évidente, comment l'intégrale dépend des paramètres du modèle et comment elle varie dans l'espace, il serait peut-être avantageux de la développer en série. On pourrait chercher un développement polynomial, mais, dans la mesure où on a affaire à la loi normale, il est probablement plus instructif et plus pratique de la développer en polynômes d'Hermite. Les polynômes sont définis par les équations suivantes :

$$\begin{aligned} H_0(z) &= 1 \\ H_n(z) g(z) &= \frac{d^n}{dz^n} g(z) \quad n > 0 \end{aligned} \quad (60)$$

Rappelons que cette famille de polynômes est orthogonale par rapport à la distribution gaussienne :

$$\int_{-\infty}^{\infty} H_n(z) H_m(z) g(z) dz = \begin{cases} 0 & m \neq n \\ n! & m = n \end{cases} \quad (61)$$

Jusqu'à ce point, la discussion est restée générale. Il faut maintenant tenir compte des résultats de l'analyse statistique sur les valeurs des paramètres, faite dans la section 5-4 et dans l'annexe. La première conclusion était que la valeur de ρ_c était constante dans des régions assez larges. De plus, on supposera que les valeurs de ρ_a et de A_{\min} sont également constantes sur la même région que ρ_c . Cela détermine la valeur de ρ_{\min} et par suite la fonction $a(x)$ ainsi que la transformation $Z = \log(X - \rho_{\min}) / (\rho_a - X)$. En conséquence, la fonction $B(z)$ est complètement déterminée, avant même le début de l'étape géostatistique.

En regardant l'équation (59), on voit que la régionalisation de l'intégrale I ne dépend que de celle de μ et de σ (qui interviennent dans l'intégrale par l'intermédiaire de $g(\frac{z-\mu}{\sigma})$), et non de $B(z)$, ce qui simplifie le problème. On peut penser à développer l'intégrale I en une série double de polynômes d'Hermite en μ et en σ , mais ce serait beaucoup plus facile si l'on pouvait réduire le problème à une seule variable. Pour cela on profite de la forte corrélation notée au cours de l'analyse statistique entre μ et σ , ce qui suggère l'hypothèse qu'il y a une relation linéaire entre les deux variables :

$$\sigma = k \mu + c \quad (63)$$

L'intégrale I ne dépend plus que d'un paramètre μ . Si on fait l'hypothèse que μ est une variable gaussienne $N(\mu_0, \sigma_0^2)$, on peut développer l'intégrale I en série de polynômes d'Hermite en μ . (On essaiera de contrôler cette hypothèse à la section 5-6).

Par conséquent, on a

$$I = \sum \frac{\psi_n}{n!} H_n\left(\frac{\mu - \mu_0}{\sigma}\right) \quad (64)$$

où

$$\begin{aligned} \psi_n &= \int_{-\infty}^{\infty} I H_n\left(\frac{\mu - \mu_0}{\sigma}\right) g\left(\frac{\mu - \mu_0}{\sigma_0}\right) d\mu \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} B(z) g\left(\frac{z - \mu}{\sigma}\right) dz H_n\left(\frac{\mu - \mu_0}{\sigma_0}\right) g\left(\frac{\mu - \mu_0}{\sigma_0}\right) d\mu \\ &= \int_{-\infty}^{\infty} B(z) \int_{-\infty}^{\infty} g\left(\frac{z - \mu}{\sigma}\right) H_n\left(\frac{\mu - \mu_0}{\sigma_0}\right) g\left(\frac{\mu - \mu_0}{\sigma_0}\right) d\mu dz \quad (65) \end{aligned}$$

Soit $\phi_n(z)$ le coefficient de $H_n\left(\frac{\mu - \mu_0}{\sigma_0}\right)$ dans le développement de $g\left(\frac{z - \mu}{\sigma}\right)$ en série de polynômes d'Hermite. Alors :

$$\phi_n(z) = \int_{-\infty}^{\infty} g\left(\frac{z - \mu}{\sigma}\right) H_n\left(\frac{\mu - \mu_0}{\sigma_0}\right) g\left(\frac{\mu - \mu_0}{\sigma_0}\right) d\mu \quad (66)$$

D'où

$$\psi_n = \int_{-\infty}^{\infty} B(z) \phi_n(z) dz \quad (67)$$

On a donc réussi à développer I en série de polynomes d'Hermite; on peut déterminer les coefficients ψ_n par une intégration numérique une fois que la fonction B(z) ainsi que les valeurs de μ_0 et de σ_0 sont connues. Il faut noter que, bien que ces coefficients dépendent des valeurs de σ_0 et de μ_0 , ils ne dépendent pas de la valeur de μ en chaque point d'échantillonnage.

L'estimation de l'intégrale I ne pose plus de nouveau problème; il suffit d'appliquer les techniques du krigeage disjonctif. Comme de bons exposés de cette méthode existent dans la littérature (voir par exemple MARECHAL (1976) et MATHERON (1976(a) et (b))), on n'a pas besoin d'entrer dans le détail ici; il suffit d'en rappeler les principes. Supposons qu'on cherche à estimer l'intégrale en un point x du gisement à partir des échantillons faits aux points x_β . Soit μ_x la valeur de μ au point x et I_x l'intégrale à estimer.

$$I_x = I(\mu_x) \quad (68)$$

L'estimateur I_x^* de I_x est donné par l'équation :

$$I_x^* = \sum_{\beta} \sum_n \frac{\psi_n}{n!} \lambda_{n\beta} H_n\left(\frac{\mu_\beta - \mu_0}{\sigma_0}\right) \quad (69)$$

où les coefficients $\lambda_{n\beta}$ sont obtenus à partir des équations :

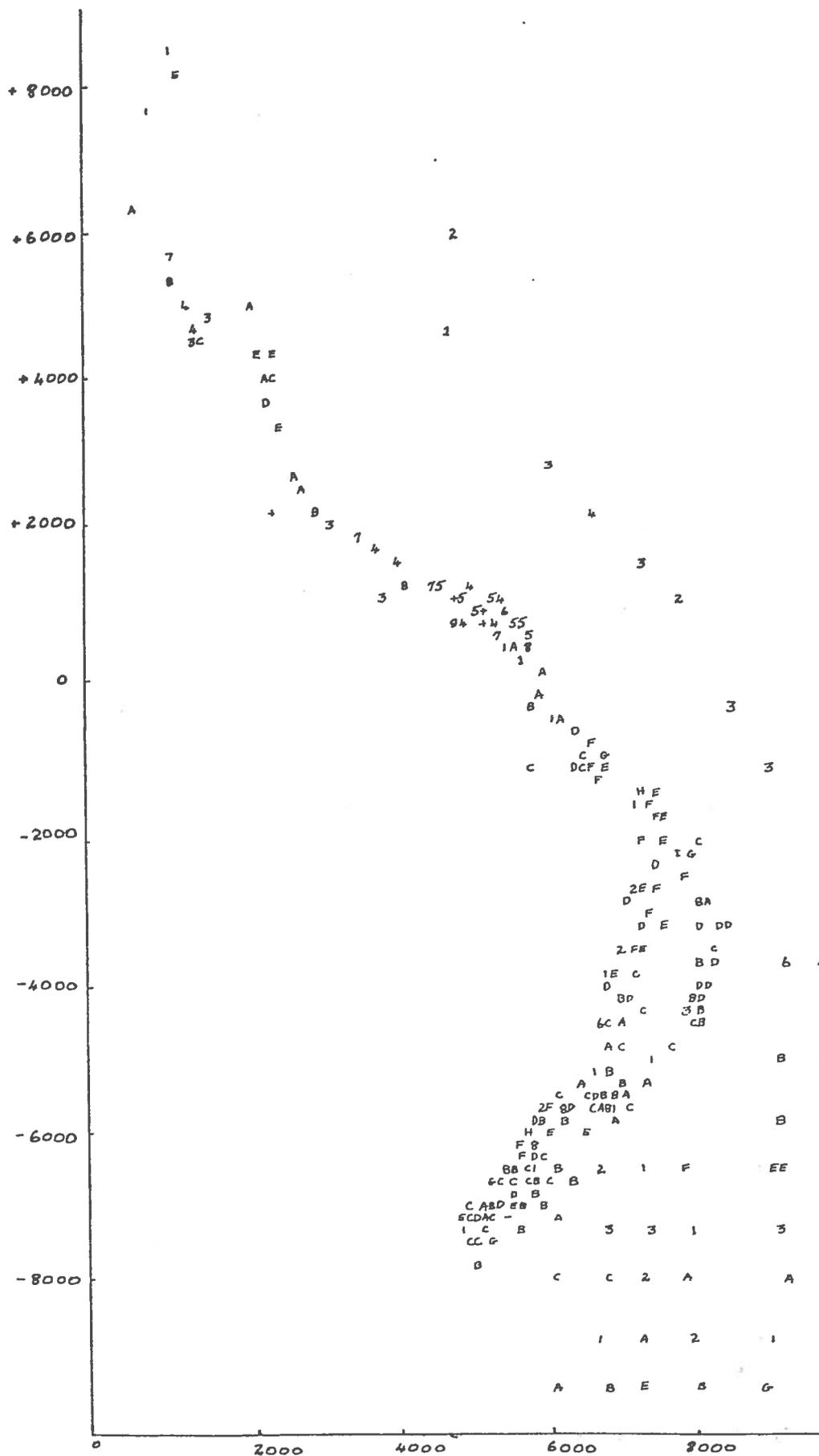
$$\sum_{\alpha} \lambda_{n\beta} \rho_{\alpha\beta}^n = \rho_{\alpha x}^n \quad (70)$$

On a donc ramené le problème de l'estimation de l'intégrale I à un domaine déjà bien connu, celui du krigeage disjonctif. Par conséquent, il n'y a aucun intérêt à développer d'avantage la procédure d'estimation; il est beaucoup plus important d'essayer de tester si la distribution de μ est, en fait, normale.

Le développement de l'intégrale I en série de polynomes d'Hermite dépend de deux hypothèses qui n'ont pas encore été vérifiées. Ce sont :

- 1) que les valeurs de ρ_a et de A_{\min} sont constantes dans la même région que ρ_c .
- 2) que la distribution de μ est gaussienne.

FIGURE 38 : Disposition des sondages



Les tentatives pour contrôler la première hypothèse (voir Annexe A4) n'ont abouti à rien parce que l'ensemble des données de lavabilité fournies par la société est peu important, mais l'hypothèse ne semble pas a priori trop forte. Par contre, la deuxième hypothèse semble moins évidente. On essaiera donc de la vérifier.

5-6 Tentative d'étude de la distribution de μ

La méthode pour étudier cette distribution n'est pas compliquée en soi. On estime les valeurs de μ en tous les points d'échantillonnage dans une région où les valeurs de ρ_c et de ρ_a sont connues, puis on utilise un test statistique tel que le test du χ^2 pour tester si les différences entre l'histogramme expérimental et la distribution gaussienne de même moyenne et de même variance que les valeurs expérimentales sont significatives.

La première étape consiste à trouver une région où les valeurs de ρ_c et de ρ_a sont connues et où on peut les supposer constantes. On a déjà estimé les valeurs de ρ_c et de ρ_a à partir des données de lavabilité provenant d'une série de sondages de 8" de diamètre, pour lesquelles les analyses ont été faites à au moins 8 densités de lavage. (Section 5-4 et annexe). On peut donc avoir confiance dans les estimateurs. La disposition de ces sondages ainsi que celle de tous les sondages de 2" de diamètre ayant été analysés à plus de deux densités est présentée à la Figure 38. On remarque que les régions où on a des estimateurs de ρ_c et de ρ_a ne coïncident pas avec la moitié Sud du gisement où se trouvent la plupart des sondages de 2" de diamètre. La région d'où proviennent les sondages N^{OS} 1155-1161 ne contient que quelques sondages tandis que les autres sondages de lavabilité (N^{OS} 901-949) proviennent d'une poche de teneurs élevées en cendre et en soufre qui n'est pas représentative du gisement. Il ne serait donc pas prudent de tester l'hypothèse dans cette région.

Au commencement de cette étude, on avait espéré trouver une région contenant un grand nombre de sondages (dont on a besoin pour estimer μ) où on connaisse les valeurs de ρ_c et de ρ_a . Finalement, on se trouve avec trois régions : la section Nord du gisement où ρ_c et ρ_a sont connus mais où il n'y a qu'une dizaine de sondages; la section médiane qui n'est pas représentative du gisement et la moitié Sud où on ne connaît pas les valeurs de ρ_c et de ρ_a . Comment peut-on procéder ?

Dans l'impossibilité d'étudier la distribution de μ dans la moitié Nord du gisement, il nous faut trouver un moyen d'estimer les valeurs de ρ_c et de ρ_a dans la moitié Sud et ensuite de tester l'hypothèse. On peut proposer trois approches possibles :

- 1) Estimer ρ_c et ρ_a pour chaque sondage, et puis calculer les valeurs moyennes.
- 2) Estimer ρ_c et ρ_a par la méthode de moindres carrés en utilisant toutes les données, et pas seulement celles provenant d'un seul sondage.
- 3) Estimer la teneur moyenne en cendre pour chaque classe de densité, et estimer ensuite les valeurs de ρ_c et de ρ_a à partir de ces moyennes.

La première de ces trois approches n'est pas réalisable parce que les analyses n'ont été effectuées qu'à 3 ou 4 densités. Les estimateurs de ces deux paramètres ne seraient pas très précis. On passe donc à la deuxième approche. Initialement, l'idée d'estimer des paramètres par moindres carrés à partir de toutes les données semble bonne; en tenant compte de tous les sondages, on minimiserait l'effet de la variabilité entre les sondages. Si on avait les teneurs en cendre pour des particules de densité connue, cette approche serait réalisable et il est probable qu'elle donnerait de bons résultats. Mais malheureusement, on ne dispose que des teneurs en cendre pour certaines classes de densité. Quand les classes sont étroites, on peut assimiler les densités des particules dans la classe à une masse de densité fixée (comme on a fait pour les données de lavabilité), mais cela n'est pas possible quand il n'y a que 3 ou 4 classes de densité.

Les deux premières approches n'étant pas réalisables, on procède à la troisième qui utilise la gamme de densités utilisée dans les analyses de lavabilité. Pendant les dix années d'exploitation de ce gisement, la compagnie a changé les densités pour les analyses de densité, les densités utilisées étant 1,3; 1,35; 1,375; 1,4; 1,45; 1,5; 1,6 et 1,8. On profiterait de cette diversité pour calculer une courbe moyenne de teneur en cendre en fonction de la densité dans la région et par suite estimer les valeurs de ρ_c et de ρ_a . Il faut d'abord décider de la manière de calculer la courbe moyenne. On a essayé deux méthodes.

Soit p_{ij} le pourcentage du charbon de l'échantillon N° j dans la $i^{\text{ème}}$ classe et a_{ij} sa teneur en cendre. La quantité de cendre dans la $k^{\text{ème}}$ classe

est $\sum_{i=1}^k p_{ij} a_{ij}$ et la quantité récupérée est $\sum_{i=1}^k p_{ij}$, ce qui implique une teneur en cendre de $\sum_{i=1}^k p_{ij} a_{ij} / \sum_{i=1}^k p_{ij}$. Après avoir pondéré par la puissance de la couche t_j , on obtient la teneur moyenne en cendre à la densité k :

$$\left. \begin{array}{l} \text{Teneur moyenne en} \\ \text{cendre à densité k} \end{array} \right\} = \frac{\sum_{j=1}^n t_j \frac{\sum a_{ij} p_{ij}}{\sum p_{ij}}}{\sum_{j=1}^n t_j} \quad (71)$$

On a calculé les teneurs moyennes en cendre pour les 8 densités. Au Tableau 21, on présente les moyennes ainsi que la somme des puissances échantillonnées dans chaque densité.

Densité	Teneur en cendre	Facteur de récupération	Puissance totale (mètres)	Teneur en cendre entre la densité et la précédente
1,3	3,776	25,075	9,17	3,776
1,35	5,781	74,704	388,77	6,793
1,375	5,979	82,151	370,19	7,974
1,4	6,774	91,220	227,71	14,009
1,45	6,979	94,578	569,41	12,438
1,5	7,451	95,799	37,25	44,032
1,6	7,396	96,555	611,12	0,442
1,8	7,975	98,676	28,08	34,348

TABLEAU 21 : Teneur moyenne en cendre pour les sondages de la moitié Sud.

On a calculé la teneur en cendre dans chacune des 8 classes de densité (1,3; 1,3 - 1,35; 1,35 - 1,375, etc...) et les valeurs sont présentées dans la dernière colonne du Tableau 21. Au lieu d'être monotones croissantes, ces valeurs décroissent pour les classes (1,4 - 1,45; 1,5 - 1,6), ce qui n'est pas du tout satisfaisant. On remarque la variabilité de la puissance totale des échantillons et on soupçonne que la

décroissance des teneurs en colonne (5) peut être dûe à cela. Après l'élimination de ces deux classes de densité, les résultats sont nettement meilleurs. (Voir Tableau 22).

classe de densité	teneur moyenne en cendre
< 1.3	3.776
1.3 - 1.35	6.793
1.35 - 1.375	7.974
1.375 - 1.45	13.580
1.45 - 1.6	27.370
1.6 - 1.8	34.348

TABLEAU 22 : Teneur en cendre (après l'élimination de deux classes de densités)

La raison des décroissances est évidente quand on trace la teneur moyenne en cendre (colonne 1) en fonction de la densité (Figure 39). Au lieu d'être régulière, la courbe est en escalier, et par suite la pente change brusquement. En regardant cette courbe on a l'impression que les échantillons aux densités (1.4, 1.5, 1.6 et 1.8) proviennent d'une région de teneurs plus faibles en cendre que ceux aux autres densités, ce qui fait que bien que la courbe de teneur en cendre en fonction de la densité croisse dans l'une ou l'autre des régions, celle pour les deux régions combinées n'a plus cette propriété (figure 40).

Ceci suggère l'existence de différences régionales. Pour étudier cette hypothèse, la moitié sud du gisement a été divisée en 4 tranches et les teneurs moyennes en cendre ont été calculées pour ces tranches (Tableau 23).

Teneur moyenne en cendre région par région				
Classe de densité	- 7500 à - 6500	- 6500 à - 5000	- 5000 à - 3.500	- 3500 à - 2000
< 1.3	} 5.73	} 5.61	3.78	} 5.51
1.3 - 1.35			7.68	
1.35 - 1.375	} 11.47	10.14	5.59	12.79
1.375 - 1.45		11.88	15.46	12.93
1.45 - 1.6	20.71	22.88	45.58	24.40
1.6 - 1.8	13.23	46.43	8.52	-
Nombre de sondages	38	56	116	92

TABLEAU 23 : Teneur moyenne en cendre région par région.

Bien que les fluctuations des teneurs suivant les classes soient notables, on ne voit pas de dérive évidente entre les régions ; par contre elle est évidente si on trace la quantité de cendre récupérée à densité donnée en fonction de la densité. (Figure 41). Les teneurs en cendre sont nettement plus élevées dans la section (- 5000 à - 3500) que dans les autres régions. Par conséquent, on sépare les données en trois régions (- 7500 à - 5000, - 5000 à - 3500, - 3500 à - 2000)?

En essayant d'améliorer les teneurs en cendre des classes de densité, on se demande si on a fait un bon choix de pondérateurs. Au lieu de la formule [71], on peut utiliser la suivante :

$$\left. \begin{array}{l} \text{Teneur moyenne en} \\ \text{cendre à densité } k \end{array} \right\} = \frac{\sum_{j=1}^n \sum_{i=1}^k a_{ij} p_{ij} t_j}{\sum_{j=1}^n \sum_{i=1}^k p_{ij} t_j} \quad [72]$$

On a recalculé les teneurs moyennes en cendre par classes de densité dans la moitié sud (-7500 à -2000). Pour les comparer aux résultats fournis par l'autre méthode on présente les deux séquences de valeur (Tableau 24).

FIGURE 39: Courbe expérimentale

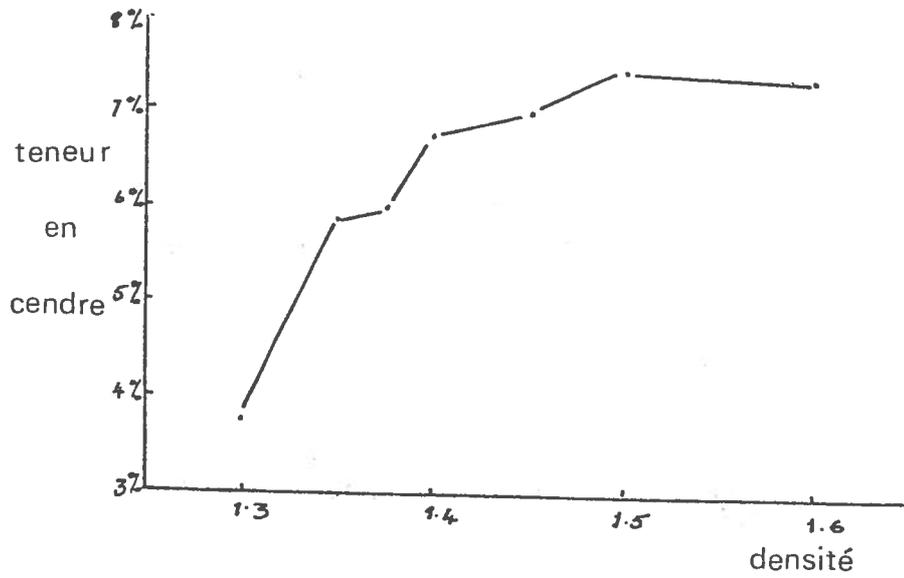


FIGURE 40:

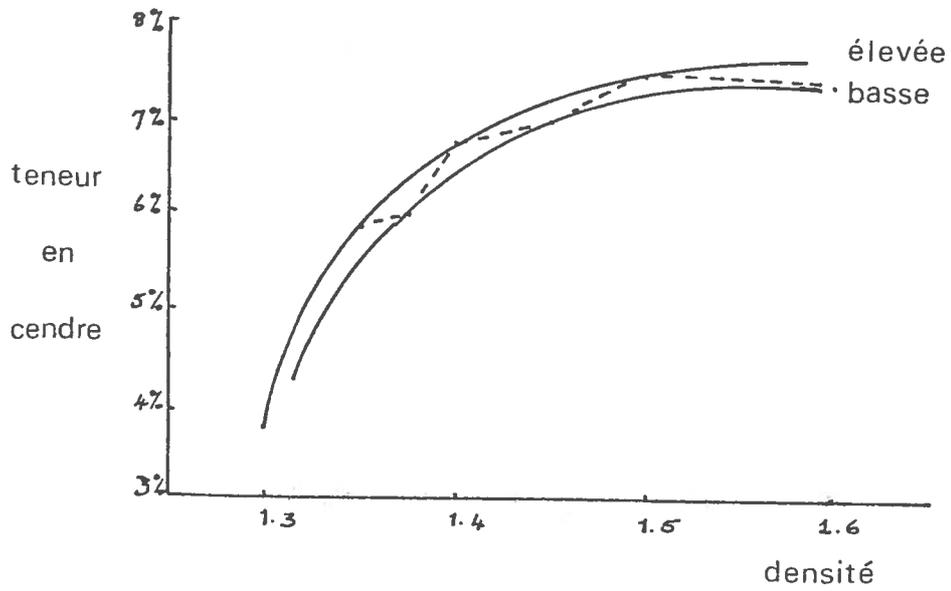
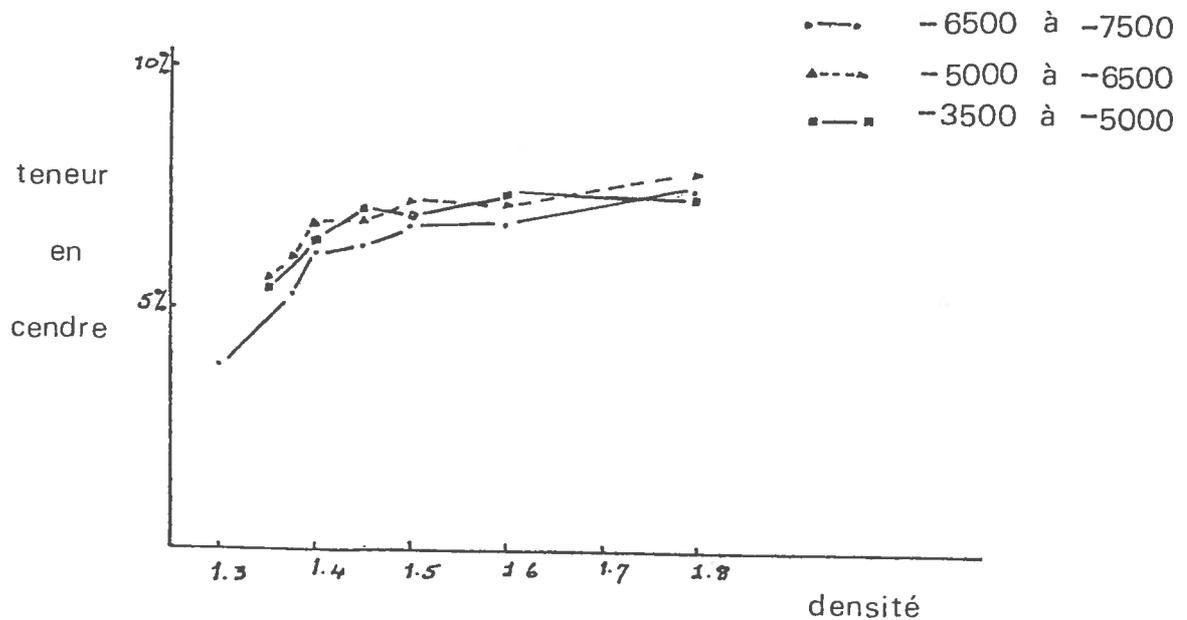


FIGURE 41:



Classe de densité	Teneur en cendre de -7500 à -2000	
	Equation [70]	Equation [71]
< 1.3	3.776	3.798
1.3 - 1.35	6.793	6.780
1.35 - 1.375	7.974	9.438
1.375 - 1.45	13.580	13.048
1.45 - 1.6	27.370	27.100
1.6 - 1.8	34.348	32.722

TABLEAU 24 : Teneurs en cendre calculées par les deux méthodes.

Puisque ces deux méthodes donnent des résultats semblables, on peut se demander si les estimateurs construits à partir de ces valeurs sont également semblables. Les valeurs de ρ_c et de ρ_a estimées par les deux méthodes sont pour ρ_c 1.266 et 1.264 respectivement, et pour ρ_a 3.480 et 3.741. Les estimateurs de ρ_c sont franchement très proches tandis que les deux autres diffèrent plus. Cependant, il n'est pas évident de savoir quelle méthode est la meilleure.

Pour élucider cette question, on a calculé la teneur en cendre région par région en utilisant la deuxième méthode. Comme les teneurs évaluées par la première méthode, elles fluctuent trop pour être utilisées pour estimer ρ_c et ρ_a . Face à l'échec de ces deux méthodes, on conclut qu'il n'est pas possible d'estimer les valeurs de ρ_c et de ρ_a à partir des données des sondages de 2" de diamètre. On se retrouve encore une fois face au problème initial qui est de trouver une région où on connaisse les valeurs de ρ_c et de ρ_a et où on ait assez de données pour étudier la distribution de μ . On en conclut donc que la disposition des données est celle qu'on n'a pas de région convenable, et qu'en conséquence on ne peut pas étudier la distribution de μ pour voir si elle est gaussienne.

5.7 - Procédure à suivre

Dans ce cas, la disposition des sondages de lavabilité par rapport aux sondages de 2" de diamètre nous a empêché de déterminer la distribution de μ . A cause de cet obstacle, on ne peut pas présenter un exemple de la procédure pour estimer le facteur de récupération et la teneur en cendre de charbon in-situ. Il est néanmoins intéressant de préciser cette procédure.

5.7.1. - Estimation de ρ_c , de ρ_a et de A_{\min}

La première étape est l'estimation des valeurs de ces trois paramètres dans une région homogène. Pour cela, il faut un ensemble de données de lavabilité semblable à celui fourni par la compagnie (c'est-à-dire, des analyses de lavabilité faites à au moins 8 densités). Il faut combiner toutes les analyses faites à densité k selon l'équation [71] ou [72]. Pour estimer la teneur moyenne à chaque densité ainsi que le facteur de récupération. Ensuite, on peut estimer les valeurs de ρ_c et de ρ_a par la méthode de moindres carrés présentée au paragraphe 5.3.1, puis on utilise la procédure du paragraphe 5.3.2 pour estimer la valeur d' A_{\min} . Le paramètre ρ_{MIN} peut être évalué à partir de ρ_c , de ρ_a et d' A_{MIN} ; ce qui détermine complètement la transformation lognormale

$$Z = \log \left(\frac{X - \rho_{\text{MIN}}}{\rho_a - X} \right)$$

Les fonctions $E(x)$ et $a(x)$ peuvent être transformées, ce qui donne la fonction $B(Z)$.

5.7.2. - Estimation de μ et de σ en chaque point d'échantillonnage

Les valeurs de ρ_c , de ρ_a et d' A_{\min} étant connues, on peut procéder à l'estimation de μ et de σ en chaque point d'échantillonnage en utilisant la méthode énoncée au paragraphe 5.3.2. On peut calculer la moyenne μ_0 et la variance σ_0^2 de la variable μ ainsi que la valeur de k et c dans la relation linéaire entre μ et σ :

$$\sigma = k \mu + c$$

Le calcul de k et c se fait par la méthode des moindres carrés.

Après avoir évalué les paramètres, μ_0 , σ_0 , k et c , on peut déterminer, à partir de l'équation [66], les coefficients $\phi_n(z)$:

$$\phi_n(z) = \int_{-\infty}^{\infty} g\left(\frac{z-\mu}{\sigma}\right) H_n\left(\frac{\mu-\mu_0}{\sigma_0}\right) g\left(-\frac{\mu-\mu_0}{\sigma_0}\right) d\mu$$

On en obtient les coefficients $\psi_n(z)$

$$\Psi_n = \int_{-\infty}^{\infty} B(z) \phi_n(z) dz$$

On a donc l'intégrale I développée en série de polynomes d'Hermite :

$$I = \sum \frac{\Psi_n}{n!} H_n \left(\frac{\mu - \mu_0}{\sigma_0} \right)$$

5.7.3 - Krigeage disjonctif

Comme on a développé I en série de polynomes d'Hermite, on peut estimer sa valeur I_x en un point x en utilisant les techniques du krigeage disjonctif soit β l'indice des points de mesure. On a donc :

$$I_x = \sum_{\beta} \sum_n \frac{\Psi_n}{n!} \lambda_{n\beta} H_n \left(\frac{\mu_{\beta} - \mu_0}{\sigma_0} \right)$$

où les coefficients $\lambda_{n\beta}$ sont obtenus à partir du système d'équations :

$$\sum_{\alpha} \lambda_{n\beta} p_{\alpha\beta}^n = p_{\alpha x}^n$$

CHAPITRE VI

CONCLUSION

CHAPITRE VI

CONCLUSION

Le but de ce projet de recherche est d'étudier quelques applications possibles de la géostatistique à l'estimation des réserves de charbon et des paramètres de qualité. Dès le début, il était évident que la question se divise en deux : la première partie est l'estimation des réserves en charbon brut, tandis que la deuxième est l'estimation du facteur de récupération et des paramètres de qualité du charbon après lavage.

Les problèmes rencontrés à la première partie sont les problèmes habituels de la géostatistique : dans un premier temps le nettoyage et l'organisation du fichier des données, puis le calcul et l'ajustement du variogramme. Puisque les données fournies par la société ont été contrôlées très soigneusement, la première phase de nettoyage n'était pas nécessaire. Du point de vue de l'organisation du fichier des données, deux petits changements dans la structure du fichier auraient facilité et amélioré l'analyse structurale. Le premier serait de remplacer le système actuel de numérotation des couches par un système de codage de sous-couches qui permettrait d'étudier la régionalisation des variables sous-couche par sous-couche au lieu de couche par couche. Etant données les fortes différences entre les sous-couches, ce système fournirait de meilleurs estimateurs ainsi qu'une variance de krigeage plus faible. Le deuxième changement serait d'ajouter un paramètre supplémentaire pour indiquer si l'échantillon provient d'une région en bordure du gisement où le charbon est oxydé. A l'heure actuelle, la qualité du charbon de chaque échantillon est codée (par exemple "charbon à cokéfier", "charbon vapeur", "charbon oxydé", etc...) mais ce système n'indique pas s'il y a d'autres échantillons provenant du même sondage et de la même couche qui sont oxydés. Ce changement permettrait d'étudier les bordures du gisement qui sont plus variables, séparément du reste du gisement, ce qui améliorerait la précision des estimateurs.

Les problèmes de calcul et d'ajustement des variogrammes (section 2.2) ont suscité le développement de deux nouveaux estimateurs plus robustes du variogramme (section 2.4) dont le variogramme de Huber, qui a fourni des résultats assez encourageants pour justifier une poursuite des recherches dans cette direction. A l'exception de ces nouveaux estimateurs, on n'a pas eu besoin de nouvelles techniques pour l'estimation des réserves de charbon brut ; les techniques géostatistiques existant à l'heure actuelle suffisent.

Cependant, ce n'est pas le cas en ce qui concerne le charbon après lavage. Pour cela il a fallu développer une nouvelle méthodologie. On a trouvé deux approches au problème de l'estimation du facteur de récupération et de la teneur en cendre (chapitre 3). La plus simple des deux consiste à estimer ces deux variables à la densité de lavage pour chaque sondage et ensuite de les kriger (Chapitre 4). Cette approche ne pose pas de problème spécial, mais elle a l'inconvénient qu'il faut répéter toute la procédure s'il faut recalculer les estimateurs ultérieurement à une nouvelle densité de lavage.

La deuxième approche qui surmonte ces obstacles consiste à développer un modèle paramétrique pour caractériser la lavabilité du charbon (Chapitre 5). On a utilisé un ensemble de données de lavabilité provenant d'une gamme de charbons australiens pour tester ce modèle. Les résultats étant satisfaisants, on a accepté ce modèle et procédé au développement d'une méthode d'estimation basée sur ce modèle. Les principes de l'application de cette méthode ont été bien établis, mais la disposition des données de lavabilité par rapport à celles des sondages de 2" de diamètre nous a empêché de la mettre en application. Bien qu'il soit regrettable de ne pas pouvoir développer cette approche aussi loin qu'on aurait voulu, cette méthodologie décrite au chapitre 5 peut fournir de bons sujets de recherche supplémentaires. Dans ce chapitre, on attire l'attention sur les deux relations fondamentales pour la caractérisation ou la lavabilité du charbon : la courbe de teneur en cendre en fonction de la densité et la distribution en masse des densités des particules de charbon ; et on montre le besoin d'améliorer notre connaissance de ces deux relations. On peut espérer que cette approche fournira un meilleur estimateur du facteur de récupération et de la teneur en cendre de charbon après lavage.

A N N E X E

ANALYSE STATISTIQUE DES PARAMETRES DU MODELE DE LAVABILITE

ANNEXE

ANALYSE STATISTIQUE DES PARAMETRES DU MODELE DE LAVABILITE

A-1 Introduction

A-1-1 Données

On dispose des données de lavabilité (teneur en cendre et pourcentage en masse) pour au moins 9 classes de densité sur un ensemble de 21 sondages de diamètre 8", réalisés par la compagnie sur la couche inférieure selon l'implantation indiquée sur la figure A1.

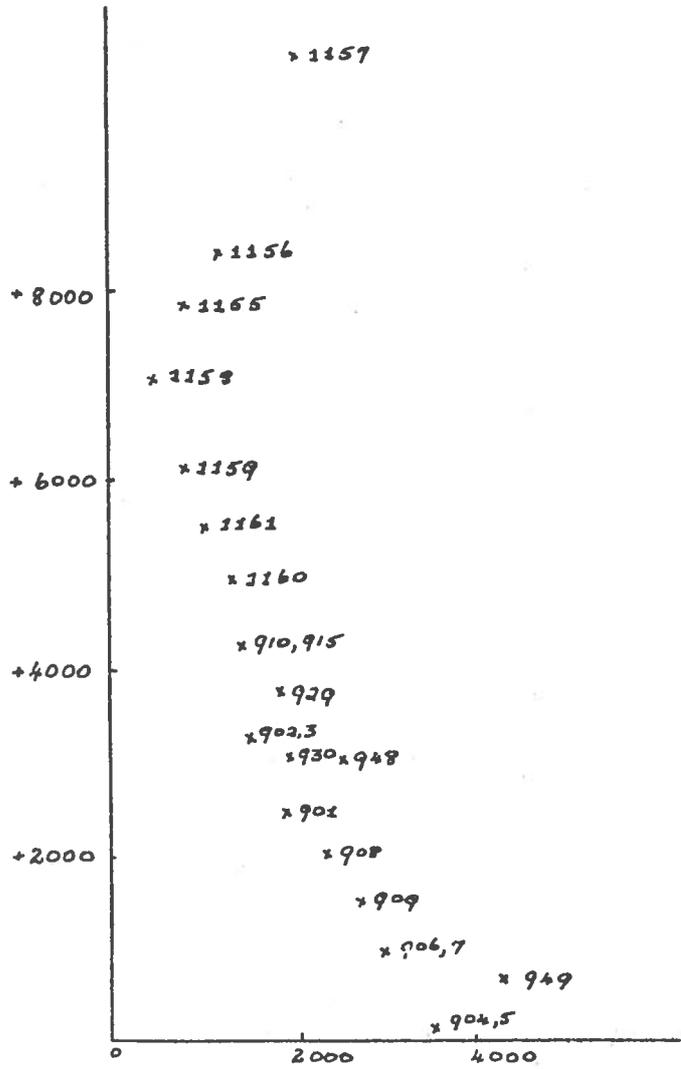
Pour 7 de ces sondages, les analyses ont été faites séparément pour les deux sous-couches, ainsi que pour la couche entière. Dans trois cas, les échantillons proviennent d'un mélange de charbon de deux sondages voisins (Nos 902, 903; 904, 905; 906, 907). Cela donne 32 ensembles de données :

Nombre de sondages	Nos des sondages	En provenance de
11	901 - 949	Couche entière
7	1155 - 1161	Couche entière
7	1155 - 1161	Sous-couche inférieure
7	1155 - 1161	Sous-couche supérieure

TABLEAU A1 : Résumé des sondages de diamètre 8".

Il est important de savoir si ces sondages sont représentatifs de la couche (ou sous-couche) ou s'ils se trouvent dans des poches de hautes-teneurs en cendre ou en soufre. En fait, c'est le cas pour les sondages Nos 901 - 949; ils se trouvent dans la région de teneurs élevées en cendre. Il ne serait donc pas surprenant de trouver des différences entre les caractéristiques de ces sondages et celles du reste des échantillons.

FIGURE A1 : Dispositions des sondages



A-1-2 Paramètres de lavabilité

Les deux modèles mathématiques proposés à la section 5-2 ont été ajustés à ces données. Rappelons que le premier modèle donne la teneur en cendre $a(x)$ d'une particule de densité x en fonction de la densité de charbon "pur", ρ_c , et celle de cendre "pure", ρ_a . (Equation 35). Les densités des particules ont, selon ces modèles, une distribution de la forme :

$$\log \left(\frac{x - \rho_{MIN}}{\rho_a - x} \right) \sim N(\mu, \sigma^2) \quad (A1)$$

où ρ_{MIN} est la vraie densité minimale des particules (correspondant à une teneur en cendre de A_{MIN} au lieu de 0 pour ρ_c). Dans le cas des mines en exploitation, le stérile provenant du toit et du mur est souvent une source de contamination des échantillons. Pour en tenir compte, on utilise un facteur de dilution p .

Cela donne un ensemble de 7 variables (ρ_c , ρ_a , ρ_{MIN} , A_{MIN} , p , μ et σ) dont les valeurs ont été estimées. (Tableau A2).

A-1-3 Réduction du nombre de paramètres.

On peut déterminer la valeur de ρ_{MIN} (resp. A_{min}) à partir de l'équation (35) une fois que les paramètres ρ_c , ρ_a et A_{min} (resp. ρ_{MIN}) sont connus, ce qui permet d'économiser un paramètre.

Comme toutes les valeurs de p sont entre 0,94 et 1, on peut attribuer à p une valeur de 1. Il y a très peu de stérile mélangé avec l'échantillon ce qui est normal dans les mines à ciel ouvert.

On a donc un ensemble de 5 paramètres (ρ_c , ρ_a , ρ_{MIN} ou A_{MIN} , μ , σ). Pour établir les relations entre eux, on a calculé des statistiques classiques.

A-2 Statistiques

A-2-1 Moyenne et variance

On a calculé les moyennes, les variances et les erreurs relatives de l'ensemble des 6 variables (Tableau A3). L'erreur relative de ρ_c étant très basse, on peut supposer que cette variable est constante. Pour voir s'il existe une dérive, on a tracé les valeurs de ρ_c en fonction de la coordonnée

N-S (Figure A2a). Pour les sondages Nos 1155-1161, il n'y a pas de dérive évidente, tandis que pour les autres sondages qui se trouvent dans la poche anormale, les valeurs semblent augmenter du Sud au Nord.

	Moyenne	Variance	Erreur relative
ρ_c	1,2623	$0,4856 \times 10^{-4}$	0,552 %
ρ_a	3,5031	0,3958	17,96 %
ρ_{MIN}	1,2698	$0,1153 \times 10^{-3}$	0,845 %
A_{MIN}	0,8611	$0,1411 \times 10^{-3}$	138 %
μ	-3,999	0,3393	14,57 %
σ	1,226	0,2556	42,24 %

TABLEAU A3 : Statistiques classiques.

On a aussi tracé les valeurs des autres paramètres mais il n'y a pas de dérive évidente. (Figures A2b-e). Pourtant, on remarque que les sondages Nos 908, 909 ont souvent des valeurs anormales.

A-2-2 Moyennes de ρ_c par sous-ensembles

En regardant la figure A2a, on remarque des différences entre les valeurs de ρ_c pour les 4 sous-ensembles. Les moyennes et les variances ont été calculées par sous-ensemble. (Tableau A4).

N° de sondage	Couche	Moyenne	Ecart-type
901 - 949	Couche entière	1,2641	$8,51 \times 10^{-3}$
1155 - 1161	Couche entière	1,2596	$2,07 \times 10^{-3}$
1155 - 1161	Sous-couche inférieure	1,2587	$5,93 \times 10^{-3}$
1155 - 1161	Sous-couche supérieure	1,2529	$7,70 \times 10^{-3}$
Moyenne des écarts-type			$6,74 \times 10^{-3}$

TABLEAU A4 : Moyennes par sous-ensembles.

FIGURE A2a : Valeurs de ρ_c

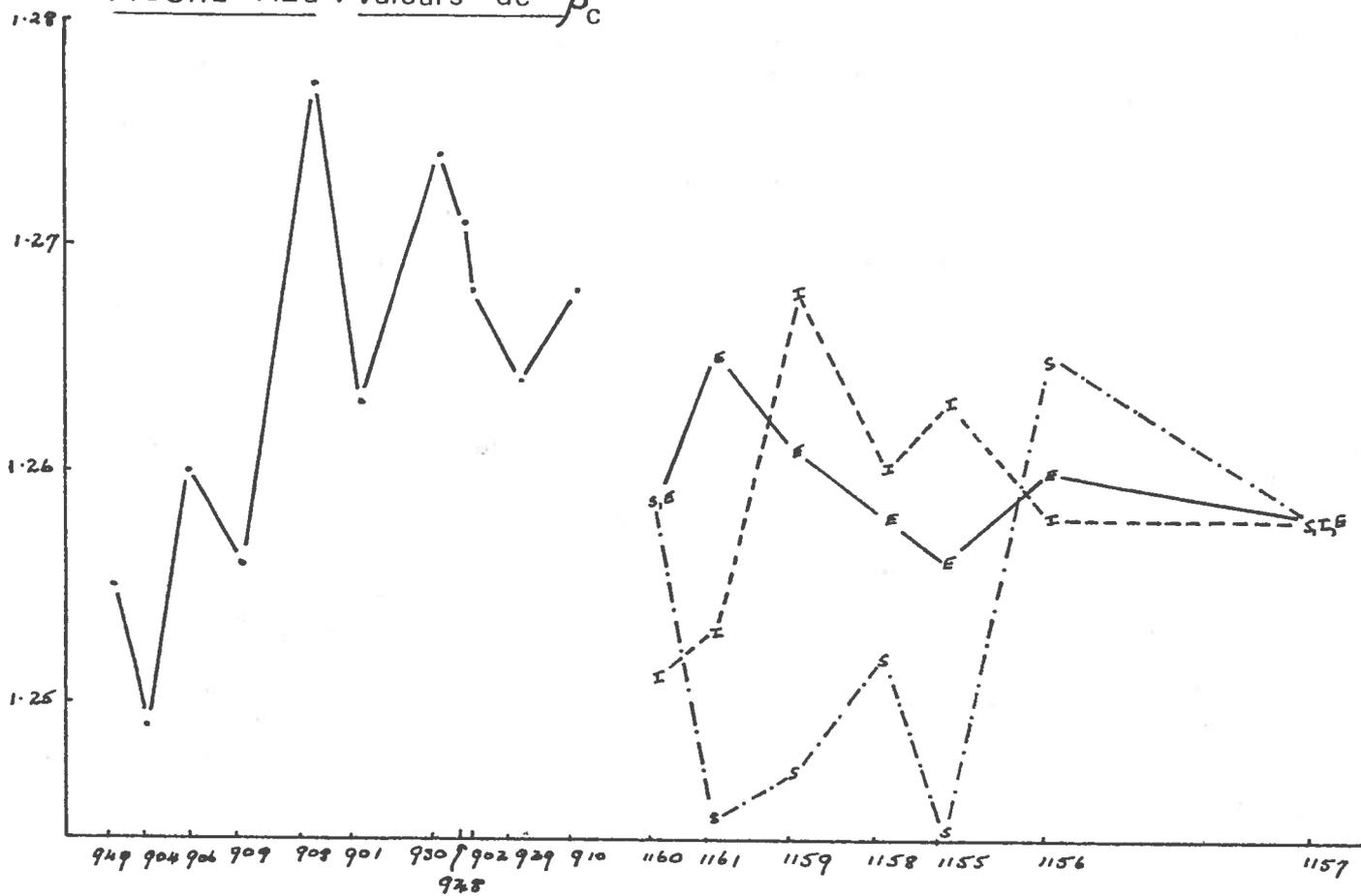


FIGURE A2b : Valeurs de ρ_a

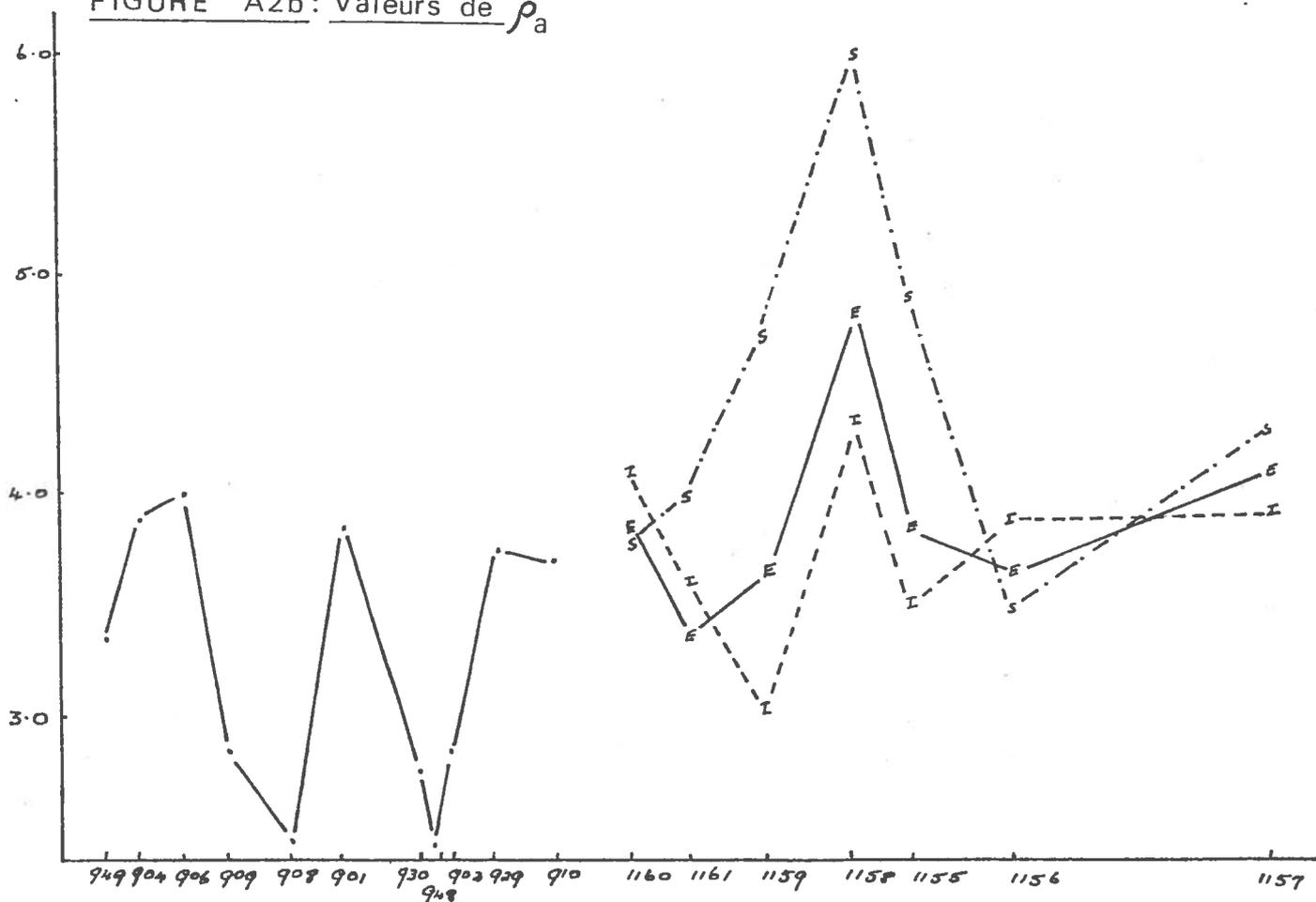


FIGURE A2c: Valeurs de μ

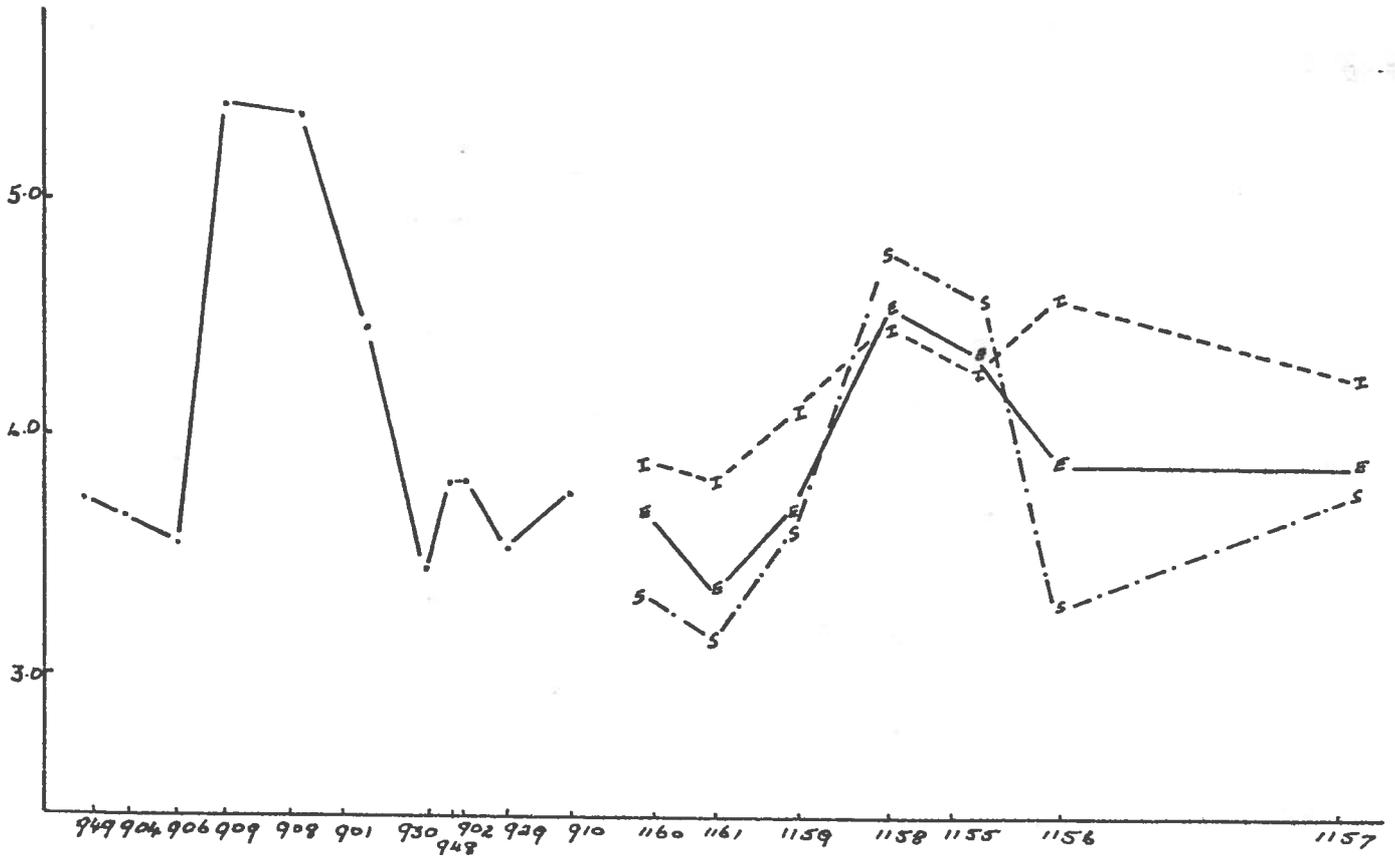


FIGURE A2d: Valeurs de σ

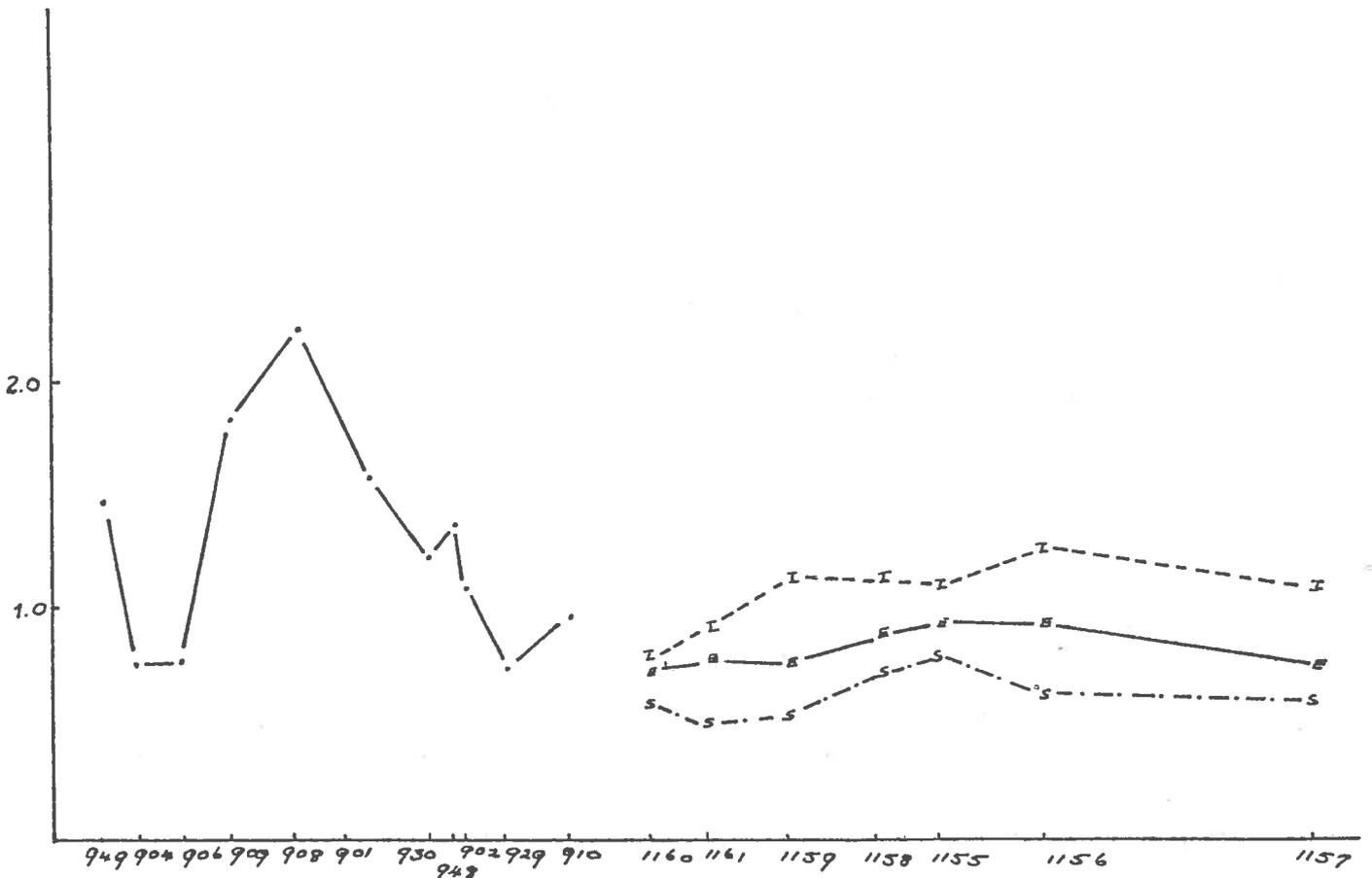


FIGURE A2e : Valeurs d'A_{min}

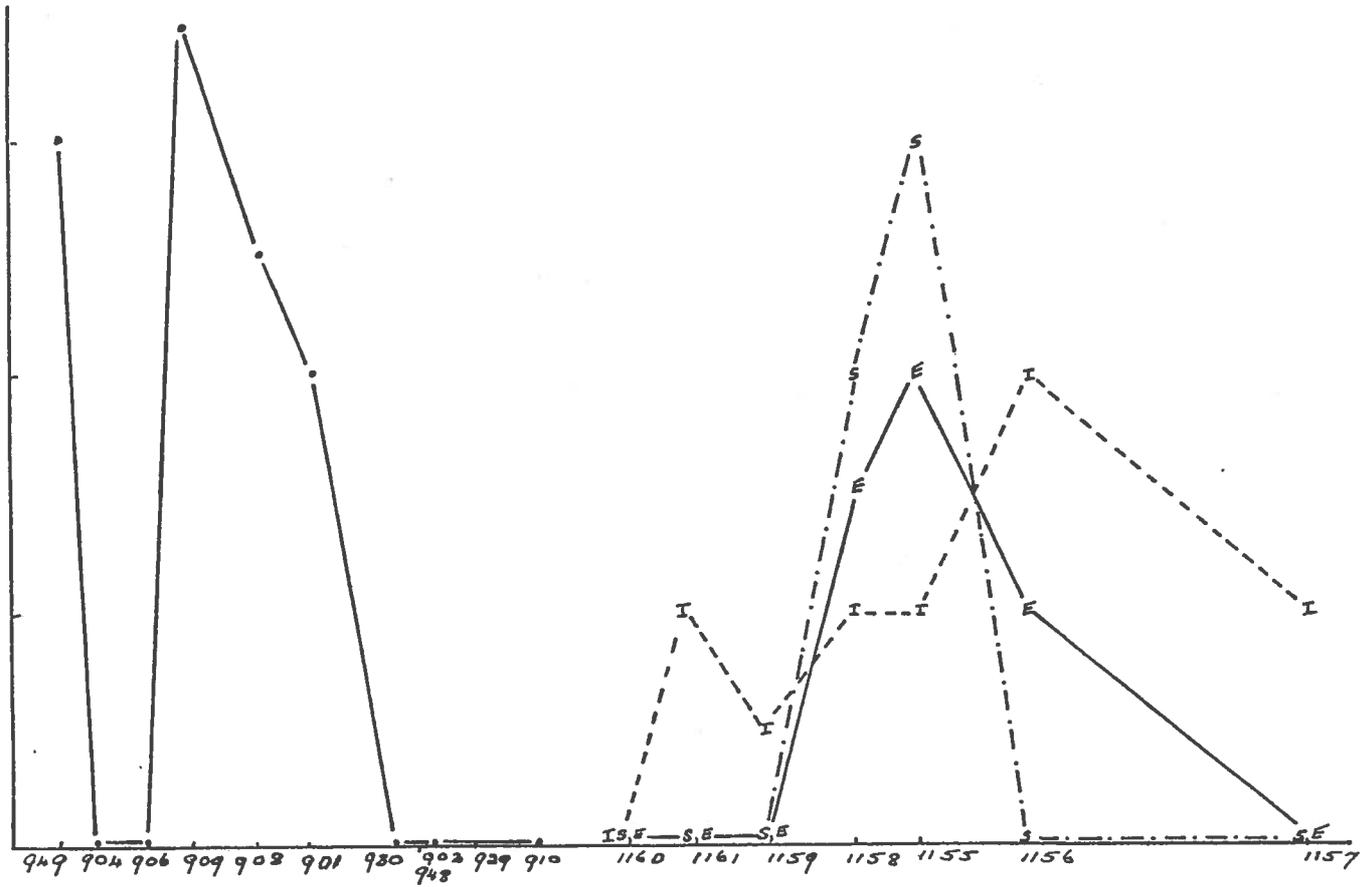
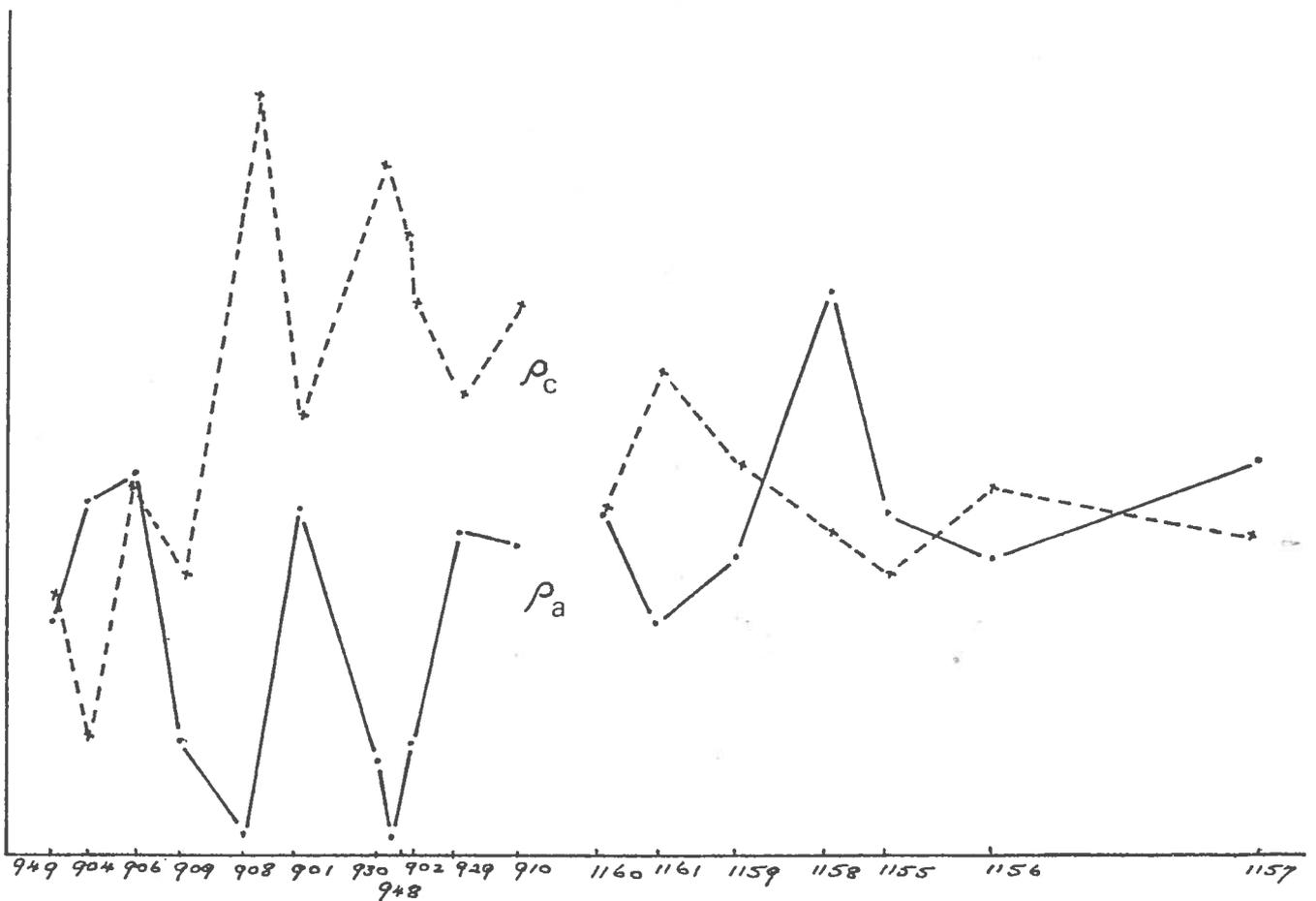


FIGURE A3 : Valeurs de ρ_c et de ρ_a



Les différences entre ces moyennes ne sont pas significatives à un niveau de confiance de 95 %.

A-2-3 Corrélations entre variables

	ρ_c	ρ_a	ρ_{MIN}	A_{MIN}	μ	σ
ρ_c	1	-0,66	0,41	-0,19	-0,04	0,51
ρ_a	-0,66	1	-0,46	-0,14	0,17	0,42
ρ_{MIN}	0,41	-0,46	1	0,81	-0,17	0,55
A_{MIN}	-0,19	-0,14	0,81	1	-0,73	-0,33
μ	-0,04	0,17	-0,73	-0,80	1	-0,56
σ	0,51	0,42	0,55	-0,33	-0,56	1

TABLEAU A5 : Corrélations entre variables.

On a calculé les corrélations entre les variables qui sont présentes dans le tableau A5. Trois aspects de ce tableau sont à noter :

- 1) la faible corrélation entre ρ_c et μ ,
- 2) la forte corrélation entre ρ_c et ρ_a , et
- 3) les fortes corrélations entre μ et A_{min} , et entre μ et σ .

Ces trois résultats suggèrent qu'on peut traiter les variables en deux groupes indépendants (ρ_c et ρ_a ; μ , σ et A_{min}). Dans ce cas, ρ_{MIN} est déterminé par les valeurs de ρ_c , ρ_a et A_{min} .

On peut voir la corrélation négative entre ρ_c et ρ_a en superposant les tracés de ces variables sur l'axe N-S (Figure A3). Les fortes corrélations entre μ , σ et A_{min} s'expliquent quand on considère la façon de les estimer. Si on trace les valeurs de μ et de σ pour une série de valeurs croissantes de A_{min} , les deux se ressemblent (Figure A4a), ce qui suggère de tracer les courbes de σ (ou σ^2) en fonction de μ pour voir si elles sont linéaires (Figure A4b et c). Un résultat semblable a été obtenu pour les 28 charbons australiens étudiés précédemment.

A-2-4 Variogrammes

Après le calcul des statistiques, on procède au calcul des variogrammes expérimentaux. Pour les comparer, on a superposé les variogrammes de ρ_c et ρ_{MIN} ainsi que ceux de μ et σ . Le variogramme de ρ_c et celui de

FIGURE A4a

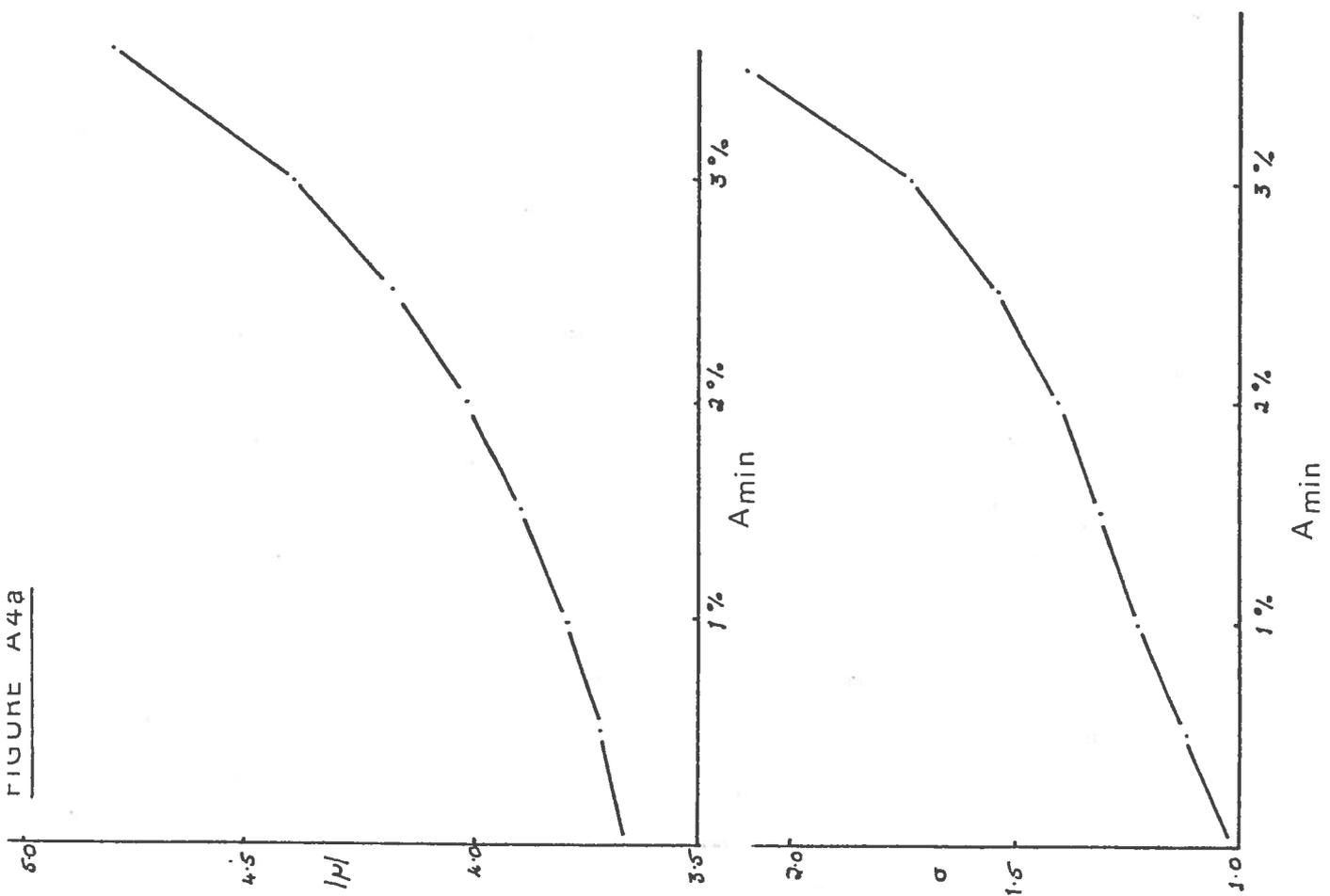
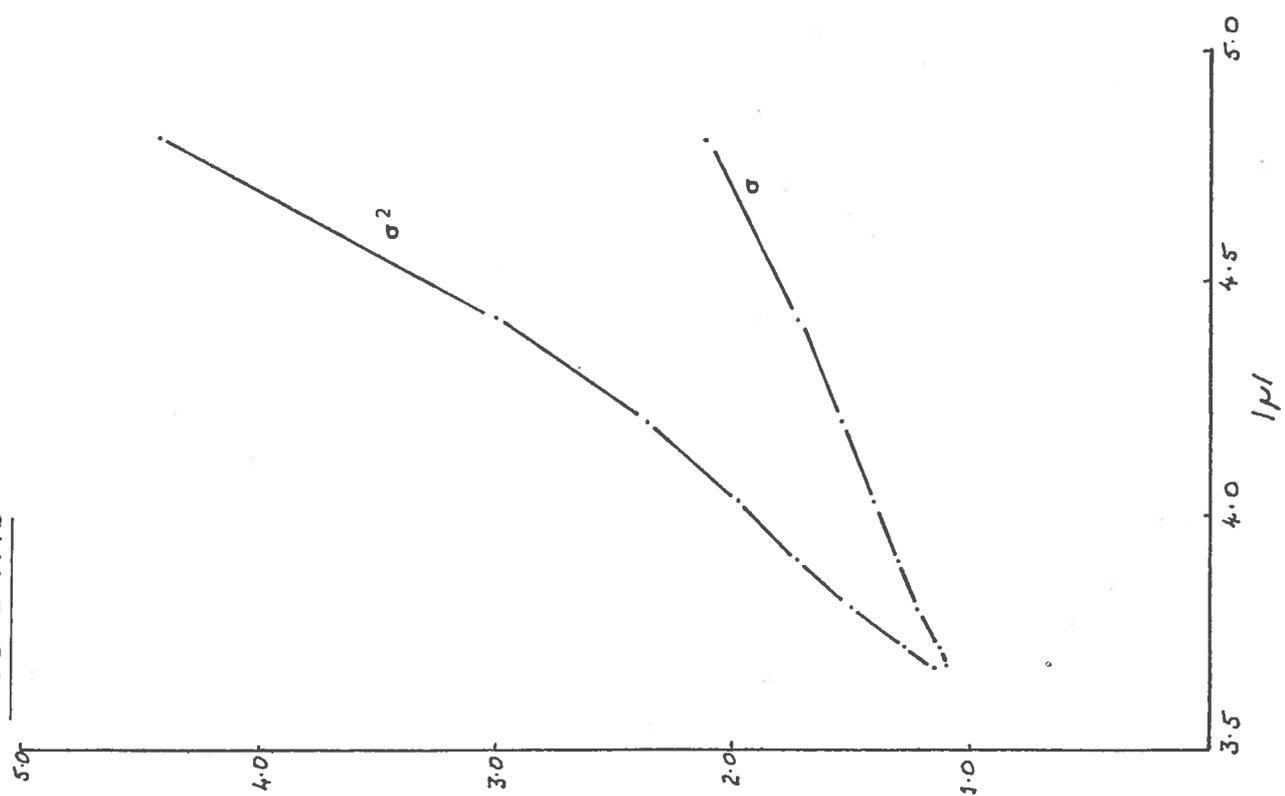


FIGURE A4b



ρ_a ont une forme sphérique classique; celui de A_{\min} présente un effet de pépite tandis que ceux de μ et de σ augmentent d'abord rapidement, puis décroissent ensuite. Les covariogrammes choisis ont également été calculés, mais ils sont sans intérêt.

A-3 Conclusion de la première partie

L'analyse statistique des paramètres du modèle de lavabilité a suggéré qu'on pouvait :

- 1) considérer ρ_c comme constant (du moins dans des domaines homogènes);
- 2) ne pas tenir compte de p et ρ_{\min} , p étant égal à 1 et ρ_{\min} étant déterminé par ρ_c , ρ_a et A_{\min}
- 3) traiter les autres paramètres comme deux groupes indépendants (ρ_c et ρ_a ; μ , σ et A_{\min}).

Compte tenu de ces conclusions, la prochaine étape de cette étude est le calcul des valeurs de ρ_a correspondant à une valeur fixée de ρ_c pour voir si ρ_a est constant, puis l'estimation de μ , σ et A_{\min} correspondant à des valeurs fixées de ρ_c et de ρ_a . Enfin, il faudrait estimer, encore une fois, la teneur en cendre et le facteur de récupération pour tester si ces nouvelles valeurs estimées sont en accord avec les valeurs réelles.

A-4 Estimation de la valeur de ρ_a pour une valeur de ρ_c fixée

A-4-1 Estimation des valeurs de ρ_a pour chaque sondage

On prend pour valeur de ρ_c la valeur moyenne du groupe auquel appartient l'échantillon (Tableau A4), puis on estime la valeur de ρ_a (Tableau A6). Les statistiques classiques ont été calculées sur ces valeurs. Pour les sondages N° 1155-1161, les erreurs relatives sont faibles et la grande partie de cette variabilité est due au sondage N° 1158. Si on exclut ce sondage des calculs, les erreurs relatives sont très faibles, ce qui suggère que la valeur de ρ_a est constante pour une couche (ou pour une sous-couche) dans une région donnée. Par contre, les résultats pour les sondages N° 901-949 ne sont pas aussi encourageants.

Il serait très commode de supposer que les valeurs de ρ_c et de ρ_a sont constantes dans une région assez large. Il est regrettable que l'ensemble de données soit trop petit pour tester de façon définitive cette hypothèse. Cependant, pour des raisons de commodité, on acceptera provisoirement l'hypothèse

sous réserve qu'il faudrait la tester sur un ensemble de données plus grand.

A-4-2 Estimation de la valeur moyenne de ρ_a

Après avoir accepté l'hypothèse, on peut procéder à l'estimation de la valeur de ρ_a dans des régions assez grandes. Mais il faut, d'abord, se demander comment calculer la valeur supposée constante de ρ_a . Faudrait-il prendre, tout simplement, la moyenne des valeurs de ρ_a du tableau A6 ? Mais cela présente un inconvénient à long terme; deux estimations sont nécessaires pour chaque sondage : une estimation préliminaire de ρ_c , puis l'estimation de la valeur de ρ_a à ρ_c fixé. Cette méthode présente un deuxième désavantage : elle ne semble pas être très stable. Par exemple, le tableau A7 donne les teneurs en cendre de chaque classe de densité de la sous-couche inférieure pour les sondages N° 1158, 1159. Les teneurs des 4 premières classes, qui contiennent plus de 80 % de la masse totale, sont peu différentes mais il y a une grande différence entre les valeurs estimées de ρ_a ($\rho_a = 4,579$ pour le sondage N° 1158 et $\rho_a = 3,839$ pour l'autre).

Classe de densité	Sondage N° 1158	Sondage N° 1159
< 1,3	3,6 %	2,7 %
1,3 - 1,35	7,3 %	7,3 %
1,35 - 1,375	10,4 %	10,8 %
1,375 - 1,4	12,0 %	12,8 %
1,4 - 1,425	14,1 %	14,8 %
1,025 - 1,45	16,0 %	17,8 %
1,45 - 1,5	18,9 %	21,0 %
1,5 - 1,55	21,7 %	26,0 %
1,55 - 1,6	24,4 %	30,7 %
> 1,6	47,8 %	56,7 %

TABLEAU A7 : Teneurs en cendre de la sous-couche inférieure pour deux sondages.

Par conséquent, il est préférable d'aborder le problème d'une autre façon. Il vaut mieux estimer la moyenne des teneurs en cendre pour chaque classe de densité ainsi que celle des pourcentages en poids, et estimer ensuite ρ_c et ρ_a . Au point de vue du calcul, cette approche est plus rapide et aussi plus facile. On a calculé ces caractéristiques de lavabilité

moyennes tout simplement comme les moyennes classe par classe (colonnes 2 et 3, Tableau A8). On pourrait également penser qu'il serait préférable de pondérer les teneurs en cendre par le pourcentage de poids dans la classe. On l'a fait pour les sondages N^{OS} 1155 - 1161 et les moyennes pondérées (qui diffèrent peu des autres) sont présentées dans la quatrième colonne (Tableau A8).

Tranche densimétrique	Nos 901-949	Sondages Nos 1155-1161	
		(moyennes simples)	(moyennes pondérées)
21,3	3,93 %	4,06 %	4,06 %
1,3 - 1,35	8,14 %	7,53 %	7,59 %
1,35 - 1,375	11,30 %	10,16 %	10,32 %
1,375 - 1,4	13,23 %	11,80 %	11,77 %
1,4 - 1,425	15,11 %	13,70 %	13,36 %
1,425 - 1,45	17,20 %	16,03 %	17,02 %
1,45 - 1,5	20,48 %	19,40 %	19,24 %
1,5 - 1,55	25,70 %	23,63 %	22,34 %
1,55 - 1,6	31,21 %	27,36 %	29,91 %

TABLEAU A8 : Teneurs en cendre moyennes (pour la couche entière).

Les valeurs estimées de ρ_c et de ρ_a (Tableau A9), ont été calculées à partir des moyennes simples (c'est-à-dire les colonnes (1) et (2) du Tableau A8). L'estimateur de ρ_a pour les sondages N^{OS} 1155-1161 semble élevé par rapport aux valeurs de ρ_a estimées pour chacun des sondages (Tableau A6).

	Sondages N° 901-949	Sondages N° 1155-1161
Valeur de ρ_c	1,266	1,246
Valeur de ρ_a	3,078	4,471

TABLEAU A9 : Valeurs estimées de ρ_c et de ρ_a .

A-5 Valeurs estimées de μ , σ et A_{\min}

Après avoir estimé les valeurs de ρ_c et de ρ_a dans une région, on peut estimer les valeurs de μ , σ et A_{\min} pour chaque sondage. On l'a fait pour tous les sondages, mais seulement sur la couche entière.

L'estimation de A_{\min} a posé quelques problèmes. Dans plusieurs cas, le minimum de la somme des carrés des écarts entre les teneurs en cendre expérimentales et celles estimées à partir du modèle n'est pas atteint pour la même valeur de A_{\min} que celle obtenue par l'ajustement du modèle lognormal. D'ailleurs, comme la valeur de cette dernière somme d'écarts reste très proche de la valeur minimale pour toute une gamme de densités il est difficile d'estimer la valeur de A_{\min} très précisément. Pour ces deux raisons, il est bien possible qu'une grande part de la variabilité de l'estimation de A_{\min} soit due à l'incertitude de l'estimation et non à sa régionalisation.

Il est évidemment très commode d'avoir un estimateur de A_{\min} constant dans la même région que ceux de ρ_c et de ρ_a ; cela déterminerait la valeur de ρ_{MIN} et en conséquence, la variable $\log((X - \rho_{\text{MIN}})/(\rho_a - X))$ ne dépendrait que de la densité X . En raison de l'incertitude sur l'estimation d' A_{\min} , on a supposé que la valeur de A_{\min} est constante et on l'a estimée pour chaque groupe de sondages comme la moyenne des valeurs (ce qui donne $A_{\min} = 0,5 \%$ pour les sondages N^{OS} 901-949 et $A_{\min} = 1,5 \%$ pour les sondages N^{OS} 1155-1161). Comme on le voit sur le Tableau A10, le changement de la valeur de A_{\min} qui minimise la somme des carrés des écarts à la moyenne régionale n'influence pas beaucoup les facteurs de récupération estimés, mais il a une influence plus forte sur les estimateurs des teneurs en cendre.

Borne supérieure de densité	Facteur de récupération expérimental	Facteur de récupération estimé correspondant à			
		$A_{\min} = 0 \%$	$A_{\min} = 0,5 \%$	$A_{\min} = 1 \%$	$A_{\min} = 1,5 \%$
1,3	39,2	40,1	39,4	38,6	37,8
1,35	67,1	67,8	68,1	68,8	69,2
1,375	74,8	74,9	75,2	75,5	76,1
1,4	80,5	80	80,2	80,5	80,8
1,425	85,3	83,6	83,4	84,1	84,1
1,45	87,5	86,6	86,6	86,6	86,8
1,5	90,9	90,5	90,5	90,5	90,5
1,55	92,5	93,2	93	92,9	92,8
1,6	93,3	95	94,8	94,7	94,4

TABLEAU A10 : Facteurs de récupération expérimentaux et estimés pour différentes valeurs de A_{\min} (Sondage N° 901)

Les valeurs de μ et σ ont été estimées en utilisant une valeur de A_{\min} de 0,5 % pour les sondages N^{os} 901-949 et de 1,5 % pour les autres. (Tableau A11). On a localisé les valeurs de μ et de σ pour voir s'il y avait une dérive, mais il n'en apparaît pas d'évidente. Par ailleurs, quand on trace les points représentatifs des couples (μ, σ) , ils sont presque alignés (Figure A5). Les droites de régression de σ sur μ ont également été calculées pour les deux groupes de sondages; bien qu'elles aient les mêmes pentes, les ordonnées à l'origine sont sensiblement différentes. De même, il apparaît également une relation linéaire entre μ et σ lorsqu'on fait varier les valeurs de A_{\min} pour un groupe de sondages fixé. (Figure A6). On peut, donc, supposer qu'il existe une relation sous-jacente de la forme

$$\sigma = k \mu + c \tag{A4}$$

Sondage N°	μ	σ
901	-3,73	1,367
902,3	-4,04	1,106
904,5	-3,67	1,178
906,7	-3,25	0,929
908	-4,59	1,615 \bar{u}
909	-4,97	1,585
910,15	-3,52	1,022
929	-3,22	1,012
930	-3,63	1,149
948	-4,13	1,287
949	-3,34	1,251

Sondage N°	μ	σ
1155	-4,28	0,918
1156	-3,94	0,873
1157	-4,10	0,938
1158	-4,17	0,861
1159	-3,98	0,825
1160	-3,95	0,818
1161	-3,75	0,741

TABLEAU A11 : Valeurs de μ et de σ .

A-6 Contrôle de la qualité de l'ajustement.

Après l'ajustement des modèles aux données, il est important de calculer les facteurs de récupération et les teneurs en cendre à partir des valeurs de ρ_{\min} , ρ_a supposées constantes, pour voir si les valeurs estimées sont en accord avec les valeurs expérimentales. Un exemple de valeurs des facteurs de récupération a déjà été présenté au Tableau A10. Les résultats sont, en général, satisfaisants mais dans quelques cas (par exemple le sondage N° 908), les

FIGURE A5

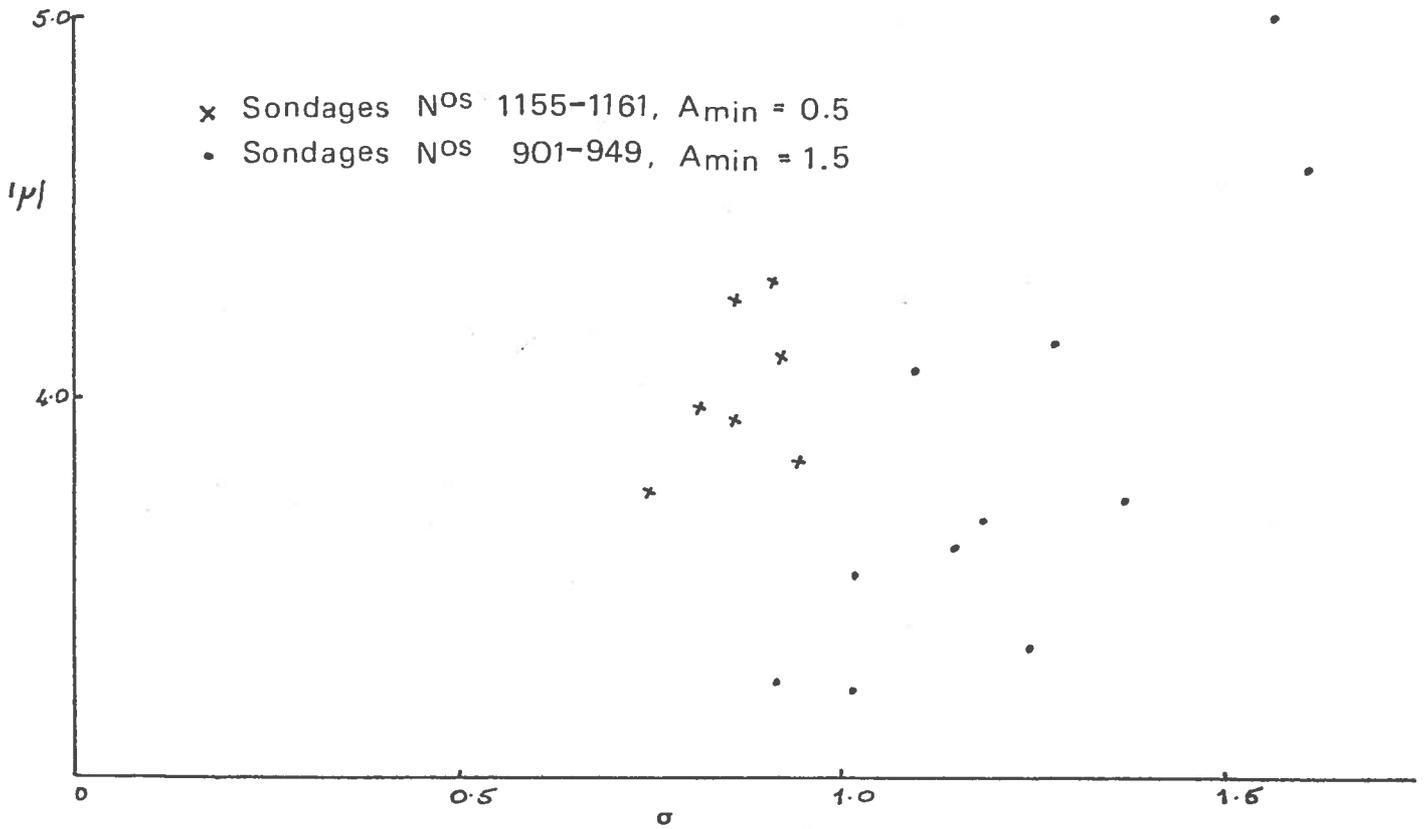
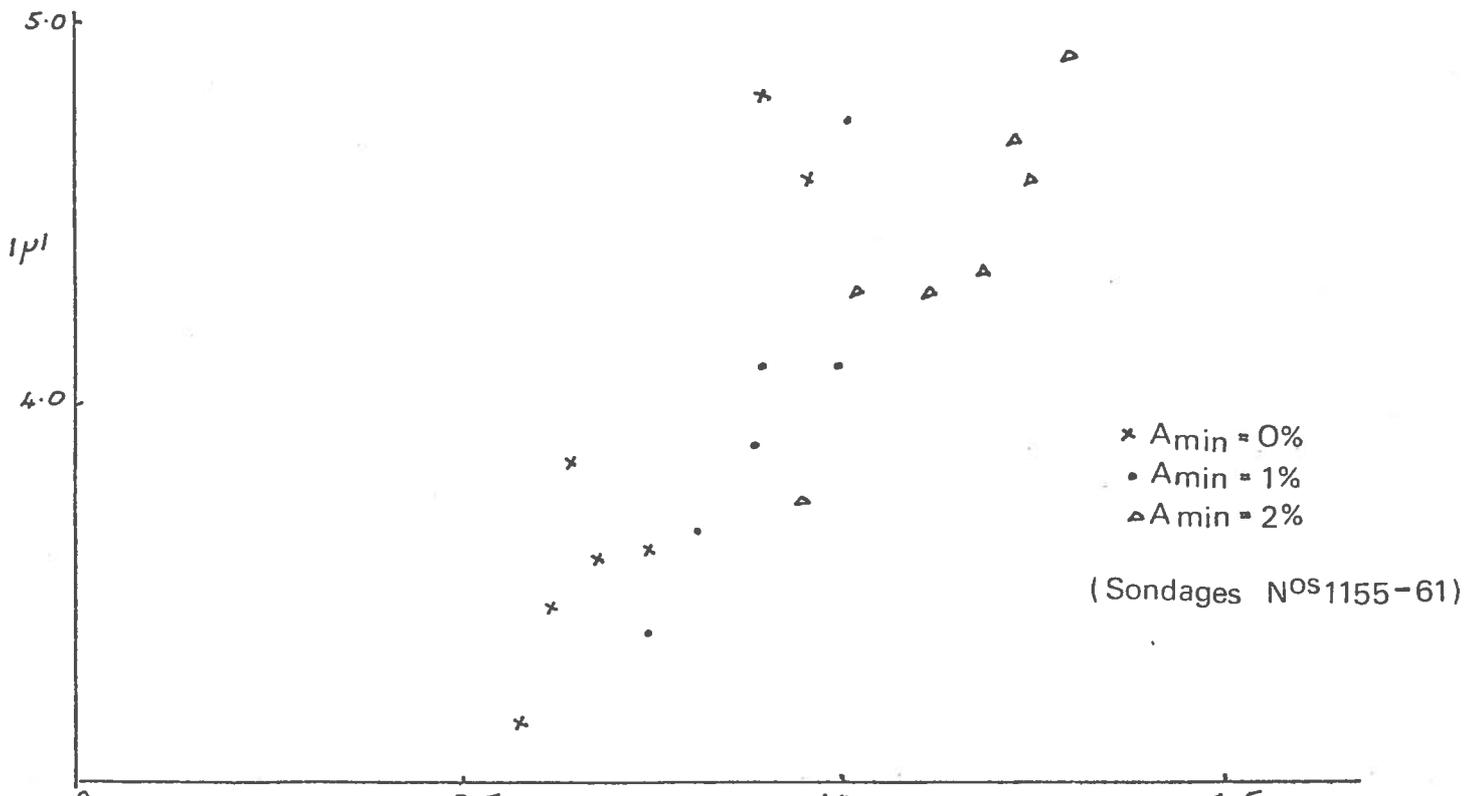


FIGURE A6



prédictions des facteurs de récupération pour une densité de 1,3 ne sont pas acceptables (voir Tableau A12). Comme la densité de lavage du charbon est, en général, entre 1,4 et 1,45, cela ne pose pas de problème sérieux. Dans la gamme des densités utilisées en pratique, les résultats sont plutôt bons.

Borne supérieure de densité	Facteur de récupération	
	expérimental	estimé ($A_{\min} = 0,5 \%$)
1,3	54,6	61,8
1,35	80,9	82,4
1,375	87,6	86,6
1,4	91,0	89,4
1,425	93,5	91,5
1,45	94,8	92,9
1,5	95,8	95,0
1,55	96,5	96,3
1,6	97,7	97,2

TABLEAU A12 : Facteurs de récupération expérimentaux et estimés (pour le sondage N° 908).

Bien que les facteurs de récupération soient en général satisfaisantes, le modèle a tendance à sous-estimer la teneur en cendre, et ceci est un défaut grave. Le tableau A13 présente le cas du sondage N° 901. On pourrait améliorer la prédiction de la teneur en cendre en augmentant la valeur de A_{\min} , mais cela baisserait la précision de l'estimateur du facteur de récupération.

Borne supérieure de densité	Teneur en cendre expérimentale	Teneur en cendre estimée	
		$A_{\min} = 0,5 \%$	$A_{\min} = 1,5 \%$
1,3	3,70	2,18	2,64
1,35	5,32	3,91	4,25
1,375	5,90	4,55	4,84
1,4	6,38	5,10	5,33
1,425	6,84	5,56	5,74
1,45	7,07	5,96	6,13
1,5	7,49	6,61	6,74
1,55	7,75	7,14	7,21
1,6	7,91	7,54	7,60

TABLEAU A13 : Teneurs en cendre, expérimentale et estimée (Sondage N° 901).

Initialement, on avait pensé que l'erreur se produisait dans la première classe de densité ($< 1,3$), mais en fait la teneur en cendre est sous-estimée pour les premières classes et surestimée pour les autres classes. (Tableau A14).

Classes de densités	Teneur en cendre expérimentale	Teneur en cendre estimée	
		$A_{\min} = 0,5 \%$	$A_{\min} = 1,5 \%$
1,3	3,70	2,18	2,64
1,3 - 1,35	7,60	6,28	6,19
1,35 - 1,375	10,95	11,13	10,71
1,375 - 1,4	12,68	13,31	13,33
1,4 - 1,425	14,55	15,70	15,66
1,425 - 1,45	15,99	18,12	18,19
1,45 - 1,5	18,30	21,24	23,64
1,5 - 1,55	22,53	25,73	25,60
1,55 - 1,60	26,4	29,44	29,80

TABLEAU A14 : Différences entre la vraie teneur en cendre de chaque classe et la valeur estimée (Sondage N° 901).

On peut se demander la raison de cette imprécision. Il est évident que le modèle ne reproduit pas assez bien la vraie courbe de la teneur en cendre en fonction de la densité mais pourquoi ? A-t-on de mauvais estimateurs de ρ_a et de ρ_c ? Ou est ce dû à un mauvais choix de la forme de la relation entre ces deux variables ? Pour élucider cette question, on a tracé les teneurs en cendre expérimentales en fonction de la densité pour les sondages N° 901 - 949 (Figure A7). La courbe est évidemment trop haute puisque l'estimateur de ρ_a pour ce sondage est 3,83, tandis que la moyenne pour les sondages N° 901 - 949 est 3,078. Ce mauvais ajustement est peut-être lié à une sous-estimation de ρ_a , mais cela n'explique pas pourquoi les teneurs en cendre sont toujours sous-estimées. Si le problème était dû à l'utilisation de la valeur moyenne de ρ_a , on devrait s'attendre à ce que les teneurs soient surestimées pour une partie des sondages, mais cela n'est pas le cas.

Une autre explication possible est que le modèle de la courbe de la teneur en cendre en fonction de la densité ne convienne pas. Au vu de la figure A7, la courbe ne semble pas être assez rapidement croissante pour les basses densités (1,25 - 1,35). Il serait donc intéressant d'étudier

le comportement de cette courbe à basses densités, mais à l'heure actuelle l'ensemble des données disponibles ne le permet pas. Pour y arriver, il faudrait faire des analyses sur des classes de densité très étroites.

A-7 Conclusions

Dans la section A5, on a établi une relation linéaire entre μ et σ . On a vu qu'il serait intéressant que la valeur de ρ_a soit constante dans le même domaine que celle de ρ_c . Mais bien que les résultats pour les sondages N^{OS} 1155 - 1161 satisfassent cette hypothèse, les résultats pour les autres sondages ne sont pas aussi encourageants. Il faudra donc tester cette hypothèse avec un jeu de données de lavabilité plus détaillé.

Il faut aussi prêter attention à un autre aspect du modèle qui n'est pas tout à fait satisfaisant; c'est la sous-estimation des teneurs en cendre. Bien que le modèle lognormal donne de bons résultats, la procédure pour estimer les teneurs en cendre a fourni des estimateurs inacceptables. La raison peut être une mauvaise estimation de ρ_a et de ρ_c , ou un mauvais choix du modèle de la relation entre la teneur en cendre d'une particule de charbon et sa densité. On espère disposer dans un proche avenir d'un jeu de données convenable.

B I B L I O G R A P H I E

BIBLIOGRAPHIE

- ABRAMOWITZ M. et STEGUN I.A. (1964), Handbook of Mathematical Functions, National Bureau of Standards, U.S. Government Printing Office, Washington, 1046 pp.
- AGTERBERG F.P. et CHUNG C.F. (1973), Geomathematical prediction of sulphur in coal, New Lingan mine area, Sydney Coalfield, Can. Min. Metall. Bull., N° 738, pp. 85-96.
- AITCHINSON J. et BROWN J.A.C. (1957), The lognormal distribution with special reference to its uses in economics. Cambridge University Press, p. 17.
- ALPERN B. (1977), Evaluation of the energy potential of carbonaceous sediments, World Coal, Vol. 8, N° 3, pp. 17-20.
- ARMSTRONG M. et WHITMORE R.L. (1979), Modelling the washability characteristics of coal, Internal report N° N-619 CGMM, Ecole des Mines de Paris, Fontainebleau. (Accepted for publication in "Sciences de la terre informatique").
- BENNETT A.J.R. et TAYLOR G.H. (1970), A petrographic basis for classifying Australian coals, Proc. Australas. Inst. Min. Metall., N° 233, pp. 1-6.
- HOURIGAN J.A. (1977), Mining and Reclamation in Queensland, International Mining and Reclamation Conference, May 1977.
- HUBER P.J. (1964), Robust estimation of a location parameter, Ann. Math. Stat. N° 35, pp. 73-101.
- JOURNEL A.G. et HUIJBREGTS C.J. (1978), "Mining Geostatistics", Academic Press, London, 600 pp.
- KOLMOGOROFF A.N. (1941), Über das logarithmisch normale Verteilungsgesetz der Dimensionen der Teilchen bei Zerstückelung, C. R. Acad. Sci. URSS, 31, 99.
- MARECHAL A. (1976), The practice of transfer functions : numerical methods and their application, in Proc. NATO ASI, Rome 1975, Advanced Geostatistics in the mining industry, ed. M. Guarascio et al., pp. 253-76, Reidel, Holland.

- MATHERON G. (1960), *Traité de Géostatistique appliquée (Tome 1)*, Mémoires de Bureau de Recherches Géologiques et Minières, N° 14, Editions Technip, Paris, pp. 308-311.
- MATHERON G. (1976 a), A simple substitute for conditional expectation : the disjunctive kriging, in Proc. NATO ASI, Rome 1975, *Advanced Geostatistics in the mining industry*, ed. M. Guarascio et al., pp. 221-36, Reidel, Hollande.
- MATHERON G. (1976 b), Forecasting block grade distributions : the transfer functions, in Proc. NATO ASI, Rome 1975, *Advanced Geostatistics in the mining industry*, ed. M. Guarascio et al., pp. 237-51, Reidel, Holland.
- SABOURIN R. (1975), Geostatistical Evaluation of Sulphur Content in Lingan coal mine, Cape Breton, Proc. 13th Int. APCOM Symp., Claustal, pp. I-II, 1-16.
- TIEMAN (1973), Chemistry of Coal, in "Elements of practical coal mining", Ed. Cassidy, S. N., A.I.M.E., New York.
- DE WIJS H.R. (1953), Statistics of ore distribution : Part II, Theory of binomial distribution applied to sampling and engineering problems, *Geologie Minj b.*, 15, pp. 12-24.