



document public

logiciels géostatistiques de GDM manuel de l'utilisateur

B. Bourgine

août 1995
R 37842

n° de référence R00501507

BRGM
DIRECTION DE LA RECHERCHE
Département Géophysique et Imagerie Géologique
SERVICE GÉOLOGIQUE NATIONAL
Département Infrastructure Géologique et Géophysique
Groupe Géophysique Structures profondes représentation 3D
B.P. 6009 - 45060 ORLEANS CEDEX 2 - France - Tél. : (33) 38.64.34.34

Mots clés : Géostatistique.

En bibliographie, ce rapport sera cité de la façon suivante :

BOURGINE B. (1995) - Logiciels géostatistiques de GDM : Manuel de l'utilisateur. Rapport BRGM R 37842, 81 p., 20 fig.

© BRGM, 1995, ce document ne peut être reproduit en totalité ou en partie sans l'autorisation expresse du BRGM.

RÉSUMÉ

Ce document est destiné aux utilisateurs des logiciels de géostatistique de GDM. Il explique les principaux concepts théoriques mis en oeuvre dans ces logiciels et donne des indications pratiques sur la façon de les utiliser. Un manuel "Logiciels géostatistiques de GDM : Cas d'Etude" vient en complément de ce manuel et l'illustre par des exemples.

Les aspects passés en revue sont les suivants :

1 - Étapes d'une étude géostatistique

Ce chapitre résume les différentes étapes à mettre en oeuvre lors d'une étude géostatistique.

2 - Préparation des données et prétraitements statistiques

Ces aspects sont examinés dans le cadre de l'utilisation du logiciel GDM.

3 - Modèles géostatistiques

Les grandes lignes des modèles géostatistiques mis en oeuvre dans GDM sont rappelées (modèles stationnaires ou intrinsèques d'ordre 0 ou k, dérive externe, modèles bivariés ponctuels).

4 - Variographie

Ce chapitre assez long donne un certain nombre d'indications sur le choix des paramètres de calcul du variogramme (direction, pas, tolérance angulaire), ainsi que sur les paramètres de l'ajustement (choix de l'effet de pépite, du palier, ajustement d'une anisotropie géométrique ou zonale). La reconnaissance automatique de structure et la validation de modèles de variogrammes sont également abordées.

5 - Krigeage

Cette partie expose les différents types de krigeage proposés par GDM et rappelle brièvement leurs propriétés et les équations mises en jeu. Le choix des paramètres du krigeage (voisinage, délimitation du domaine à estimer, ...) est également discuté.

6 - Cas multivariable

Les possibilités de traitement multivariable sont présentées.

7 - Simulations

Le principe des simulations conditionnelles est exposé. Des indications sont données sur les paramètres à introduire dans les programmes correspondant.

8 - Réserves récupérables globales et anamorphose

Les aspects théoriques du calcul des réserves et de l'anamorphose sont rappelés. Des commentaires sont fournis sur l'effet de support et sa modélisation.



TABLE DES MATIÈRES

INTRODUCTION.....	9
1. ÉTAPES D'UNE ÉTUDE GÉOSTATISTIQUE.....	11
2. PRÉPARATION DES DONNÉES ET PRÉTRAITEMENTS STATISTIQUES	13
2.1. Étude en 2D.....	13
2.2. Étude en 3D.....	13
2.3. Traitements complémentaires en présence de failles.....	13
2.4. STATISTIQUES DE BASE.....	13
3. MODÈLES GÉOSTATISTIQUES	15
3.1. Modèles stationnaires à moyenne connue	15
3.2. Modèles stationnaires à moyenne inconnue et modèles intrinsèques d'ordre 0	15
3.3. Modèles intrinsèques d'ordre k ; krigeage universel	15
3.3.1. Modèle du krigeage universel	15
3.3.2. Modèle des fonctions aléatoires d'ordre k (FAI-k)	15
3.4. Modèles avec dérive externe	19
3.5. Modèles bivariables ponctuels	19
3.5.1. Fonction d'anamorphose.....	20
3.5.2. Densité des lois bivariées	20
3.5.3. Modèle mosaïque.....	20
3.5.4. Modèle pur (bigaussien ou bigamma)	21
3.5.5. Modèle barycentrique.....	21
3.5.6. Modèle bêta	21
3.5.7. Comparaison et choix des modèles	21

4. VARIOGRAPHIE.....	23
4.1. Principes du calcul du variogramme.....	24
4.1.1. Choix de la distance de calcul.....	24
4.1.2. Choix des directions de calcul	24
4.1.3. Choix de la tolérance angulaire.....	25
4.1.4. Variogramme omnidirectionnel.....	25
4.2. Calcul du variogramme en 2D	25
4.3. Calcul du variogramme en 3D	26
4.3.1. Quand calculer un variogramme horizontal ou vertical ?	28
4.3.2. Quand calculer un variogramme le long des sondages ?.....	28
4.4. Variogrammes robustes	29
4.5. Ajustement du variogramme.....	30
4.5.1. Types de modèles autorisés	30
4.5.2. Choix du palier total du variogramme.....	34
4.5.3. Jusqu'à quelle distance ajuster le variogramme ?	34
4.5.4. Choix de l'effet de pépite	34
4.5.5. Ajustement d'une anisotropie géométrique.....	36
4.5.6. Ajustement d'une anisotropie zonale	37
4.5.7. Cas d'un variogramme sur des données régularisées	40
4.6. Variographie en présence d'une dérive.....	41
4.6.1. Variogramme des résidus	42
4.6.2. Variogramme généralisé.....	42
4.6.3. Variogramme avec dérive externe.....	43
4.7. Reconnaissance automatique.....	43
4.8. Validation du variogramme	44
5. KRIGEAGE	47
5.1. Krigeage simple.....	47
5.2. Krigeage ordinaire	48
5.3. Krigeage universel	50
5.4. Krigeage en FAI-K	50
5.5. Krigeage avec dérive externe.....	50

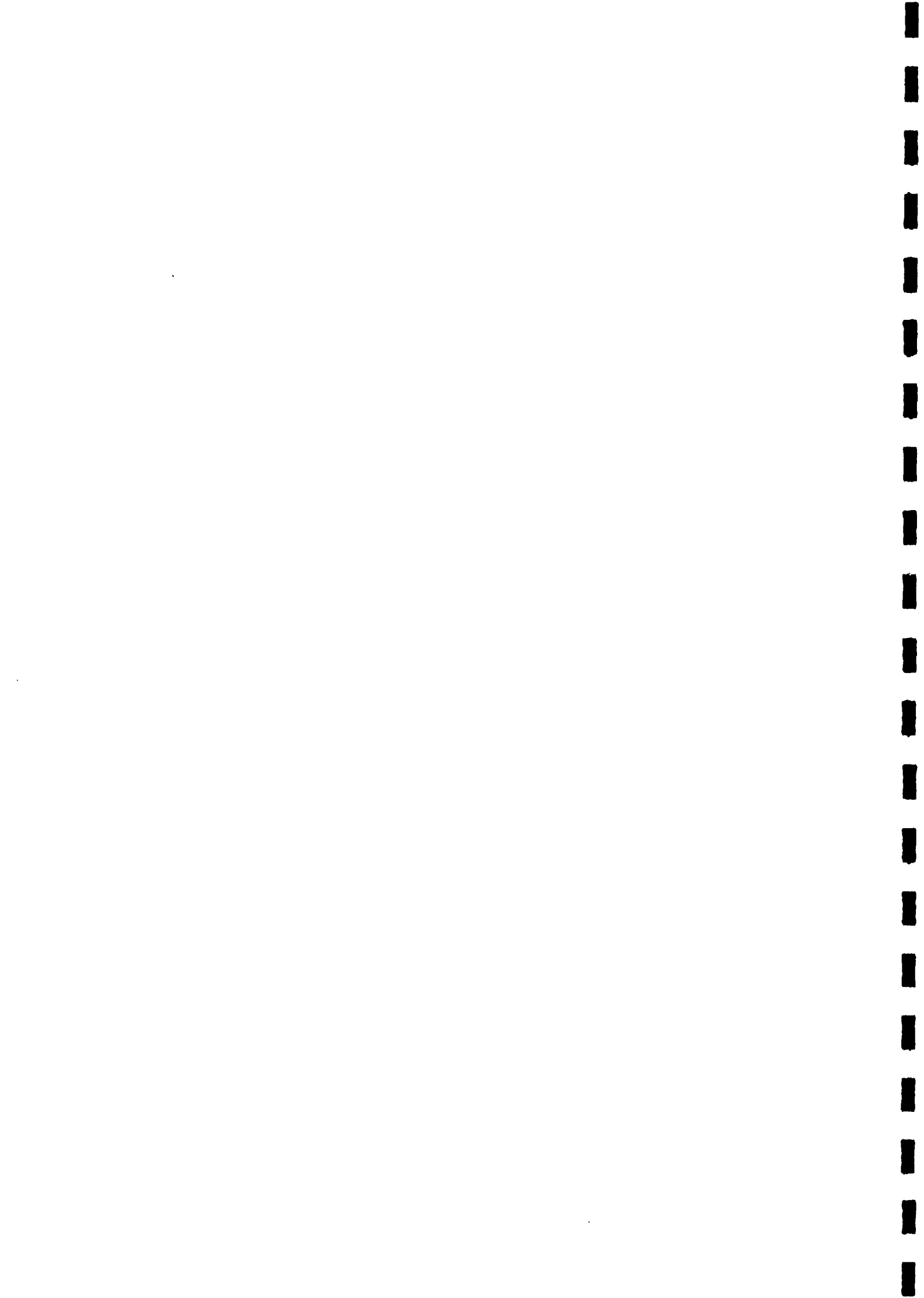
5.6. Mise en oeuvre d'un krigeage en 2 ou 3 dimensions	51
5.6.1. Krigeage en 2 dimensions.....	54
5.6.2. Krigeage en 3 dimensions.....	54
5.7. Exploitation des résultats, calcul de réserves.....	55
6. CAS MULTIVARIABLE	57
7. SIMULATIONS	59
7.1. Introduction	59
7.2. Construction d'une simulation non conditionnelle	59
7.3. Construction d'une simulation conditionnelle.....	61
7.4. Simulation de variables corrélées.....	62
7.5. Prise en compte de l'histogramme.....	62
7.6. Fichiers nécessaires pour une simulation	63
8. RÉSERVES RÉCUPÉRABLES GLOBALES ET ANAMORPHOSE.....	65
8.1. Courbes tonnage-teneur	65
8.1.1. Tonnage minéral récupéré en fonction de la coupure : $T(z_c)$	65
8.1.2. Quantité de métal récupérée en fonction de la coupure : $Q(z_c)$	65
8.1.3. Teneur moyenne récupérée en fonction de la coupure : $m(z_c)$	66
8.1.4. Bénéfice conventionnel en fonction de la coupure : $B(z_c)$	66
8.1.5. Quantité de métal récupérée en fonction du tonnage minéral récupéré : $Q(T)$	68
8.2. Effet de support et changement de support. Réserves idéales	70
8.2.1. Correction affine de support.....	71
8.2.2. Modèle continu	73
8.2.3. Modèle discret	73
8.3. Effet d'information. Réserves prévues et récupérées.....	73
8.3.1. Réserves prévues	74
8.3.2. Réserves récupérées, effet d'information	74
8.4. Anamorphose	75
8.4.1. Anamorphose graphique.....	78
8.4.2. Anamorphose polynômiale	80
RÉFÉRENCES BIBLIOGRAPHIQUES	81

LISTE DES FIGURES

Figure 1	: Ellipse d'anisotropie dans le plan XY	26
Figure 2	: Définition des tolérances pour le calcul du variogramme horizontal en 3 dimensions (K3GAM)	27
Figure 3	: Définition des tolérances pour le calcul du variogramme vertical en 3 dimensions (K3GAM)	27
Figure 4	: Modèles sphérique, exponentiel, cubique, gaussien	32
Figure 5	: Modèles gravimétrique, magnétique et h^α	33
Figure 6	: Variogramme expérimental : incertitude sur l'effet de pépite	35
Figure 7	: Croix de sondages	35
Figure 8	: Variogramme 2D présentant une anisotropie géométrique.....	39
Figure 9	: Variogramme 3D présentant une anisotropie zonale selon la verticale.....	39
Figure 10	: Tonnage récupéré en fonction de la teneur de coupure.....	67
Figure 11	: Quantité de métal récupérée en fonction de la teneur de coupure	67
Figure 12	: Teneur moyenne récupérée en fonction de la teneur de coupure	68
Figure 13	: Bénéfice conventionnel en fonction de la teneur de coupure	69
Figure 14	: Quantité de métal récupérée en fonction du tonnage minéral récupéré.....	69
Figure 15	: Densités de probabilités de variables mesurées sur des supports différents	72
Figure 16	: Fonctions de répartition de variables mesurées sur des supports différents	72
Figure 17	: Principe de l'anamorphose	77
Figure 18	: Transformation de l'histogramme par anamorphose	77
Figure 19	: Allure de la fonction anamorphose	79
Figure 20	: Anamorphose graphique.....	79

INTRODUCTION

Ce document est destiné aux utilisateurs des logiciels de géostatistique de GDM. Il vient en complément des documentations utilisateur de ces logiciels, et présente les aspects théoriques qui ne sont pas toujours abordés dans ces documentations. Il donne également des indications pratiques sur la façon d'utiliser les logiciels de géostatistique, depuis la préparation des données jusqu'au choix des principaux paramètres des programmes en fonction du résultat recherché. Un manuel "Logiciels géostatistiques de GDM : Cas d'Etude" vient en complément de ce manuel et l'illustre par des exemples.



1. ÉTAPES D'UNE ÉTUDE GÉOSTATISTIQUE

Schématiquement une étude géostatistique comporte les phases suivantes :

1 - Définition du problème : variables d'étude, travail en 2D ou en 3D, support

Le problème sera d'autant plus facile à résoudre qu'il est bien posé dès le départ. Ne pas hésiter donc à passer du temps au début pour bien définir le problème.

Un des premiers choix consiste à décider si le traitement doit être effectué en 2 dimensions (2D) ou en 3 dimensions (3D). Dans la mesure du possible, il est préférable de se ramener au cas 2 dimensions, car les traitements sont alors plus simples.

Il est important également de bien choisir dès le départ les variables d'étude, que l'on travaille en 2D ou en 3D. Il faut se souvenir que les traitements géostatistiques ne peuvent s'appliquer qu'à des variables additives, c'est-à-dire pour lesquelles des moyennes ont un sens. Ainsi la blancheur d'un matériau comme le talc ou le kaolin n'est pas en général une variable additive ; de même une porosité mesurée sur des échantillons de tailles différentes n'est pas additive.

Les logiciels de géostatistique supposent implicitement que les données sont relatives à un même support. On ne doit donc pas mélanger indistinctement des teneurs provenant de carottes et d'autres provenant de puits ou de rainurages. Si les teneurs proviennent de carottes de longueur variable, on se ramènera d'abord à un support de longueur constante, ou alors on se ramènera à 2D en travaillant avec la puissance et l'accumulation.

D'autre part, d'autres variables ou des codes (code géologique, par exemple) peuvent être utiles et servir de masques de sélection pour les traitements à effectuer sur les variables d'étude proprement dites. Il faut donc également prévoir ces variables suffisamment tôt.

2 - Saisie et contrôle des données

Cette phase n'est pas à négliger car elle peut représenter parfois près de 50% de la durée de l'étude. Le contrôle des données nécessite la réalisation de nombreuses cartes, coupes, listings, statistiques, ainsi que des corrections si nécessaire.

3 - Statistiques de base

Ces statistiques servent à la fois au contrôle des données et à la caractérisation des variables et données à étudier. Elles permettent également de classer les données en populations homogènes sur lesquelles un traitement différent pourra être effectué.

4 - Analyse structurale

L'analyse structurale regroupe le calcul du variogramme, son ajustement et éventuellement la validation de l'ajustement. C'est une phase délicate puisque d'elle va dépendre la qualité des estimations qui seront faites par la suite.

5 - Krigeage (et/ou simulations et calcul de réserves récupérables)

Le krigeage ne pose en principe pas de difficulté (sauf peut-être le choix du voisinage), si l'on dispose d'un modèle géostatistique bien ajusté.

Les simulations et les calculs de réserves récupérables sont des techniques géostatistiques plus délicates qui nécessitent un minimum de connaissances théoriques avant de pouvoir être appliquées.

6 - Calcul de réserves

Cette phase concerne les applications minières, et fait intervenir des programmes plus aval comme K3STAT ou PITCAD.

2. PRÉPARATION DES DONNÉES ET PRÉTRAITEMENTS STATISTIQUES

2.1. ÉTUDE EN 2D

Le fichier à construire doit contenir au minimum les coordonnées X,Y, ainsi que les variables d'étude. On peut également inclure des variables complémentaires ou des codes permettant de classer les données en différents groupes et d'opérer des sélections sur les points de données à considérer pour les traitements.

Si on étudie un gisement filonien ou stratiforme, il est possible de se ramener de données de sondages en 3 dimensions à des données en 2 dimensions grâce au programme MPROJ (projection sur un plan) ou MCUTOF (sélection sur sondage). On travaille alors sur la puissance minéralisée et l'accumulation métal (puissance*teneur moyenne).

2.2. ÉTUDE EN 3D

Le fichier à construire doit contenir au minimum les 3 coordonnées X,Y,Z, ainsi que les variables d'étude. On peut également inclure des variables complémentaires ou des codes permettant de classer les données en différents groupes.

Des opérations complémentaires sont nécessaires pour le calcul du variogramme horizontal ou vertical et le krigeage : il faut classer les données dans un réseau de blocs (programme K3CLAS). S'il s'agit de sondages et que les passes ne sont pas de longueur constante, il faut avant K3CLAS découper les sondages en passes ou en bancs de hauteur constante (programme MSLICE).

2.3. TRAITEMENTS COMPLÉMENTAIRES EN PRÉSENCE DE FAILLES

Les failles ou discontinuités peuvent être prises en compte en 2 dimensions dans les programmes K2REC et K2KRI (et GDISP et GISOF pour la visualisation). Elles sont définies comme des lignes brisées et sont considérées comme des écrans.

Pour cela il faut créer un fichier GDM de type triplet (.HED, .BSD, .IND), l'enregistrement tête comprenant les coordonnées du premier point de la faille, ainsi qu'un code d'identification de la faille, les enregistrements base étant les coordonnées des autres points de la faille.

En trois dimensions les failles ne peuvent être prises en compte.

2.4. STATISTIQUES DE BASE

Les statistiques de base sont indispensables avant tout traitement géostatistique. Elles permettent le contrôle et la meilleure connaissance des données.

Les statistiques sur les coordonnées (X,Y,Z) permettent de localiser les points aberrants (erreur de saisie). Les nuages de corrélation (X,Y), (X,Z), (Y,Z) qui peuvent être calculés à l'aide du programme MSTAT permettent d'obtenir rapidement une carte des données en projection sur un des plans de coordonnées. Le programme GDISP permet également de tracer des cartes de position des données, avec superposition de valeurs de variables, en 2D ou 3D.

Les statistiques sur les variables (histogrammes, nuages de corrélation - programmes MSTAT, HISTOC, BINAIR) peuvent également permettre de détecter des valeurs anormales. Les très fortes valeurs provoquant une déstructuration du variogramme, il est important de les répertorier et de localiser les points expérimentaux correspondants. D'autre part, si plusieurs populations apparaissent sur les histogrammes, il faut en analyser la cause, déterminer si ces populations peuvent être localisées et faire l'objet de traitements séparés.

Les nuages de corrélation (teneur,coordonnée) ou (teneur,profondeur) permettent de mettre en évidence d'éventuels enrichissements dans différentes directions de l'espace.

3. MODÈLES GÉOSTATISTIQUES

3.1. MODÈLES STATIONNAIRES À MOYENNE CONNUE

Dans les modèles stationnaires on fait l'hypothèse que le phénomène étudié est homogène dans l'espace, c'est-à-dire que :

1 - $E[Z(x)] = m$

La moyenne existe et est indépendante de la position x du point considéré

2 - La covariance entre $Z(x)$ et $Z(x+h)$ existe et ne dépend que de la distance h séparant les points x et $x+h$.

On a alors : $C(h) = \text{Cov}(Z(x), Z(x+h)) = E[Z(x+h)Z(x)] - m^2$.

En particulier pour $h = 0$, $C(0) = E[Z(x)^2] - m^2 = \text{Var}[Z(x)]$: la variance existe, mais son calcul nécessite la connaissance de m .

Ces modèles supposent une grande homogénéité du phénomène et un grand nombre de données pour que la moyenne m puisse être connue, ce qui se rencontre rarement en pratique. Ils sont en général utilisés dans le cadre de la recherche pour tester et valider des méthodes et modèles géostatistiques. La méthode d'estimation utilisée dans ce cas est le krigeage simple.

Certaines techniques particulières de géostatistique peuvent utiliser le krigeage simple et les modèles stationnaires (simulations de fonctions aléatoires, calcul de réserves récupérables).

3.2. MODÈLES STATIONNAIRES À MOYENNE INCONNUE ET MODÈLES INTRINSÈQUES D'ORDRE 0

Dans ces modèles la moyenne de la variable aléatoire sous-jacente est inconnue.

Lorsque le modèle est intrinsèque, la variance du phénomène peut ne pas exister (c'est-à-dire être infinie).

Le phénomène est donc beaucoup moins stable dans l'espace. Par contre ses accroissements $Z(x+h) - Z(x)$ le sont, et on fait l'hypothèse qu'ils sont stationnaires, c'est-à-dire que :

1 - $E[Z(x+h) - Z(x)]$ existe et est indépendante de x .

On suppose en fait $E[Z(x+h) - Z(x)] = 0$, c'est-à-dire $E[Z(x+h)] = E[Z(x)] = m$, m pouvant être inconnue, voire indéfinie.

2 - la variance des accroissements existe et est indépendante de x :

$\text{Var}[(Z(x+h) - Z(x))] = E[(Z(x+h) - Z(x))^2] = 2\gamma(h)$.

La fonction $\gamma(h)$ est appelée demi-variogramme, ou plus simplement variogramme. Son calcul ne nécessite pas de connaître la moyenne m .

Les modèles intrinsèques d'ordre 0 sont en pratique les modèles utilisés lorsqu'il n'y a pas de dérive ou tendance. Les calculs dans ce modèle nécessitent la détermination expérimentale et l'ajustement de la fonction variogramme $\gamma(h)$. Ils conduisent au krigeage dit "ordinaire".

3.3. MODÈLES INTRINSÈQUES D'ORDRE K ; KRIGEAGE UNIVERSEL

3.3.1. Modèle du krigeage universel

Cette fois la variable étudiée n'est plus stationnaire ni intrinsèque d'ordre 0, mais présente une dérive qui dépend des coordonnées : par exemple la profondeur d'une couche qui augmente vers le sud-est du domaine considéré.

Dans ce cas le principe consiste à décomposer la variable en deux :

$$Z(x) = Y(x) + m(x)$$

$Y(x)$ est une variable stationnaire appelée **résidu**

$m(x)$ est la **dérive** pouvant être exprimée sous la forme d'un polynôme de degré n , dépendant des coordonnées (u,v,w) des points. Avec deux coordonnées (u,v) et un degré 2, $m(x)$ prend la forme : $m(x) = a + bu + cv + duv + eu^2 + fv^2$

En pratique on ne peut pas effectuer cette décomposition de façon exacte et on procède de la façon suivante :

1. La dérive (c'est-à-dire les coefficients du polynôme $m(x)$) est estimée par moindres carrés à partir de l'ensemble des $Z(x_i)$ disponibles ; on obtient ainsi l'estimation $m^*(x)$.
2. On construit alors $Y^*(x_i) = Z(x_i) - m^*(x_i)$ qui représente une estimation du résidu au point x_i , une fois retranchée de la valeur vraie $Z(x_i)$ l'estimation de la dérive $m^*(x_i)$. Le variogramme de Y^* peut alors être calculé et ajusté par un modèle stationnaire.
3. On estime alors par krigeage "universel" la valeur $Z^*(x)$. Ceci revient à estimer simultanément $Y^*(x)$ et les coefficients de la dérive $a^*, b^*, \text{etc.}$ En théorie, on devrait utiliser le variogramme des résidus vrais $Y(x)$. En pratique, celui-ci n'est pas calculable, et on lui substitue le variogramme des estimations des résidus $Y^*(x_i)$. Ce variogramme est biaisé par rapport au variogramme vrai des résidus, particulièrement aux grandes distances.

Pour cette raison, le krigeage universel est délicat à manipuler et nécessite une certaine expérience.

3.3.2. Modèle des fonctions aléatoires d'ordre k (FAI-k)

La théorie des FAI-k (fonctions aléatoires intrinsèques d'ordre k) permet de résoudre les problèmes non stationnaires, présentant une dérive.

La variable $Z(x)$, comme dans le cas du krigeage universel, peut se mettre sous la forme :

$$Z(x) = Y(x) + m(x)$$

$m(x)$ est une **dérive polynômiale de degré k** dépendant des coordonnées,
 $Y(x)$ est un **résidu** dont les accroissements, ou au moins les accroissements d'ordre k , sont stationnaires.

Mais contrairement au krigeage universel, on ne cherche pas à calculer le variogramme des résidus $Y(x)$, et à séparer l'étude de $Y(x)$ de celle de $m(x)$.

L'idée est de travailler directement sur la variable $Z(x)$, et plus précisément sur les **accroissements généralisés** de $Z(x)$.

Sans entrer dans la théorie des FAI- k , on peut faire une analogie avec le krigeage ordinaire :

Accroissements d'ordre 1

Dans le krigeage ordinaire, on met en jeu le variogramme, qui correspond à des accroissements d'ordre 1 : $(Z(x) - Z(x+h))$.

$Z(x)$ est dite fonction aléatoire intrinsèque d'ordre $k = 0$ (FAI-0) car ses accroissements d'ordre $k+1 = 1$ filtrent une dérive de degré $k = 0$.

Ainsi, si $Z(x) = Y(x) + a_0$, c'est à dire $m(x) = a_0$ (monôme de degré 0), on a :
 $Z(x) - Z(x+h) = Y(x) - Y(x+h)$ qui est stationnaire puisque Y est stationnaire.

Cet accroissement d'ordre 1 au point x : $A_1(x) = Z(x) - Z(x+h)$ est obtenu en attribuant un poids 1 à $Z(x)$ et un poids -1 à $Z(x+h)$.

Accroissements d'ordre 2

Si l'on calcule maintenant les accroissements d'ordre 2 par $A_2(x) = A_1(x) - A_1(x+h)$ on obtient :

$$\begin{aligned} A_2(x) &= (Z(x) - Z(x+h)) - (Z(x+h) - Z(x+2h)) \\ A_2(x) &= Z(x) - 2Z(x+h) + Z(x+2h) \end{aligned}$$

Cette nouvelle combinaison linéaire filtre une dérive de degré $k = 1$.

Ainsi, en une dimension, si $Z(x) = Y(x) + a_0 + a_1x$, on peut vérifier que :

$$\begin{aligned} A_2(x) &= Y(x) - 2Y(x+h) + Y(x+2h) + a_0 - 2a_0 + a_0 + a_1(x - 2(x+h) + x+2h) \\ &= Y(x) - 2Y(x+h) + Y(x+2h) \end{aligned}$$

$A_2(x)$ est donc stationnaire.

La combinaison linéaire $A_2(x) = Z(x) - 2Z(x+h) + Z(x+2h)$ est dite "accroissement généralisé" d'ordre 2 : elle filtre tous les monômes de degré inférieur ou égal à 1. On a ainsi :

$$\sum_{i=1,3} \lambda_i f^m(x_i) = 0 \quad \text{pour } m = 0 \text{ et } 1$$

avec :

$$\begin{aligned} \lambda_1 &= 1 & x_1 &= x \\ \lambda_2 &= -2 & x_2 &= x+h \\ \lambda_3 &= 1 & x_3 &= x+2h \\ f^0(x) &= 1 \\ f^1(x) &= x \end{aligned}$$

$Z(x) = Y(x) + a_0 + a_1 x$ est alors appelée fonction aléatoire intrinsèque d'ordre 1 (FAI-1).

Accroissements généralisés d'ordre $k + 1$

Ce qui vient d'être montré peut être généralisé à un ordre quelconque :

Soit une configuration de points $\{x_i ; i=1,j\}$

Un accroissement généralisé d'ordre $k+1$ est une combinaison linéaire : $\sum_{i=1,j} \lambda_i Z(x_i)$

telle que : $\sum_{i=1,j} \lambda_i f^m(x_i) = 0$ pour $m = 0, \dots, k$

Dans ce cas : $\sum_{i=1,j} \lambda_i Z(x_i) = \sum_{i=1,j} \lambda_i Y(x_i)$: la dérive de degré k est filtrée.

$$Z(x) = Y(x) + \sum_{m=0,k} a_m f^m(x) \quad \text{est une FAI-}k$$

Covariance généralisée $K(h)$

De la même façon qu'on avait défini le variogramme $\gamma(h)$ à l'ordre 1 par :

$$E\{ (Z(x) - Z(x+h))^2 \} = 2\gamma(h),$$

on définit, à l'ordre $k+1$, la covariance généralisée $K(h)$ par la variance de la combinaison linéaire :

$$E\left[\sum_i \lambda_i Z(x_i) \right]^2 = \sum_i \sum_j \lambda_i \lambda_j K(h_{ij})$$

avec : $\sum_i \lambda_i Z(x_i)$ accroissement généralisé d'ordre $k+1$: filtre une dérive de degré k , et

h_{ij} = distance entre x_i et x_j .

$K(h)$ ne peut être calculée expérimentalement car il est en général impossible de trouver parmi les données expérimentales x_i une configuration de j points $\{x_i, i=1,j\}$ pouvant former une combinaison linéaire $\sum_i \lambda_i Z(x_i)$ telle que : $\sum_{i=1,j} \lambda_i f^m(x_i) = 0$ pour $m = 0, \dots, k$

Le calcul n'est possible en pratique que sur des grilles.

$K(h)$ est déterminé numériquement par approximations successives, dans des programmes dits de "reconnaissance automatique de structure". Les modèles choisis sont de la forme :

$$K(h) = C_0 - b_0.h + b_1.h^3 \dots$$

D'autres modèles sont toutefois possibles.

3.4. MODÈLES AVEC DÉRIVE EXTERNE

La dérive ne dépend pas toujours des coordonnées. Elle peut dépendre d'une ou plusieurs autres variables appelées variables ou composantes de la dérive externe, de telle façon que :

$$Z(x) = a + b.E(x) + Y(x)$$

$E(x)$ étant la dérive externe, et $Y(x)$ un résidu stationnaire.

Le krigeage avec dérive externe est utilisé dans différents cas :

- il existe une relation entre différentes variables.
Exemple : géothermie : $T(x) = a + b.P(x) + Y(x)$, avec $T(x)$ température en un point x , b gradient géothermique, et $P(x)$ = profondeur (dérive externe).
- estimation d'une surface qui fluctue autour d'une surface de référence connue (ou fixée).
Soit $Z(x)$ la cote de la surface et $Z_0(x)$ la cote de la surface de référence autour de laquelle $Z(x)$ varie. On a : $Z(x) = a + b.Z_0(x) + Y(x)$

Un exemple d'application est l'estimation d'une surface piézométrique $Z(x)$ à partir d'un réseau de surveillance réduit, lorsqu'on dispose d'une piézométrie de référence $Z_0(x)$ reconnue par un réseau de mesures plus dense.

3.5. MODÈLES BIVARIABLES PONCTUELS

Le calcul des réserves récupérables avec changement de support nécessite la connaissance des lois bivariées ponctuelles.

3.5.1. Fonction d'anamorphose

Les modèles simples de lois bivariées ne sont connus que pour des lois monovariées particulières telles que la loi de Gauss ou la loi Gamma. C'est pourquoi on est amené à effectuer une transformation préalable de la variable brute $Z(x)$ en une loi $Y(x)$ (gaussienne ou gamma par exemple). La fonction de transformation Φ est appelée fonction d'anamorphose de $Z(x)$ et est définie par :

$$Z(x) = \Phi(Y(x))$$

Elle est mise sous la forme d'une décomposition en polynômes $H_n(Y(x))$ de degré n , orthonormés et liés à la loi $Y(x)$ (polynômes d'Hermite pour la loi de Gauss et polynômes de Laguerre pour la loi Gamma) :

$$Z(x) = \Phi(Y(x)) = \sum_{n=0}^{+\infty} [\Phi_n \cdot H_n(Y(x))]$$

Les Φ_n étant les coefficients du développement (la section 8.4 donne des éléments complémentaires sur l'anamorphose).

3.5.2. Densité des lois bivariées

La densité bivariée de la loi $Y(x)$ liée à $Z(x)$ est alors exprimée sous la forme :

$$f_{xx'}(y, y') = \sum_{n=0}^{+\infty} \rho_n(x, x') H_n(y) H_n(y') g(y) g(y')$$

où :

- x et x' sont deux points de l'espace
- y et y' sont les valeurs de $Y(x)$ aux points x et x'
- $g(y)$ est la densité de probabilité marginale de $Y(x)$, $g(y')$ celle de $Y(x')$
- $\rho_n(x, x')$ est le coefficient de corrélation entre $H_n(y)$ et $H_n(y')$.

La densité bivariée $f_{xx'}(y, y')$ est calculable si on sait calculer les $\rho_n(x, x')$.

Pour permettre le calcul des $\rho_n(x, x')$, quatre modèles sont proposés ; ces modèles permettent de calculer de façon plus ou moins complexe la valeur de $\rho_n(x, x')$ en fonction de ρ , ρ étant le coefficient de corrélation entre $Y(x)$ et $Y(x')$:

3.5.3. Modèle mosaïque

Pour ce modèle, tous les $\rho_n(x, x')$ sont confondus.

Si on note $h = x - x'$ (distance séparant x et x'), on a :

$$\rho_n(h) = \rho(h)$$

Ce modèle revient à effectuer une partition aléatoire de l'espace R^n , dans laquelle les teneurs sont constantes à l'intérieur d'un même compartiment, et indépendantes d'un compartiment à l'autre. La probabilité pour que $Z(x)$ et $Z(x+h)$ tombent dans le même compartiment est $\rho(h)$ et on a alors $Z(x) = Z(x+h)$. Inversement, pour x et $x+h$ dans des compartiments différents (ce qui a lieu avec la probabilité $1-\rho(h)$), $Z(x)$ et $Z(x+h)$ sont indépendants (mais de même loi).

3.5.4. Modèle pur (bigaussien ou bigamma)

Pour ce modèle on a :

$$\rho_n(h) = \rho^n(h)$$

3.5.5. Modèle barycentrique

Il s'agit d'une combinaison entre les deux précédents et il dépend d'un paramètre complémentaire r compris entre 0 et 1 :

$$\rho_n(h) = r \cdot \rho^n(h) + (1-r) \cdot \rho(h)$$

3.5.6. Modèle bêta

C'est le modèle le plus complexe.

Il est lié à la loi bêta et dépend d'un paramètre β compris entre 0 et $+\infty$:

$$\rho_n(h) = \frac{\Gamma(\beta)}{\Gamma(\beta+n)} \cdot \frac{\Gamma(\beta \cdot \rho(h) + n)}{\Gamma(\beta \cdot \rho(h))}$$

3.5.7. Comparaison et choix des modèles

Parmi ces modèles les extrêmes sont le modèle mosaïque (le plus discontinu) et le modèle pur (le plus continu).

Les deux autres modèles varient entre ces deux extrêmes en fonction des paramètres r ou β .

Le choix d'un des 4 modèles peut être guidé par l'examen du variogramme d'ordre 1 de la variable $Y(x)$. Ce variogramme est défini par :

$$\gamma_Y^1(h) = \frac{1}{2} E|Y(x) - Y(x+h)|$$

Ainsi, pour le modèle mosaïque, le variogramme d'ordre 1 et le variogramme d'ordre 2 sont proportionnels.

D'autres relations existent pour les autres modèles.

Le lecteur intéressé par des compléments d'information pourra se reporter à LIAO (1991), qui s'appuie lui-même sur MATHERON (1982) et HU (1988).



4. VARIOGRAPHIE

Définition

Le variogramme expérimental $\gamma(h)$ pour le vecteur $h=(r,\theta)$ est défini par :

$$\gamma(h) = \frac{1}{2 \cdot N_h} \sum_i [(Z(x_i) - Z(x_i+h))^2]$$

avec :

r	distance de calcul
θ	direction de calcul (θ n'apparaît pas explicitement dans la formule, mais le calcul du variogramme se fait en général selon une ou plusieurs directions particulières, sauf en cas d'isotropie).
x_i	point expérimental
x_i+h	point expérimental situé à la distance r de x_i selon la direction θ
$Z(x_i)$	valeur mesurée au point x_i
$Z(x_i+h)$	valeur mesurée au point x_i+h
N_h	nombre de couples de points (x_i, x_i+h) pouvant être formés pour la distance r et la direction θ .

La somme est étendue aux N_h couples possibles.

Le calcul pratique du variogramme nécessite donc :

- 1 - de disposer des coordonnées des points expérimentaux (pour le calcul de h),
- 2 - de choisir les distances et les directions de calcul. Ce choix doit être dicté par les considérations suivantes :
 - disposer de suffisamment de couples de calcul pour assurer que le résultat traduit bien un phénomène réel,
 - mettre en évidence d'éventuelles anisotropies (choisir les directions de calcul en fonction des directions d'anisotropie supposées).

Les principes du calcul du variogramme expérimental, en 2 dimensions ou en 3 dimensions, sont exposés dans les chapitres suivants.

4.1. PRINCIPES DU CALCUL DU VARIOGRAMME

4.1.1. Choix de la distance de calcul

Les données n'étant généralement pas situées à des distances régulières, le calcul n'est possible que par classes de distance.

Ces classes sont déterminées en fonction du pas p .

- . Classe 1 : distance = 0 (points doubles)
- . Classe 2 : distance de 0 à $0,5 p$
- . Classe 3 : distance de $0,5 p$ à $1,5 p$
- etc.

Le choix du pas p est laissé à la discrétion de l'utilisateur. Il faut choisir ce paramètre en fonction des considérations suivantes :

- 1 - le nombre de couples doit être suffisant, ceci pour chaque distance de calcul. Ainsi une valeur du variogramme reposant sur moins de 50 couples risque d'être peu représentative. Certains auteurs préconisent également de ne pas considérer les valeurs calculées avec un nombre de couples inférieur à $1/3$ du nombre total des données.
- 2 - il est inutile de calculer le variogramme à des distances supérieures à la moitié du champ d'observation. En effet, le nombre de couples possibles risque d'être faible pour les grandes distances. D'autre part, le calcul du variogramme aux grandes distances ne fait intervenir que des points situés dans la périphérie du domaine d'étude. Le résultat obtenu ainsi n'est pas représentatif de l'ensemble du domaine.
- 3 - pour gagner du temps calcul, on peut également limiter le calcul à des distances inférieures ou égales à la taille du voisinage qui sera utilisé lors du krigeage.
- 4 - le meilleur moyen pour obtenir un maximum de couples sans perdre en finesse est de choisir un pas voisin de la distance moyenne entre les données. Lorsque les données ne sont pas à maille régulière, on peut utiliser le programme GDISP pour visualiser les positions des données, et mesurer leur espacement. Des nuages de corrélation (X,Y), (X,Z) ou (Y,Z) obtenus avec le programme MSTAT permettent également de visualiser rapidement la position des données en projection sur un plan.

4.1.2. Choix des directions de calcul

Dans GEORES, la direction horizontale est comptée selon la convention géologique : 0° = Nord = axe des Y ; 90° = Est = axe des X.

Pour vérifier l'existence d'une éventuelle anisotropie, il faut calculer le variogramme expérimental dans plusieurs directions (minimum 4) : les deux directions principales d'anisotropie (ou du moins celles supposées), et deux directions intermédiaires orthogonales. Si les variogrammes sont voisins quelle que soit la direction on peut en conclure l'isotropie du phénomène (dans ce cas on ne retiendra que le variogramme omnidirectionnel calculé toutes directions confondues). Si les paliers sont voisins et que la portée varie, il s'agit d'une anisotropie géométrique. Si les paliers sont différents, l'anisotropie est dite zonale.

4.1.3. Choix de la tolérance angulaire

Lorsque la maille n'est pas rigoureusement régulière, il n'existe que peu de couples de points alignés selon une direction θ donnée.

Pour que le calcul du variogramme soit possible, il faut donc admettre une tolérance $d\theta$ autour de la direction θ . Pour une direction θ donnée, on considérera ainsi les couples de points pouvant être reliés par un segment dont la direction varie de $\theta-d\theta$ à $\theta+d\theta$.

Il est évident que pour mieux discriminer différentes directions de calcul, surtout lorsqu'on recherche une direction principale d'anisotropie, il faut choisir angles et tolérances angulaires de façon à ce que les secteurs angulaires $[\theta_i-d\theta_i; \theta_i+d\theta_i]$ ne se recouvrent pas. Ceci est vérifié automatiquement par le programme.

4.1.4. Variogramme omnidirectionnel

C'est le variogramme obtenu toutes directions confondues. Ce variogramme n'a de sens que si le phénomène est isotrope.

On le calcule également lorsqu'on dispose de trop peu de données pour espérer obtenir un variogramme représentatif selon plusieurs directions.

Le variogramme omnidirectionnel est obtenu en choisissant les paramètres de calcul suivants (clique GAM2) :

- nombre de directions (dans le plan horizontal) : 1,
- direction : 0° ,
- tolérance horizontale : 90° .

4.2. CALCUL DU VARIOGRAMME EN 2D

En cas d'anisotropie dans le plan horizontal, il faut déterminer les deux directions principales d'anisotropie en effectuant plusieurs calculs selon plusieurs directions, et en recherchant les directions donnant les portées minimales et maximales (ces deux directions devant être orthogonales).

Dans K2GAM, le programme propose par défaut une tolérance angulaire calculée de telle sorte que les secteurs angulaires soient jointifs et ne se chevauchent pas.

Par exemple, avec 4 directions de calcul: 0° , 45° , 90° , 135° , la tolérance angulaire est de 22.5° .

Les secteurs angulaires correspondant à chaque direction sont ainsi :

- direction 0° : de -22.5° à 22.5° ,
- direction 45° : de 22.5° à 67.5° ,
- direction 90° : de 67.5° à 112.5° ,
- direction 135° : de 112.5° à 157.5° .

Remarque

Dans GEORES les anisotropies géométriques sont modélisées par des ellipsoïdes d'anisotropie : la portée selon une direction correspond au rayon de l'ellipsoïde dans cette direction. Un des axes de l'ellipsoïde est obligatoirement l'axe vertical des coordonnées. Les deux autres axes X' et Y' se situent dans le plan horizontal et peuvent être tournés d'un angle Φ par rapport aux axes X, Y des coordonnées (cf. figure 1).

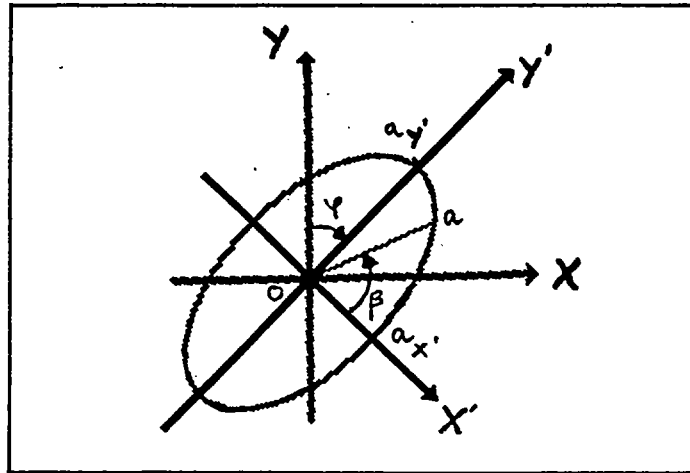


Fig. 1 - Ellipse d'anisotropie dans le plan XY
 $a_{X'}$ et $a_{Y'}$ sont les portées selon les directions principales

4.3. CALCUL DU VARIOGRAMME EN 3D

Dans le cas de jeux de données en 3 dimensions, la verticale joue en général un rôle particulier. Aussi K3GAM propose-t-il de calculer le variogramme vertical ou le variogramme horizontal selon différentes directions. L'horizontale et la verticale dont il s'agit ici sont relatives au repère de travail.

Remarque

Si l'on soupçonne que l'axe "vertical" de l'ellipsoïde d'anisotropie ne coïncide pas avec l'axe vertical Z du repère de travail, il faut au préalable effectuer une rotation du repère (programme MPROT) de façon à ce que le nouvel axe Z du repère de travail coïncide avec une des 3 directions de l'ellipsoïde. En effet, les programmes d'ajustement de variogrammes supposent que s'il y a anisotropie, une des directions principales d'anisotropie est obligatoirement l'axe vertical du repère de travail.

D'autre part, la tolérance angulaire dans la direction de calcul d'un variogramme horizontal ne suffit pas car les points ne sont pas tous à la même cote Z . Il faut également définir une tolérance dans le plan perpendiculaire à la direction de calcul.

Dans K3GAM, le calcul du variogramme nécessite de définir cette tolérance.

Pour le calcul du variogramme dans un plan horizontal, il faut ainsi définir :

- les directions de calcul (dans le plan horizontal),
- la tolérance sur la direction de calcul $d\theta$,

- la tolérance verticale DZ (dans le plan vertical),
- la tolérance tangentielle DT .

Ainsi dans le cas du schéma figure 2, deux points $A(x_a, y_a, z_a)$ et $B(x_b, y_b, z_b)$ ne seront pris en compte pour la direction θ que si :

1. $H_z = z_b - z_a < DZ$.
- 2 - la direction du vecteur AB' est comprise entre $\theta - d\theta$ et $\theta + d\theta$, avec $B'(x_b, y_b, z_a)$ = projection verticale de B dans le plan horizontal contenant A .
- 3 - le rapport H_z / H_{xy} est inférieur à DT .

Pour le calcul du variogramme vertical, les conditions suivantes doivent être respectées (cf. fig. 3) :

1. $H_{xy} < DH$ (tolérance horizontale),
2. $H_{xy} / H_z < DT$ (tolérance tangentielle).

D'autre part le choix est laissé entre la distance réelle (AB) ou la distance projetée (AB'). La distance projetée peut être utile pour étudier une anisotropie zonale.

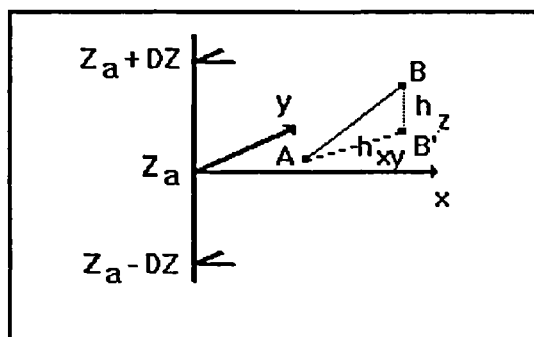


Fig. 2 - Définition des tolérances pour le calcul du variogramme horizontal en 3 dimensions (K3GAM)

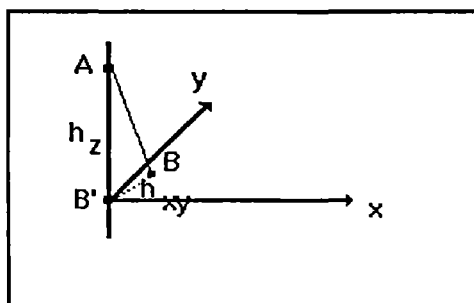


Fig. 3 - Définition des tolérances pour le calcul du variogramme vertical en 3 dimensions (K3GAM)

4.3.1. Quand calculer un variogramme horizontal ou vertical ?

Les variogrammes selon la verticale et l'horizontale sont ceux utilisés pour l'ajustement et le krigeage. Il faut donc les calculer si on mène l'étude en 3 dimensions et que l'objectif est l'estimation d'une grille ou d'un réseau de blocs 3D.

Pour calculer un variogramme vertical ou horizontal, il faut obligatoirement disposer d'un index sur les blocs, car la classification des données dans un réseau de blocs permet un calcul plus rapide.

Cet index sur les blocs est généré automatiquement par le programme K3CLAS, qui doit donc être exécuté avant K3GAM.

Si les données ne sont pas mesurées sur des supports (carottes) de longueur constante, ou si les sondages ne sont pas tous verticaux, il est également recommandé d'exécuter le programme MSLICE avant K3CLAS, en choisissant l'option "régularisation par bancs de hauteur constante".

4.3.2. Quand calculer un variogramme le long des sondages ?

Le calcul du variogramme le long des sondages est plus rapide car il ne met en jeu que des couples de points appartenant à un même sondage. Il ne nécessite pas de pilotage bloc. Si les supports sur lesquels sont mesurées les variables ne sont pas de longueur constante, il faut au préalable régulariser les sondages avec l'option "découpage le long des sondages" (programme MSLICE).

Le calcul du variogramme le long des sondages, dans le cas où les sondages sont de directions différentes, n'a de sens que dans le cas où il y a isotropie. S'il n'y a pas isotropie, il faut effectuer un calcul par classes de directions en sélectionnant les sondages à prendre en compte (ce qui suppose que les deux variables direction existent dans le fichier tête).

4.4. VARIOGRAMMES ROBUSTES

La présence de données fortement erronées ou anormales se traduit sur un histogramme par une queue de distribution qui n'a pas de sens réel dans le cas où elle provient de données erronées, et qui est difficile à modéliser s'il s'agit d'anomalies. Ces données rendent délicate l'estimation d'une valeur moyenne. Un variogramme, qui est une moyenne de carrés d'écarts, est encore plus sensible à de telles données. Il existe des variantes du variogramme qui sont moins sensibles aux perturbations apportées par des données extrêmes. Ces variogrammes sont appelés des variogrammes robustes, par référence aux estimateurs robustes définis en statistique. Mentionnons ceux qui figurent dans GDM :

- Variogramme relatif : Au lieu de diviser la somme des carrés des écarts par deux fois le nombre de couples, on la divise par la somme des carrés des valeurs utilisées:

$$\gamma(h) = \frac{\sum_i [Z(x_i + h) - Z(x_i)]^2}{\sum_i [Z(x_i + h)^2 + Z(x_i)^2]}$$

les sommes étant étendues aux couples (x_i, x_i+h) pouvant être formés pour la distance h . Ce type de variogramme ne doit bien évidemment être utilisé que pour une variable positive. Il permet de limiter les variations parasites dues à un effet proportionnel (faible variabilité dans les zones à faible teneur et forte variabilité dans les zones à forte teneur).

- Variogramme d'ordre 1 : Au lieu de prendre la moyenne des demi carrés des écarts, on prend la moyenne des demi valeurs absolues des écarts. L'effet des fortes valeurs est alors atténué.
- D'autres façons de rendre le calcul du variogramme plus robuste consistent à travailler non pas sur les données de base mais sur des transformées :
 - données régularisées (teneurs moyennes sur des passes de 10 m plutôt que sur des passes métriques par exemple) ;
 - variable écrêtée (les valeurs dépassant un seuil donné sont limitées à la valeur du seuil) ;
 - variable indicatrice (transformation en 0 ou 1) ;
 - passage au logarithme (pour des variables positives à distribution de type lognormal) ;
 - transformation par anamorphose gaussienne.

Lorsqu'on utilise des variogrammes robustes, il ne faut pas oublier que si au niveau du krigeage on utilise les données de départ, c'est le variogramme ordinaire qui doit être utilisé. Il faut donc établir un lien entre le variogramme robuste calculé et le modèle de variogramme ordinaire. En pratique on peut mettre à profit les remarques qui suivent.

- Le variogramme de données régularisées se déduit de celui des données de base par une régularisation.
- Si les valeurs écrêtées sont réparties "au hasard", on peut admettre que le variogramme théorique peut être pris comme un variogramme ajusté sur le variogramme expérimental, augmenté d'un effet de pépite à caler à partir de la variance expérimentale des données de base.

- Le variogramme d'une indicatrice est proportionnel à celui de la variable de départ dans le cas très particulier où celle-ci est constante dans des compartiments indépendants partitionnant l'espace (modèle mosaïque) ; il en va de même pour le variogramme d'ordre 1. Sinon la relation entre les deux variogrammes n'est pas évidente, sauf par exemple dans le cas multigaussien.
- Le lien entre le variogramme d'une variable lognormale et celui de son logarithme est bien connu (dans le cas où le logarithme est une variable bigaussienne). Il en va de même dans le cas de transformations plus complexes comme les anamorphoses gaussiennes ou gamma.
- En cas d'effet proportionnel, on considère que le variogramme local a partout la même forme. Il est alors proportionnel au variogramme moyen, ainsi qu'au variogramme relatif, dont le calcul est plus stable. Les pondérateurs de krigeage étant invariants lorsqu'on change le variogramme d'un facteur multiplicatif, on peut effectuer le krigeage avec le variogramme relatif. Si on s'intéresse à l'écart-type de krigeage, il faudra néanmoins caler ce facteur multiplicatif localement (par exemple par une validation zone par zone).

Les outils disponibles dans GDM, notamment dans le logiciel KFIGAM, permettent de modéliser les variogrammes robustes expérimentaux à partir du modèle de variogramme de la variable de base en intégrant les relations entre les deux types de variogramme

4.5. AJUSTEMENT DU VARIOGRAMME

Une fois le variogramme expérimental calculé, la phase suivante de l'analyse structurale consiste à ajuster un modèle théorique sur la courbe expérimentale.

Le programme d'ajustement de GEORES est le programme KFIGAM. Il est également possible d'ajuster un variogramme dans K2REC, mais seulement pour des données en deux dimensions.

4.5.1. Types de modèles autorisés

Chaque variogramme peut être ajusté par un effet de pépité C_0 plus 1 à 4 modèles emboîtés avec ou sans anisotropie.

La plupart des variogrammes dérivent d'une covariance $C(h)$ à laquelle ils sont liés par la relation :

$$\gamma(h) = C(0) - C(h)$$

On note :

- C** le palier d'une composante (c'est à dire la valeur à l'origine $C(0)$ de la covariance correspondante)
- a** sa portée,
- r** le rapport : $r=h/a$

Les modèles de covariance isotrope possibles sont les suivants (voir aussi figures 4 et 5) :

S :	Sphérique	$C(h) = C (1 - 3/2 r + 1/2 r^3)$ $C(h) = 0$	$r \leq 1$ $r \geq 1$
E :	Exponentiel	$C(h) = C \exp(-r)$	
C :	Cubique	$C(h) = C (1 - 7 r^2 + 8,75 r^3 - 35 r^5 + 0,75 r^7)$ $C(h) = 0$	$r \leq 1$ $r \geq 1$
G :	Gaussien	$C(h) = C \exp (-r^2/2)$	
R :	Gravimétrie	$C(h) = C (1 + r^2)^{-1/2}$	
M :	Magnétisme	$C(h) = C (1+r^2)^{-3/2}$	
L :	Linéaire ou h^α	variogramme $\gamma(h) = \omega \cdot h^\alpha$ le variogramme linéaire correspond à $\alpha = 1$ ω = pente ou coefficient multiplicateur covariance généralisée $K(h) = \pm \omega \cdot h^\alpha$ où k = degré de la dérive.	$0 < \alpha < 2$ $0 < \alpha < 2k+2$

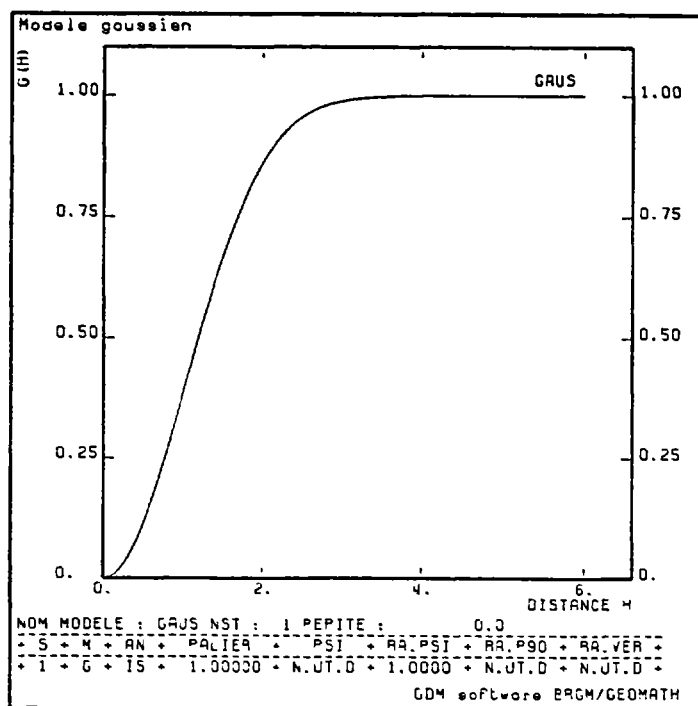
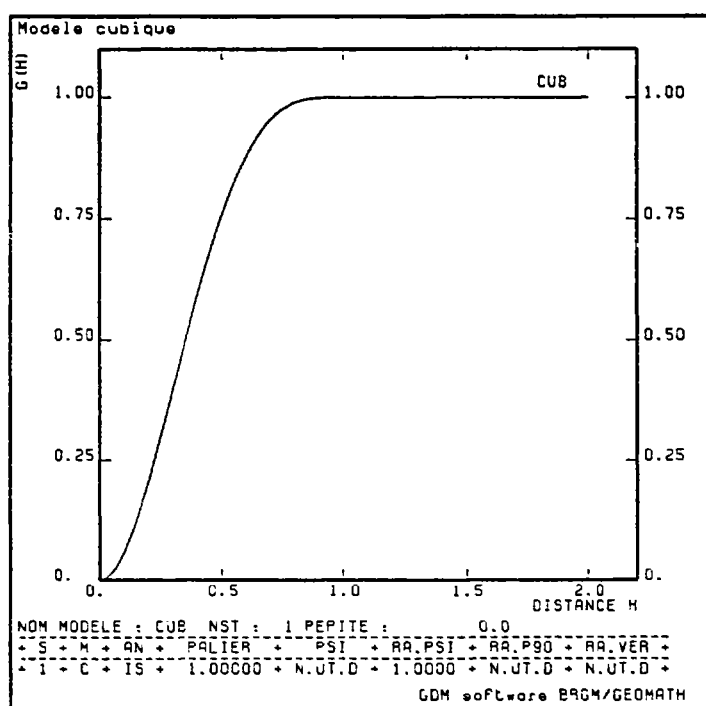
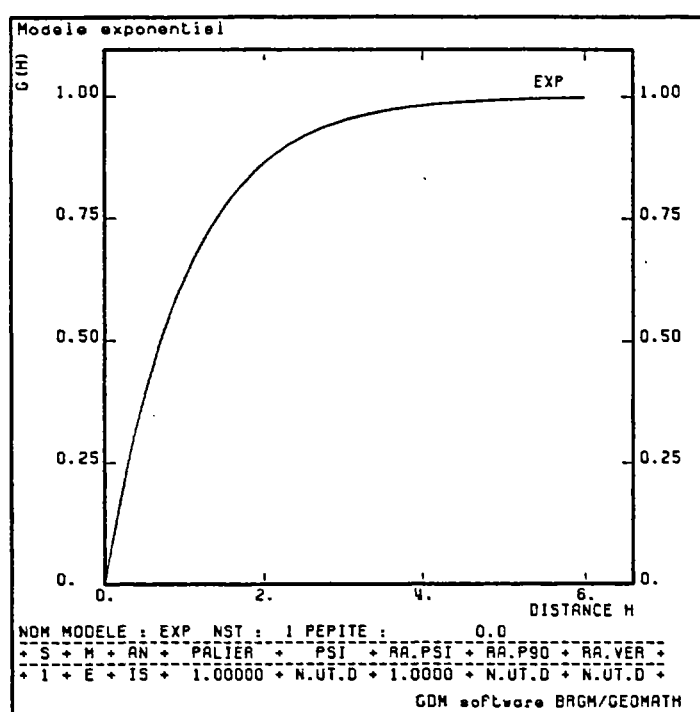
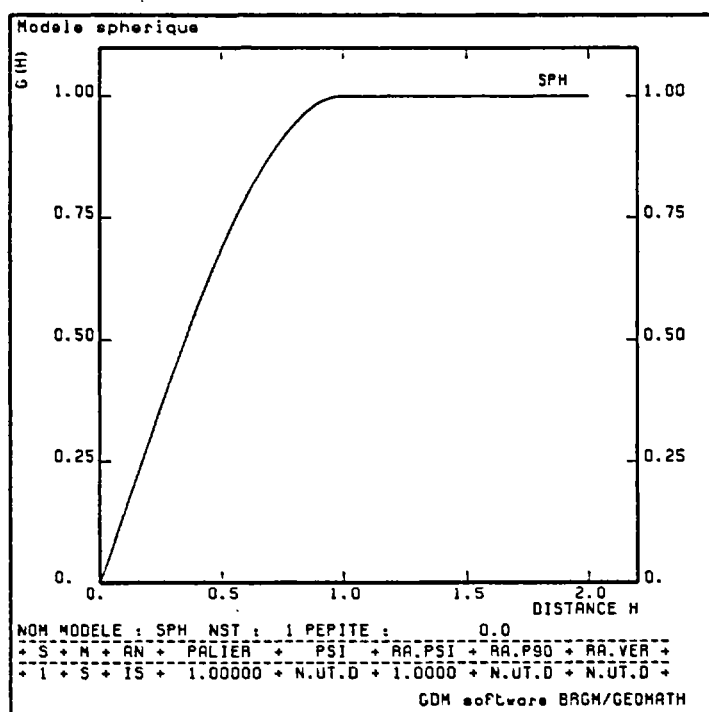


Fig. 4 - Modèles sphérique, exponentiel, cubique, gaussien

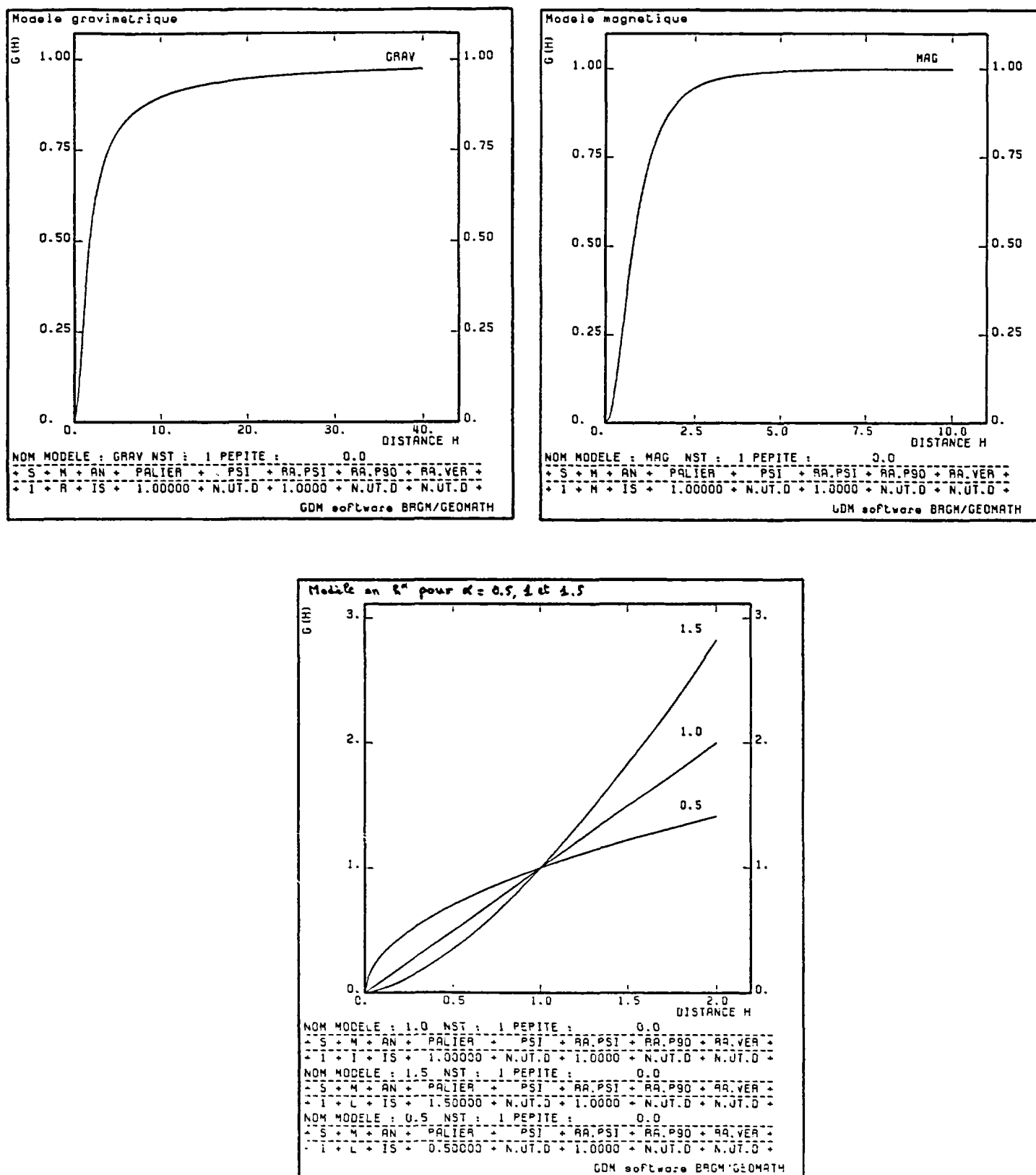


Fig. 5 - Modèles gravimétrique, magnétique, et h^α

4.5.2. Choix du palier total du variogramme

En théorie, le palier du variogramme est égal à la variance "théorique" de la variable aléatoire étudiée.

La seule variance disponible en pratique est la variance expérimentale des données, qui n'est voisine de la variance théorique que dans le cas d'une portée courte par rapport au champ. On peut alors ajuster le palier du variogramme à la variance expérimentale.

La variance expérimentale est (en théorie) inférieure à la variance théorique. Toutefois, en raison des fluctuations statistiques, la variance expérimentale peut fluctuer autour de la valeur théorique selon le nombre, la valeur et la position des données.

Par conséquent, il est tout à fait possible que la variance expérimentale se situe en-dessous ou au-dessus de la variance théorique, c'est-à-dire du palier du variogramme.

Dans le cas où il y a une bonne stationnarité (palier stable, pas de dérive, et portée petite vis-à-vis de la taille du domaine), les 2 variances expérimentales et théoriques sont voisines.

4.5.3. Jusqu'à quelle distance ajuster le variogramme ?

Il n'est pas nécessaire que l'ajustement soit parfait de l'origine aux grandes distances. En revanche il est important que l'ajustement soit fiable dans la zone "utile" du variogramme, c'est-à-dire pour toutes les distances qui seront mises en jeu dans les phases de calcul ultérieures.

Par exemple, si le but est d'effectuer une estimation par krigeage, avec un voisinage de dimension maximale 300 m, il n'est pas indispensable d'ajuster le modèle avec précision pour des distances supérieures à 300 m. Il est alors préférable d'obtenir un bon ajustement pour les distances inférieures à 300 m.

Si inversement le but est une étude de variabilité à grande échelle, il n'est pas nécessaire d'affiner l'ajustement des structures de très courte portée.

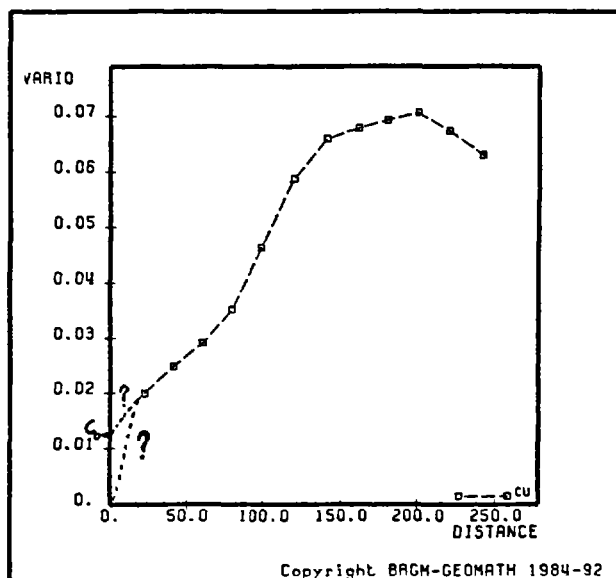
4.5.4. Choix de l'effet de pépité

En théorie on a $\gamma(0) = 0$. Lorsque ceci ne s'observe pas expérimentalement, on dit qu'on est en présence d'un effet de pépité, c'est-à-dire d'une discontinuité à l'origine.

Cette discontinuité peut avoir plusieurs causes (éventuellement combinées) :

- 1 - une erreur de mesure. Si les erreurs de mesure sont indépendantes de la teneur et de variance V_e , l'effet de pépité résultant sera d'amplitude V_e ,
- 2 - une structure de petite échelle, dont la portée est inférieure à (ou voisine de) la maille de reconnaissance : par exemple une structure de portée 50 m ne pourra être détectée par une maille de reconnaissance de 50 m (le variogramme expérimental ne pouvant être calculé pour des distances inférieures à 50 m),

- 3 - une structure de portée inférieure à la taille des échantillons (microstructure) : c'est l'effet de pépité proprement dit ; il provient d'une composante dont la portée est égale à la dimension des pépites, et qui macroscopiquement, pour des teneurs mesurées sur un support de volume v , se traduit sur le variogramme par une discontinuité de la forme $C_0=A/v$, donc inversement proportionnelle au volume v ,
- 4 - une incertitude de localisation : le cas typique concerne des données recueillies le long de profils effectués en mer; au croisement de deux profils, on est censé se trouver au même point ; en fait, vu les erreurs de positionnement il n'en est rien et les deux mesures de la variable étudiée donnent deux valeurs différentes.



Fi. 6 - Variogramme expérimental. Incertitude sur l'effet de pépité

Dans le schéma figure 6, rien ne permet a priori de savoir quelle est la bonne façon de prolonger le variogramme à l'origine. De plus, si on opte pour la présence d'un effet de pépité, il faut choisir son amplitude.

Un moyen d'évaluer l'amplitude de l'effet de pépité ainsi que le comportement du variogramme au voisinage de l'origine est de réaliser une "croix de sondages", c'est-à-dire d'implanter des sondages à des distances inférieures à la maille de sondage: par exemple des sondages espacés de 10 m pour une maille de 50 m (fig. 7).

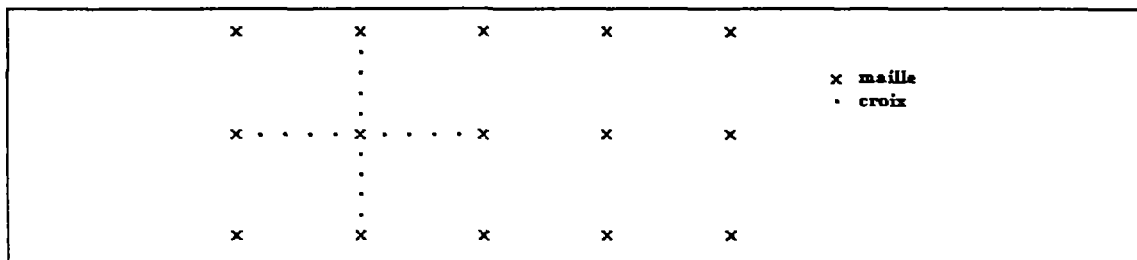


Fig. 7 - Croix de sondages

Cette croix permet le calcul du variogramme aux petites distances (avec cependant peu de couples). Ceci peut aider à lever l'indétermination sur la présence d'un effet de pépité ou d'une structure de courte portée.

En l'absence de points expérimentaux, seule une analogie avec d'autres variogrammes mieux connus sur des données du même type, ou des hypothèses sur le comportement du phénomène aux courtes distances, permettent d'estimer le comportement du variogramme au voisinage de l'origine. De telles hypothèses ne peuvent cependant être validées en l'absence de données expérimentales.

La connaissance de l'effet de pépité est importante lorsqu'on cherche à effectuer des estimations à de courtes distances, et que l'on s'intéresse à la précision de l'estimation : celle-ci sera d'autant meilleure que l'effet de pépité est faible.

Dans le cas où l'on dispose de données régularisées le long de sondages, on peut cependant effectuer certains contrôles pour préciser l'effet de pépité (cf. plus loin).

4.5.5. Ajustement d'une anisotropie géométrique

Il y a anisotropie géométrique lorsque le palier reste constant alors que la portée a varie selon la direction de calcul, selon l'équation d'une ellipse (en 2 dimensions) ou d'un ellipsoïde (en 3 dimensions).

Par exemple, en 2 dimensions, pour une anisotropie selon les directions X et Y :

$$\frac{\cos^2(\theta)}{a_y^2} + \frac{\sin^2(\theta)}{a_x^2} = \frac{1}{a^2(\theta)}$$

avec :

θ	direction de calcul par rapport à l'axe des y (Nord)
a_y	portée selon la direction y
a_x	portée selon la direction x
$a(\theta)$	portée dans la direction θ

Dans le cas où le variogramme est de type $\omega \cdot h^\alpha$, la portée a doit être remplacée par $\omega^{-\alpha}$.

La figure 8 donne un exemple d'anisotropie géométrique en 2 dimensions, avec un modèle sphérique de portée 1 dans une des 2 directions principales, et de portée 2 dans l'autre direction principale. Le palier est constant et égal à 1.

Pour ajuster un modèle d'anisotropie géométrique, il suffit de connaître le variogramme expérimental selon les deux directions principales d'anisotropie. Mais ces directions ne sont pas toujours connues a priori. Il faut alors calculer le variogramme selon au moins quatre directions (deux directions orthogonales et les deux directions diagonales) pour caractériser l'anisotropie.

Le principe de l'ajustement d'une anisotropie géométrique est alors le suivant :

- visualiser sur un même graphique les variogrammes selon les différentes directions (au moins 4) et déterminer la direction correspondant à la portée maximale. La direction de portée minimale doit lui être orthogonale. Déterminer les 2 portées correspondantes. Vérifier que le palier est sensiblement constant dans toutes les directions,
- ajuster le modèle de variogramme sur les 2 directions principales,
- visualiser les variogrammes expérimentaux des différentes directions et le modèle associé dans chacune de ces directions. Si la concordance est bonne, le modèle peut être accepté, sinon procéder par ajustements successifs en faisant varier direction d'anisotropie et portées.

Un exemple d'ajustement d'anisotropie géométrique est fourni dans le manuel "Logiciels géostatistiques de GDM : Cas d'étude".

4.5.6. Ajustement d'une anisotropie zonale

Lorsque le palier n'est pas le même dans toutes les directions, on se trouve le plus souvent dans le cas d'une anisotropie zonale.

Le principe d'ajustement consiste à ajouter une composante au modèle de variogramme, composante qui intervient dans la direction perpendiculaire à la direction de palier minimum.

Soit par exemple un phénomène dans l'espace (X,Y,Z) présentant une anisotropie zonale selon la direction verticale : le palier est plus important selon la verticale, et est minimum dans le plan (X,Y). Un tel variogramme est ajusté par : $\gamma(h) = \gamma_1(h) + \gamma_2(h_z)$

La deuxième composante $\gamma_2(h_z)$ ne dépend que de la composante verticale h_z du vecteur h et ne joue donc pas dans l'horizontale. Le premier terme $\gamma_1(h)$ dépend des 3 composantes (h_x , h_y , h_z) de h , et peut être un modèle isotrope ou présentant une anisotropie géométrique. $\gamma_1(h)$ peut très bien comporter si nécessaire plusieurs structures et un effet de pépité.

Dans le plan (X,Y), la composante h_z est nulle, et le variogramme est égal à $\gamma(h) = \gamma_1(h)$. Son palier est celui de γ_1 , soit C_1 .

Dans toute autre direction, la composante h_z est non nulle et le variogramme vaut :

$$\gamma(h) = \gamma_1(h) + \gamma_2(h_z).$$

Le palier à l'infini est donc égal à $C_1 + C_2$ (C_2 étant le palier de γ_2).

On peut aussi avoir une anisotropie zonale du type $\gamma(h) = \gamma_1(h) + \gamma_2(h_{xy})$ où verticalement seule la composante $\gamma_1(h)$ intervient, alors que dans les autres directions s'ajoute une composante $\gamma_2(h_{xy})$ qui dépend seulement de la composante horizontale de h . En ce cas le palier est minimum dans la direction verticale. On peut bien sûr avoir des modèles plus complexes en superposant plusieurs composantes.

La figure 9 donne un exemple de modèle d'anisotropie zonale : selon l'horizontale le modèle MD1 est un modèle sphérique de palier 1 et de portée 1 ; selon la verticale, le modèle MD3 est égal au modèle MD1, auquel s'ajoute un modèle sphérique de palier 0.5 et de portée 2.

Le principe d'ajustement est le même que pour l'anisotropie géométrique. Il faut ainsi :

- calculer le variogramme selon plusieurs directions,
- visualiser les résultats sur un même graphique, afin de :
 - déterminer la direction D1 (ou le plan P1) pour laquelle le palier est minimum. Ce palier correspond à la première composante γ_1 du variogramme, qui peut être éventuellement composée de plusieurs structures. Ces structures peuvent le cas échéant présenter une anisotropie géométrique,
 - ajuster le modèle γ_1 selon cette direction (ou ce plan),

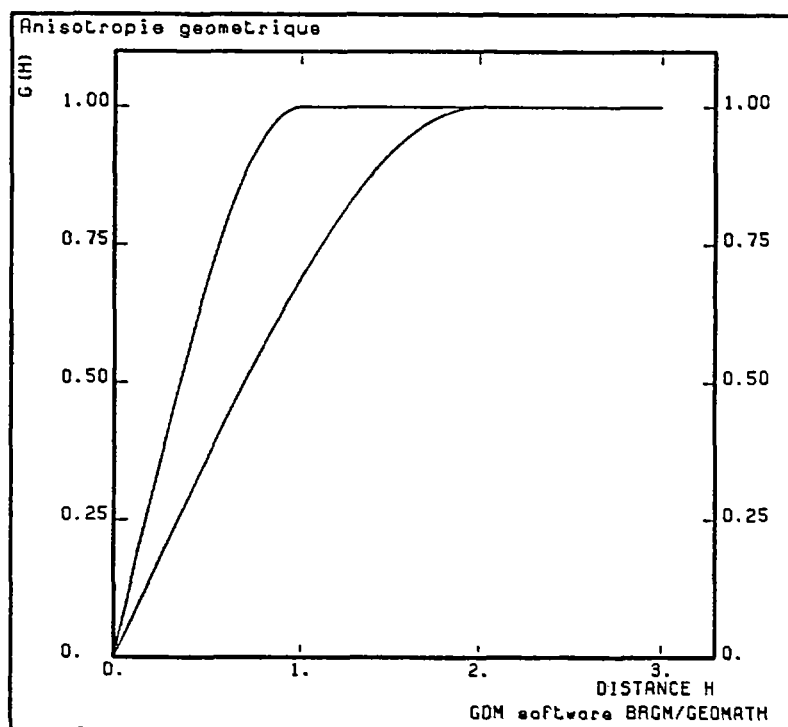


Fig. 8 - Variogramme 2D présentant une anisotropie géométrique

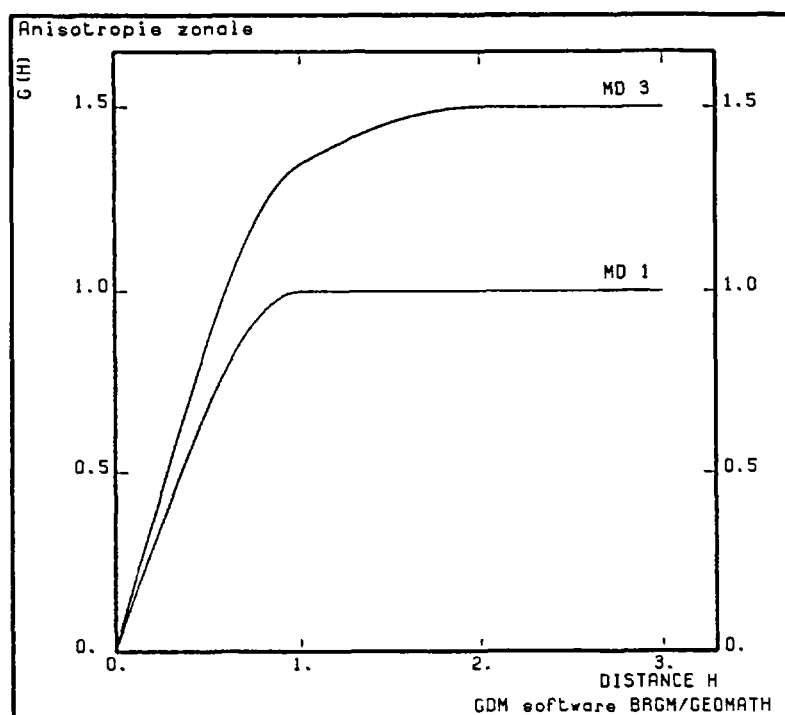


Fig. 9 - Variogramme 3D présentant une anisotropie zonale selon la verticale

- définir une deuxième composante γ_2 présentant une anisotropie zonale, de palier C2 = {palier des autres directions - C1} et s'appliquant dans le plan (respectivement la direction) orthogonal à la direction (respectivement le plan) de palier minimum,
- tester l'ajustement dans des directions intermédiaires.

Un exemple d'ajustement d'anisotropie zonale est donné dans le manuel "Logiciels géostatistiques de GDM : Cas d'étude".

4.5.7. Cas d'un variogramme sur des données régularisées

Lorsqu'on dispose de données régularisées le long de sondages ou sur des bancs de hauteur constante, il est possible de déterminer avec plus de précision l'effet de pépité et les portées des structures à petite échelle, en comparant les variogrammes calculés par régularisation sous différentes longueurs.

Il existe en effet une relation théorique entre certains des paramètres de ces variogrammes.

Par exemple, si C_{ov} est l'effet de pépité observé sur un support v (par exemple des carottes de longueur L) et C_{oV} l'effet de pépité observé sur un support V (par exemple carottes de longueur $2L$), la relation suivante peut être démontrée :

$$C_{oV} = C_{ov} * v / V$$

(dans notre exemple $C_{o2L} = 1/2 C_{oL}$)

En d'autres termes, l'effet de pépité est inversement proportionnel à la taille du support de mesure.

De la même manière, la portée du variogramme le long des sondages augmente d'une valeur égale à la différence des longueurs de régularisation. Par exemple : $a_{3L} = a_L + 2L$.

Par contre, la portée du variogramme ne change pas dans la direction perpendiculaire.

Enfin, dans tous les cas, le palier du variogramme (effet de pépité exclu) diminue d'une valeur qui peut être calculée en fonction de la taille des volumes.

Ces relations peuvent être utilisées pour vérifier manuellement ou automatiquement l'ajustement du variogramme aux petites distances.

Pour illustrer ceci, considérons le cas où on disposerait de sondages subverticaux, avec des teneurs mesurées par passes de 2,5 m en moyenne.

A l'aide du programme MSLICE, il est possible de découper les sondages en passes régulières ou en bancs de 2,5 m, ou 5 m, ou 10 m, etc.

Il est ensuite possible de calculer, pour chacun de ces cas de figure, le variogramme expérimental le long des sondages, puis d'ajuster le variogramme obtenu :

- 1- si l'on observe un effet de pépité de 20 sur le variogramme des passes de 2,5 m, on doit obligatoirement constater un effet de pépité de 10 sur les passes de 5 m, et de 5 sur les passes de 10 m (si l'effet de pépité observé sur les variogrammes des passes de 5 et 10 m est supérieur

à la valeur prévue par la théorie, c'est qu'une partie de l'effet de pépité imputé au variogramme des passes de 2,5 m est en réalité une micro-structure de portée inférieure à 2,5 m),

- 2 - de la même manière, si la portée est de 8 m pour les variogrammes sur des passes de 2,5 m, elle sera de 10,5 m pour des passes de 5 m et de 15,5 m pour des passes de 10 m. Cette portée de 15,5 m n'aurait pas pu être mise en évidence en travaillant directement sur les données régularisées sur 10 m, puisque le premier point disponible du variogramme aurait été le point pour la distance $h = 10$.

Ces relations peuvent être vérifiées graphiquement assez rapidement, en reportant sur un même tracé les variogrammes des données régularisées sous différentes hauteurs.

Il est également possible de contrôler les paramètres d'ajustement en utilisant la clique **SPECIAL** du programme KFIGAM.

Pour illustrer ceci en reprenant l'exemple précédent avec seulement les variogrammes des passes de 2.5 et 5m pour simplifier, on procéderait de la sorte :

- 1 - lecture des variogrammes régularisés le long des sondages, pour $L = 2,5$ et 5 m,
- 2 - ajustement du variogramme des passes de 2,5 m et tracé simultané du variogramme et du modèle. Soit MD1 le modèle retenu,
- 3 - utilisation de la clique **SPECIAL** pour définir le variogramme expérimental pour $L = 5$ m comme étant une régularisation du variogramme pour $l = 2,5$ m. Il faut pour cela entrer les paramètres suivants : direction de régularisation : 0 0 1 (si les sondages sont verticaux) ; longueur de régularisation : 5 ; nombre de points de discrétisation : 2 (il y a 2 passes de 2,5 m dans une de 5 m),
- 4 - association du variogramme expérimental $L = 5$ m avec le modèle MD1 et demande de tracé de la courbe expérimentale et du modèle associé (clique **LKGM**),
- 5 - tracé des variogrammes. Le programme dessinera alors :
 - le variogramme expérimental $L = 2,5$ m,
 - le modèle associé MD1,
 - le variogramme expérimental $L = 5$ m,
 - le modèle MD1 régularisé sous $L = 5$ m.

Si l'ajustement est valide, le modèle régularisé doit bien ajuster le variogramme pour $L = 5$ m.

Un exemple est donné dans le manuel "Logiciels géostatistiques de GDM : Cas d'étude".

4.6. VARIOGRAPHIE EN PRÉSENCE D'UNE DÉRIVE

En présence d'une dérive, le phénomène ne peut être modélisé à l'aide des modèles stationnaires ou intrinsèques. Il faut alors faire appel soit au modèle de krigeage universel, ce qui nécessite la calcul du variogramme des résidus, soit au modèle des FAI-K, ce qui suppose le calcul du variogramme généralisé.

4.6.1. Variogramme des résidus

Dans ce cas le principe consiste à décomposer la variable en deux (Cf. paragraphe sur les modèles géostatistiques) :

$$Z(x) = Y(x) + m(x)$$

$Y(x)$ est une variable stationnaire appelée résidu

$m(x)$ est la dérive pouvant être exprimée sous la forme d'un polynôme de degré n , dépendant des coordonnées (u,v,w) des points. Avec deux coordonnées (u,v) et un degré 2, $m(x)$ prend la forme :

$$m(x) = a + bu + cv + duv + eu^2 + fv^2$$

Le variogramme des résidus est le variogramme du résidu $Y(x)$ obtenu à partir de $Z(x)$ en retranchant la dérive $m(x)$.

Son calcul suppose de connaître exactement $m(x)$, c'est-à-dire le degré de la dérive et les coefficients a, b, \dots du polynôme.

En pratique le degré de la dérive est choisi par l'utilisateur et les coefficients du polynôme $m(x)$ sont estimés par moindres carrés à partir de l'ensemble des $Z(x_i)$ disponibles. On obtient ainsi l'estimation $m^*(x) = a^* + b^*u + c^*v \dots$

On construit alors $Y^*(x_i) = Z(x_i) - m^*(x_i)$ qui représente une estimation du résidu au point x_i , une fois retranchée de la valeur vraie $Z(x_i)$ l'estimation de la dérive $m^*(x_i)$. Le variogramme de Y est approché par le variogramme de Y^* . Il y a donc un biais, qui est toutefois négligeable aux petites distances.

4.6.2. Variogramme généralisé

Au lieu d'essayer d'estimer la dérive $m(x)$, on peut travailler sur des accroissements généralisés qui permettent de filtrer la dérive.

Ainsi, comme cela a été exposé au chapitre 3.3, la combinaison linéaire:

$$A_2(x) = Z(x) - 2Z(x+h) + Z(x+2h)$$

est un "accroissement généralisé" d'ordre 2 : elle filtre les dérives d'ordre inférieur ou égal à 1. $A_2(x)$ est donc stationnaire si le degré de dérive de $Z(x)$ est inférieur ou égal à 1. On définit le variogramme généralisé d'ordre 1 de $Z(x)$ par : $\Gamma(h) = 1/6 \text{Var}(A_2(x))$. D'autres expressions similaires existent pour des dérives d'ordre supérieur à 1. Le variogramme généralisé n'est toutefois accessible expérimentalement que pour des données à maille régulière (grille, profils) puisqu'il fait intervenir des distances $h, 2h$, etc.

Lorsque les données ne sont pas à maille régulière, on a recours à des méthodes indirectes utilisées dans des algorithmes dits de "reconnaissance automatique".

4.6.3. Variogramme avec dérive externe

En cas de dérive externe la décomposition $Z(x) = Y(x) + m(x)$ est remplacée par une décomposition du type :

$$Z(x) = a + b \cdot E(x) + Y(x)$$

où la dérive externe $E(x)$ est une autre variable et non un monôme. Si $E(x)$ est connue aux points expérimentaux où $Z(x)$ l'est, on peut déterminer une estimation a^* , b^* des coefficients a, b et calculer le variogramme des résidus comme ci-dessus.

4.7. RECONNAISSANCE AUTOMATIQUE

Dans le cas de phénomènes avec dérive (modèle intrinsèque d'ordre supérieur ou égal à 1), le variogramme est remplacé par un variogramme (ou une covariance) généralisé(e). Ce variogramme n'est pas calculable expérimentalement, sauf dans le cas de données sur une grille. Pour traiter de tels phénomènes, il faut utiliser des logiciels dits de "reconnaissance automatique de structure" qui identifient le degré de la dérive ainsi que la covariance généralisée associée. Le programme de GEORES qui réalise ces opérations est le programme K2REC, utilisable pour des données en 2 dimensions.

La reconnaissance automatique se fait en deux temps : choix du degré de dérive et choix de la covariance généralisée.

Choix du degré k

K2REC met en concurrence les valeurs $k=0$ (pas de dérive), $k=1$ (dérive localement linéaire) et $k=2$ (dérive localement quadratique). Il construit des voisinages, qui sont subdivisés en une auréole intérieure et une auréole extérieure. Chaque point de l'auréole extérieure est estimé à partir des points de l'auréole intérieure par moindres carrés, pour chacune des valeurs possibles du degré k . La vraie valeur étant connue, on obtient ainsi trois erreurs d'estimation. Le degré k conduisant aux erreurs les plus faibles est retenu. Le critère utilisé n'est pas l'erreur quadratique moyenne, car celle-ci serait très sensible à une donnée anormale, mais le rang moyen : pour chaque point estimé, les trois solutions sont classées 1, 2, 3 de l'erreur la plus faible (en valeur absolue) à l'erreur la plus forte ; on retient le degré k qui conduit au meilleur classement moyen.

Choix de la covariance généralisée

Le choix est fait parmi les covariances généralisées avec effet de pépite. Autrement dit, le modèle, en fonction du degré k considéré, est de la forme suivante :

- si $k=0$: $K(h) = C_0 - b_0 \cdot h$
- si $k=1$: $K(h) = C_0 - b_0 \cdot h + b_1 \cdot h^3$
- si $k=2$: $K(h) = C_0 - b_0 \cdot h + b_1 \cdot h^3 - b_2 \cdot h^5$

avec des coefficients C_0, b_0, b_1, b_2 positifs ou nuls. La variance de l'erreur d'estimation d'un point à partir des données voisines s'exprime linéairement en fonction de ces paramètres. Pour les déterminer, on réestime des points expérimentaux à partir de leurs voisins (on procède par krigeage, avec un variogramme linéaire), et on calcule par régression linéaire les coefficients qui fournissent la meilleure reconstitution des carrés des erreurs d'estimation par leur variance

théorique. La régression n'est pas faite sous contrainte que les coefficients soient positifs, et peut donc aboutir à une solution qui n'est pas admissible. Aussi calcule-t-on de plus les quatre solutions à trois coefficients, les six solutions à deux coefficients, et les quatre solutions à un seul coefficient. Ces dernières donnent par construction des solutions admissibles. Et on retient, parmi toutes les solutions admissibles, celle qui fournit les meilleurs résultats.

Pour affiner le choix de la solution, on retient en fait les trois meilleures solutions, et on compare leurs performances dans une phase de validation croisée similaire à celle qui est décrite ci-après, mais simplifiée (on ne réestime que 50 à 100 données). On retient celle qui conduit aux meilleurs résultats.

Ces modèles identifiés automatiquement doivent ensuite de préférence être validés (cf. chapitre suivant).

Il est possible d'effectuer la reconnaissance automatique de phénomènes stationnaires, mais cela présente un intérêt moindre.

4.8. VALIDATION DU VARIOGRAMME

Une fois le variogramme calculé puis ajusté, il est recommandé d'effectuer un test de "validation croisée" afin de vérifier si l'ajustement est correct.

La validation est surtout utile dans le cas des phénomènes non stationnaires avec dérive, pour lesquels le calcul et l'ajustement manuel du variogramme n'est pas possible. Pour ces phénomènes, le programme K2REC permet (en deux dimensions) la détermination automatique d'un modèle structural qui n'est pas un variogramme, mais une "covariance généralisée". Ce modèle doit ensuite être validé avant d'être utilisé.

Le principe de la validation croisée est expliqué ci-dessous :

Soit x_i un point de donnée. Le test de validation croisée consiste à :

- faire "comme si" la teneur $Z(x_i)$ au point x_i était inconnue,
- calculer par krigeage une estimation $Z^*(x_i)$ de cette teneur en utilisant les informations voisines et le variogramme des $Z(x_i)$,
- calculer expérimentalement l'erreur d'estimation : $e(x_i) = Z^*(x_i) - Z(x_i)$,
- calculer la variance de krigeage correspondante : $S_k^2(x_i)$,
- répéter cette opération sur un ensemble de points.

Il est alors possible de calculer expérimentalement :

- la moyenne des erreurs : $E[e(x_i)]$,
- la variance des erreurs : $\text{Var}[e(x_i)]$.

Par définition du krigeage, la moyenne théorique des erreurs d'estimation est nulle (le krigeage est un estimateur non biaisé) et la variance des erreurs est égale à la variance de krigeage S_k^2 moyenne.

La loi des erreurs se rapprochant généralement d'une loi gaussienne, on doit avoir environ 95 % des erreurs dans la fourchette $0 \pm 2 \cdot S_k$.

Pour s'affranchir de la valeur de S_k dans l'examen des résultats, il est commode de travailler sur l'erreur réduite :

$$e_r(x_i) = \frac{e(x_i)}{S_k(x_i)} = \frac{Z^*(x_i) - Z(x_i)}{S_k(x_i)}$$

Dans ces conditions, on doit avoir :

$$\begin{aligned} E[e_r] &= 0 \\ \text{Var}[e_r] &= 1 \end{aligned}$$

Pour une distribution gaussienne des erreurs, 95 % des erreurs réduites doivent être comprises entre -2 et +2, et 99 % entre -2.5 et +2.5. Les erreurs réduites supérieures à 2.5 (en valeur absolue), surtout si elles sont supérieures à cette limite, proviennent de données anormales, soit erronées, soit correctes mais hétérogènes par rapport à leur environnement.

Le test de validation permet donc de détecter les points mal estimés et de rejeter ou confirmer l'ajustement retenu selon le nombre de ces points.

D'autres résultats sont également possibles :

- une carte des erreurs d'estimation : cette carte permet de vérifier si les points les moins bien estimés sont situés dans des zones géographiques préférentielles ou non, et d'éventuellement modifier l'ajustement du variogramme dans ces zones ;
- des nuages de corrélation.

Le nuage de corrélation (Z^*, Z) doit être aussi proche que possible de la bissectrice.

La corrélation entre valeur estimée (Z^*) et erreur d'estimation ($Z^* - Z$) doit être la plus faible possible.

Dans GEORES, la validation d'un modèle peut se faire à l'aide du programme K2REC pour des données en 2 dimensions, et du programme K3VAL pour des données en 3 dimensions. K2REC permet l'évaluation des performances d'un modèle, mais aussi la comparaison de plusieurs modèles. K3VAL ne permet de tester qu'un modèle à la fois.

Pour comparer plusieurs modèles, on compte le nombre de fois où un modèle a été "meilleur" qu'un autre en comparant les valeurs absolues des erreurs ($Z^* - Z$) en chaque point réestimé. Le modèle qui obtient le meilleur score est retenu. Il est possible également de comparer les variances d'estimation expérimentales obtenues avec les différents modèles et de retenir celui qui donne la variance la plus faible.

Pour atténuer l'influence de quelques éventuelles très fortes erreurs, les résultats peuvent également être fournis en terme "d'estimation robuste". Dans ce cas, on recherche les points pour lesquels, avec tous les modèles testés, on obtient $|Z^* - Z|/S_k > 2.5$ en valeur absolue, et on remplace pour ces points la valeur Z^* par la valeur Z^{**} solution de $|Z^{**} - Z|/S_k = 2.5$. Ceci permet de comparer les modèles sans trop donner de poids aux données anormales.

Si les résultats de la validation sont jugés satisfaisants, le modèle peut être utilisé ensuite pour une estimation à l'aide de K2KRI.

5. KRIGEAGE

Le krigeage est utilisé pour tout problème d'estimation ou d'interpolation de variables stationnaires ou non (teneurs, profondeur ou épaisseur d'un horizon géologique, temps sismique, température, etc.)

Les différents types de krigeage disponibles dans GEORES sont les suivants :

- krigeage "simple" d'une variable d'espérance connue (et en général nulle) : K2KRI ;
- krigeage "ordinaire" sous hypothèse stationnaire ou intrinsèque : K2KRI ou K3KRI ;
- krigeage universel (en présence d'une dérive polynômiale) : K2KRI ;
- krigeage dans le cadre des FAI-k (fonctions aléatoires intrinsèques d'ordre k) : K2KRI ;
- krigeage avec dérive externe : K2KRI.

Tous ces types de krigeage peuvent être employés pour estimer des données ponctuelles (points isolés ou grille) ou des valeurs moyennes sur des blocs parallélépipédiques (en 3 dimensions) ou rectangulaires (en 2 dimensions).

En définissant des variables auxiliaires, on peut également effectuer un krigeage des indicatrices ou un krigeage lognormal.

5.1. KRIGEAGE SIMPLE

Le krigeage simple est utilisé dans le cas des modèles stationnaires à moyenne connue. Il consiste à estimer chaque point ou bloc à partir des informations voisines par combinaison linéaire :

$$Z_v^*(x) = m + \sum_{i=1,n} \lambda_i (Z(x_i) - m)$$

avec :

$Z_v^*(x)$	estimation de la teneur moyenne du volume v (point ou bloc) situé au point x
$Z(x_i)$	valeurs aux n points de mesure x_i
λ_i	pondérateurs affectés aux points x_i
m	moyenne globale (dans K2KRI, la moyenne m est supposée nulle. Si ce n'est pas le cas, travailler avec $Z(x)-m$).

Les λ_i sont calculés de façon à minimiser l'erreur d'estimation (plus précisément la variance de cette erreur).

La mise en oeuvre de ce type de krigeage nécessite la connaissance du variogramme qui doit être un variogramme admettant un palier (fonction aléatoire stationnaire).

Les équations du krigeage simple sont les suivantes (le système de krigeage n'est valable qu'exprimé en covariance car la somme des pondérateurs est différente de 1) :

$$\sum_{j=1,n} \lambda_j C(x_i, x_j) = C(x_i, v) \quad \forall i \text{ (n équations)}$$

avec :

n	=	nombre d'informations disponibles
i, j	=	indices des points de mesure ; i et j varient de 1 à n
x_i, x_j	=	informations disponibles pour l'estimation, mesurées aux points x_i et x_j
v	=	volume à estimer (point ou bloc)
λ_i (ou λ_j)	=	pondérateur affecté à la valeur mesurée en x_i
$C(x_i, x_j)$	=	valeur de la covariance $C(h)$ pour h =distance entre x_i et x_j
$C(x_i, v)$	=	valeur moyenne de la covariance $C(h)$ pour h =distance entre x_i et x , x étant un point décrivant le volume v

La variance de l'erreur d'estimation, appelée **variance de krigeage**, est :

$$S_k^2 = \sum_{i=1,n} \lambda_i C(x_i, v) - \bar{C}(v, v)$$

Pour une estimation d'un point x à partir de deux informations ponctuelles x_1 et x_2 , le système s'écrit :

$$\begin{aligned} \lambda_1 C(x_1, x_1) + \lambda_2 C(x_1, x_2) &= C(x_1, x) \\ \lambda_1 C(x_2, x_1) + \lambda_2 C(x_2, x_2) &= C(x_2, x) \end{aligned}$$

$C(x_1, x_1)$ est la valeur de $C(h)$ lorsque le premier point du vecteur h est positionné en x_1 et le deuxième également, c'est à dire pour une distance $h = 0$. De même $C(x_1, x_2)$ est la valeur de $C(h)$ pour h = distance entre x_1 et x_2 .

5.2. KRIGEAGE ORDINAIRE

Le krigeage ordinaire est utilisé dans le cas de modèles stationnaires à moyenne inconnue ou dans le cas de modèles intrinsèques d'ordre 0.

Il consiste à estimer chaque point ou bloc à partir des informations voisines par combinaison linéaire :

$$Z^*(x) = \sum_{i=1,n} \lambda_i Z(x_i)$$

La moyenne n'étant pas connue, le non biais ne peut être assuré qu'à la condition complémentaire que :

$$\sum_{i=1,n} \lambda_i = 1$$

La mise en oeuvre de ce type de krigeage nécessite la connaissance du variogramme qui doit être :

- soit un variogramme admettant un palier (fonction aléatoire stationnaire),
- soit un variogramme non borné, mais pouvant être ajusté par un modèle en h^α , avec $\alpha < 2$ (fonction aléatoire intrinsèque d'ordre 0)

La plupart des problèmes miniers peuvent être traités avec ce type de krigeage.

Les équations du krigeage ordinaire sont les suivantes :

$$\sum_{j=1,n} \lambda_j \gamma(x_i, x_j) + \mu = \gamma(x_i, v) \quad \forall i \text{ (n équations)}$$

$$\sum_{i=1,n} \lambda_i = 1$$

avec :

n	=	nombre d'informations disponibles
i, j	=	indices des points de mesure ; i et j varient de 1 à n
x_i, x_j	=	informations disponibles pour l'estimation, mesurées aux points x_i et x_j
v	=	volume à estimer (point ou bloc)
λ_i (ou λ_j)	=	pondérateur affecté à la valeur mesurée en x_i
μ	=	paramètre de Lagrange
$\gamma(x_i, x_j)$	=	valeur du variogramme $\gamma(h)$ pour h =distance entre x_i et x_j
$\gamma(x_i, v)$	=	valeur moyenne du variogramme $\gamma(h)$ pour h =distance entre x_i et x , x étant un point décrivant le volume v

Pour une estimation d'un point x à partir de deux informations ponctuelles x_1 et x_2 , le système s'écrit:

$$\begin{aligned} \lambda_1 \gamma(x_1, x_1) + \lambda_2 \gamma(x_1, x_2) + \mu &= \gamma(x_1, x) \\ \lambda_1 \gamma(x_2, x_1) + \lambda_2 \gamma(x_2, x_2) + \mu &= \gamma(x_2, x) \\ \lambda_1 + \lambda_2 &= 1 \end{aligned}$$

$\gamma(x_1, x_2)$ est la valeur de $\gamma(h)$ lorsque le premier point du vecteur h est positionné en x_1 et le deuxième également, c'est à dire pour une distance $h = 0$. De même $\gamma(x_1, x_2)$ est la valeur de $\gamma(h)$ pour h = distance entre x_1 et x_2 .

Enfin la variance de l'erreur d'estimation, appelée variance de krigeage, est :

$$S_k^2 = \sum_{i=1,n} \lambda_i \gamma(x_i, v) - \bar{\gamma}(v, v) + \mu$$

$\bar{\gamma}(v, v)$ étant la valeur moyenne du variogramme sur le volume v .

5.3. KRIGEAGE UNIVERSEL

Les équations du krigeage universel sont similaires à celles du krigeage ordinaire : on introduit en plus les termes liés aux fonctions de base de la dérive (les monômes $f^m(x_i)$) :

$$\sum_{j=1,n} \lambda_j \gamma(x_i, x_j) + \sum_{m=0,k} \mu_m f^m(x_i) = \gamma(x_i, v) \quad \forall i \text{ (n équations)}$$

$$\sum_{i=1,n} \lambda_i f^m(x_i) = f^m(x) \quad \text{pour } m = 0, \dots, k \text{ (k+1 équations)}$$

Dans ce cas, le variogramme $\gamma(h)$ est le variogramme des résidus $Y(x)$ (Cf. chapitres 3.3 et 4.6). En pratique, on utilise le variogramme des résidus estimés qui est biaisé par rapport au variogramme des résidus vrais, puisqu'il fait intervenir une dérive estimée. La deuxième série d'équations, ou conditions d'universalité, assure une estimation exacte de la composante due à la dérive.

5.4. KRIGEAGE EN FAI-K

Les équations du krigeage en FAI-k sont similaires à celles du krigeage universel : la covariance généralisée $K(h)$ y remplace simplement le variogramme :

$$\sum_{j=1,n} \lambda_j K(x_i, x_j) + \sum_{m=0,k} \mu_m f^m(x_i) = K(x_i, v) \quad \forall i \text{ (n équations)}$$

$$\sum_{i=1,n} \lambda_i f^m(x_i) = f^m(x) \quad \text{pour } m = 0, \dots, k \text{ (k+1 équations)}$$

5.5. KRIGEAGE AVEC DÉRIVE EXTERNE

Dans ce modèle (cf. section 3.4), la variable d'étude $Z(x)$ peut se décomposer en :

$$Z(x) = a + b.E(x) + Y(x)$$

$E(x)$ étant la dérive externe, et $Y(x)$ un résidu stationnaire.

La méthode d'interpolation nécessite 2 krigeages successifs :

- 1 - Un premier krigeage de la dérive externe, qui permet d'obtenir un estimateur $E^*(x)$ en tout point de la grille à estimer. Ce krigeage utilise le variogramme de $E(x)$ qu'il faut calculer et ajuster au préalable.

- 2 - Un deuxième krigeage de la variable finale avec prise en compte de la dérive externe, qui est analogue à un krigeage universel où on remplacerait les monômes $f^m(x)$ par les deux fonctions de base 1 et $E(x)$ (ou 1, $E_1(x)$, $E_2(x)$, ... si la dérive externe a plusieurs composantes).

Ce deuxième krigeage utilise le variogramme des résidus $Y(x) = Z(x) - a - b.E(x)$. Ce variogramme est calculable directement si $Z(x)$ et $E(x)$ (c'est à dire la variable brute et la dérive externe) sont connues simultanément aux points expérimentaux. Le calcul du variogramme des résidus se fait comme indiqué à la section 4.6.1.

5.6. MISE EN OEUVRE D'UN KRIGEAGE EN 2 OU 3 DIMENSIONS

Les programmes de krigeage de GEORES sont K2KRI pour les données en 2 dimensions et K3KRI pour les données en 3 dimensions. Ces programmes offrent de nombreuses possibilités concernant :

- le support estimé (points isolés, grille de points, réseau de blocs 3D ou de panneaux 2D) ;
- le choix des données à prendre en compte pour l'estimation et le type de ces données ;
- la délimitation du domaine à estimer (limitation par un contour, par un code...) ;
- la méthode d'interpolation à utiliser (méthode géostatistique ou autre) ;
- le choix du voisinage d'estimation (méthode de recherche des données servant à l'estimation) ;
- la discrétisation des blocs.

Ces différents aspects sont examinés ci-dessous.

5.6.1. Krigeage en 2 dimensions

a. Support estimé, choix de la taille des blocs

K2KRI permet d'estimer :

- des valeurs ponctuelles en des points isolés ou répartis sur une grille 2D,
- des valeurs moyennes sur des panneaux en 2D formant une grille régulière ou un maillage écossais.

Dans le cas d'une estimation sur une grille 2D, le choix des dimensions de la grille doit se faire sur la base des considérations suivantes :

- une grille trop fine conduit à un grand nombre de points, à des temps de calcul plus longs et à des tailles de fichiers plus importantes pour une amélioration de la précision qui est illusoire,
- si la grille est très fine par rapport à la densité d'information, on risque d'obtenir des artefacts lors des changements de voisinage,
- il est conseillé de choisir comme taille horizontale des panneaux une valeur voisine de la maille de sondages (ou de la distance moyenne entre sondages). Si possible, ne pas prendre une taille inférieure à la moitié de la maille de sondages.

b. Préparation des données d'entrée

Le fichier des données à utiliser doit contenir au minimum les coordonnées X et Y des points expérimentaux, ainsi que les valeurs des variables mesurées. Il peut s'agir d'un fichier de points isolés, d'un fichier tête ou base, ou d'un fichier grille.

S'il s'agit de données en profils et qu'on souhaite utiliser des options particulières à ce type de données, le fichier base doit contenir un code définissant le nom du profil auquel appartient la donnée.

Si l'on souhaite effectuer un krigeage avec dérive externe, les valeurs des composantes de la dérive externe doivent également figurer dans le fichier.

Enfin, d'autres variables optionnelles sont possibles (Cf. manuel d'utilisation de K2KRI) :

- variable somme des poids de krigeage (à ajouter si nécessaire avec MEXTEN),
- variable variance de l'erreur de mesure de chaque donnée ou profil.

c. Délimitation du domaine à estimer.

Il est parfois utile de réaliser l'estimation en plusieurs fois (exemple: différents types de minerais, ou variogrammes différents selon les zones), ou de limiter géographiquement l'extension du domaine estimé (exemple: limitation par des failles ou par des contours géologiques).

Ceci est possible dans K2KRI, à condition de préparer un fichier "forme", différent du fichier des données d'entrée. Ce fichier décrit la zone à estimer.

Deux cas sont possibles selon qu'on estime des points isolés ou une grille. La clique WINDOW permet également de limiter le domaine estimé.

1 - points isolés

Dans ce cas la liste des points à estimer est contenue dans un fichier qui donne les coordonnées (X,Y) des points à estimer. Ce fichier peut également contenir une variable permettant de définir quels points estimer, par l'intermédiaire d'un intervalle de sélection appliqué à cette variable. Le fichier, la variable et l'intervalle de sélection sont définis dans la clique DOMAIN (qui sert par ailleurs à choisir le type de données à estimer : points isolés ou grille). Il est différent du fichier des données d'entrée et du fichier résultat.

La clique FORMKRIG permet ensuite d'opérer une sélection complémentaire par l'intermédiaire de 2 options :

2. polygones,
3. enveloppe convexe des données.

Dans le cas 2, les coordonnées des sommets des polygones sont lues au clavier. Il est possible d'estimer à l'intérieur ou à l'extérieur des polygones. Dans le cas 3, seuls les points situés à l'intérieur de l'enveloppe convexe des données sont estimés.

2 - grille de points, de blocs, ou maillage écossais

La définition du domaine à estimer se fait dans la clique FORMKRIG qui permet 3 types de délimitation :

1. variable indicatrice (pour une grille seulement),
2. polygones,
3. enveloppe convexe des données.

Dans le cas 1, la variable de sélection est lue dans un fichier grille ayant la même structure que celle définie à la clique DOMAIN. Ce fichier doit être différent du fichier résultat.

Les options 2 et 3 sont disponibles également dans le cas d'une estimation sur des points isolés.

Dans le cas d'une estimation d'une grille de blocs ou de maillage écossais, la surface du bloc prise en compte pour l'estimation peut également être stockée sur le fichier résultat.

3 - clique WINDOW

L'estimation peut être faite à l'intérieur d'une fenêtre rectangulaire définie dans cette clique. Cette délimitation se superpose à celles définies dans DOMAIN ou FORMKRIG.

d. recherche du voisinage d'estimation

Pour estimer un point ou un bloc, les programmes utilisent les données environnantes (voisinage). La méthode de recherche de ces données peut avoir une importance dans certains cas.

Le choix des paramètres de voisinage se fait dans les cliques NEIGHBOR (estimation de points isolés ou de grilles) ou PATNEIG (estimation de maillages écossais). Deux types de voisinages sont possibles :

1 - voisinage rectangulaire

Ce type de voisinage est conseillé lorsque les données sont réparties de façon homogène. Il permet de limiter la recherche des voisins à l'intérieur d'un rectangle. On l'utilise surtout lorsqu'on estime une valeur moyenne par maille.

2 - recherche par octants

Cette méthode est conseillée lorsque les données sont disposées en profils ou en paquets. Elle permet d'obtenir une répartition plus homogène des données utilisées pour l'estimation, en faisant en sorte que ces dernières soient recherchées dans les différentes directions de l'espace (voir le manuel d'utilisation de K2KRI pour de plus amples informations).

Dans tous les cas il est recommandé que les voisinages comportent au moins une douzaine de points, voire plus lorsqu'il y a une dérive.

5.6.2. Krigeage en 3 dimensions

a. Support estimé, choix de la taille des blocs

K3KRI permet d'estimer des valeurs ponctuelles aux noeuds d'un réseau 3D, ou des valeurs moyennes sur des blocs réguliers.

Le choix des dimensions horizontales de la grille 3D doit se faire sur la base de considérations similaires au cas 2D. D'autre part, la hauteur des blocs doit correspondre à la hauteur des gradins ou à un sous multiple de cette hauteur (pour les applications minières). Il est généralement souhaitable, avant le krigeage, (et avant K3CLAS) de découper les sondages en bancs de hauteur égale à celle des blocs (ou à un sous multiple - cf. point e. ci-dessous).

b. Préparation des données d'entrée

Le fichier de données d'entrée doit obligatoirement avoir été au préalable indexé sur les blocs du réseau à estimer, à l'aide du programme K3CLAS. Il doit contenir au minimum les 3 coordonnées milieu de passe XMID, YMID, ZMID calculées par K3CLAS (ou à défaut XXXX,YYYY,ZZZZ) et les valeurs des variables aux points expérimentaux.

c. Délimitation du domaine à estimer

La définition du domaine à estimer est réalisée à l'aide des cliques FORMKRIG et WINDOW.

• clique FORMKRIG

La délimitation du domaine à estimer se fait par l'intermédiaire d'un fichier grille contenant une variable sur laquelle on peut appliquer un intervalle de sélection. Ce fichier doit avoir la même structure que le fichier grille résultat. Il peut avoir été créé et valorisé à l'aide du programme K3FORM.

• clique WINDOW

Le traitement peut être limité à un parallélépipède défini dans cette clique. Si FORMKRIG a été définie, c'est l'intersection des 2 domaines qui est estimée.

d. recherche du voisinage d'estimation

Pour estimer un point ou un bloc, les programmes utilisent les données environnantes (voisinage). La méthode de recherche de ces données peut avoir une importance dans certains cas.

Pour accélérer le temps calcul et faciliter la recherche des voisins, le voisinage est défini par un certain nombre de blocs entourant le bloc à estimer. Toutes les données tombant à l'intérieur de ces blocs sont utilisées.

Selon le degré de précision souhaité et selon la disposition plus ou moins régulière des données, il est possible ou non de prendre en compte des symétries.

La définition du voisinage se fait à l'aide d'une des cliques STANDNEI et COMPLNEI.

STANDNEI permet de définir le voisinage en donnant le nombre de blocs en X et Y sur le niveau du bloc à estimer, puis sur les niveaux -1,+1, puis sur les niveaux -2,+2, etc. autour du bloc. Il y a symétrie par rapport au niveau auquel appartient le bloc krigé.

COMPLNEI permet une description fine du voisinage, en définissant bloc par bloc (à l'intérieur d'un grand voisinage parallélépipédique) ceux qui appartiennent au voisinage, avec possibilité d'affecter un pondérateur commun à un groupe de blocs en fonction des symétries.

La clique MININEIG permet de plus de définir un voisinage restreint, défini par un nombre de blocs en X, Y et Z : l'estimation ne sera pas faite si il n'y a pas au moins une donnée dans ce voisinage. Ceci permet d'éviter d'estimer des blocs à partir de données qui seraient toutes éloignées.

Par défaut le programme choisit un voisinage correspondant à STANDNEI, avec 3*3 blocs sur le niveau du bloc à estimer et 1 bloc en dessus et en dessous du bloc à estimer.

e. Discrétisation

Dans le cas de krigeage de blocs, la valeur moyenne du bloc est la moyenne des krigeages de ses points de discrétisation (effectués avec le même voisinage). Plus on augmente le nombre de points de discrétisation, meilleure est la précision. Selon l'horizontale, l'expérience montre qu'une discrétisation de 3*3 suffit en général. Si les données sont ponctuelles, on peut prendre aussi une discrétisation de 3 selon la verticale. Mais les données concernent fréquemment des passes de sondages verticaux. La solution la plus simple consiste à redécouper les sondages en passes de longueur égale à la hauteur des blocs, et par conséquent à prendre une discrétisation de 1 selon la verticale. Le variogramme utilisé représente alors directement les données régularisées par la hauteur des blocs. Si par contre on utilise directement les données de passes de longueur L alors que les blocs ont une hauteur H, on prendra une discrétisation de l'ordre du rapport H/L.

5.7. EXPLOITATION DES RÉSULTATS, CALCUL DE RÉSERVES

Les résultats du krigeage peuvent être valorisés par les programmes suivants :

- GDISP : dessin d'une grille 2D ou d'un banc de grille 3D, avec report de la variable estimée, d'un symbole, ou hachurage des blocs avec différents figurés ou couleurs ;
- GISOF : tracé de courbes isovaleurs d'une variable 2D ;
- GDUNE : visualisation en 3 dimensions d'une variable 2D ;
- K3LIST : édition sur listing des variables contenues dans les blocs ;
- K3STAT : calcul de moyennes par bancs, calcul de réserves, avec pondération possible par des variables volumes ou tonnage.

Un exemple d'utilisation de K3STAT, GDISP et GISOF est donné dans le cas d'étude "analyse structurale en 3D, krigeage ordinaire, calcul des réserves" (voir le manuel "Logiciels géostatistiques de GDM : cas d'étude").

6. CAS MULTIVARIABLE

GEORES permet l'analyse structurale de données multivariées :

- les statistiques de base (MSTAT, HISTOG, BINAIR) permettent de comparer les variables 2 à 2 (calcul de coefficients et de nuages de corrélation) ;
- les variogrammes directs ou croisés peuvent être calculés (K2GAM, K3GAM) et ajustés (KFIGAM).

Au niveau du krigeage, plusieurs variables peuvent être traitées simultanément si elles sont définies aux mêmes points expérimentaux et si elles ont le même variogramme (ou des variogrammes directs et croisés proportionnels). GEORES ne comporte pas de module de cokrigeage, mais le krigeage avec dérive externe remplace le cokrigeage dans certaines applications.

7. SIMULATIONS

7.1. INTRODUCTION

Comme toute méthode d'interpolation, le krigeage présente l'inconvénient bien connu de lisser les valeurs. Lorsque l'objectif est de modéliser les fluctuations d'un phénomène, le krigeage n'est donc pas adapté. C'est pour cette raison que les techniques de simulation ont été mises au point dans les années 1970.

Plus précisément, les simulations permettent de construire une variable qui :

- 1 - respecte l'histogramme des données initiales ;
- 2 - respecte le variogramme des données initiales ;
- 3 - soit un interpolateur exact : la variable simulée sera égale à la valeur réelle de la variable au droit des points de données.

Si les points 1 et 2 sont vérifiés, la simulation est dite **non conditionnelle** : la variable simulée reproduit la structure spatiale de la variable initiale, mais ne passe pas par les mêmes valeurs aux points de données.

Si le point 3 est également vérifié, la simulation est dite **conditionnelle** : elle passe par les points expérimentaux.

Dans le domaine minier, les simulations sont utilisées pour étudier différentes méthodes d'exploitation et évaluer leurs performances, en tenant compte de leurs impacts possibles sur la sélectivité. On peut par exemple simuler les teneurs sur des supports correspondant au volume de sélection (godet de pelle, volée, camion...). En couplant le modèle des teneurs avec un algorithme (manuel ou automatique) permettant de calculer les coûts d'exploitation, les recettes, ... en fonction des engins et de la méthode d'exploitation envisagée et de la sélectivité possible avec cette méthode, chaque méthode peut ainsi être évaluée. La comparaison de plusieurs tests permet soit d'effectuer des analyses de sensibilité, soit d'aider à appréhender la solution qui donne les meilleurs résultats.

Dans des domaines non miniers, les simulations peuvent être utiles chaque fois que la variabilité du phénomène étudié doit être prise en compte (exemple : simulation de la transmissivité en vue de modéliser des écoulements dans un massif rocheux).

7.2. CONSTRUCTION D'UNE SIMULATION NON CONDITIONNELLE

La méthode de construction utilisée dans LUCY est celle dite de "dilution de germes poissonniens par des sphères" :

- 1 - des germes ponctuels sont implantés dans l'espace au hasard selon une loi uniforme (les coordonnées des germes sont tirées au hasard). Le nombre de germes générés est fonction de la taille du domaine à simuler et de la portée du variogramme à reproduire (les germes sont implantés dans le domaine à simuler dilaté du diamètre des sphères) ;

- 2 - en chaque germe on plante une sphère (ou une ellipsoïde si anisotropie) de diamètre égal à la portée du variogramme à simuler, affectée d'un poids aléatoire de +1 ou -1 (pour un variogramme sphérique) ;
- 3 - la valeur simulée en un point P est la somme des poids attribués à toutes les sphères qui le contiennent.

Les conséquences pratiques de ce mode opératoire sont de 3 ordres :

1. Les variables simulées sont (pseudo)gaussiennes

L'histogramme des valeurs simulées ne peut pas être exactement gaussien. Il s'en rapproche toutefois si le nombre de sphères engendré est grand.

Si la variable d'étude n'est pas gaussienne, il faut faire appel à des algorithmes de transformation de loi non gaussienne en loi gaussienne ("anamorphose inverse"), ou inversement ("anamorphose") - cf. section 8.4.

2. Choix du nombre de germes. Conséquences sur les structures à simuler

Supposons que l'on simule des valeurs aux noeuds d'une grille régulière. Pour que deux noeuds voisins n'aient pas systématiquement la même valeur simulée, il faut qu'en moyenne il y ait au moins une frontière de sphère entre ces deux noeuds. Lorsque la portée est grande, cette condition peut conduire à un nombre important de sphères. Lorsque la portée est petite, cette condition revient à choisir au moins autant de sphères que de noeuds. Par ailleurs, il est souhaitable que le volume occupé par les sphères couvre le domaine simulé. Ceci peut conduire à un très grand nombre de sphères lorsque la portée est très faible. On risque donc d'avoir un grand nombre de sphères lorsque :

- la portée est grande par rapport au champ simulé,
- la portée est faible par rapport à la maille.

On peut donc conseiller de :

- éviter de simuler une composante de très grande portée,
- ramener la simulation d'une composante de courte portée (inférieure à la maille de la grille ou des données) à la simulation d'un effet de pépite. Celui-ci sera simulé directement sans passer par la dilution de sphères, par le tirage au sort d'une gaussienne indépendante en chaque point à simuler.

3. générateur de nombres au hasard

Le programme utilise des séquences de valeurs au hasard :

- pour les coordonnées des germes ;
- pour la simulation de l'effet de pépite.

Le nombre de valeurs au hasard dépend du nombre de valeurs à simuler (grille+points expérimentaux) et du variogramme à simuler. A chaque nouvelle exécution du programme, la séquence de valeurs générée est la même, ce qui garantit d'obtenir les mêmes résultats si la même simulation est effectuée à des instants différents. L'utilisateur peut néanmoins se positionner ailleurs dans la séquence.

4. Normation des valeurs

Comme indiqué précédemment, les valeurs simulées sont pseudo-gaussiennes. Leur moyenne doit être nulle et leur variance égale à la variance théorique (dans le cas de la simulation d'une grille par exemple, la variance des données doit être égale à la variance de dispersion dans une grille), cette valeur étant calculable théoriquement à partir du variogramme.

En pratique on n'observe jamais cette égalité pour différentes raisons : nombres au hasard pas réellement au hasard, nombre de sphères non idéal, fluctuations statistiques... Il est donc possible, si on le demande, de centrer et normer les valeurs simulées de façon à ce que leur moyenne et leur variance soient égales aux valeurs prévues par la théorie. Cette normation est faite de façon à reproduire la moyenne et la variance théoriques sur la grille simulée.

7.3. CONSTRUCTION D'UNE SIMULATION CONDITIONNELLE

La conditionnalisation se base sur les propriétés du krigeage simple sous hypothèse stationnaire. On peut en effet écrire que :

$$Y(x) = Y^*(x) + [Y(x) - Y^*(x)] = Y^*(x) + e(x)$$

avec $e(x) = (Y(x) - Y^*(x))$ (erreur de krigeage) indépendante de la valeur krigée $Y^*(x)$. Le krigeage donne Y^* et la variance de $e(x)$ mais bien sûr ne permet pas de connaître la valeur exacte de $e(x)$.

Le principe de construction d'une simulation conditionnelle consiste à remplacer l'erreur de krigeage inconnue $e(x)$ par une simulation de cette erreur. Pour ce faire on engendre une simulation non conditionnelle $Y_s(x)$ dont on calcule la valeur aux points à simuler et aux points expérimentaux. En tout point x à simuler, on effectue le krigeage $Y_s^*(x)$ à partir des valeurs de la simulation aux seuls points expérimentaux (dans les mêmes conditions que pour les données). On obtient ainsi une simulation de l'erreur de krigeage $e_s(x) = Y_s(x) - Y_s^*(x)$. La simulation conditionnelle est :

$$Y_{sc}(x) = Y^*(x) + e_s(x) = Y^*(x) + [Y_s(x) - Y_s^*(x)] \quad [1]$$

Elle passe par les points expérimentaux car $e_s(x) = 0$ en ces points en raison du non biais du krigeage, et elle a même variogramme que $Y(x)$.

L'expression [1] peut se réécrire sous la forme :

$$Y_{sc}(x) = Y_s(x) + [Y^*(x) - Y_s^*(x)]$$

La simulation conditionnelle finale est donc la somme :

- d'une simulation non conditionnelle $Y_s(x)$,
- d'un terme $Y^*(x) - Y_s^*(x)$ correspondant à la différence des krigeages de la variable initiale $Y(x)$ et de sa simulation $Y_s(x)$. Comme ces deux variables ont le même variogramme et qu'elles sont connues aux mêmes points (pour cela $Y_s(x)$ doit être calculée également aux points expérimentaux), ces deux krigeages seront donc identiques et donneront les mêmes poids λ_i . Ils peuvent donc être faits en parallèle (pour que cela soit possible d'un point de vue informatique, il faut stocker $Y(x)$ et $Y_s(x)$ dans le même fichier).

Il apparaît en fin de compte qu'une simulation conditionnelle comporte les étapes suivantes :

- 1 - calcul de la simulation non conditionnelle $Y_s(x)$ aux points expérimentaux et aux noeuds de la grille à simuler,
- 2 - krigeage simultané avec le même variogramme de $Y(x)$ et $Y_s(x)$ pour obtenir $Y^*(x)$ et $Y_s^*(x)$,
- 3 - calcul de $Y_{sc}(x) = Y_s(x) + Y^*(x) - Y_s^*(x)$ aux noeuds de la grille à simuler. Ce calcul est effectué par le programme CONDIT (voir exemple d'utilisation dans le manuel : "Logiciels géostatistiques de GDM : Cas d'étude").

7.4. SIMULATION DE VARIABLES CORRÉLÉES

Lorsque l'on simule plusieurs variables simultanément, les simulations obtenues sont indépendantes car les nombres au hasard utilisés sont indépendants. Une étape est donc nécessaire pour rendre les simulations corrélées si les variables de départ le sont. Cette étape est réalisée par le programme GAUCOR (voir exemple d'utilisation dans le manuel : "Logiciels géostatistiques de GDM : Cas d'étude").

7.5. PRISE EN COMPTE DE L'HISTOGRAMME

Les valeurs simulées suivent une loi gaussienne. Si l'on souhaite obtenir une simulation conditionnelle $Z_{sc}(x)$ d'une variable $Z(x)$ non gaussienne, il faut :

- 1 - transformer $Z(x)$ en une variable gaussienne $Y(x)$. Ceci fait appel à une "anamorphose inverse", définie par des coefficients d'anamorphose,
- 2 - simuler $Y_{sc}(x)$,
- 3 - effectuer la transformation inverse $Y_{sc}(x) \rightarrow Z_{sc}(x)$, avec les mêmes coefficients que ceux utilisés à l'étape 1.

Le principe de l'anamorphose est expliqué section 8.4

7.6. FICHIERS NÉCESSAIRES POUR UNE SIMULATION

Pour une simulation conditionnelle on a besoin de :

- 1 fichier base (points isolés) contenant au minimum :
 - les coordonnées des points expérimentaux ;
 - les valeurs de la (des) variable(s) initiale(s) après transformation éventuelle en gaussienne(s) ;
 - deux variables pour chaque variable simulée : la première reçoit les simulations indépendantes (créées par LUCY) et la seconde les simulations corrélées (GAUCOR) en cas d'étude de variables corrélées.

Ce fichier doit exister avant de lancer le programme LUCY.

- 3 fichiers grille
 - le premier contient pour chaque variable à simuler deux variables : la première pour les simulations indépendantes et la seconde pour les simulations corrélées. Ce fichier peut être créé par LUCY ;
 - le deuxième contient les résultats du krigeage conditionnant, c'est à dire deux variables pour chaque variable à simuler: le krigeage des valeurs initiales et le krigeage des valeurs simulées. Ce fichier est créé par K2KRI (ou K3KRI en 3 dimensions) ;
 - le troisième fichier contient les valeurs finales de la simulation conditionnelle et est valorisé par CONDIT.

Un exemple complet de simulation conditionnelle de 2 variables corrélées est fourni dans le manuel : "Logiciels géostatistiques de GDM : Cas d'étude".

8. RÉSERVES RÉCUPÉRABLES GLOBALES, ANAMORPHOSE

Ce chapitre concerne plus particulièrement les applications minières où l'on cherche à calculer les réserves récupérables lorsqu'on applique une teneur de coupure z_c pour distinguer le minerai du stérile. Des applications non minières sont également possibles, par exemple dans le domaine de l'environnement, si l'on cherche à calculer la proportion de valeurs supérieures à un seuil donné (seuil de contamination par exemple).

8.1. COURBES TONNAGE-TENEUR

Si l'on applique une teneur de coupure z_c à la variable $Z(x)$ et que l'on ne retient que les valeurs supérieures ou égales à z_c , on obtient les fonctions suivantes (ces fonctions sont dites courbes tonnage-teneur, ou courbes de sélectivité) :

8.1.1. Tonnage minéral récupéré en fonction de la coupure : $T(z_c)$

On l'obtient en multipliant le tonnage total T_{tot} par la proportion de valeurs supérieures ou égales à la coupure z_c :

$$T(z_c) = T_{tot} \cdot (1 - F(z_c))$$

$F(z_c)$ étant la fonction de répartition de Z , soit : $F(z_c) = P(Z(x) < z_c)$

$T(z_c)$ est une fonction décroissante de z_c . Son allure est représentée figure 10.

Pour $z_c = 0$, on récupère l'ensemble du gisement, et pour z_c supérieur à la plus grande valeur, on récupère un tonnage nul. Par convention on norme $T(z_c)$ par T_{tot} de façon à ramener sa valeur dans l'intervalle $[0 ; 1]$.

Si on utilise la densité de probabilité $f(z)$, $T(z_c)$ s'exprime par :

$$T(Z_c) = T_{tot} \cdot \int_{Z_c}^{+\infty} f(z) dz$$

8.1.2. Quantité de métal récupérée en fonction de la coupure : $Q(z_c)$

La quantité de métal présente dans un élément de volume s'exprime en multipliant ce volume par sa teneur. Avec les mêmes notations que précédemment, on obtient :

$$Q(Z_c) = T_{tot} \cdot \int_{Z_c}^{+\infty} z \cdot f(z) dz$$

$Q(z_c)$ est une fonction décroissante de la teneur de coupure z_c . Son allure est représentée figure 11.

Pour $z_c = 0$, on récupère l'ensemble du métal du gisement, qui est égal au tonnage total T_{tot} multiplié par la teneur moyenne m . Par convention on norme $Q(z_c)$ par T_{tot} de façon à ramener sa valeur dans l'intervalle $[0 ; 1]$.

8.1.3. Teneur moyenne récupérée en fonction de la coupure: $m(z_c)$

Par définition la teneur moyenne de la partie récupérée s'exprime par :

$$m(z_c) = \frac{Q(z_c)}{T(z_c)}$$

$m(z_c)$ est une fonction croissante de z_c . Pour une coupure $z_c = 0$, la teneur moyenne récupérée est la teneur moyenne de l'ensemble du gisement. Lorsqu'on augmente la coupure, la teneur moyenne augmente jusqu'à la teneur maximale observée. L'allure de la fonction $m(z_c)$ est visible figure 12.

8.1.4. Bénéfice conventionnel en fonction de la coupure: $B(z_c)$

Imaginons une exploitation sélective où l'on peut extraire indépendamment un bloc donné, où la valeur d'un bloc est proportionnelle au métal qu'il contient, et où les frais d'extraction et de traitement, proportionnels au tonnage, sont les mêmes pour chaque bloc. Le bénéfice maximal est obtenu en retenant chacun des blocs dont le métal paye ces frais, c'est-à-dire dont la teneur dépasse une certaine coupure z_c . Pour chacun de ces blocs, le bénéfice qu'il rapporte est égal, à un facteur constant près, au "bénéfice conventionnel" défini par :

$$B(z_c) = Q(z_c) - z_c \cdot T(z_c)$$

$B(z_c)$ est une fonction décroissante convexe de z_c . Son allure est donnée figure 13.

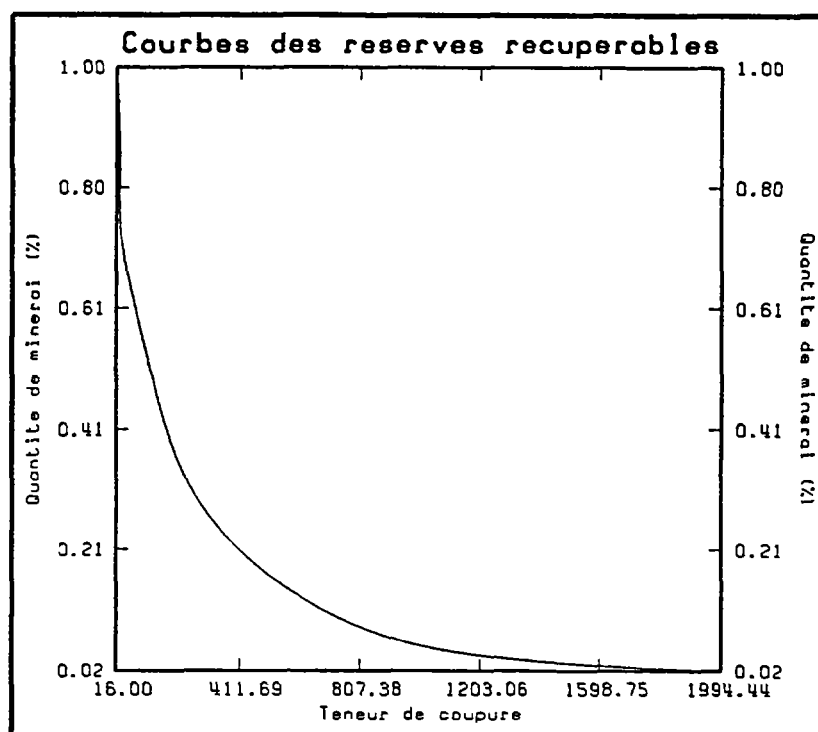


Fig. 10 - Tonnage minéral récupéré en fonction de la teneur de coupure

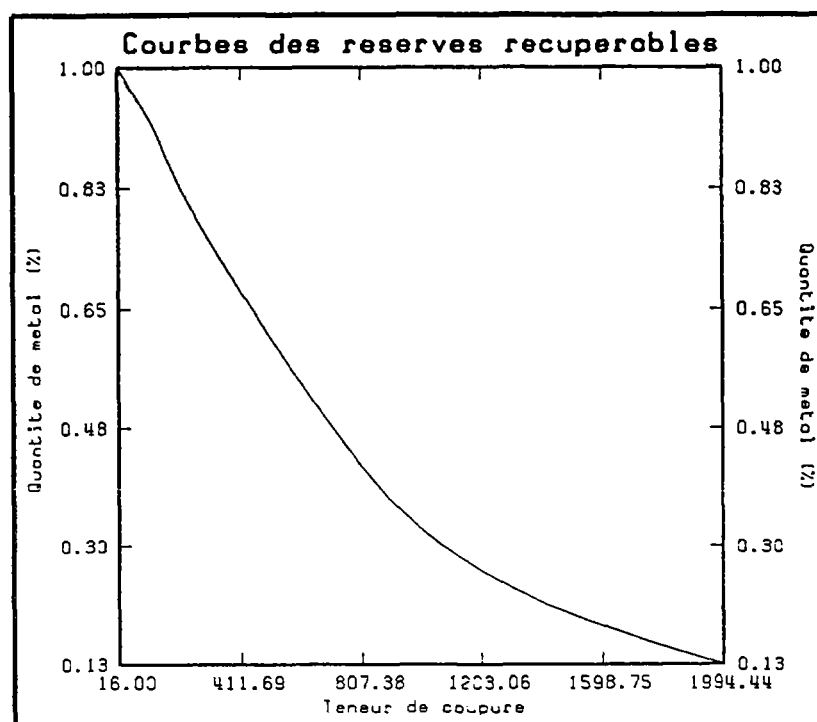


Fig. 11 - Quantité de métal récupérée en fonction de la teneur de coupure

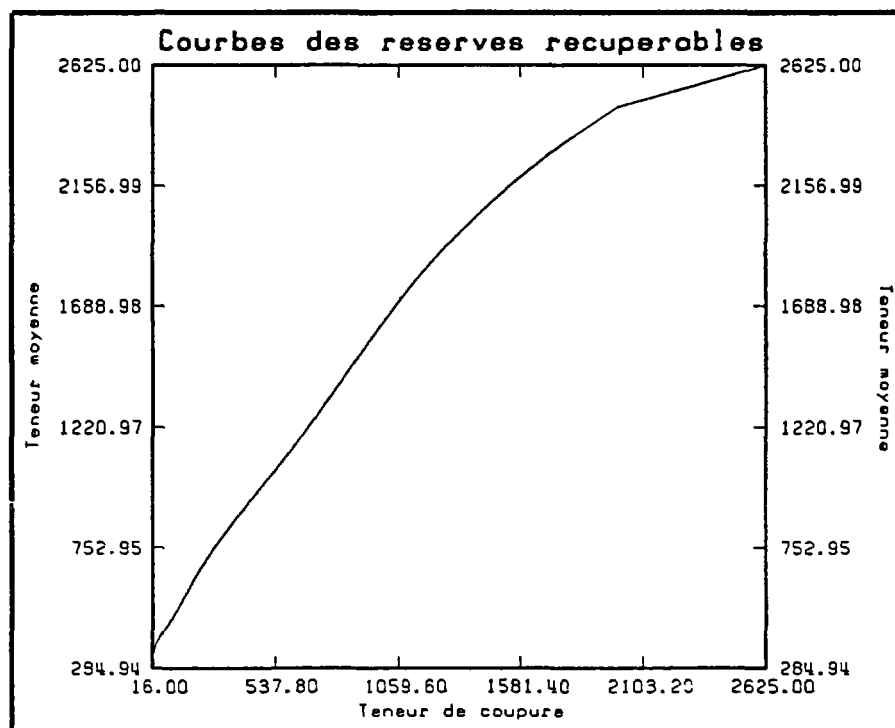


Fig. 12 - Teneur moyenne récupérée en fonction de la teneur de coupure

8.1.5. Quantité de métal récupérée en fonction du tonnage minéral récupéré : $Q(T)$

Pour comparer Q et T directement sans passer par les courbes $Q(z_c)$ et $T(z_c)$, on construit la courbe $Q(T) = f(T)$ qui donne la fraction de quantité métal récupérée lorsqu'on récupère la fraction de tonnage minéral T .

Chaque point de la courbe $(T, Q(T))$ correspond à la même coupure z_c . Cette courbe est croissante et concave (figure 14).

$T = 0$ correspond à une coupure supérieure à la coupure maximale et on ne récupère rien. $T = 1$ correspond à la coupure $z_c = 0$ où on récupère l'ensemble du gisement et donc la quantité métal totale.

Toutes les relations précédentes correspondent au cas idéal où on connaît la teneur ponctuelle z de chaque point de façon à pouvoir la comparer à la coupure z_c . Elles concernent de plus les réserves en place, c'est-à-dire en excluant toute contrainte technique (accessibilité, dilution, salissage), ou économique.

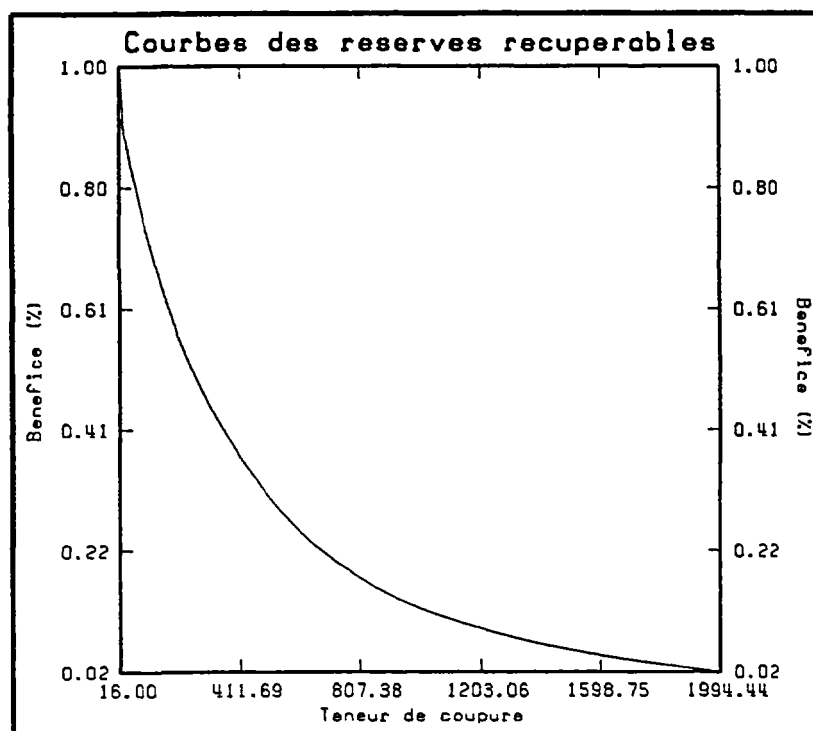


Fig.13 - Bénéfice conventionnel en fonction de la teneur de coupure

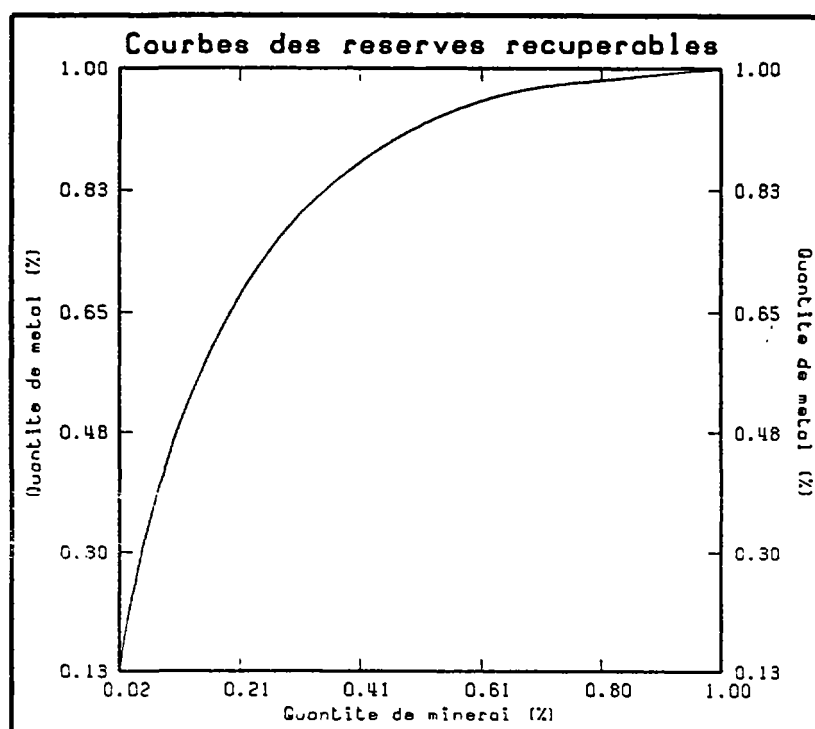


Fig. 14 - Quantité de métal récupérée en fonction du tonnage minéral récupéré

8.2. EFFET DE SUPPORT ET CHANGEMENT DE SUPPORT. RÉSERVES IDÉALES

En pratique les données sont mesurées sur des supports quasi-ponctuels, et les réserves calculées sur des supports v plus grands (par exemple une volée). Or la distribution d'une variable, et en particulier sa variance, dépendent du support sur lequel est définie la variable.

Les fonctions de répartition des variables "ponctuelles" $Z(x)$ et des variables définies sur des blocs $Z_v(x)$ ne sont pas égales comme le montrent les figures 15 et 16 : c'est l'effet de support. En particulier les teneurs "ponctuelles" peuvent prendre plus facilement des valeurs faibles ou fortes, alors que sur un support v les teneurs sont moyennées et ont tendance à prendre des valeurs plus proches des teneurs intermédiaires.

Les courbes tonnage-teneur dépendent donc du support considéré, et les réserves récupérables sont différentes selon que l'on considère une sélection sur des teneurs ponctuelles ou sur des teneurs de blocs. C'est la même chose si on s'intéresse à la probabilité de dépassement d'un seuil.

Un modèle de changement de support est donc nécessaire pour passer de la distribution (supposée connue) des points à celle (inconnue) des blocs.

Les réserves récupérables, si on sélectionne sur des teneurs de blocs, sont données par :

$$T_v(Z_c) = T_{tot} \cdot \int_{Z_c}^{+\infty} f_v(Z_v) dZ_v$$

$$Q_v(Z_c) = T_{tot} \cdot \int_{Z_c}^{+\infty} Z_v \cdot f_v(Z_v) dZ_v$$

Ces réserves sont dites "idéales" dans la mesure où elles supposent que l'on connaît la teneur vraie Z_v d'un bloc.

Pour les calculer, il faut disposer de la distribution des blocs : densité f_v ou fonction de répartition F_v . Un modèle de changement de support est donc nécessaire pour passer de la distribution supposée connue des points à celle inconnue des blocs. Des modèles simples ne sont disponibles que pour des lois monovariées particulières telles que la loi de gauss ou la loi gamma. C'est pourquoi on est amené à effectuer une transformation de la variable brute initiale $Z(x)$ en une variable de loi particulière $Y(x)$ (gaussienne ou gamma). Une telle transformation se fait par l'intermédiaire d'une fonction d'anamorphose $Z(x) = \Phi(Y(x))$ que l'on écrit sous forme d'un développement polynômial (cf. section 8.4) :

$$Z(x) = \Phi(Y(x)) = \sum_0^{+\infty} \psi_n H_n(Y(x))$$

De la même manière que l'on peut obtenir $Y(x)$ à partir de $Z(x)$ par $Y(x) = \Phi^{-1}(Z(x))$, on peut calculer $Y_v(x) = \Phi_v^{-1}(Z_v(x))$ au niveau des blocs.

Mais si on peut calculer $Z_v(x)$ à partir de $Z(x)$ par $Z_v(x) = 1/v \cdot \int_v Z(x) \cdot dx$, la relation n'est plus valable au niveau de Y_v : $Y_v(x) \neq 1/v \cdot \int_v Y(x) \cdot dx$. La détermination de la distribution de Y_v ne peut donc se faire simplement à partir de celle de $Y(x)$ (ou de $Z(x)$). On utilise alors des hypothèses supplémentaires sur la loi bivariable point-bloc ($Y(x), Y_v(x)$) pour traduire la relation entre $Y(x)$ et $Y_v(x)$. Ces hypothèses permettent schématiquement d'exprimer les densités bivariées point-bloc en fonction d'un coefficient de corrélation point-bloc, des densités monovariées de $Y(x)$ et $Y_v(x)$ (elles mêmes liées à celles de $Z(x)$ et $Z_v(x)$) et de paramètres liés aux anamorphoses Φ et Φ_v . Le coefficient de corrélation point-bloc est lui-même calculé à partir du coefficient de corrélation point-point (Cf. section 3.5). Le coefficient de corrélation point-point peut être calculé à partir du corrélogramme $\rho(h)$ de $Y(x)$ (ou de son variogramme).

En résumé, les modèles de changement de support utilisés font appel à deux modèles différents :

- un modèle de changement de support proprement dit, qui indique comment passer de la distribution des points à celle des blocs et qui permet de calculer les coefficients de corrélation point-bloc en fonction des coefficients de corrélation point-point ;
- un modèle bivariable ponctuel qui permet de calculer les coefficients de corrélation point-point à partir des lois monovariées (en utilisant le corrélogramme ou le variogramme).

On peut alors calculer les densités bivariées point-point ou point-bloc et en déduire les réserves récupérables.

Trois modèles de changement de support sont utilisés dans GEORES :

- la correction affine de support,
- le modèle continu,
- le modèle discret.

Ces modèles sont décrits ci-après.

8.2.1. Correction affine de support

Ce modèle très grossier ne permet que d'obtenir des ordres de grandeur. Il consiste à faire l'hypothèse que les distributions des points et blocs (densité, fonction de répartition) ont même forme et que l'on passe de l'une à l'autre par une affinité destinée à reproduire la réduction de variance tout en conservant la moyenne.

On écrit alors que $\frac{Z(x) - m}{\sigma}$ et $\frac{Z_v(x) - m}{\sigma_v}$ ont même loi.

Dans la mesure où la loi de Z_v est explicitement donnée par ce modèle, il n'est pas nécessaire de faire appel à un modèle bivariable ponctuel. L'avantage de ce modèle est sa grande simplicité. Son inconvénient est qu'il gomme tout effet 0 (il y a "effet 0" quand la proportion de valeurs nulles est significative). Il est de plus en plus approximatif au fur et à mesure que la taille du support augmente.

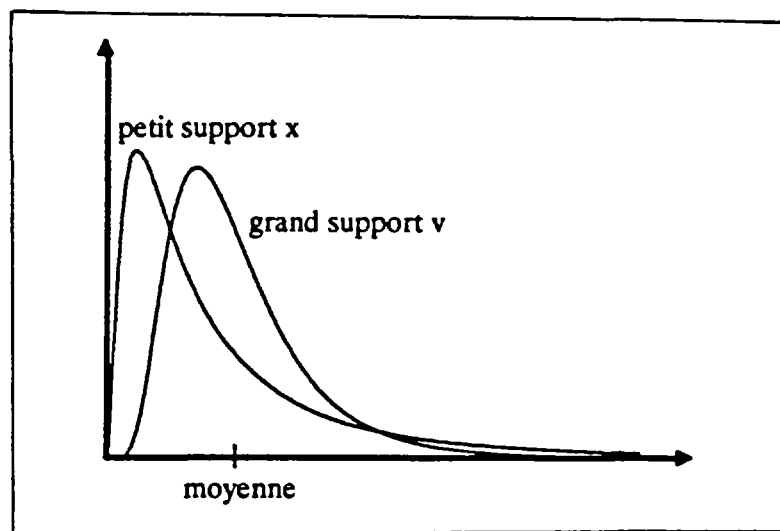


Fig. 15 - Densités de probabilités de variables mesurées sur des supports différents (d'après RIVOIRARD, 1991, page 64)

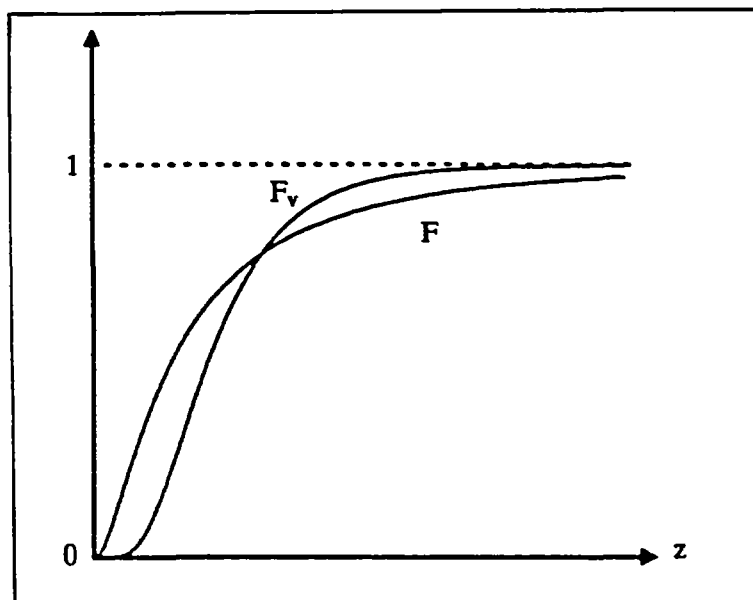


Fig. 16 - Fonctions de répartition de variables mesurées sur des supports différents (d'après RIVOIRARD, 1991, page 65)

8.2.2. Modèle continu

La position x de la variable $Z(x)$ est supposée bien fixée dans le bloc v . On peut alors calculer les coefficients de corrélation point-bloc $\rho_n^v(x,v)$ par discrétisation du volume v et moyenne (à un facteur près) du coefficient point-point $\rho_n(x,x')$. Les $\rho_n(x,x')$ sont eux-mêmes déterminés à partir du variogramme de la variable $Y(x)$ liée aux teneurs ponctuelles, et des hypothèses sur la loi bivariable ponctuelle (Cf. section 3.5). Dans le cas du modèle continu, le programme demande donc les paramètres du bloc v (taille) et ceux du variogramme des anamorphosées Y .

N.B. : n est le degré du polynôme dans le développement polynômial de l'anamorphose (section 8.4).

8.2.3. Modèle discret

\underline{x} étant un point aléatoire tiré avec une densité uniforme dans v , $Y(\underline{x})$ et $Y_v(\underline{x})$ étant les lois particulières (gauss, gamma) associées à $Z(\underline{x})$ et $Z_v(\underline{x})$, le modèle discret consiste à faire une hypothèse sur la loi bivariable point-bloc $(Y(\underline{x}), Y_v(\underline{x}))$, qui est supposée bigaussienne ou bigamma, de coefficient de corrélation ρ .

Le coefficient de corrélation point-bloc $\rho_n^v(x,v)$ ne dépend plus de x , mais est constant sur tout le volume. Il peut se noter ρ_n^v . Le calcul de ρ_n^v en fonction de ρ se fait selon les hypothèses retenues pour la loi bivariable ponctuelle, et ρ est déterminé lui-même numériquement à partir de la formule :

$$Var(Z_v) = \sum_{n=1}^{+\infty} (\psi_n^v)^2 (\rho_n^v)^2$$

les ψ_n^v étant les coefficients de l'anamorphose sur les blocs.

$Var(Z_v)$ est calculée à partir du variogramme de $Z(x)$. Dans le modèle discret, le programme demande directement la valeur de $Var(Z_v)$. Celle-ci peut être calculée en utilisant les abaques de géostatistique donnant la valeur moyenne d'un variogramme sur un volume : $\bar{\gamma}(v,v)$, sachant que :

$$Var(Z_v) = Var(Z(x)) - \bar{\gamma}(v,v)$$

8.3. EFFET D'INFORMATION. RÉSERVES PRÉVUES ET RECUPÉRÉES

En pratique les teneurs de blocs Z_v ne sont pas connues mais estimées par un estimateur Z_v^* construit à partir des teneurs ponctuelles $Z(x)$.

8.3.1. Réserves prévues

A partir de l'estimateur Z_v^* il est possible de calculer les réserves que l'on prévoit de récupérer si on ne considère que les blocs v dont la teneur Z_v^* est supérieure à Z_c :

$$T_v^*(Z_c) = T_{tot} \cdot \int_{Z_c}^{+\infty} f_v^*(Z_v^*) dZ_v^*$$

$$Q_v^*(Z_c) = T_{tot} \cdot \int_{Z_c}^{+\infty} Z_v^* \cdot f_v^*(Z_v^*) dZ_v^*$$

Ces quantités sont appelées les **prévisions**.

Leur calcul nécessite des hypothèses sur la loi bivariable $(Y(x), Y_v^*(x))$, de la même façon que pour le changement de support $(Y(x), Y_v(x))$. Il s'agit donc d'un deuxième type de changement de support. Le calcul de T_v^* et Q_v^* est similaire à celui de T_v et Q_v et ne pose pas de problème supplémentaire.

8.3.2. Réserves récupérées, effet d'information

Le problème se complique lorsqu'on cherche à calculer non pas les prévisions, mais les quantités effectivement récupérées: en effet les teneurs estimées Z_v^* peuvent être supérieures ou inférieures à la teneur réelle Z_v . On peut ainsi être amené à conserver des blocs pauvres et à éliminer des blocs riches. Les quantités récupérées ont donc toutes les chances d'être différentes des prévisions. Comment alors estimer ces quantités récupérées ?

Les blocs effectivement récupérés sont ceux que l'on a estimé au dessus de la coupure. Le tonnage récupéré $T_v^{**}(Z_c)$ est donc égal au tonnage prévu :

$$T_v^{**}(Z_c) = T_{tot} \cdot \int_{Z_c}^{+\infty} f_v^*(Z_v^*) dZ_v^*$$

Par contre la quantité de métal récupérée sera différente de la quantité de métal prévue. En effet, la teneur réelle d'un bloc récupéré car estimé à $Z_v^* \geq z_c$ peut être supérieure ou inférieure à z_c , du fait que l'estimation est inévitablement entachée d'une certaine erreur. En moyenne, la teneur récupérée est en fait la quantité :

$$h(z) = E[Z_v / Z_v^* = z]$$

$h(z)$ est l'espérance conditionnelle, et $E[Z_v / Z_v^* = z]$ peut se traduire par "teneur moyenne réelle des blocs dont l'estimation a donné la valeur $Z_v^* = z$ ".

La quantité de métal $Q_v^{**}(Z_c)$ effectivement récupérée s'écrit alors :

$$Q_v^{**}(Z_c) = T_{tot} \cdot \int_{Z_c}^{+\infty} h(Z_v^*) f_v^*(Z_v^*) dZ_v^*$$

Les quantités effectivement récupérées sont aussi appelées "réalisations".

Lorsque l'estimateur est moins dispersé que les teneurs réelles, on a en général :

- $h(Z_v^*) > Z_v^*$ si Z_v^* est grand : les teneurs réelles sont en moyenne plus fortes que les teneurs estimées quand la teneur estimée est forte. En d'autres termes, les teneurs fortes sont sous-estimées,

- $h(Z_v^*) < Z_v^*$ si Z_v^* est faible: les teneurs faibles sont surestimées.

On observe le phénomène inverse si l'estimateur est plus dispersé que les teneurs réelles.

Si on exprime la quantité métal comme une fonction du tonnage récupéré T, on peut montrer qu'à T constant :

$$Q(T) \geq Q_v(T) \geq Q_v^{**}(T)$$

La perte de quantité métal lorsqu'on passe d'un support ponctuel à un support de bloc ($Q_v(T) \geq Q(T)$) s'explique par l'effet de support.

La perte de quantité métal récupérée lorsque la sélection s'effectue sur un estimateur, par rapport à la quantité idéale $Q_v(T)$ sur le même support est liée à l'effet d'information.

L'effet d'information joue donc un rôle comparable à celui de l'effet de support. Sa modélisation nécessite des hypothèses sur la loi bivariable $(Y_v(x), Y_v^*(x))$ et est traitée comme un effet de support.

Enfin, il faut signaler que du fait que $h(Z_v^*) \neq Z_v^*$, les réalisations peuvent être supérieures ou inférieures aux prévisions $Q_v^{**}(T) \geq Q_v^*(T)$ ou $Q_v^{**}(T) \leq Q_v^*(T)$. Par contre les réalisations sont toujours inférieures au cas idéal : $Q_v(T) \geq Q_v^{**}(T)$

8.4. ANAMORPHOSE

L'anamorphose consiste à transformer une variable Z quelconque en une variable Y de loi particulière (gauss ou gamma par exemple). L'intérêt de cette transformation réside dans le fait que la loi de la variable Y présente certaines propriétés qui permettent d'appliquer certains modèles, ou d'opérer des calculs qui nécessitent des hypothèses sur le type de loi.

L'anamorphose est utilisée par exemple :

- dans le domaine des simulations : les simulations génèrent en général des lois gaussiennes qu'il faut ensuite transformer dans la loi des données expérimentales,
- dans le domaine des réserves récupérables : les changement de support et le calcul des réserves mettent en jeu des modèles dans le cadre de lois particulières.

La démarche est alors la suivante :

- transformation de la variable brute Z en variable de loi particulière Y (gausienne ou gamma) : anamorphose dite inverse,
- opérations sur la variable Y (simulation, changement de support, calcul de réserves ...),
- retour à la variable brute (anamorphose dite directe).

La fonction d'anamorphose Φ utilisée est définie par :

$$Z(x) = \Phi(Y(x))$$

Sur les histogrammes cumulés (fonctions de répartition $F(Z)$ et $G(Y)$) elle consiste à associer à chaque valeur de z la valeur de y correspondant à la même fréquence cumulée: $F(z) = G(y)$ (fig. 17).

Ceci a pour conséquence de déformer l'histogramme de $Z(x)$ en l'histogramme de $Y(x)$ (par exemple une gaussienne sur la fig. 18).

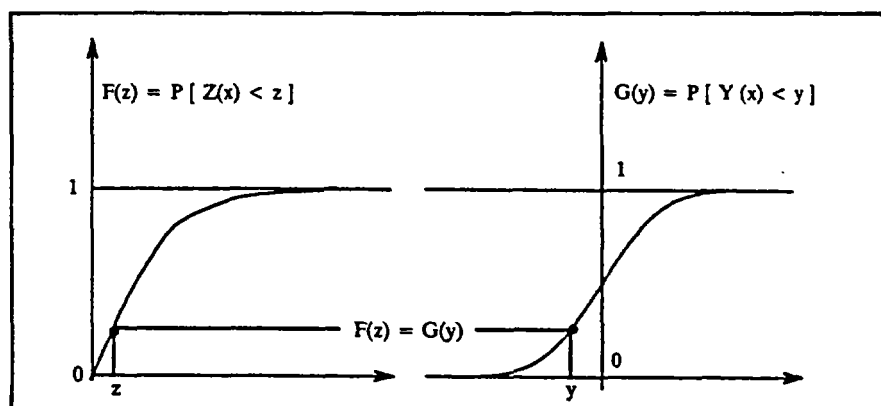


Fig. 17 - Principe de l'anamorphose
(d'après RIVOIRARD, 1991, page 40)

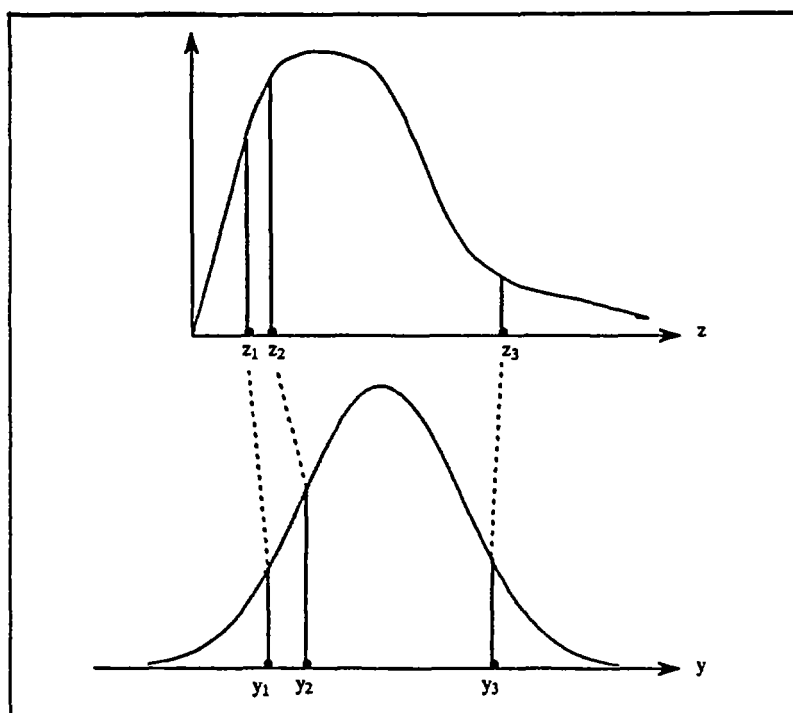


Fig. 18 - Transformation de l'histogramme par anamorphose
(d'après RIVOIRARD, 1991, page 39)

La fonction anamorphose est une fonction non décroissante de Y dont l'aspect est donné par la figure 19.

Dans GEORES, la fonction d'anamorphose Φ peut être déterminée de deux manières :

- anamorphose graphique,
- anamorphose polynômiale.

8.4.1. Anamorphose graphique

L'anamorphose graphique, ou empirique, consiste à discrétiser l'histogramme de Z en une série de points (Z_i, F_i) donnant les fréquences cumulées F_i associées aux Z_i .

A chaque Z_i on associe la variable Y_i qui a la même fréquence cumulée $G_i = G(Y_i)$ (Cf. figure 20). L'anamorphose Φ est alors définie de façon discrète par la série de couples de valeurs (Z_i, Y_i) dans l'intervalle $[Z_{\min}, Z_{\max}]$.

Différentes variantes peuvent être envisagées selon :

- la façon de discrétiser l'histogramme des Z . On peut en effet :
 - . soit discrétiser l'histogramme en classes de largeur constante (histogramme par classes),
 - . soit ranger les données en ordre croissant et construire autant de classes qu'il y a de données différentes (histogramme par classement),
- la façon de calculer les fréquences associées à chaque borne de classe: la fréquence associée peut être la fréquence de la classe précédente, de la classe suivante, ou la fréquence moyenne comme sur la figure 20.

Dans le programme K3RES, l'anamorphose graphique est définie à partir de la fonction $P(Z \leq z) = F(z)$. Elle est caractérisée par 3 types d'informations :

- le pourcentage de données nulles P_0 que l'on associe à la valeur minimale de Z non nulle Z_{\min} ,
- une série de points (Z_i, F_i) , avec $F_i = P(Z \leq z_i)$,
- la valeur maximale Z_{\max} à laquelle on associe la fréquence 1 ($P(Z \leq Z_{\max}) = 1$).

Aux fréquences ainsi définies on associe les variables gaussiennes ou gamma Y_i telles que $G(Y_i) = F_i$.

L'anamorphose obtenue de cette façon présente l'avantage de bien respecter l'histogramme de la variable Z , lors du retour $Y \rightarrow Z$. Par contre, du fait qu'il s'agit d'une méthode empirique, elle n'est pas utilisable si on veut effectuer un changement de support s'appuyant sur les modèles discret ou continu. Le changement de support n'est possible qu'avec le modèle rudimentaire de correction affine de support.

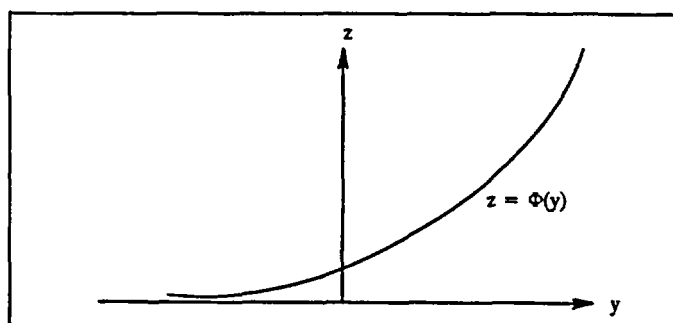


Fig. 19 - Allure de la fonction anamorphose
(d'après RIVOIRARD, 1991, page 40)

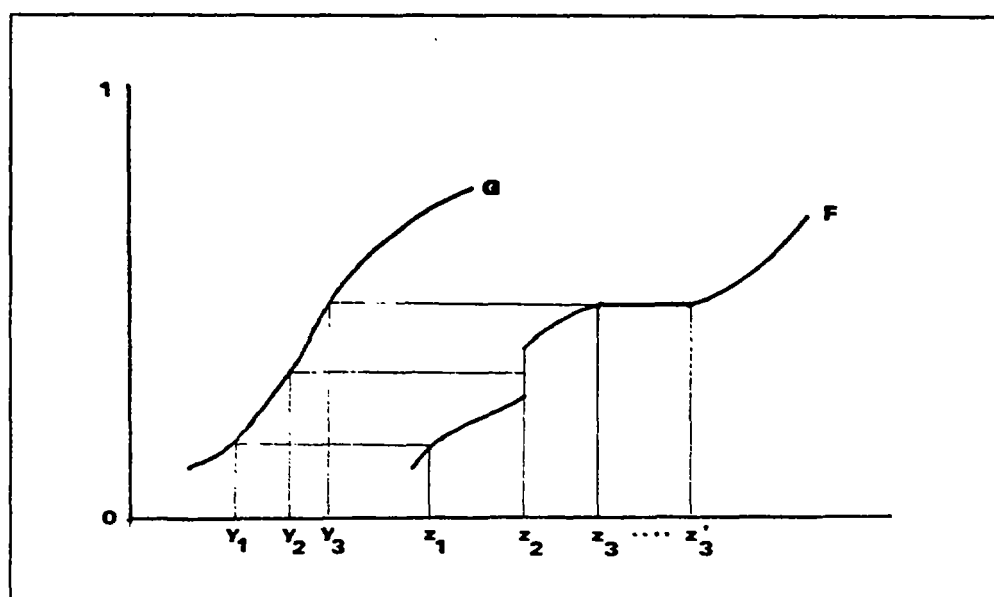


Fig. 20 - Anamorphose graphique
(d'après LANTUEJOUL et RIVOIRARD, 1984)

8.4.2. Anamorphose polynômiale

Cette technique consiste à exprimer la fonction Φ comme un développement illimité en polynômes de degré n affectés de coefficients ψ_n :

$$Z(x) = \Phi(Y(x)) = \sum_0^{+\infty} \psi_n H_n(Y(x))$$

Les ψ_n sont les coefficients d'anamorphose, et les $H_n(Y(x))$ sont des polynômes choisis pour leurs propriétés particulières lorsque $Y(x)$ suit une loi particulière :

- pour la loi de gauss, les polynômes H_n sont les polynômes d'Hermite :

$$H_0(y) = 1; \quad H_1(y) = -y; \quad H_2(y) = (y^2 - 1)/\sqrt{2}; \text{ etc.}$$

- pour la loi gamma il s'agit des polynômes de Laguerre :

$$H_0(y) = 1; \quad H_1(y) = (1 - y/\alpha) \cdot \sqrt{\alpha}; \text{ etc. } (\alpha = \text{paramètre de la loi gamma}).$$

La propriété la plus utile de ces polynômes est leur orthogonalité : $E[H_p(Y(x)) \cdot H_n(Y(x))] = 0$ si $p \neq n$.

Cette orthogonalité permet notamment le calcul des coefficients ψ_n . On a en particulier :

$$\psi_0 = m \quad \text{et} \quad \sum_1^{+\infty} \psi_n^2 = \text{Var}(Z(x)) \quad (m = \text{moyenne de } Z(x))$$

En pratique la nombre de polynômes utilisés n'est pas infini mais limité à un entier k (quelques dizaines) choisi de façon que :

$$\sum_1^k \psi_n^2 \approx \text{Var}(Z(x))$$

L'anamorphose peut également être effectuée sur des variables définies sur un support v non ponctuel. On a alors :

$$Z_v(x) = \Phi_v(Y_v(x)) = \sum_0^{+\infty} \psi_n^v H_n(Y_v(x))$$

Les coefficients ψ_n^v sont liés aux coefficients de l'anamorphose ponctuelle par : $\psi_n^v = \psi_n \cdot T_n$

Les facteurs T_n dépendent du type de modèle de changement de support :

- dans le cas du modèle continu on les calcule à partir du variogramme de $Y(x)$ par discrétisation du bloc v ,
- dans le cas du modèle discret ils sont déterminés à partir de $\text{Var}(Z_v)$ seulement.

RÉFÉRENCES BIBLIOGRAPHIQUES

DAVID M. (1977) - Geostatistical ore reserve estimation, Elsevier.

HU, L.Y. (1988) - Mise en oeuvre du modèle gamma pour l'estimation des distributions spatiales. Thèse de Docteur en Géostatistique, ENS des Mines de Paris, 142 p.

JOURNEL A. (1977) - Géostatistique minière, Thèse de doctorat d'Etat - Centre de Géostatistique - Fontainebleau.

LANTUEJOUL Ch., RIVOIRARD J. (1984) - Une méthode de détermination d'anamorphose. Note N916 - Centre de Géostatistique - Fontainebleau.

LANTUEJOUL Ch. (1990) - Cours de sélectivité. Note C-140 - Centre de Géostatistique - Fontainebleau.

LIAO, H.T. (1991) - Estimation des réserves récupérables de gisements d'or. Comparaison entre krigeage disjonctif et krigeage des indicatrices. Thèse de Docteur en Géostatistique, Université d'Orléans. Editions du BRGM, document n° 202, 198 p.

MATHERON, G. (1982) - La déstructuration des hautes teneurs et le krigeage des indicatrices. Note N.761 - Centre de Géostatistique - Fontainebleau.

RIVOIRARD J. (1991) - Introduction au krigeage disjonctif et à la géostatistique non linéaire. Note C-139 - Centre de Géostatistique - Fontainebleau.

