

Ecole des Mines
de Paris



COURS DE GEOSTATISTIQUE MULTIVARIABLE

Jacques RIVOIRARD

Cours C-17 2

Février 2003

CENTRE DE GÉOSTATISTIQUE

35 rue Saint-Honoré

77305 FONTAINEBLEAU (France)

Tél. : 33-1 64 69 47 81 Fax : 33-1 64 69 47 05

<http://cg.ensmp.fr>

Cours de géostatistique multivariable

Table des matières

0 Introduction	5
0.1 Avertissement	5
0.1.1 Contenu, références	5
0.1.2 Abus de langage commodes	5
0.1.3 Notations et indices	6
0.2 Pourquoi la géostatistique multivariable ?	6
0.3 Exemples de covariables	6
1 Rappels de statistique monovariante	7
1.1 Nuage de corrélation entre 2 variables	7
1.2 Distribution par classes (lois conditionnelles)	7
1.3 Moyennes par classes (régression)	7
1.4 Coefficient de corrélation	7
1.5 Régressions linéaires entre 2 variables	8
1.6 Remarque	9
2 Outils structuraux multivariés	10
2.1 Rappel: le cas monovariante	10
2.1.1 Modèle de FA stationnaire d'ordre 2	10
2.1.2 Modèle de FA intrinsèque	10
2.2 La covariance croisée	11
2.3 Covariance croisée stationnaire	11
2.4 Le modèle multivariable stationnaire d'ordre 2	12
2.4.1 Exercice: dérivée	13
2.5 Variogramme croisé	13
2.5.1 Définition	13
2.5.2 Relation entre variogramme et covariance croisés	14
2.6 Modèle multivariable intrinsèque	15
2.6.1 Définition	15
2.6.2 Variance de Combinaisons Linéaires Autorisées	15
2.6.3 Remarque: la relation de symétrie	17
2.6.4 Caractérisation du modèle dans le cas symétrique	17

2.7	Le pseudo-variogramme croisé	17
2.8	Exercice: Calcul d'une structure bivariable	18
2.9	Support	19
2.9.1	Régularisation.....	19
2.9.2	Co-dispersion	19
2.9.3	Exercice : corrélation à différents supports.....	20
2.10	Structure de combinaisons linéaires.....	21
2.11	Fermeture	21
3	Quelques modèles courants	23
3.1	Le modèle de corrélation intrinsèque	23
3.1.1	Définition	23
3.1.2	Matrice des coefficients	24
3.1.3	Factorisation	25
3.1.4	Exemples	26
3.1.5	Exercice : exemples de modèles de corrélation intrinsèque.....	27
3.2	Le modèle linéaire de corégionalisation.....	28
3.2.1	Composantes gigognes en monovariable : le modèle linéaire de régionalisation.....	28
3.2.2	Cas multivariable: le modèle linéaire de corégionalisation	29
3.2.3	Exemple.....	31
3.2.4	Exercice : exemples de modèles linéaires de corégionalisation.....	32
3.3	Exercice : modèle d'erreur	32
3.4	Exercice : multi-support	33
4	Le cokrigage	34
4.1	Généralités sur le cokrigage d'une ou plusieurs variables	34
4.2	Cokrigage simple (ou cokrigage à moyennes connues)	34
4.2.1	Exercice: cohérence entre valeurs krigées.....	35
4.2.2	Exercice: cohérence entre valeurs cokrigées.....	36
4.2.3	Exercice: cokrigage simple.....	36
4.3	Cokrigage ordinaire (ou cokrigage à moyennes inconnues).....	36
4.3.1	Exercice: cohérence entre valeurs cokrigées.....	37
4.4	Cokrigage intrinsèque.....	38
4.5	Cokrigage à moyennes inconnues mais liées	38
4.5.1	Exercice: filtrage d'erreurs.....	38

4.6 Remarques	39
4.6.1 Variance de cokrigeage	39
4.6.2 Poids de cokrigeage.....	39
5 Simplification du cokrigeage	40
5.1 Remarques préliminaires.....	40
5.1.1 Cohérence des estimations	40
5.1.2 Données redondantes.....	40
5.2 Variables sans corrélation spatiale	41
5.2.1 Exercice: variable composite	41
5.2.2 Exercice: moyennes (ou dérivées) liées	41
5.3 Modèles factorisés ou de type « isofactoriel »	41
5.3.1 Recherche des facteurs	42
5.4 Variables autokrigeables	42
5.4.1 Exercice préliminaire: effet d'écran en cokrigeage.....	42
5.4.2 Autokrigeabilité.....	42
5.5 Modèle à résidu entre 2 variables.....	43
5.5.1 Caractérisation.....	43
5.5.2 Simplification du cokrigeage	44
5.5.3 Krigeage du résidu ou cokrigeage collocalisé.....	44
5.5.4 Exercice : « dislocated cokriging »	44
5.5.5 Exercice: Modèle à résidu et corrélation intrinsèque	45
5.5.6 Exercice: Exemples de modèles à résidu	45
5.5.7 Passage à un modèle à dérive.....	45
5.5.8 Krigeage en dérive externe.....	46
5.5.9 Remarques sur krigeage du résidu et krigeage en dérive externe	47
6 L'analyse krigeante.....	48
6.1 Cas monovariante stationnaire	48
6.2 Cas monovariante intrinsèque	49
6.3 Cas multivariante.....	49
7 Synthèse et autres remarques	51
7.1 Modèles multivariante	51
7.2 Simulation	51
7.3 Simplification du cokrigeage	51

7.4 Voisinage de cokrigeage	52
7.5 Usage du cokrigeage	53
8 Indicatrices	54
8.1 Indicatrice d'un ensemble	54
8.1.1 Variable indicatrice d'un ensemble	54
8.1.2 Modèle probabiliste	54
8.1.3 Krigeage d'une indicatrice	56
8.2 Cas d'une indicatrice et d'une autre variable	56
8.2.1 Cas général	56
8.2.2 Indicatrice et variable simultanément nulles	56
8.2.3 Variables utiles	57
8.3 Indicatrices de plusieurs ensembles	59
8.3.1 Variographie croisée	59
8.3.2 Ensembles indépendants	59
8.3.3 Ensembles disjoints	59
8.3.4 Ensembles emboîtés	60
8.3.5 Remarque : relation entre ensembles disjoints et emboîtés	60
8.3.6 Cokrigeage d'indicatrices	61
8.4 Quelques modèles de base	61
8.4.1 Modèle sans effet de bord	61
8.4.2 Modèle mosaïque à valuations indépendantes	62
8.4.3 Modèle de type diffusif	62
8.4.4 Vers la géostatistique non-linéaire	63
9 Références	64
10 Index	65
Figures	

0 Introduction

0.1 Avertissement

0.1.1 Contenu, références

Ces notes ont été rédigées à l'occasion du cours de géostatistique multivariée donné au Centre de Géostatistique de Fontainebleau, dans le cadre du CFSG (Cycle de Formation Spécialisée en Géostatistique) en 2000-2001 et 2001-2002 ainsi que des formations « hivernales » de 2002. Elles ont vocation à présenter les bases de la géostatistique multivariée (essentiellement la géostatistique linéaire, dans les cas stationnaire et intrinsèque). Pour des développements plus poussés, on pourra se reporter à des ouvrages plus étoffés, par exemple Wackernagel (1993), Wackernagel (1995) et Chilès and Delfiner (1999), où l'on trouvera de nombreuses références.

Le chapitre consacré à la simplification du cokrigage est cependant en grande partie original, de même que le dernier chapitre, qui traite du cas particulier des indicatrices et ouvre la voie à la géostatistique non-linéaire.

0.1.2 Abus de langage commodes

Quelques abus de langage seront d'une grande commodité :

Variables et Fonctions Aléatoires corrélées

Les variables régionalisées étudiées, variables dépendant de la localisation, seront représentées, comme le plus souvent en géostatistique, par un (modèle de) Fonctions Aléatoires (FA en abrégé). Bien que cette distinction soit fondamentale, nous parlerons en général, par commodité, de « variables », même s'il s'agit des FA les représentant, lorsque cela ne prête pas à ambiguïté.

La référence explicite à des FA sera cependant utilisée pour souligner une différence. Ainsi des FA « non corrélées » représenteront des variables qui sont sans corrélation, non seulement au même point, mais également entre points quelconques, c'est-à-dire des variables « sans corrélation spatiale ». Le mot « théorique » se référera aussi au modèle de FA, comme par exemple la variance théorique dans le cas stationnaire (palier du variogramme, différent de la variance expérimentale des échantillons).

Dans le cas de FA intrinsèques, on parlera abusivement de variables sans corrélation spatiale, alors qu'il s'agit de non corrélation entre les accroissements des variables - plus exactement d'ailleurs, entre les accroissements des FA correspondantes...

Structures identiques, structures différentes

Toujours par abus de langage, on dira, si cela ne prête pas à confusion, que deux variables ont « même structure » ou des structures « identiques », si leurs variogrammes, par exemple, sont les mêmes à un facteur multiplicatif près, autrement dit si ils sont proportionnels : ainsi $4 \text{ sph}(h/100)$ et $9 \text{ sph}(h/100)$.

De la même façon, on dira que deux variables ont des structures différentes, si elles ne sont pas proportionnelles, par exemple $4 \text{ sph}(h/100)$ et $4 \text{ pépite}(h) + 4 \text{ sph}(h/100)$.

Remarques sur le mot « intrinsèque »

Enfin on fera attention à la signification de l'adjectif « intrinsèque », qui signifie « propre à l'objet lui-même, indépendant de facteurs externes ».

Le variogramme a été un temps appelé fonction de dispersion intrinsèque, car contrairement à la covariance, il décrit directement la structure spatiale, indépendamment des moyennes et des problèmes posés par l'estimation de ces moyennes. Le modèle de FA dite intrinsèque est précisément caractérisé par ce seul variogramme.

0.1.3 Notations et indices

La multiplication des variables, et en particulier des indices, est une difficulté première lorsqu'on aborde le multivariable. Pour s'y retrouver facilement, nous avons choisi de ne faire figurer en exposant (à ne pas confondre alors avec des puissances) que des indices représentant des composantes d'échelles différentes. Ainsi la 1^{ère} d'un ensemble de variables, $Z_1(x)$, aura un variogramme qui pourra s'écrire : $\gamma_1(h)$ ou $\gamma_{11}(h) = b_{11}^1 \gamma^1(h) + b_{11}^2 \gamma^2(h)$

où l'exposant 1 se référera, par exemple, à une composante de courte portée, et l'exposant 2 à une composante de longue portée.

0.2 Pourquoi la géostatistique multivariable ?

La géostatistique multivariable sert essentiellement à :

- Mettre en évidence les relations structurales entre variables ;
- Améliorer l'estimation d'une variable à l'aide d'autres variables, échantillonnées aux mêmes points (cas « isotopique ») ou non (cas « hétérotopique ») ;
- Améliorer la cohérence entre les estimations de différentes variables ;
- Simuler conjointement plusieurs variables.

0.3 Exemples de covariables

Voici quelques exemples :

- Cotes d'horizons géologiques:
 - toit et mur d'une couche (ex: bedrock et recouvrement), ainsi que la puissance correspondante
 - plusieurs couches
 - cote d'un horizon + données de pendage
 - cote d'un horizon connue aux puits + sismique connue en beaucoup de points
- teneurs en différents métaux, concentrations en différents éléments
- puissance et accumulation-métal d'un gisement 2D, ainsi que leur rapport, la teneur 2D
- différents types de mesure, erreurs
- indicatrices de différents faciès
- etc.

1 Rappels de statistique monovariante

Dans ce chapitre, nous rappelons quelques notions de base sur la corrélation et les régressions entre deux variables.

1.1 Nuage de corrélation entre 2 variables

On ne saurait exagérer l'utilité du nuage de corrélation. Il permet :

- de distinguer des populations différentes, d'identifier facilement des valeurs extrêmes, ou même certains artefacts ;
- de visualiser la relation entre variables.

Attention cependant à la distribution variable d'effectifs pouvant compromettre la lisibilité et l'interprétation d'un nuage de corrélation (une transformation logarithmique, par exemple, peut améliorer la lisibilité de variables à distribution fortement dissymétrique).

1.2 Distribution par classes (lois conditionnelles)

De même que le nuage de corrélation représente la version expérimentale d'une loi statistique bivariable, la distribution des valeurs d'une variable, disons Z_2 , par classe de Z_1 , est la version expérimentale de la loi conditionnelle de Z_2 connaissant $Z_1 = z_1$.

1.3 Moyennes par classes (régression)

La moyenne de Z_2 par classe de Z_1 correspond à la version expérimentale de l'espérance conditionnelle $E[Z_2|Z_1]$ de Z_2 connaissant Z_1 .

1.4 Coefficient de corrélation

Soit deux variables Z_1 et Z_2 de :

- moyennes :

$$m_1 = E[Z_1] \text{ et } m_2 = E[Z_2]$$

- variances :

$$\begin{aligned}\sigma_1^2 &= \text{var } Z_1 = E[(Z_1 - m_1)^2] \\ &= E(Z_1^2) - 2E(Z_1)m_1 + m_1^2 = E(Z_1^2) - m_1^2\end{aligned}$$

$$\text{et } \sigma_2^2$$

- écarts-types :

$$\sigma_1 = \sqrt{\sigma_1^2} \text{ et } \sigma_2$$

- covariance :

$$\begin{aligned}C_{12} &= \text{cov}(Z_1, Z_2) = E[(Z_1 - m_1)(Z_2 - m_2)] \\ &= \dots = E(Z_1 Z_2) - m_1 m_2\end{aligned}$$

Par définition, le coefficient de corrélation est égal à :

$$\rho = \frac{C_{12}}{\sigma_1 \sigma_2}$$

Il vérifie: $-1 \leq \rho \leq 1$. (Mais des incohérences peuvent apparaître - un coefficient de corrélation supérieur à 1, par exemple - si les variables ne sont pas échantillonnées toutes deux en même temps, et si moyennes, variances et covariance ne sont pas calculées sur les mêmes échantillons.)

Le coefficient de corrélation mesure la dépendance linéaire entre variables. Il est nul si les variables sont indépendantes, mais peut être nul sans qu'elles le soient. Il est très sensible aux valeurs extrêmes.

1.5 Régressions linéaires entre 2 variables

Rappelons que la régression linéaire de Z_2 sur Z_1 est le meilleur estimateur linéaire de Z_2 à partir de Z_1 .

$$Z_2^* = aZ_1 + b$$

(meilleur au sens d'estimateur sans biais et de variance d'erreur minimale).

L'erreur commise s'écrit:

$$R = Z_2 - Z_2^* = Z_2 - aZ_1 - b$$

L'estimateur est sans biais si l'erreur est de moyenne nulle, ce qui donne b:

$$E(R) = m_2 - am_1 - b = 0 \Leftrightarrow b = m_2 - am_1$$

d'où l'écriture de l'estimateur à l'aide des variables centrées:

$$Z_2^* - m_2 = a(Z_1 - m_1)$$

ce qui montre au passage que la droite de régression passe par les moyennes.

La variance de l'erreur se développe:

$$\begin{aligned} \text{Var}(R) &= \text{Var}(Z_2) + a^2 \text{Var}(Z_1) - 2a \text{Cov}(Z_1, Z_2) \\ &= \sigma_2^2 + a^2 \sigma_1^2 - 2a C_{12} \end{aligned}$$

C'est une fonction de a, dont le minimum s'obtient en annulant la dérivée par rapport à a, d'où la pente de la régression:

$$a = \frac{\text{Cov}(Z_1, Z_2)}{\text{Var}(Z_1)} = \frac{C_{12}}{\sigma_1^2} = \rho \frac{\sigma_2}{\sigma_1}$$

L'expression de la régression sur variables réduites est facile à mémoriser :

$$\frac{Z_2^* - m_2}{\sigma_2} = \rho \frac{Z_1 - m_1}{\sigma_1}$$

Résidu et variable connue sont non corrélés:

$$\text{Cov}(Z_1, R) = \text{Cov}(Z_1, Z_2 - aZ_1 - b) = C_{12} - a\sigma_1^2 = 0$$

(Comme $E(R) = 0$, on a également: $E(Z_1 R) = 0$: Z_1 et R sont dites "orthogonales")

On peut donc décomposer Z_2 en deux variables non corrélées:

$$Z_2 = a Z_1 + b + R$$

Le carré de la corrélation donne la proportion de variance d'une quelconque des deux variables Z_1 ou Z_2 , expliquée par la dépendance linéaire à l'autre variable. On a en effet par exemple :

$$\frac{Var(Z_2^*)}{Var(Z_2)} = \frac{Var(aZ_1 + b)}{Var(Z_2)} = \frac{a^2 \sigma_1^2}{\sigma_2^2} = \rho^2$$

La valeur de la corrélation peut être trompeuse. Ainsi une corrélation de 0.7 n'explique que 49% de la variance, une corrélation de 0.5, 25%, et une corrélation de 0.1, seulement 1%.

Attention: les régressions linéaires de Z_2 sur Z_1 et de Z_1 sur Z_2 sont deux droites différentes (les pentes ne sont pas inverses l'une de l'autre), se croisant sur les moyennes :

$$Z_2^* - m_2 = \rho \frac{\sigma_2}{\sigma_1} (Z_1 - m_1)$$

$$Z_1^* - m_1 = \rho \frac{\sigma_1}{\sigma_2} (Z_2 - m_2)$$

On remarquera également que les droites de régression n'indiquent pas les "directions principales" du nuage, selon lesquelles la variabilité est maximale ou minimale.

1.6 Remarque

Les outils statistiques, particulièrement les outils simples de visualisation et d'exploration, comme le nuage de corrélation, sont extrêmement utiles.

S'agissant de variables régionalisées, on notera cependant que les statistiques, comme les lois de distribution, le coefficient de corrélation, les régressions, ne sont pas des statistiques « spatiales »: elles ne dépendent pas de la position des échantillons dans l'espace (intervertir les valeurs des variables entre 2 points ne change pas la corrélation par exemple).

De plus ces statistiques n'ont pas de caractère « intrinsèque », au sens où elles dépendent du support (et du champ) sur lequel est définie la variable : ainsi le coefficient de corrélation entre 2 variables ne mesure en général la dépendance linéaire entre ces variables qu'au support utilisé, non à d'autres supports, ce qui restreint leur signification.

Enfin on notera que l'absence de corrélation entre variables au même point x (par exemple entre le résidu d'une régression et la variable conditionnante) n'entraîne pas l'absence de corrélation « spatiale », i.e. entre points x et y différents.

2 Outils structuraux multivariés

2.1 Rappel: le cas monovarié

On considère une variable (représentée par une FA) $Z(x)$. En géostatistique linéaire, on s'intéresse aux seuls moments d'ordre 1 et 2 (FA dite d'ordre 2):

$m(x) = E[Z(x)]$ moyenne au point x (appelée en général dérive lorsqu'elle dépend de x)

$E[Z(x)Z(y)]$ covariance non-centrée entre $Z(x)$ et $Z(y)$

$Cov[Z(x), Z(y)] = E[Z(x)-m(x)][Z(y)-m(y)] = E[Z(x)Z(y)] - m(x)m(y)$ covariance (habituelle, centrée) entre $Z(x)$ et $Z(y)$.

On fait alors des hypothèses de stationnarité, i.e. invariance par translation, permettant l'inférence de tels moments à partir d'une réalisation unique.

2.1.1 Modèle de FA stationnaire d'ordre 2

Il est défini par :

- moyenne constante : $E[Z(x)] = m$

- covariance: $Cov[Z(x), Z(x+h)] = E[Z(x)-m][Z(x+h)-m] = C(h)$ fonction de la distance h entre les points

- variance constante: $var Z(x) = E\{[Z(x)-m]^2\} = C(0)$

On a $|C(h)| \leq C(0)$ d'où :

$\rho(h) = C(h)/C(0)$, corrélogramme, ou corrélation entre $Z(x)$ et $Z(x+h)$.

La covariance permet de calculer la variance de toute combinaison linéaire :

$$Z = \sum_{\alpha} \lambda_{\alpha} Z(x_{\alpha}) = \sum_{\alpha} \lambda_{\alpha} Z_{\alpha}$$

de $Z(x)$ en des points quelconques:

$$E(Z) = m \sum_{\alpha} \lambda_{\alpha}$$

$$\begin{aligned} Var(Z) &= \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} Cov(Z_{\alpha}, Z_{\beta}) \\ &= \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} C(x_{\beta} - x_{\alpha}) \geq 0 \end{aligned}$$

2.1.2 Modèle de FA intrinsèque

Il est défini par des accroissements $Z(x+h) - Z(x)$, d'espérance nulle :

$$E[Z(x+h) - Z(x)] = 0$$

et de variance ne dépendant que de la distance h entre les points :

$$0.5 E\{[Z(x+h) - Z(x)]^2\} = \gamma(h) \text{ (variogramme)}$$

Ceci permet le calcul de l'espérance et de la variance de toute Combinaison Linéaire Autorisée (CLA), autrement dit, de toute combinaison:

$$\sum_{\alpha} \lambda_{\alpha} Z(x_{\alpha}) = \sum_{\alpha} \lambda_{\alpha} Z_{\alpha}$$

satisfaisant à $\sum_{\alpha} \lambda_{\alpha} = 0$:

$$E \left[\sum_{\alpha} \lambda_{\alpha} Z_{\alpha} \right] = 0$$

$$Var \left[\sum_{\alpha} \lambda_{\alpha} Z_{\alpha} \right] = - \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} \gamma(x_{\beta} - x_{\alpha})$$

Une Fonction Aléatoire stationnaire est aussi intrinsèque, et on a :

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

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

Le palier du variogramme coïncide avec la variance théorique $C(0)$, représentant, dans le modèle, la variance de $Z(x)$ dans un grand champ (et qui peut être assez différente de la variance expérimentale des échantillons).

Mais inversement, une Fonction Aléatoire intrinsèque n'est pas nécessairement stationnaire (cas d'un variogramme sans palier).

En géostatistique (linéaire) multivariable, ces diverses caractéristiques vont être complétées par les structures, variogrammes ou covariances, « croisées » entre variables, et éventuellement par les relations pouvant exister entre les moyennes.

2.2 La covariance croisée

Par définition, la covariance croisée entre 2 FA $Z_1(x)$ et $Z_2(x)$ est:

$$Cov[Z_1(x), Z_2(y)]$$

pour tout couple de points (x, y) .

Il est important de bien la distinguer de la covariance en un même point :

$$Cov[Z_1(x), Z_2(x)]$$

Si la covariance croisée est nulle pour tous x et y , on dira que les FA sont sans corrélation, ou encore que les variables sont sans corrélation spatiale, propriété beaucoup plus forte que l'absence de corrélation entre variables au même point.

Dans certains cas (notamment les indicatrices ou le cas non-stationnaire), il est avantageux de considérer la covariance croisée non-centrée :

$$E[Z_1(x)Z_2(y)]$$

2.3 Covariance croisée stationnaire

Dans ce cas la covariance croisée est invariante par translation, i.e. ne dépend que de la distance h entre les points x et $x+h$. D'où :

- covariance croisée non-centrée stationnaire:

$$E[Z_1(x)Z_2(x+h)] = K_{12}(h)$$

laquelle peut s'estimer sans biais à partir des données de Z_1 et Z_2 distantes de h ,

- ou covariance croisée (habituelle, centrée) stationnaire:

$$Cov[Z_1(x), Z_2(x+h)] = C_{12}(h)$$

dont l'estimation est délicate.

En effet, celle-ci s'écrit :

$$E[(Z_1(x) - E[Z_1(x)])(Z_2(x+h) - E[Z_2(x+h)])] = E[Z_1(x)Z_2(x+h)] - E[Z_1(x)]E[Z_2(x+h)]$$

Son estimation à partir des données nécessite les moyennes $E[Z_1(x)]$ et $E[Z_2(x+h)]$. Si celles-ci doivent faire l'objet d'une estimation, on sera, dans les cas les plus simples, conduit à les supposer stationnaires : $E[Z_1(x)] = m_1$ et $E[Z_2(x+h)] = m_2$ (voir plus loin le modèle multivariable stationnaire). On peut alors utiliser une même estimation globale de m_1 et m_2 pour toutes les distances h (covariance dite ergodique) ou une estimation basée sur les couples $(Z_1(x), Z_2(x+h))$ effectivement utilisés pour la distance h (covariance dite non-ergodique). Dans le premier cas, et plus encore dans le second, se posent cependant de sérieux problèmes de biais (venant de la difficulté à séparer, dans les variables, la partie « moyenne » et la partie « écart par rapport à la moyenne »).

On peut également définir la corrélation croisée, ou corrélogramme croisé :

$$\rho_{12}(h) = corr[Z_1(x), Z_2(x+h)] = \frac{C_{12}(h)}{\sigma_1 \sigma_2}$$

laquelle nécessite moyennes et variances de $Z_1(x)$ et $Z_2(x+h)$. Ici encore on peut distinguer les corrélations croisées ergodique et non ergodique, cette dernière (basée sur les moyennes et écarts-types des couples utilisés pour la distance h) assurant par construction l'inégalité $|\rho_{12}(h)| \leq 1$, mais avec des problèmes de biais encore plus redoutables.

La covariance croisée généralise la covariance simple (faire $Z_2(x) = Z_1(x)$). Elle n'a cependant pas de raison d'être positive, maximale, ou minimale, en $h = 0$, et n'est pas nécessairement paire, $C_{12}(h) = C_{21}(-h)$ pouvant différer de $C_{21}(h) = C_{12}(-h)$. Elle peut en particulier montrer une corrélation maximale pour un vecteur distance non nul, et indiquer par exemple un effet de retard, ou de décalage, entre les variables, selon cette direction. Ainsi la covariance croisée entre une FA stationnaire $Z_1(x)$ et $Z_2(x) = Z_1(x-t)$ est maximale pour une distance égale au décalage entre les variables :

$$C_{12}(h) = Cov[Z_1(x), Z_1(x+h-t)] = C_1(h-t)$$

2.4 Le modèle multivariable stationnaire d'ordre 2

Un ensemble de Fonctions Aléatoires $(Z_1(x), \dots, Z_i(x), \dots, Z_p(x))$ est stationnaire d'ordre 2 si tous les moments d'ordre 1 et 2 sont invariants par translation:

- moyennes constantes :

$$E[Z_i(x)] = m_i$$

- covariances ne dépendant que du vecteur distance entre points, par exemple pour les covariances centrées:

$$\text{Cov}[Z_i(x), Z_j(x+h)] = C_{ij}(h)$$

Ceci implique la stationnarité d'ordre 2 de chaque FA (moyennes et covariances simples $C_i(h) = C_{ii}(h)$ pour $j = i$), et doit être complété par la stationnarité des covariances croisées ($i \neq j$).

Dans un tel modèle, toute combinaison linéaire des $(Z_1(x), \dots, Z_i(x), \dots, Z_p(x))$, en des points quelconques :

$$Z = \sum_i \sum_\alpha \lambda_{i\alpha} Z_i(x_\alpha) = \sum_i \sum_\alpha \lambda_{i\alpha} Z_{i\alpha}$$

a pour moyenne et variance :

$$E(Z) = \sum_i \sum_\alpha \lambda_{i\alpha} m_i$$

$$\begin{aligned} \text{Var}(Z) &= \sum_i \sum_j \sum_\alpha \sum_\beta \lambda_{i\alpha} \lambda_{j\beta} \text{Cov}(Z_{i\alpha}, Z_{j\beta}) \\ &= \sum_i \sum_j \sum_\alpha \sum_\beta \lambda_{i\alpha} \lambda_{j\beta} C_{ij}(x_\beta - x_\alpha) \geq 0 \end{aligned}$$

On notera que la variance met en jeu simultanément covariances simples et croisées des variables.

2.4.1 Exercice: dérivée

Soit $Z(x)$ une FA stationnaire dérivable de covariance $C(h)$.

Montrer que la covariance croisée entre Z et Z' est impaire, nulle en $h=0$:

$$\text{Cov}[Z(x), Z'(x+h)] = C'(h)$$

et que la covariance de $Z'(x)$ s'écrit:

$$\text{Cov}[Z'(x), Z'(x+h)] = -C''(h)$$

2.5 Variogramme croisé

2.5.1 Définition

Soit 2 Fonctions Aléatoires $Z_1(x)$ et $Z_2(x)$ intrinsèques (accroissements $Z_1(x+h) - Z_1(x)$ et $Z_2(x+h) - Z_2(x)$ d'espérance nulle et de semi-variance stationnaire, égale au variogramme de chaque variable). On définit le variogramme croisé par la demi-covariance des accroissements entre x et $x+h$, supposée stationnaire (invariante par translation):

$$\gamma_{12}(h) = \frac{1}{2} E [Z_1(x+h) - Z_1(x)][Z_2(x+h) - Z_2(x)]$$

Son estimation nécessite la connaissance, en des points distants de h (à la tolérance habituelle près), de chacune des variables.

On en déduit que :

$$\frac{\gamma_{12}(h)}{\sqrt{\gamma_1(h)\gamma_2(h)}}$$

représente la corrélation entre accroissements des deux variables.

Par ailleurs le rapport :

$$\frac{\gamma_{12}(h)}{\gamma_1(h)}$$

représente la pente de la régression linéaire

$$[Z_2(x+h) - Z_2(x)]^* = a [Z_1(x+h) - Z_1(x)]$$

de $Z_2(x+h) - Z_2(x)$ connaissant $Z_1(x+h) - Z_1(x)$.

Le variogramme croisé généralise le variogramme simple (faire $Z_2(x) = Z_1(x)$), mais peut présenter des formes beaucoup plus variées. Il peut être négatif (ex: substitution d'un métal ou élément par un autre). Il est nul en 0, et est symétrique :

$$\gamma_{12}(0) = 0$$

$$\gamma_{12}(h) = \gamma_{12}(-h)$$

2.5.2 Relation entre variogramme et covariance croisés

Dans le modèle multivariable stationnaire de FA, on a :

$$\gamma_{12}(h) = K_{12}(0) - \frac{K_{12}(h) + K_{12}(-h)}{2}$$

ou

$$\gamma_{12}(h) = C_{12}(0) - \frac{C_{12}(h) + C_{12}(-h)}{2}$$

Autrement dit, le variogramme croisé ne « voit » que la partie paire de la covariance :

$$C_{12}(h) = \frac{C_{12}(h) + C_{12}(-h)}{2} + \frac{C_{12}(h) - C_{12}(-h)}{2}$$

= partie paire + partie impaire

La covariance croisée est donc un outil plus riche que le variogramme croisé; en contrepartie, elle nécessite des hypothèses de stationnarité plus fortes.

On note en outre:

$$\text{palier du variogramme croisé: } \gamma_{12}(\infty) = C_{12}(0) = \text{Cov}[Z_1(x), Z_2(x)]$$

$$C_{12}(h) \equiv 0 \quad \forall h \Rightarrow \gamma_{12}(h) \equiv 0 \quad \forall h$$

2.6 Modèle multivariable intrinsèque

Cette section est délicate. On commencera par une définition générale du modèle, assez lourde, et on finira par une caractérisation usuelle plus simple, mais limitée au seul cas symétrique.

2.6.1 Définition

De façon générale, on peut dire qu'un ensemble de Fonctions Aléatoires ($Z_1(x), \dots, Z_i(x), \dots, Z_p(x)$) est intrinsèque si tous les accroissements des FA sont d'espérance nulle et si tous leurs moments d'ordre 2 (simples et croisés) sont invariants par translation (stationnaires). On doit donc avoir:

$$E[Z_i(x+h) - Z_i(x)] = 0 \quad \text{pour tous } h \text{ et } i$$

$$\text{et} \quad \text{Cov}[Z_i(x+h) - Z_i(x), Z_j(x+l+h') - Z_j(x+l)] \quad [1]$$

$$\text{soit} \quad E\{[Z_i(x+h) - Z_i(x)][Z_j(x+l+h') - Z_j(x+l)]\}$$

ne dépendant que de $h, h',$ et l , non de x .

On retrouve donc en particulier:

- pour $l = 0, h = h', i = j$, le variogramme:

$$\gamma_i(h) = \gamma_{ii}(h) = \frac{1}{2} E\left([Z_i(x+h) - Z_i(x)]^2\right)$$

de sorte que chaque FA prise individuellement est intrinsèque;

- pour $l = 0, h = h', i \neq j$, le variogramme croisé:

$$\gamma_{ij}(h) = \frac{1}{2} E[Z_i(x+h) - Z_i(x)][Z_j(x+h) - Z_j(x)]$$

Cependant, en général, variogrammes simples et croisés ne suffisent pas à calculer une quantité comme [1]. A titre d'exemple, un modèle multivariable stationnaire de FA est également un modèle multivariable intrinsèque, et [1], qui s'écrit:

$$C_{ij}(l+h'-h) + C_{ij}(l) - C_{ij}(l+h') - C_{ij}(l-h)$$

peut s'écrire également en fonction de $\gamma_{ij}(h)$ si $C_{ij}(h)$ est paire:

$$\gamma_{ij}(l+h') + \gamma_{ij}(l-h) - \gamma_{ij}(l+h'-h) - \gamma_{ij}(l)$$

mais non, en général, si $C_{ij}(h)$ n'est pas paire.

2.6.2 Variance de Combinaisons Linéaires Autorisées

De même, c'est seulement sous certaines conditions que variogrammes simples et croisés permettent de calculer la variance des Combinaisons Linéaires Autorisées, autrement dit de combinaisons :

$$Z = \sum_i \sum_\alpha \lambda_{i\alpha} Z_{i\alpha}$$

satisfaisant à:

$$\sum_{\alpha} \lambda_{i\alpha} = 0 \quad \forall i$$

ce qui implique d'ailleurs $E(Z)=0$ (noter que chacune des variables n'est pas nécessairement considérée en chacun des points x_{α} : il suffit de prendre $\lambda_{i\alpha} = 0$).

Ecrivons en effet, avec un point x_0 quelconque:

$$Z = \sum_i \sum_{\alpha} \lambda_{i\alpha} Z_{i\alpha} = \sum_i \sum_{\alpha} \lambda_{i\alpha} (Z_{i\alpha} - Z_{i0})$$

d'où:

$$Var[Z] = \sum_i \sum_j \sum_{\alpha} \sum_{\beta} \lambda_{i\alpha} \lambda_{j\beta} Cov(Z_{i\alpha} - Z_{i0}, Z_{j\beta} - Z_{j0}) \quad [2]$$

qui fait apparaître les termes:

$$Cov(Z_{i\alpha} - Z_{i0}, Z_{j\beta} - Z_{j0})$$

Or:

$$E[(Z_{i\beta} - Z_{i\alpha})(Z_{j\beta} - Z_{j\alpha})] = E[(Z_{i\beta} - Z_{i0}) - (Z_{i\alpha} - Z_{i0})][(Z_{j\beta} - Z_{j0}) - (Z_{j\alpha} - Z_{j0})]$$

soit

$$2\gamma_{ij\alpha\beta} = 2\gamma_{ij0\beta} + 2\gamma_{ij0\alpha} - Cov(Z_{i\alpha} - Z_{i0}, Z_{j\beta} - Z_{j0}) - Cov(Z_{i\beta} - Z_{i0}, Z_{j\alpha} - Z_{j0})$$

ou encore

$$Cov(Z_{i\alpha} - Z_{i0}, Z_{j\beta} - Z_{j0}) + Cov(Z_{i\beta} - Z_{i0}, Z_{j\alpha} - Z_{j0}) = 2\gamma_{ij0\beta} + 2\gamma_{ij0\alpha} - 2\gamma_{ij\alpha\beta}$$

Si les $\lambda_{i\alpha} \lambda_{j\beta} Cov(Z_{i\alpha} - Z_{i0}, Z_{j\beta} - Z_{j0})$ ne changent pas par permutation de i et j , ou par permutation de α et β , on aura:

$$Var[Z] = - \sum_i \sum_j \sum_{\alpha} \sum_{\beta} \lambda_{i\alpha} \lambda_{j\beta} \gamma_{ij} (x_{\beta} - x_{\alpha})$$

(même formule que dans le cas stationnaire, en remplaçant les C_{ij} par $-\gamma_{ij}$)

Ce sera notamment le cas si :

- Soit $\lambda_{i\alpha} = k_i \lambda_{\alpha}$ (poids similaires pour toutes les variables) et donc: $\lambda_{i\alpha} \lambda_{j\beta} = \lambda_{i\beta} \lambda_{j\alpha}$

- Soit $Cov(Z_{i\alpha} - Z_{i0}, Z_{j\beta} - Z_{j0}) = Cov(Z_{i\beta} - Z_{i0}, Z_{j\alpha} - Z_{j0})$ [3]

(alors égal à $\gamma_{ij0\beta} + \gamma_{ij0\alpha} - \gamma_{ij\alpha\beta}$).

Dans le cas général, c'est seulement sous cette dernière hypothèse de symétrie entre points quelconques que variogrammes simples et croisés permettent le calcul de la variance d'une CLA.

2.6.3 Remarque: la relation de symétrie

Cette relation de symétrie à l'aide de 3 points est équivalente à la relation de symétrie suivante à partir de 4 points:

$$\begin{aligned} & Cov[Z_i(x+h) - Z_i(x), Z_j(x+l+h') - Z_j(x+l)] \\ &= Cov[Z_j(x+h) - Z_j(x), Z_i(x+l+h') - Z_i(x+l)] \end{aligned} \quad [4]$$

Ces covariances, caractérisant le modèle multivariable intrinsèque, sont alors égales à:

$$\begin{aligned} & Cov[Z_i(x+h) - Z_i(x), Z_j(x+l+h') - Z_j(x) + Z_j(x) - Z_j(x+l)] \\ &= Cov[Z_i(x+h) - Z_i(x), Z_j(x+l+h') - Z_j(x)] \\ &\quad - Cov[Z_i(x+h) - Z_i(x), Z_j(x+l) - Z_j(x)] \end{aligned}$$

soit: $\gamma_{ij}(l+h') + \gamma_{ij}(l-h) - \gamma_{ij}(l+h'-h) - \gamma_{ij}(l)$

Une telle relation de symétrie, on l'a vu, est nécessairement satisfaite dans le cas d'un modèle multivariable stationnaire si les covariances croisées sont paires, mais non si elles ne sont pas symétriques. On reconnaît là l'impuissance du variogramme croisé à rendre compte d'une dissymétrie dans la structure croisée.

On retiendra que variogrammes simples et croisés ne caractérisent un modèle multivariable intrinsèque que dans le cas symétrique, d'où la caractérisation suivante. (Dans le cas non stationnaire plus général, il faut faire appel à un autre outil, la covariance croisée généralisée, difficile à appréhender, et hors du présent propos.)

2.6.4 Caractérisation du modèle dans le cas symétrique

Soit un ensemble de Fonctions Aléatoires ($Z_1(x), \dots, Z_i(x), \dots, Z_p(x)$) intrinsèque satisfaisant les relations de symétrie [3] ou [4] pour tous points. Ce modèle est entièrement caractérisé par:

- des espérances d'accroissements nulles:

$$E[Z_i(x+h) - Z_i(x)] = 0$$

- les variogrammes simples:

$$\gamma_i(h) = \gamma_{ii}(h) = \frac{1}{2} E\left([Z_i(x+h) - Z_i(x)]^2\right)$$

- les variogrammes croisés:

$$\gamma_{ij}(h) = \frac{1}{2} E[Z_i(x+h) - Z_i(x)][Z_j(x+h) - Z_j(x)]$$

Variogrammes simples et croisés permettent alors de calculer la variance de toute combinaison linéaire autorisée, selon [2].

2.7 Le pseudo-variogramme croisé

Définition:

$$\psi_{12}(h) = \frac{1}{2} E [Z_2(x+h) - Z_1(x)]^2$$

Il peut se calculer entre variables connues en des points différents (contrairement à $\gamma_{12}(h)$).

Il exige cependant:

- que la différence entre variables ait un sens (unités en particulier), par exemple une même variable mesurée sous des conditions différentes ;
- que $Z_2(x) - Z_1(x)$ (ou plus exactement l'espérance de son carré) soit stationnaire, car

$$\psi_{12}(0) = \frac{1}{2} E [Z_2(x) - Z_1(x)]^2$$

ne dépend pas de x.

On peut avoir : $\psi_{12}(h) \neq \psi_{12}(-h)$, sa partie paire étant rattachée au variogramme croisé :

$$\frac{\psi_{12}(h) + \psi_{12}(-h)}{2} = \frac{1}{2} E [Z_2(x) - Z_1(x)]^2 + \gamma_{12}(h)$$

Dans le cas stationnaire, on a:

$$\psi_{12}(h) = \frac{1}{2} [K_1(0) + K_2(0)] - K_{12}(h)$$

où $K_{12}(h) = E [Z_1(x)Z_2(x+h)]$ est la covariance croisée non-centrée.

Remarque : le pseudo-variogramme croisé, bien que non symétrique en général, ne correspond pas à la covariance croisée généralisée évoquée à la fin de la section 2.6 (Künsch et al. 1997).

2.8 Exercice: Calcul d'une structure bivariable

On considère les teneurs Z_1 et Z_2 en deux métaux (en %), tous les mètres le long d'un sondage:

Z_1	Z_2
0	0
1	0
0	0
0	1
0	1
0	0
0	0
0	0
1	0
0	1
0	1

Moyenne, moyenne des carrés, variance, écart-type, de chaque variable, ainsi que coefficient de corrélation?

Calculer et tracer, jusqu'à $h = 4$:

- le variogramme simple de chaque variable;

- le variogramme croisé;
- la covariance non-centrée de chaque variable;
- la covariance de chaque variable.

Calculer et tracer, de $h = -4$ jusqu'à $h = 4$:

- la covariance croisée non-centrée;
- la covariance croisée;
- la corrélation croisée.

Que devient les structures croisées lorsque la dernière valeur de Z_2 est égale à 3 au lieu de 1 ? lorsqu'elle est inconnue ?

Remarque : Dans le cas hétérotopique (variables connues en des points éventuellement différents), des incohérences peuvent apparaître entre les structures expérimentales simples et croisées, lorsque celles-ci sont calculées sur des ensembles de points différents.

2.9 Support

Les formules de régularisation et de variances du cas monovarié se généralisent au cas multivarié, à l'aide des covariances et variogrammes croisés (cas stationnaire et intrinsèque respectivement).

2.9.1 Régularisation

Covariance croisée régularisée sur support v :

$$\begin{aligned} C_{ij,v}(h) &= \text{Cov}[Z_i(v), Z_j(v_h)] = \frac{1}{|v|^2} \int_v \int_{v_h} C_{ij}(x-y) dx dy \\ &= \bar{C}_{ij}(v, v_h) \end{aligned}$$

Variogramme croisé régularisé sur support v :

$$\gamma_{ij,v}(h) = \bar{\gamma}_{ij}(v, v_h) - \bar{\gamma}_{ij}(v, v)$$

2.9.2 Co-dispersion

Variance de dispersion de v dans V :

$$\text{Var}[Z_i(\underline{v}) - Z_i(V)] = \bar{\gamma}_i(V, V) - \bar{\gamma}_i(v, v)$$

\underline{v} parcourant uniformément les différents volumes de même support v partitionnant le champ V (soit $\bar{C}_i(v, v)$ dans le cas stationnaire sur un grand champ).

Covariance de dispersion de v dans V :

$$\text{Cov}[Z_i(\underline{v}) - Z_i(V), Z_j(\underline{v}) - Z_j(V)] = \bar{\gamma}_{ij}(V, V) - \bar{\gamma}_{ij}(v, v)$$

(soit $\bar{C}_{ij}(v, v)$ dans le cas stationnaire sur un grand champ).

Corrélation de v dans V :

$$\rho_{ij}(v|V) = \frac{\bar{\gamma}_{ij}(V, V) - \bar{\gamma}_{ij}(v, v)}{\sqrt{(\bar{\gamma}_i(V, V) - \bar{\gamma}_i(v, v))(\bar{\gamma}_j(V, V) - \bar{\gamma}_j(v, v))}}$$

(soit $\frac{\bar{C}_{ij}(v, v)}{\sqrt{C_i(v, v)C_j(v, v)}}$ dans le cas stationnaire sur un grand champ).

La régression linéaire:

$$[Z_j(\underline{v}) - Z_j(V)]^* = a[Z_i(\underline{v}) - Z_i(V)]$$

de $Z_j(v)$ sur $Z_i(v)$, pour v parcourant V , a pour pente:

$$\frac{\bar{\gamma}_{ij}(V, V) - \bar{\gamma}_{ij}(v, v)}{\bar{\gamma}_i(V, V) - \bar{\gamma}_i(v, v)}$$

(soit $\frac{\bar{C}_{ij}(v, v)}{C_i(v, v)}$ dans le cas stationnaire sur un grand champ).

2.9.3 Exercice : corrélation à différents supports

On considère un gisement de plomb-zinc-argent dont les teneurs d'échantillons $Z_1(x)$, $Z_2(x)$ et $Z_3(x)$ ont pour moyennes :

$$m_1 = 13.2 \text{ (\%)}$$

$$m_2 = 10.3 \text{ (\%)}$$

$$m_3 = 3.72 \text{ oz/ton (1 once par ton } \approx 30 \text{ g/t)}$$

et pour structures simples et croisées :

$$\gamma_1(h) = 11 \text{ Pep}(h) + 39 \text{ Sph}(h / 60)$$

$$\gamma_2(h) = 9 \text{ Pep}(h) + 15 \text{ Sph}(h / 60)$$

$$\gamma_3(h) = 1.1 \text{ Pep}(h) + 1.8 \text{ Sph}(h / 60)$$

$$\gamma_{12}(h) = 14.5 \text{ Sph}(h / 60)$$

$$\gamma_{13}(h) = 5 \text{ Sph}(h / 60)$$

$$\gamma_{23}(h) = 3.8 \text{ Sph}(h / 60)$$

(d'après P. Dowd, cité dans Mining geostatistics, Journel et Huijbregts, Academic Press, 1978)

Que valent, dans le modèle, les coefficients de corrélation entre les différentes teneurs d'échantillons ? (On suppose le champ grand par rapport à la portée.)

Que valent approximativement les coefficients de corrélation entre les différentes teneurs pour de petits blocs ?

2.10 Structure de combinaisons linéaires

La connaissance des variogrammes (ou covariances) simples et croisées permet de calculer la structure simple d'une nouvelle variable obtenue par combinaison linéaire de variables, ou la structure croisées entre de telles combinaisons.

Soit par exemple: $H(x) = Z_2(x) - Z_1(x)$ la puissance d'une couche entre toit Z_2 et mur Z_1 (de moyenne $m_H = m_2 - m_1$ dans le cas stationnaire).

La structure de H s'écrit, en covariance:

$$\begin{aligned} C_H(h) &= \text{Cov}[H(x), H(x+h)] = \text{Cov}[Z_2(x) - Z_1(x), Z_2(x+h) - Z_1(x+h)] \\ &= C_1(h) + C_2(h) - C_{12}(h) - C_{12}(-h) \end{aligned}$$

ou en variogramme:

$$\begin{aligned} \gamma_H(h) &= \frac{1}{2} E[H(x+h) - H(x)]^2 \\ &= \gamma_1(h) + \gamma_2(h) - 2\gamma_{12}(h) \end{aligned}$$

La structure croisée de H avec Z_1 par exemple vaut :

$$\text{Cov}[H(x), Z_1(x+h)] = C_{12}(-h) - C_1(h)$$

ou en variogramme:

$$\gamma_{12}(h) - \gamma_1(h)$$

Autre exemple : soit $Z(x) = m + Z_1(x) + Z_2(x)$. On a:

$$\begin{aligned} m_Z &= m + m_1 + m_2 \\ C_Z(h) &= C_1(h) + C_2(h) + C_{12}(h) + C_{12}(-h) \\ \gamma_Z(h) &= \gamma_1(h) + \gamma_2(h) + 2\gamma_{12}(h) \end{aligned}$$

Si Z_1 et Z_2 sont de moyenne nulle et sans corrélation spatiale, on obtient :

$$\begin{aligned} m_Z &= m \\ C_Z(h) &= C_1(h) + C_2(h) \\ \gamma_Z(h) &= \gamma_1(h) + \gamma_2(h) \end{aligned}$$

Inversement une structure gigogne dans le cas monovarié (superposition de plusieurs composantes structurales, par exemple à différentes échelles) peut être interprétée comme la décomposition de la variable en composantes à différentes échelles. Ce dernier exemple contient en germe le modèle linéaire de corégionalisation étudié plus loin.

2.11 Fermeture

Soit la relation de fermeture (somme des variables constante):

$$Z(x) = \sum_i Z_i(x) = 1$$

avec alors:

$$\sum_i m_i = 1$$

Croisons $Z_i(x)$ avec la constante $Z(x)$. On déduit des covariances :

$$\sum_{j \neq i} C_{ij}(h) = -C_i(h)$$

qui est négatif pour $h = 0$ (opposé de la variance $C_i(0)$). Quant aux variogrammes :

$$\sum_{j \neq i} \gamma_{ij}(h) = -\gamma_i(h) \leq 0$$

Ainsi, la relation de fermeture est responsable de valeurs nécessairement négatives sur au moins quelques unes des covariances et variogrammes croisés.

3 Quelques modèles courants

Les structures croisées peuvent présenter des allures beaucoup plus variées que les structures simples dont elles sont une généralisation (ainsi des combinaisons linéaires, à coefficients positifs et négatifs, de structures de base pourront représenter des structures croisées, dans le modèle linéaire de corégionalisation). Cependant les structures croisées entre variables ne sont pas indépendantes (et ne doivent donc pas être modélisées indépendamment) des structures simples de ces variables.

On a vu en effet que les variances de combinaisons linéaires se développaient, en fonction des covariances simples et croisées dans le cas stationnaire, et (sous certaines conditions) en fonction des variogrammes simples et croisés dans le cas intrinsèque. De telles variances devant être positives ou nulles, la modélisation des structures croisées doit se faire conjointement avec les structures simples, par un modèle multivariable autorisé. Voici quelques modèles d'emploi courant. D'autres modèles intéressants seront décrits dans le chapitre 5 sur la simplification du cokrigage.

3.1 Le modèle de corrélation intrinsèque

3.1.1 Définition

C'est le plus simple de tous les modèles : toutes les structures, simples et croisées, y sont identiques (proportionnelles) :

$$\gamma_{ij}(h) = b_{ij} \gamma(h) \quad \forall i, j$$

ce qu'il peut être commode d'écrire matriciellement:

$$(\gamma_{ij}(h)) = (b_{ij}) \gamma(h)$$

soit :

$$\begin{bmatrix} \gamma_{11}(h) & \gamma_{12}(h) & \cdots & \gamma_{1p}(h) \\ \gamma_{21}(h) & \gamma_{22}(h) & \cdots & \gamma_{2p}(h) \\ \cdots & \cdots & \ddots & \vdots \\ \gamma_{p1}(h) & \gamma_{p2}(h) & \cdots & \gamma_{pp}(h) \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \cdots & \cdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pp} \end{bmatrix} \gamma(h)$$

On en déduit immédiatement que la corrélation entre accroissements est constante (indépendante de h) :

$$\frac{\gamma_{ij}(h)}{\sqrt{\gamma_i(h)\gamma_j(h)}} = \frac{b_{ij}}{\sqrt{b_{ii}b_{jj}}}$$

Dans ce modèle, la corrélation entre $Z_i(v)$ et $Z_j(v)$ dans un champ V (section 2.9.2) vaut :

$$\rho_{ij}(v|V) = \frac{b_{ij}}{\sqrt{b_{ii}b_{jj}}} \text{ constante}$$

Elle est dite « intrinsèque » car elle ne dépend ni du support v, ni du champ V (d'où le nom du modèle).

Dans le cas stationnaire, on peut écrire:

$$C_{ij}(h) = b_{ij}C(h)$$

ou mieux (ce que l'on supposera par la suite):

$$C_{ij}(h) = b_{ij}\rho(h)$$

avec $\rho(0)=1$ (on notera au passage que les covariances croisées sont nécessairement paires).

3.1.2 Matrice des coefficients

On a (cas stationnaire):

$$b_{ij} = C_{ij}(0) = \text{cov}[Z_i(x), Z_j(x)]$$

et $b_{ii} = C_{ii}(0) = \text{var}[Z_i(x)]$

de sorte que la matrice des paliers (symétrique car $b_{ij} = b_{ji}$):

$$(b_{ij}) = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \cdots & \cdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pp} \end{bmatrix}$$

représente, dans le modèle, la matrice des variances-covariances des variables en un même point (comme en monovarié, les paliers du modèle peuvent différer des variances - covariances expérimentales).

Une telle matrice ne peut être quelconque : elle doit être (semi-) définie positive, ce qui signifie que la variance de toute combinaison linéaire des variables en un même point est positive, ou éventuellement nulle :

$$\begin{aligned} \text{var}\left(\sum_i \lambda_i Z_i(x)\right) &= \sum_{ij} \lambda_i \lambda_j \text{Cov}[Z(x_i), Z(x_j)] \\ &= \sum_{ij} \lambda_i \lambda_j b_{ij} \geq 0 \end{aligned}$$

Remarque : dans le cas purement intrinsèque, (b_{ij}) doit aussi être semi-définie positive. En effet $2(b_{ij}) \gamma(h)$ représente la matrice des variances-covariances des accroissements $Z_i(x+h) - Z_i(x)$ entre deux points, d'où:

$$\begin{aligned} \text{var}\left(\sum_i \lambda_i [Z_i(x+h) - Z_i(x)]\right) &= 2 \sum_{ij} \lambda_i \lambda_j \gamma_{ij}(h) \\ &= 2 \left(\sum_{ij} \lambda_i \lambda_j b_{ij}\right) \gamma(h) \geq 0 \end{aligned}$$

Revenons au cas stationnaire, plus facile à exposer. La matrice (b_{ij}) étant (semi-) définie positive, on a nécessairement :

$$b_{ii} = \text{var}[Z_i(x)] \geq 0 \text{ pour tout } i$$

et $b_{ii}b_{jj} - b_{ij}^2 \geq 0$ pour toute paire (i, j)

ce qui correspond (si leurs variances sont non nulles) à un coefficient de corrélation

$\rho_{ij} = \frac{b_{ij}}{\sqrt{b_{ii}b_{jj}}}$ entre les variables $Z_i(x)$ et $Z_j(x)$ compris entre -1 et $+1$ comme il se doit.

Ces conditions, qui doivent nécessairement être satisfaites pour tous i et j, ne sont cependant pas suffisantes pour assurer la définie positivité de la matrice (b_{ij}) au delà de deux variables. Celle-ci sera automatiquement assurée s'il est possible de faire correspondre à la matrice une factorisation des variables.

3.1.3 Factorisation

Commençons par énoncer deux propriétés élémentaires, mais importantes, du modèle de corrélation intrinsèque :

- Il est facile de voir que la corrélation intrinsèque entre variables s'étend à leurs combinaisons linéaires : les variogrammes simples et croisés de toutes combinaisons linéaires sont en effet proportionnels à $\gamma(h)$.

- Par ailleurs :

$$C_{ij}(0) = 0 \Rightarrow b_{ij} = 0 \Rightarrow C_{ij}(h) = 0$$

Ainsi, dans le modèle de corrélation intrinsèque, la non-corrélation de deux variables (ou de deux combinaisons linéaires) au même point entraîne leur non-corrélation spatiale.

Supposons maintenant qu'à la matrice (b_{ij}) , nous puissions faire correspondre une factorisation des variables, i.e. une décomposition des variables $Z_i(x)$ en « facteurs » $Y_q(x)$ sans corrélation au même point, par exemple dans le cas stationnaire et avec $E[Y_q(x)] = 0$:

$$Z_i(x) = \sum_q a_{iq} Y_q(x) + m_i$$

et qu'inversement on puisse exprimer les $Y_q(x)$ en fonction des $Z_i(x)$. Une telle décomposition garantit la positivité de la matrice (b_{ij}) : en effet toute combinaison linéaire des $Z_i(x)$ en un même point x est aussi une combinaison linéaire des $Y_q(x)$ non corrélés, et a donc une variance positive ou nulle :

$$\text{var} \left(\sum_i \lambda_i Z_i(x) \right) = \sum_q \left(\sum_i \lambda_i a_{iq} \right)^2 \text{var}(Y_q) \geq 0$$

Par ailleurs, le développement des variances et covariances des variables donne la relation suivante (si les $Y_q(x)$ partagent le même variogramme $\gamma(h)$, palier ou échelle verticale compris):

$$b_{ij} = \sum_q a_{iq} a_{jq}$$

Dans le modèle de corrélation intrinsèque, de tels facteurs $Y_q(x)$, étant non corrélés en un même point, sont de plus sans corrélation spatiale. Ainsi une factorisation statistique quelconque (par

résidus successifs, ou bien composantes principales, par exemple) fournit une factorisation spatiale du modèle. Le nombre de facteurs utiles est égal en général au nombre de variables (mais peut être inférieur si les variables sont liées). Bien que conventionnels (puisque dépendant de la méthode choisie pour la factorisation), ces facteurs sont objectifs dans le sens où leur valeur en un point x est déterminée dès que les valeurs des Z_i y sont connues.

Finalement, le modèle de corrélation intrinsèque est entièrement spécifié par la structure de base $\gamma(h)$, qui doit être autorisée, et par la matrice de variances-covariances (b_{ij}) , qui doit être (semi-)définie positive, ceci étant assuré par une factorisation. Les variables peuvent donc être interprétées comme des combinaisons linéaires de facteurs ayant même variogramme et sans corrélation spatiale. En pratique, la structure de base étant choisie, il est possible de procéder à un ajustement automatique fournissant les b_{ij} et a_{iq} .

3.1.4 Exemples

a) Soit deux variables Z_1 et Z_2 de structures :

$$\gamma_1(h) = 100 \text{ sph}(h/20)$$

$$\gamma_2(h) = 25 \text{ sph}(h/20)$$

$$\gamma_{12}(h) = -40 \text{ sph}(h/20)$$

Il s'agit d'un modèle de corrélation intrinsèque, car on peut écrire $\gamma_{ij}(h) = b_{ij}\gamma(h)$ avec:

- la structure de base $\gamma(h) = \text{sph}(h/20)$

- la matrice de variances-covariance : $(b_{ij}) = \begin{pmatrix} 100 & -40 \\ -40 & 25 \end{pmatrix}$

Cette matrice est bien définie positive car elle correspond à deux variables de variances positives et de corrélation -0.8 . Mais on peut le voir également en exhibant une factorisation possible du modèle. Ecrivons par exemple, avec Y_1 et Y_2 non corrélées de variance unité (et en faisant abstraction des moyennes), d'abord:

$$Z_1 = 10 Y_1$$

puis:

$$Z_2 = -4Y_1 + 3 Y_2$$

(ici Y_1 n'est autre que la variable Z_1 réduite, et $3Y_2$ représente le résidu de la régression linéaire de Z_2 sachant Z_1 ou Y_1). D'où inversement les facteurs:

$$Y_1 = 0.1 Z_1$$

$$Y_2 = (4/30) Z_1 + (1/3) Z_2$$

b) De même le modèle pour trois variables $\gamma_{ij}(h) = b_{ij}\gamma(h)$ avec $\gamma(h) = \text{sph}(h/20)$ et

$(b_{ij}) = \begin{pmatrix} 100 & -40 & -30 \\ -40 & 25 & 6 \\ -30 & 6 & 17 \end{pmatrix}$ est également un modèle de corrélation intrinsèque.

La matrice est bien définie positive, car on peut factoriser le modèle. En effet, en introduisant Y_1 , Y_2 , et Y_3 , non corrélées de variance unité, de façon à respecter successivement b_{11} , b_{12} , b_{22} , b_{13} , b_{23} , b_{33} , on peut écrire par exemple:

$$Z_1 = 10 Y_1$$

$$Z_2 = -4Y_1 + 3 Y_2$$

$$Z_3 = -3Y_1 - 2 Y_2 + 2 Y_3$$

les facteurs étant alors donnés inversement par:

$$Y_1 = 0.1 Z_1$$

$$Y_2 = (4/30) Z_1 + (1/3) Z_2$$

$$Y_3 = (17/60) Z_1 + (1/3) Z_2 + (1/2) Z_3$$

A titre d'exercice, on exhibera autres factorisations possibles: par exemple par la même méthode de résidus successifs, mais en prenant les variables dans un ordre différent; par analyse en composantes principales, ou bien encore par analyse en composantes principales sur variables normées.

(On pourra vérifier que si on avait $b_{33} = 13$, deux facteurs seraient suffisants pour assurer la semi-définie positivité, témoignant d'une relation linéaire entre les trois variables; mais qu'avec $b_{33} = 10$, la factorisation n'est plus possible: le modèle n'est pas autorisé, bien que les corrélations entre variables soient comprises entre -1 et $+1$.)

c) Si on rajoute, à l'intérieur du modèle (b), une composante pépitique sur le variogramme simple de la 3^e variable, seules Z_1 et Z_2 sont en corrélation intrinsèque.

3.1.5 Exercice : exemples de modèles de corrélation intrinsèque

Les modèles suivants sont-ils des modèles de corrélation intrinsèque?

$$\gamma_1(h) = 64 \text{ sph}(h/20)$$

$$\gamma_2(h) = 64 \text{ sph}(h/20)$$

$$\gamma_{12}(h) = 38.4 \text{ sph}(h/20)$$

$$\gamma_1(h) = 36 \text{ pep}(h) + 64 \text{ sph}(h/20)$$

$$\gamma_2(h) = 36 \text{ pep}(h) + 64 \text{ sph}(h/20)$$

$$\gamma_{12}(h) = 38.4 \text{ sph}(h/20)$$

$$\gamma_1(h) = 36 \text{ pep}(h) + 64 \text{ sph}(h/20)$$

$$\gamma_2(h) = 36 \text{ pep}(h) + 64 \text{ sph}(h/20)$$

$$\gamma_{12}(h) = 21.6 \text{ pep}(h) + 38.4 \text{ sph}(h/20)$$

$$\gamma_1(h) = 36 \text{ pep}(h) + 64 \text{ sph}(h/20)$$

$$\gamma_2(h) = 36 \text{ pep}(h) + 64 \text{ sph}(h/20)$$

$$\gamma_{12}(h) = 18 \text{ pep}(h) + 38.4 \text{ sph}(h/20)$$

$$\gamma_1(h) = 64 \text{ sph}(h/20)$$

$$\gamma_2(h) = 16 \text{ sph}(h/20)$$

$$\gamma_{12}(h) = 38.4 \text{ sph}(h/20)$$

3.2 Le modèle linéaire de corégionalisation

Dans ce modèle, toutes les structures simples et croisées sont des combinaisons linéaires des mêmes composantes structurales de base, et ceci peut s'interpréter par une décomposition des variables elles-mêmes en combinaisons linéaires de composantes correspondantes. Ceci peut être vu comme une généralisation, au cas multivariable, des composantes gigognes du cas monovariable.

3.2.1 Composantes gigognes en monovariable : le modèle linéaire de régionalisation

Prenons l'exemple (voir section 2.9.3) de la teneur en plomb $Z(x)$, de moyenne 13.2 (%) et de variogramme composés de 2 structures gigognes, l'une étant un pépite de 11, l'autre étant sphérique de palier 39 et de portée 60 :

$$\gamma(h) = 11 \text{ Pep}(h) + 39 \text{ Sph}(h/60)$$

soit: $\gamma(h) = b^1 \gamma^1(h) + b^2 \gamma^2(h)$

Cette variable peut s'interpréter comme la somme de deux composantes $Z^1(x)$ et $Z^2(x)$ sans corrélation spatiale entre elles, de moyennes nulles et de structures respectives

$$b^1 \gamma^1(h) = 11 \text{ Pep}(h) \text{ et } b^2 \gamma^2(h) = 39 \text{ Sph}(h/60) :$$

$$Z(x) = Z^1(x) + Z^2(x) + m$$

Il est souvent pratique d'utiliser des composantes normés (de variance 1), ici $Y^1(x)$ et $Y^2(x)$, de variogrammes respectifs $\gamma^1(h) = \text{Pep}(h)$ et $\gamma^2(h) = \text{Sph}(h/60)$, d'où finalement:

$$Z(x) = 3.32 Y^1(x) + 6.24 Y^2(x) + 13.2$$

soit: $Z(x) = a^1 Y^1(x) + a^2 Y^2(x) + m$

avec $a^1 = \sqrt{b^1} = \sqrt{11} = 3.32$ et $a^2 = \sqrt{b^2} = \sqrt{39} = 6.24$ (attention : les indices en exposant ne sont pas des puissances).

Plus généralement, une variable de structure gigogne :

$$\gamma(h) = \sum_k b^k \gamma^k(h)$$

pourra s'interpréter comme une décomposition de la forme :

$$Z(x) = \sum_k Z^k(x) + m$$

ou bien, en utilisant des composantes normées:

$$Z(x) = \sum_k a^k Y^k(h) + m$$

avec $b^k = (a^k)^2$ (les facteurs sont pris ici de moyenne nulle, ou de moyenne d'accroissements nulle, la moyenne m disparaissant dans le cas intrinsèque pur, i.e. non stationnaire).

3.2.2 Cas multivariable: le modèle linéaire de corégionalisation

Décomposition des structures

Pour introduire les notations, partons de l'exemple des deux variables, teneurs en plomb et zinc:

$$\begin{aligned}\gamma_1(h) &= 11 \text{Pep}(h) + 39 \text{Sph}(h/60) \\ \gamma_2(h) &= 9 \text{Pep}(h) + 15 \text{Sph}(h/60) \\ \gamma_{12}(h) &= 14.5 \text{Sph}(h/60)\end{aligned}$$

qu'on peut noter:

$$\begin{aligned}\gamma_1(h) &= b_{11}^1 \gamma^1(h) + b_{11}^2 \gamma^2(h) \\ \gamma_2(h) &= b_{22}^1 \gamma^1(h) + b_{22}^2 \gamma^2(h) \\ \gamma_{12}(h) &= b_{12}^1 \gamma^1(h) + b_{12}^2 \gamma^2(h)\end{aligned}$$

soit matriciellement :

$$[\gamma_{ij}(h)] = [b_{ij}^1] \gamma^1(h) + [b_{ij}^2] \gamma^2(h)$$

avec $\gamma^1(h) = \text{Pep}(h)$, $\gamma^2(h) = \text{Sph}(h/60)$, et les matrices $b^1 = (b_{ij}^1) = \begin{pmatrix} 11 & 0 \\ 0 & 9 \end{pmatrix}$ et

$$b^2 = (b_{ij}^2) = \begin{pmatrix} 39 & 14.5 \\ 14.5 & 15 \end{pmatrix}.$$

Avec les trois variables, plomb-zinc-argent, nous aurons:

$$\begin{aligned}\gamma_1(h) &= 11 \text{Pep}(h) + 39 \text{Sph}(h/60) \\ \gamma_2(h) &= 9 \text{Pep}(h) + 15 \text{Sph}(h/60) \\ \gamma_3(h) &= 1.1 \text{Pep}(h) + 1.8 \text{Sph}(h/60) \\ \gamma_{12}(h) &= 14.5 \text{Sph}(h/60) \\ \gamma_{13}(h) &= 5 \text{Sph}(h/60) \\ \gamma_{23}(h) &= 3.8 \text{Sph}(h/60)\end{aligned}$$

avec maintenant $b^1 = (b_{ij}^1) = \begin{pmatrix} 11 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 1.1 \end{pmatrix}$ et $b^2 = (b_{ij}^2) = \begin{pmatrix} 39 & 14.5 & 5 \\ 14.5 & 15 & 3.8 \\ 5 & 3.8 & 1.8 \end{pmatrix}$

De façon générale, un modèle linéaire de corégionalisation peut s'écrire:

- en variogramme:

$$\gamma_{ij}(h) = \sum_k b_{ij}^k \gamma^k(h)$$

soit matriciellement:

$$[\gamma_{ij}(h)] = \sum_k [b_{ij}^k] \gamma^k(h)$$

- ou en covariances, dans le cas stationnaire:

$$C_{ij}(h) = \sum_k b_{ij}^k C^k(h)$$

$$C_{ij}(h) = \sum_k b_{ij}^k \rho^k(h)$$

avec

$$\rho^k(0) = 1$$

où par exemple les $\gamma^k(h)$ représentent des structures de base d'échelles différentes (ainsi $\gamma^1(h) = \text{Pep}(h)$, $\gamma^2(h) = \text{Sph}(h/60)$, ..., mais on évitera de prendre des structures liées, comme de rajouter la structure $\text{Pep}(h) + \text{Sph}(h/60)$ aux deux structures précédentes).

On remarque que dans un modèle linéaire de corégionalisation, les covariances croisées sont nécessairement paires.

Décomposition des variables en différentes échelles

Dans un tel modèle les variables peuvent se décomposer ainsi (par exemple dans le cas stationnaire, avec m_i moyenne de Z_i et des facteurs de moyenne 0):

$$Z_i(x) = \sum_k Z_i^k(x) + m_i$$

Les composantes d'échelle k , $Z_i^k(x)$, de même structure proportionnelle à $\gamma^k(h)$ ou $\rho^k(h)$, sont en corrélation intrinsèque, et (b_{ij}^k) représente leur matrice de variances-covariances. Chaque matrice $b^k = (b_{ij}^k)$ doit donc être définie positive. En particulier on doit avoir:

$$|b_{ij}^k| \leq \sqrt{b_{ii}^k b_{jj}^k}$$

Ainsi une composante structurale (pépite, courte portée, ou même longue portée) ne peut figurer dans une structure croisée sans figurer sur chacune des deux structures simples correspondantes. Cependant une composante peut figurer sur deux structures simples sans figurer sur le croisé correspondant.

Le variogramme croisé, qui se présente comme une combinaison linéaire à coefficients positifs ou négatifs des structures de base, se situe nécessairement à l'intérieur de l'enveloppe :

$$-\sum_k \sqrt{b_{ii}^k b_{jj}^k} \gamma^k(h) \leq \gamma_{ij}(h) \leq \sum_k \sqrt{b_{ii}^k b_{jj}^k} \gamma^k(h)$$

Factorisation de chaque échelle

En allant plus loin, on peut factoriser les composantes en corrélation intrinsèque de chaque échelle k (voir section 3.1.3). On obtient alors la décomposition en facteurs:

$$Z_i(x) = \sum_k \sum_q a_{iq}^k Y_q^k(x) + m_i$$

où les $Y_q^k(x)$ de même k partagent la structure $\gamma^k(h)$ et avec $b_{ij}^k = \sum_q a_{iq}^k a_{jq}^k$. Une telle

décomposition garantit la (semi-)définie positivité des matrices b_{ij}^k .

En pratique, les structures de base étant choisies, il est possible de procéder à un ajustement automatique fournissant les b_{ij}^k et a_{iq}^k .

3.2.3 Exemple

Considérons le modèle linéaire de corégionalisation suivant :

$$\begin{aligned} \gamma_{11}(h) &= 25\gamma^1(h) + 100\gamma^2(h) + 16\gamma^3(h) \\ \gamma_{12}(h) &= -40\gamma^2(h) + 12\gamma^3(h) \\ \gamma_{22}(h) &= 16\gamma^1(h) + 25\gamma^2(h) + 13\gamma^3(h) \\ \gamma_{13}(h) &= -30\gamma^2(h) \\ \gamma_{23}(h) &= 6\gamma^2(h) \\ \gamma_{33}(h) &= 9\gamma^1(h) + 17\gamma^2(h) \end{aligned}$$

soit matriciellement:

$$\begin{pmatrix} \gamma_{11}(h) & \gamma_{12}(h) & \gamma_{13}(h) \\ \gamma_{12}(h) & \gamma_{22}(h) & \gamma_{23}(h) \\ \gamma_{13}(h) & \gamma_{23}(h) & \gamma_{33}(h) \end{pmatrix} = \begin{pmatrix} 25 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & 9 \end{pmatrix} \gamma^1(h) + \begin{pmatrix} 100 & -40 & -30 \\ -40 & 25 & 6 \\ -30 & 6 & 17 \end{pmatrix} \gamma^2(h) + \begin{pmatrix} 16 & 12 & 0 \\ 12 & 13 & 0 \\ 0 & 0 & 0 \end{pmatrix} \gamma^3(h)$$

où $\gamma^1(h) = \text{Pep}(h)$, $\gamma^2(h) = \text{Sph}(h/20)$, et $\gamma^3(h)$ est un schéma de grande portée (ou même intrinsèque, linéaire par exemple). C'est bien un modèle autorisé. En effet, on peut écrire par exemple (aux moyennes éventuelles près):

$$\begin{aligned} Z_1 &= 5Y_1^1 && +10Y_1^2 && +4Y_1^3 \\ Z_2 &= 4Y_2^1 && -4Y_1^2 & +3Y_2^2 && +3Y_1^3 & +2Y_2^3 \\ Z_3 &= 3Y_3^1 && -3Y_1^2 & -2Y_2^2 & +3Y_3^2 \end{aligned}$$

où les Y_q^1 ont la structure $\gamma^1(h) = \text{Pep}(h)$, les Y_q^2 la structure $\gamma^2(h) = \text{Sph}(h/20)$, et les Y_q^3 la structure de grande portée. Dans cet exemple, 3 facteurs sont nécessaires pour factoriser chacune des deux premières composantes, mais 2 suffisent pour la troisième (qui ne figure pas dans Z_3).

Exercice: proposer d'autres factorisations possibles de ce modèle.

De façon générale, il est possible de factoriser un modèle linéaire de corégionalisation en prenant un nombre de facteurs égal au maximum au produit du nombre de variables par le nombre de composantes.

3.2.4 Exercice : exemples de modèles linéaires de corégionalisation

Les modèles suivants sont-ils des modèles linéaires de corégionalisation admissibles?

$$\gamma_1(h) = 16 \text{ sph}(h/20) + 4 \text{ sph}(h/100)$$

$$\gamma_2(h) = 25 \text{ sph}(h/20)$$

$$\gamma_{12}(h) = 12 \text{ sph}(h/20)$$

$$\gamma_1(h) = 4 \text{ sph}(h/100)$$

$$\gamma_2(h) = 25 \text{ sph}(h/20)$$

$$\gamma_{12}(h) = 12 \text{ sph}(h/20)$$

$$\gamma_1(h) = 4 \text{ sph}(h/20)$$

$$\gamma_2(h) = 25 \text{ sph}(h/20)$$

$$\gamma_{12}(h) = 12 \text{ sph}(h/20)$$

$$\gamma_1(h) = 16 \text{ sph}(h/20) + 4 \text{ sph}(h/100)$$

$$\gamma_2(h) = 25 \text{ sph}(h/20) + 9 \text{ sph}(h/100)$$

$$\gamma_{12}(h) = 12 \text{ sph}(h/20) - 3 \text{ sph}(h/100)$$

3.3 Exercice : modèle d'erreur

Soit $Z_1(x)$ (moyenne m_1 , covariance $C_1(h)$, variogramme $\gamma_1(h)$) connue en certains points et :

$$Z_2(x) = Z_1(x) + R(x)$$

une mesure de $Z_1(x)$ (éventuellement en des points différents), affectée d'une erreur $R(x)$ sans corrélation spatiale avec $Z_1(x)$ et pépitique.

On montrera les relations suivantes :

Moyennes :

$$m_2 = m_1 + m_R$$

Variogramme de Z_2 :

$$\gamma_2(h) = \gamma_1(h) + \text{var}(R) \text{Pep}(h)$$

Variogramme croisé de Z_1 et Z_2 :

$$\gamma_{12}(h) = \gamma_1(h)$$

Pseudo-variogramme croisé de Z_1 et Z_2 :

$$\psi_{12}(h) = \gamma_1(h) + 0.5 E[R^2]$$

3.4 Exercice : multi-support

Une même variable étudiée simultanément sous différents supports peut être considérée comme un cas multivariable particulier. Bien que la géostatistique linéaire soit théoriquement en mesure de faire le lien structural entre différents supports, il est souvent plus pratique d'utiliser le modèle approximatif suivant (modèle « discret » ou « discrétisé »).

On choisit un support de référence (disons, un bloc v), de variogramme connu $\gamma_v(h)$, en supposant que:

- un support plus grand peut être considéré comme un support multiple de v (formules de régularisation classiques);
- le champ lui-même est partitionné (“discrétisé”) en unités de support v ;
- un support plus fin que v (par exemple, quasi-ponctuel x) est localisé de manière aléatoire, uniforme et indépendante dans le support v auquel il appartient (notons-le alors \underline{x}).

Montrer (dans le cas stationnaire) que la régression linéaire de $Z(\underline{x})$ connaissant $Z(v)$ de son bloc coïncide avec $Z(v)$ (ce qui ne serait pas le cas pour un point x fixé), d'où:

$$E[Z(\underline{x}) | Z(v)] = Z(v)$$

On a la décomposition:

$$Z(\underline{x}) = Z(v) + [Z(\underline{x}) - Z(v)]$$

Se basant sur l'hypothèse que par exemple, conditionnellement à son bloc $Z(v)$, $Z(\underline{x})$ est indépendant de tout autre bloc ou échantillon ponctuel, montrer que l'on a (avec $\underline{x} \in v$, $\underline{y} \in v_h$, et v_h désignant v translaté de h):

- structure croisée point-bloc:

$$\frac{1}{2} E\{[Z(\underline{y}) - Z(\underline{x})][Z(v_h) - Z(v)]\} = \gamma_v(h)$$

- structure ponctuelle discrétisée :

$$\begin{aligned} \frac{1}{2} E\{[Z(\underline{y}) - Z(\underline{x})]^2\} &= \frac{1}{2} E\{[(Z(\underline{y}) - Z(v_h)) + (Z(v_h) - Z(v)) + (Z(v) - Z(\underline{x}))]^2\} \\ &= \gamma_v(h) + \text{var}[Z(\underline{x}) - Z(v)] \end{aligned}$$

(y compris pour $h = 0$, pour $\underline{x} \neq \underline{y} \in v$)

Le modèle est donc défini par la seule donnée du variogramme sur le support de référence $\gamma_v(h)$ et par $\text{var}[Z(\underline{x}) - Z(v)]$. On remarquera l'analogie avec l'exercice précédent: le terme $Z(\underline{x}) - Z(v)$ n'est autre que l'erreur d'estimation de $Z(v)$ par $Z(\underline{x})$.

4 Le cokrigage

4.1 Généralités sur le cokrigage d'une ou plusieurs variables

Pour simplifier l'écriture, et sans nuire à la généralité, nous considérerons le cas de 2 variables $Z_1(x)$ et $Z_2(x)$, informées respectivement sur deux ensembles de points S_1 et S_2 identiques (cas isotopique) ou non (cas hétérotopique). Le cokrigage permet de tenir compte des informations sur les deux variables en estimant au mieux, linéairement en fonction des données sur Z_1 et Z_2 : $Z_1(x_0)$, ou $Z_2(x_0)$, ou d'ailleurs chacune des deux variables, ou une régularisée de ces variables, ou encore une combinaison linéaire comme $Z_2(x_0) - Z_1(x_0)$. Le cokrigage assure la cohérence entre estimations :

$$[Z_2(x_0) - Z_1(x_0)]^{CK} = Z_2(x_0)^{CK} - Z_1(x_0)^{CK}$$

ce que ne fait pas le krigeage en général (voir exercices 4.2.1, 4.2.2 et 4.3.1).

Dans la suite nous commencerons par développer les calculs en considérant l'estimation d'une variable cible notée $Z(x_0)$ (qui pourra être $Z_1(x_0)$, $Z_2(x_0)$, ou par exemple $Z_2(x_0) - Z_1(x_0)$...) de la manière la plus générale. On considère donc un estimateur constitué par une combinaison linéaire des données:

$$Z^*(x_0) = \sum_{S_1} \lambda_{1\alpha} Z_1(x_\alpha) + \sum_{S_2} \lambda_{2\alpha} Z_2(x_\alpha) + \lambda_0$$

ce qui correspond à l'erreur d'estimation :

$$\varepsilon = Z_0 - Z_0^* = Z_0 - \sum_{S_1} \lambda_{1\alpha} Z_1(x_\alpha) - \sum_{S_2} \lambda_{2\alpha} Z_2(x_\alpha) - \lambda_0$$

On cherchera ensuite les poids tels que l'erreur d'estimation soit une combinaison linéaire autorisée (ce qui permet d'en calculer moyenne et variance), soit d'espérance nulle (estimateur sans biais) et enfin de variance minimale. En terme de covariance, et avec des notations abrégées, la variance qu'il s'agit de minimiser s'écrit :

$$\begin{aligned} Var(\varepsilon) &= C(Z_0, Z_0) - 2 \sum_{S_1} \lambda_{1\alpha} C(Z_0, Z_{1\alpha}) - 2 \sum_{S_2} \lambda_{2\alpha} C(Z_0, Z_{2\alpha}) \\ &\quad + 2 \sum_{S_1} \sum_{S_2} \lambda_{1\alpha} \lambda_{2\beta} C(Z_{1\alpha}, Z_{2\beta}) \\ &\quad + \sum_{S_1} \sum_{S_1} \lambda_{1\alpha} \lambda_{1\beta} C(Z_{1\alpha}, Z_{1\beta}) \\ &\quad + \sum_{S_2} \sum_{S_2} \lambda_{2\alpha} \lambda_{2\beta} C(Z_{2\alpha}, Z_{2\beta}) \end{aligned}$$

Il est clair que si $Z(x_0)$ figure dans les données, le cokrigage de $Z(x_0)$, meilleur estimateur linéaire, s'identifie à lui-même: comme le krigeage: le cokrigage est un interpolateur exact.

4.2 Cokrigage simple (ou cokrigage à moyennes connues)

Nous nous plaçons ici dans le cas stationnaire. Dans ce cas il n'y a aucune condition à imposer pour pouvoir calculer moyenne et variance de l'erreur d'estimation. Le non-biais s'écrit:

$$E(\varepsilon) = m - m_1 \sum_{S_1} \lambda_{1\alpha} - m_2 \sum_{S_2} \lambda_{2\alpha} - \lambda_0 = 0$$

d'où :

$$\lambda_0 = m - m_1 \sum_{S_1} \lambda_{1\alpha} - m_2 \sum_{S_2} \lambda_{2\alpha}$$

ce qui revient à considérer les variables centrées:

$$Z^*(x_0) - m_0 = \sum_{S_1} \lambda_{1\alpha} [Z_1(x_\alpha) - m_1] + \sum_{S_2} \lambda_{2\alpha} [Z_2(x_\alpha) - m_2]$$

Pour minimiser la variance de l'erreur, on annule sa dérivée par rapport au poids de chaque donnée. Chaque donnée donne lieu à une équation, formant le système linéaire :

(système de cokrigage simple)

$$\begin{cases} \sum_{\beta \in S_1} \lambda_{1\beta} C(Z_{1\alpha}, Z_{1\beta}) + \sum_{\beta \in S_2} \lambda_{2\beta} C(Z_{1\alpha}, Z_{2\beta}) = C(Z_0, Z_{1\alpha}) & \forall x_\alpha \in S_1 \\ \sum_{\beta \in S_1} \lambda_{1\beta} C(Z_{1\beta}, Z_{2\alpha}) + \sum_{\beta \in S_2} \lambda_{2\beta} C(Z_{2\alpha}, Z_{2\beta}) = C(Z_0, Z_{2\alpha}) & \forall x_\alpha \in S_2 \end{cases}$$

soit matriciellement :

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{12}^t & C_{22} \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} C_{01} \\ C_{02} \end{bmatrix}$$

covariance covariance
entre données cible-données

La résolution du système conduit aux poids optimaux $\lambda_{1\alpha}$ et $\lambda_{2\alpha}$ (poids de cokrigage simple).

La variance l'erreur (variance de cokrigage) peut s'écrire alors:

$$\sigma_{CKS}^2 = \text{var}(\varepsilon) = C(Z_0, Z_0) - \sum_{S_1} \lambda_{1\alpha} C(Z_0, Z_{1\alpha}) - \sum_{S_2} \lambda_{2\alpha} C(Z_0, Z_{2\alpha})$$

soit matriciellement :

$$\sigma_{CKS}^2 = \text{var}(\varepsilon) = C(Z_0, Z_0) - \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}^t \begin{bmatrix} C_{01} \\ C_{02} \end{bmatrix}$$

4.2.1 Exercice: cohérence entre valeurs krigées

Soit Z_1 et Z_2 deux variables de variogrammes simples et croisé $\gamma_1(h)$, $\gamma_2(h)$ et $\gamma_{12}(h)$, connues aux mêmes points. Montrer que si $\gamma_2(h)$ et $\gamma_{12}(h)$ sont proportionnelles à $\gamma_1(h)$, le krigeage de la différence des deux variables est égal à la différence de leur krigeage (ce qui n'est pas le cas en général).

4.2.2 Exercice: cohérence entre valeurs cokrigées

Montrer que le cokrigage de $Z_2(x_0) - Z_1(x_0)$ (ou plus généralement d'une combinaison linéaire de $Z_1(x_0)$ et $Z_2(x_0)$) est égal à la différence entre les cokrigages de $Z_1(x_0)$ et $Z_2(x_0)$.

(On utilisera la linéarité des poids de cokrigage en fonction du second membre du système de cokrigage.)

4.2.3 Exercice: cokrigage simple

Soit Z_1 et Z_2 stationnaires de moyennes m_1 et m_2 . On connaît $Z_1(A)$ et $Z_2(B)$ en deux points A et B. Cokrigage simple de $Z_1(O)$ au point O ?

4.3 Cokrigage ordinaire (ou cokrigage à moyennes inconnues)

Plaçons-nous dans le cas stationnaire. Nous sommes donc en mesure de calculer moyenne et variance de l'erreur d'estimation. Cependant les moyennes sont ici supposées inconnues. Il va alors falloir spécifier la variable à estimer, par exemple $Z_1(x_0)$ (le cas général sera examiné à l'exercice 4.3.1). Le non-biais s'écrit alors:

$$E(\varepsilon) = m_1 \left(1 - \sum_{S_1} \lambda_{1\alpha} \right) - m_2 \sum_{S_2} \lambda_{2\alpha} - \lambda_0 = 0$$

Pour qu'il soit vérifié quelles que soient les valeurs de m_1 et m_2 (lesquelles seront supposées ici non liées), on impose les conditions suivantes :

$$\lambda_0 = 0 \quad \text{qui disparaît définitivement des équations}$$

$$\sum_{S_1} \lambda_{1\alpha} = 1 \quad (\text{somme des poids pour la variable à estimer})$$

$$\sum_{S_2} \lambda_{2\alpha} = 0 \quad (\text{somme des poids pour la variable complémentaire, ou le cas échéant, pour chacune des variables complémentaires})$$

On recherche donc l'estimateur:

$$Z_1^*(x_0) = \sum_{S_1} \lambda_{1\alpha} Z_1(x_\alpha) + \sum_{S_2} \lambda_{2\alpha} Z_2(x_\alpha)$$

pour lequel la variance d'estimation est minimum, compte tenu des deux conditions ci-dessus. Ceci conduit au système linéaire d'équations suivant, où μ_1 et μ_2 sont des paramètres de Lagrange introduits pour satisfaire ces conditions :

(système de cokrigage ordinaire)

$$\left\{ \begin{array}{l} \sum_{\beta \in S_1} \lambda_{1\beta} C(Z_{1\alpha}, Z_{1\beta}) + \sum_{\beta \in S_2} \lambda_{2\beta} C(Z_{1\alpha}, Z_{2\beta}) + \mu_1 = C(Z_0, Z_{1\alpha}) \quad \forall x_\alpha \in S_1 \\ \sum_{\beta \in S_1} \lambda_{1\beta} C(Z_{1\beta}, Z_{2\alpha}) + \sum_{\beta \in S_2} \lambda_{2\beta} C(Z_{2\alpha}, Z_{2\beta}) + \mu_2 = C(Z_0, Z_{2\alpha}) \quad \forall x_\alpha \in S_2 \\ \sum_{\alpha \in S_1} \lambda_{1\alpha} = 1 \\ \sum_{\alpha \in S_2} \lambda_{2\alpha} = 0 \end{array} \right.$$

soit matriciellement:

$$\begin{bmatrix} C_{11} & C_{12} & 1 & 0 \\ C_{12}' & C_{22} & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} C_{01} \\ C_{02} \\ 1 \\ 0 \end{bmatrix}$$

La variance de l'erreur (variance de cokrigage) peut s'écrire:

$$\sigma_{CKO}^2 = \text{var}(\varepsilon) = C(Z_0, Z_0) - \sum_{S_1} \lambda_{1\alpha} C(Z_0, Z_{1\alpha}) - \sum_{S_2} \lambda_{2\alpha} C(Z_0, Z_{2\alpha}) - \mu_1$$

soit matriciellement :

$$\sigma_{CKO}^2 = \text{var}(\varepsilon) = C(Z_0, Z_0) - \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \mu_1 \\ \mu_2 \end{bmatrix}' \begin{bmatrix} C_{01} \\ C_{02} \\ 1 \\ 0 \end{bmatrix}$$

Dans le cas où les covariances croisées sont paires, on vérifie facilement que, grâce aux conditions imposées sur les poids (et contrairement au cokrigage simple), le système peut s'écrire presque à l'identique en variogramme (il suffit de changer chaque terme de covariance, simple ou croisée, par l'opposé du variogramme correspondant : le système est donc identique, au signe près en ce qui concerne les paramètres de Lagrange). En fait, comme on va le voir, le cokrigage ordinaire coïncide alors avec un cokrigage intrinsèque.

4.3.1 Exercice: cohérence entre valeurs cokrigées

Soit à estimer, par cokrigage ordinaire, une combinaison linéaire de $Z_1(x_0)$ et $Z_2(x_0)$, par exemple $Z_2(x_0) - Z_1(x_0)$, à partir de données sur $Z_1(x)$ et $Z_2(x)$.

- Quelles sont les conditions de non-biais à introduire?
- Etablir le système de cokrigage.
- Vérifier que le cokrigage de $Z_2(x_0) - Z_1(x_0)$ est bien la différence des cokrigages de $Z_2(x_0)$ et $Z_1(x_0)$.

4.4 Cokrigage intrinsèque

En effet, dans le cas intrinsèque, les conditions :

$$\sum_{S_1} \lambda_{1\alpha} = 1, \quad \sum_{S_2} \lambda_{2\alpha} = 0$$

sont précisément celles pour lesquelles l'erreur d'estimation :

$$\begin{aligned} \varepsilon &= Z_1(x_0) - Z_1^*(x_0) \\ &= Z_1(x_0) - \sum_{S_1} \lambda_{1\alpha} Z_1(x_\alpha) - \sum_{S_2} \lambda_{2\alpha} Z_2(x_\alpha) \end{aligned}$$

est une CLA. Sa moyenne est nulle (ce qui assure le non-biais). Sa variance se développe à l'aide des variogrammes simples et croisés, et conduit à un système de cokrigage identique au précédent, mais écrit en variogrammes (au signe près pour les paramètres de Lagrange).

4.5 Cokrigage à moyennes inconnues mais liées

Le cokrigage va changer si les moyennes, supposées inconnues, sont liées. Prenons l'exemple de deux variables dont les moyennes sont inconnues mais égales : $m_1 = m_2$.

Le non-biais:

$$E(\varepsilon) = m_1 \left(1 - \sum_{S_1} \lambda_{1\alpha} - \sum_{S_2} \lambda_{2\alpha} \right) - \lambda_0 = 0$$

devant être respecté quelle que soit la moyenne, donne:

$$\begin{aligned} \lambda_0 &= 0 \\ \sum_{S_1} \lambda_{1\alpha} + \sum_{S_2} \lambda_{2\alpha} &= 1 \end{aligned}$$

On montrera que cela conduit au système de cokrigage suivant:

$$\left\{ \begin{array}{l} \sum_{\beta \in S_1} \lambda_{1\beta} C(Z_{1\alpha}, Z_{1\beta}) + \sum_{\beta \in S_2} \lambda_{2\beta} C(Z_{1\alpha}, Z_{2\beta}) + \mu = C(Z_0, Z_{1\alpha}) \quad \forall x_\alpha \in S_1 \\ \sum_{\beta \in S_1} \lambda_{1\beta} C(Z_{1\beta}, Z_{2\alpha}) + \sum_{\beta \in S_2} \lambda_{2\beta} C(Z_{2\alpha}, Z_{2\beta}) + \mu = C(Z_0, Z_{2\alpha}) \quad \forall x_\alpha \in S_2 \\ \sum_{\alpha \in S_1} \lambda_{1\alpha} + \sum_{\alpha \in S_2} \lambda_{2\alpha} = 1 \end{array} \right.$$

4.5.1 Exercice: filtrage d'erreurs

On se place dans le cadre du modèle d'erreur de l'exercice 3.3.

Les erreurs sont supposées systématiques $E[R] \neq 0$ et les moyennes inconnues: krigeage et système de krigeage de $Z_1(x_0)$?

Les erreurs sont supposées non systématiques $E[R]=0$ et m_1 inconnue: krigeage et système de krigeage de $Z_1(x_0)$?

4.6 Remarques

4.6.1 Variance de cokrigage

La prise en compte supplémentaire des données sur Z_2 , pour estimer la variable Z_1 , ne peut qu'augmenter la précision: la variance de cokrigage ne peut qu'être inférieure ou égale à la variance de krigeage à partir des seules données sur Z_1 (en effet le krigeage revient à donner un poids nul, en général non optimal, aux données de Z_2).

4.6.2 Poids de cokrigage

Attention à l'amplitude des poids: les poids de la variable à estimer Z_1 sont sans unité, mais ceux d'une co-variable Z_2 sont en (unité Z_1 / unité Z_2)

En cokrigage ordinaire, les poids d'une co-variable Z_2 somment à 0. Les poids négatifs, s'ils sont associés à des valeurs fortes, peuvent rendre facilement négative l'estimation d'une variable positive Z_1 .

5 Simplification du cokrigeage

Le cokrigeage est une opération d'autant plus lourde que le nombre de variables et de points échantillonnés est élevé. D'où l'intérêt de se placer dans des modèles permettant des simplifications, lorsque cela est compatible avec structure et implantation des données.

5.1 Remarques préliminaires

5.1.1 Cohérence des estimations

Ainsi qu'il a été vu, le cokrigeage assure la cohérence entre estimations de différentes variables (le cokrigeage d'une combinaison linéaire étant la combinaison linéaire des cokrigeages). Ainsi le cokrigeage de $Z_2 - Z_1$ (à partir de données sur Z_1 et Z_2 par exemple) est égal à la différence des cokrigeages de Z_2 et Z_1 .

Par ailleurs, le cokrigeage d'une quantité étant la meilleure combinaison linéaire des données, on peut remplacer ces données par toute famille de combinaisons linéaires équivalente (généralisant le même espace). Ainsi il est théoriquement équivalent de faire un cokrigeage (que ce soit de Z_1 , Z_2 ou $Z = Z_2 - Z_1$, etc) à partir de données équivalentes sur Z_1 et Z_2 , sur Z_1 et Z , sur Z_2 et Z , ou encore sur Z et $Z_1 + Z_2$. En pratique certains choix peuvent être préférables pour les simplifications qu'ils apportent au cokrigeage.

5.1.2 Données redondantes

Commençons par voir ce qui se passe en cas de redondance dans un exemple très simple de krigeage.

Le krigeage simple de $Z(x)$ à partir d'un doublon de valeurs au point $x+h$ (deux données égales à $Z(x+h)$) s'écrit :

$$Z(x)^k = \lambda Z(x+h) + \lambda' Z(x+h) + (1 - \lambda - \lambda')m$$

Le système de krigeage se compose de deux équations identiques à :

$$\lambda C(0) + \lambda' C(0) = C(h)$$

d'où :

$$\lambda + \lambda' = \frac{C(h)}{C(0)}$$

Le doublon étant cohérent (même valeur), le krigeage est parfaitement déterminé :

$$Z(x)^k = (\lambda + \lambda')Z(x+h) + (1 - \lambda - \lambda')m$$

et coïncide avec le krigeage à partir d'une seule des données (en cas d'incohérence - deux valeurs différentes au même point de données - le krigeage ne serait plus défini.)

En pratique on évitera cependant les redondances, car la solution du système n'est théoriquement pas unique (indétermination) et en pratique peut se révéler instable.

Il en est de même en cokrigeage. On évitera par exemple d'utiliser simultanément comme données les valeurs de Z_1 , Z_2 , et de $Z = Z_2 - Z_1$ en un même point. De même, si l'on dispose de p variables de somme constante, connues aux mêmes points, on n'en retiendra que $p-1$. En théorie,

le cokrigeage ne dépendra pas du choix des variables retenues, mais certains choix sont susceptibles d'apporter des simplifications au cokrigeage - ce qui est toujours utile en pratique - et donc peuvent se révéler plus judicieux.

5.2 Variables sans corrélation spatiale

Soit un ensemble de variables:

$$Z_1, Z_2, \dots, Z_p$$

sans corrélation spatiale:

$$C_{ij}(h) = \text{Cov}[Z_i(x), Z_j(x+h)] = 0 \quad \forall h$$

ou

$$\gamma_{ij}(h) = 0 \quad \forall h$$

Alors le cokrigeage (simple ou ordinaire) d'une des variables, par exemple Z_1 , se réduit à son krigeage propre:

$$Z_i(x_0)^{CK} = Z_i(x_0)^K$$

On pourra vérifier que les poids de krigeage sont en effet solution du système de cokrigeage. Ceci suppose cependant que les données (éventuellement hétérotopiques) sont constituées par les valeurs de ces variables. De plus la simplification disparaît si les moyennes sont inconnues mais liées. Voir les exercices suivants pour ces exceptions.

5.2.1 Exercice: variable composite

Soit $Z_1(x)$ et $Z_2(x)$ deux variables sans corrélation spatiale, et $Z(x) = Z_2(x) - Z_1(x)$.

On connaît $Z(A)$ et $Z_2(B)$.

Cokrigeage simple de $Z_1(O)$?

Cokrigeage ordinaire de $Z_1(O)$?

5.2.2 Exercice: moyennes (ou dérivées) liées

Soit $Z_1(x)$ et $Z_2(x)$ deux variables sans corrélation spatiale, mais de même moyenne m inconnue.

On connaît $Z_1(A)$ et $Z_2(B)$.

Cokrigeage de $Z_1(O)$?

5.3 Modèles factorisés ou de type « isofactoriel »

Dans ces modèles, on passe des variables Z_1, Z_2, \dots, Z_p à des "facteurs" Y^k sans corrélation spatiale, et inversement :

$$Y^k(x) = \sum_i h_i^k Z_i(x)$$

$$Z_i(x) = \sum_k a_i^k Y^k(x)$$

le nombre de Y^k n'excédant pas celui des Z_i .

On a:

$$\gamma_{ij}(h) = \sum_k a_i^k a_j^k \gamma^k(h)$$

Si la structure de l'information s'y prête (par exemple dans le cas isotopique, où les Z_i , donc les Y^k , sont connus aux mêmes points), on est ramené au cas précédent de variables sans corrélation spatiale (les Y^k), d'où l'on déduit par combinaison linéaire le cokrigage des Z_i :

$$[Z_i(x)]^{CK} = \sum_k a_i^k [Y^k(x)]^K$$

Les modèles isofactoriels sont très utilisés en géostatistique non linéaire (krigeage disjonctif).

5.3.1 Recherche des facteurs

Il est important de noter que les facteurs issus d'une méthode statistique de factorisation (composantes principales, par exemple), sans corrélation au même point x , n'ont a priori aucune raison d'être sans corrélation spatiale. Par ailleurs tout modèle de corégionalisation multivariable ne se décompose pas forcément en facteurs sans corrélation spatiale. Il est cependant possible de déterminer une famille de facteurs sans corrélation à distance 0 (i.e. au même point) ainsi qu'à une autre distance préalablement choisie, disons h_0 : c'est ce que proposent Desbarats et Dimitrakopoulos (2000), se référant à Switzer et Green (1984), en recherchant les facteurs expliquant une continuité spatiale (corrélogramme) de moins en moins forte pour cette distance (« min/max autocorrelation factors »). Reste ensuite à vérifier dans quelle mesure la corrélation spatiale entre ces facteurs aux autres distances peut être négligée. Cette méthode est très attrayante lorsque le nombre de variables est élevé, car elle permet de retenir les facteurs les mieux structurés spatialement, au détriment des facteurs quasi-pépitiques (on notera que les facteurs présentant la meilleure continuité spatiale n'expliquent pas forcément le plus de variabilité statistique, ce qui est une différence essentielle avec les composantes principales).

5.4 Variables autokrigeables

5.4.1 Exercice préliminaire: effet d'écran en cokrigage

Cokrigage simple de $Z_1(x)$ connaissant $Z_1(x+h)$ et $Z_2(x+h)$?

Condition pour que le poids de Z_2 soit nul?

5.4.2 Autokrigeabilité

Soit $Z_1(x), \dots, Z_i(x), \dots, Z_p(x)$ un groupe de variables. Une variable, par exemple Z_1 , est dite autokrigeable si, dans le cas isotopique, son cokrigage coïncide avec son krigeage propre.

Une condition pour cela est que sa structure croisée avec les autres variables soit identique (plus exactement : proportionnelle) à sa structure propre:

$$C_{1j}(h) \equiv C_1(h)$$

ou

$$\gamma_{1j}(h) \equiv \gamma_1(h)$$

pour tout j . (A partir de l'exercice préliminaire, on pourra montrer qu'il s'agit bien d'une condition nécessaire et suffisante dans le cas stationnaire à moyennes connues ; le cas intrinsèque est plus complexe.)

Supposons qu'il existe deux variables autokrigeables, par exemple Z_1 et Z_2 . Alors, soit elles ont même structure, et sont donc en corrélation intrinsèque ainsi que leurs combinaisons linéaires (voir section 3.1.3) :

$$\gamma_1(h) \equiv \gamma_{12}(h) \equiv \gamma_2(h)$$

soit elles ont des structures différentes, et sont alors sans corrélation spatiale:

$$\gamma_{12}(h) = 0 \quad \forall h$$

On en déduit plus généralement que les variables (ou combinaisons linéaires de variables) autokrigeables se répartissent en groupes, sans corrélation spatiale entre groupes, chaque groupe étant constitué de variables en corrélation intrinsèque avec une structure différente.

(Mathématiquement : les FA autokrigeables appartient à des sous-espaces orthogonaux caractérisés chacun par une structure propre.)

Deux cas sont spécialement intéressants :

- on peut ramener, par transformation linéaire, l'ensemble des variables à de nouvelles variables sans corrélation spatiale et de structures en général différentes : ce sont les modèles factorisés vus plus haut ;
- toutes les variables (ainsi que leurs combinaisons linéaires) sont autokrigeables et dans un seul groupe : c'est le modèle de corrélation intrinsèque déjà vu, caractérisé par des structures simples et croisées toutes identiques.

5.5 Modèle à résidu entre 2 variables

Ce modèle, encore appelé parfois modèle de Markov, est un modèle bivariable factorisé dans lequel une des variables est autokrigeable.

5.5.1 Caractérisation

Considérons deux variables $Z_1(x)$ et $Z_2(x)$ dont la structure croisée est identique à celle de l'une d'entre elles, par exemple $Z_1(x)$:

$$C_{12}(h) = a C_1(h)$$

ou:

$$\gamma_{12}(h) = a \gamma_1(h)$$

(Autrement dit, $Z_1(x)$ est autokrigeable.) Alors le résidu de la régression linéaire de $Z_2(x)$ sur $Z_1(x)$:

$$R(x) = Z_2(x) - a Z_1(x) - b$$

(par construction de moyenne 0 et non corrélé à $Z_1(x)$ au même point) est sans corrélation spatiale avec $Z_1(x)$:

$$Cov[Z_1(x), R(x+h)] = C_{12}(h) - aC_1(h) = 0$$

Le modèle se factorise donc à l'aide des facteurs $Z_1(x)$ et $R(x)$:

$$Z_2(x) = a Z_1(x) + b + R(x)$$

$$\gamma_2(h) = a^2 \gamma_1(h) + \gamma_R(h)$$

A noter que la seule connaissance de la structure de la variable $Z_1(x)$ ne donne aucune indication sur la structure résiduelle. (Rien n'empêche d'ailleurs que la structure du résidu soit plus continue que celle de Z_1 .)

Remarque : Dans le cas intrinsèque pur (variogramme mais pas de covariance stationnaire), on doit raisonner en accroissements : b disparaît, et ce sont les accroissements $R(x+h)-R(x) = [Z_2(x+h) - Z_2(x)] - a [Z_1(x+h) - Z_1(x)]$ qui sont sans corrélation avec $[Z_1(x+h) - Z_1(x)]$.

5.5.2 Simplification du cokrigage

Supposons que $Z_2(x)$ soit connu sur le même ensemble (cas isotopique) ou sur un sous-ensemble des points de $Z_1(x)$. Alors la connaissance de Z_1 et de Z_2 en ces différents points est équivalente à celle de Z_1 et de R . On a donc, grâce à la factorisation :

$$\begin{cases} Z_1^{CK} = Z_1^K \\ Z_2^{CK} = a Z_1^K + b + R^K \end{cases}$$

Il peut s'agir du cokrigage simple (à partir du krigeage simple de Z_1 et de R), ou du cokrigage ordinaire. Remplaçant R par sa définition, on vérifiera que, dans le cokrigage ordinaire de Z_2 , la somme des poids de Z_2 vaut bien 1, celle des poids de Z_1 , 0.

5.5.3 Krigeage du résidu ou cokrigage collocalisé

Plaçons-nous dans le cas hétérotopique où $Z_1(x)$ est connu de façon dense, i.e. en tout point voulu. Le krigeage (ou cokrigage) étant un interpolateur exact, sa valeur en un point x_0 coïncide avec la valeur vraie de $Z_1(x_0)$. Le cokrigage de $Z_2(x_0)$ correspond à l'estimation par krigeage du résidu de $Z_2(x_0)$ sur $Z_1(x_0)$, que l'on peut écrire :

$$\begin{aligned} Z_2^{CK}(x_0) &= a Z_1(x_0) + b + R^K(x_0) \\ &= a Z_1(x_0) + b + \sum_{\alpha} \lambda_{\alpha} [Z_{2\alpha} - a Z_{1\alpha} - b] \\ &= \sum_{\alpha} \lambda_{\alpha} Z_{2\alpha} + a \left(Z_1(x_0) - \sum_{\alpha} \lambda_{\alpha} Z_{1\alpha} \right) + b \left(1 - \sum_{\alpha} \lambda_{\alpha} \right) \end{aligned}$$

C'est un cokrigage « collocalisé » (« collocated cokriging ») car il utilise Z_1 seulement en x_0 et là où Z_2 est connu : l'ensemble des autres points où Z_1 est disponible reçoivent un poids nul et n'apparaissent donc pas.

Il peut s'agir d'un cokrigage simple (correspondant au krigeage simple du résidu), ou d'un cokrigage ordinaire (correspondant au krigeage ordinaire du résidu, avec la condition $\sum \lambda_{\alpha} = 1$ filtrant b).

5.5.4 Exercice : « dislocated cokriging »

Considérons, dans le cas stationnaire, un modèle à résidu, par exemple :

$$Z_2(x) = Z_1(x) + R(x)$$

où $R(x)$ est sans corrélation spatiale avec $Z_1(x)$ et de structure pépétique. Par exemple $Z_2(x)$ peut représenter une mesure de $Z_1(x)$ affectée d'une erreur $R(x)$ (voir exercice 3.3). Il s'agit d'un modèle à résidu avec $\gamma_{12}(h) = \gamma_1(h)$ et $\gamma_2(h) = \gamma_1(h) + \gamma_R(h)$.

On se place dans un cas hétérotopique quelconque, où Z_1 et Z_2 sont connus en des points éventuellement différents.

Montrer que :

- le cokrigeage simple de $Z_1(x_0)$ ne dépend de $Z_2(x)$ qu'aux points où $Z_1(x)$ est inconnu (c'est un cokrigeage dis-loqué - « dislocated cokriging » - à l'opposé du collocated cokriging) ;
- et que le cokrigeage simple de $Z_2(x_0)$, là où il est inconnu, ne dépend de $Z_2(x)$ également qu'aux points où $Z_1(x)$ est inconnu.

5.5.5 Exercice: Modèle à résidu et corrélation intrinsèque

Soit deux variables $Z_1(x)$ et $Z_2(x)$ avec $Z_1(x)$ autokrigeable. Autrement dit, nous sommes dans un modèle à résidu.

Montrer qu'il s'agit d'un modèle de corrélation intrinsèque si et seulement si la structure du résidu est identique à celle de Z_1 .

Z_1 , Z_2 , ou toute combinaison linéaire est alors autokrigeable.

5.5.6 Exercice: Exemples de modèles à résidu

Est-ce des modèles à résidus? Décomposition des variables? Structure du résidu?

$$\gamma_1(h) = 16 \text{ sph}(h/20)$$

$$\gamma_2(h) = 25 \text{ sph}(h/20) + 25 \text{ sph}(h/100)$$

$$\gamma_{12}(h) = 16 \text{ sph}(h/20)$$

$$\gamma_1(h) = 16 \text{ sph}(h/20) + 16 \text{ sph}(h/100)$$

$$\gamma_2(h) = 25 \text{ sph}(h/20) + 25 \text{ sph}(h/100)$$

$$\gamma_{12}(h) = 16 \text{ sph}(h/20)$$

$$\gamma_1(h) = 16 \text{ sph}(h/20) + 16 \text{ sph}(h/100)$$

$$\gamma_2(h) = 25 \text{ sph}(h/20)$$

$$\gamma_{12}(h) = 16 \text{ sph}(h/20)$$

5.5.7 Passage à un modèle à dérive

Plaçons-nous dans le modèle à résidu, et supposons, comme pour le krigeage du résidu, que Z_2 est connue de façon dense. Supposons de plus que Z_1 et R sont, non seulement sans corrélation spatiale, mais encore indépendants. En conditionnant le modèle par les valeurs z_1 de Z_1 , cette variable devient déterministe et on a :

$$Z_2(x) = a z_1(x) + b + R(x) = E[Z_2(x)] + R(x)$$

ce qui permet de considérer $a z_1(x) + b$ comme la dérive de $Z_2(x)$:

$$E[Z_2(x)] = a z_1(x) + b$$

définie de façon externe à partir de $z_1(x)$.

Attention: z_1 étant déterministe, la covariance (usuelle, centrée) de Z_2 à utiliser est maintenant celle du résidu:

$$\begin{aligned} \text{Cov}[Z_2(x), Z_2(x+h)] &= E\{[Z_2(x) - a z_1(x) - b][Z_2(x+h) - a z_1(x+h) - b]\} \\ &= \text{cov}[R(x), R(x+h)] \end{aligned}$$

5.5.8 Krigeage en dérive externe

Considérons la forme de l'estimateur de cokrigeage, mais cette fois-ci avec z_1 déterministe. Ceci ne change en rien l'estimation par krigeage du résidu. Si b est inconnu, l'estimation s'en affranchit en imposant la condition $\sum \lambda_\alpha = 1$ (krigeage ordinaire). Si maintenant a et b sont tous deux inconnus, l'estimation est réalisable en imposant les conditions :

$$\sum_{\alpha} \lambda_{\alpha} = 1 \quad \text{et} \quad \sum_{\alpha} \lambda_{\alpha} z_{1\alpha} = z_{10}$$

C'est le krigeage en dérive externe. On montrera que les poids sont solutions du système linéaire suivant :

(système de krigeage en dérive externe)

$$\left\{ \begin{array}{l} \sum_{\beta} \lambda_{\beta} \text{Cov}(R_{\alpha}, R_{\beta}) + \mu_0 + \mu_1 z_{1\alpha} = \text{Cov}(R_0, R_{\alpha}) \quad \forall x_{\alpha} \\ \sum_{\alpha} \lambda_{\alpha} = 1 \\ \sum_{\alpha} \lambda_{\alpha} z_{1\alpha} = z_{10} \end{array} \right.$$

soit matriciellement:

$$\begin{bmatrix} \underline{C} & 1 & \underline{z}_1 \\ 1 & 0 & 0 \\ \underline{z}_1^t & 0 & 0 \end{bmatrix} \times \begin{bmatrix} \underline{\lambda} \\ \mu_0 \\ \mu_1 \end{bmatrix} = \begin{bmatrix} \underline{C}_{01} \\ 1 \\ z_{10} \end{bmatrix}$$

La variance de krigeage s'écrit:

$$\sigma_{KDE}^2 = \text{var}(\varepsilon) = \text{Cov}(R_0, R_0) - \sum_{\alpha} \lambda_{\alpha} \text{Cov}(R_0, R_{\alpha}) - \mu_0 - \mu_1 z_{10}$$

soit matriciellement :

$$\sigma_{KDE}^2 = \text{var}(\varepsilon) = \text{Cov}(R_0, R_0) - \begin{bmatrix} \lambda \\ \mu_0 \\ \mu_1 \end{bmatrix}^t \begin{bmatrix} C_{01} \\ 1 \\ z_{10} \end{bmatrix}$$

5.5.9 Remarques sur krigeage du résidu et krigeage en dérive externe

- La variable Z_1 , connue partout, donc ne faisant pas l'objet d'une estimation, ne requiert en fait aucune hypothèse de stationnarité.

- L'hypothèse de base est la non-corrélation spatiale, ou même l'indépendance, entre la variable dense et le résidu par rapport à celle-ci.

- L'estimation ne dépend que de la structure du résidu. La détermination directe de cette structure pose cependant problème en dérive externe (les accroissements du résidu, et donc son variogramme, dépendent a priori de a , considéré comme inconnu); on peut alors avoir recours à une validation croisée pour estimer la structure.

On peut étendre les méthodes de krigeage du résidu et de krigeage en dérive externe à plusieurs variables conditionnantes. Seule importe alors la structure du résidu de la régression linéaire, en un point donné, de la variable cible à partir des variables conditionnantes. Ce résidu est supposé sans corrélation spatiale avec les variables conditionnantes.

6 L'analyse krigeante

L'analyse krigeante permet, dans un modèle linéaire de régionalisation ou de corégionalisation, de filtrer et de cartographier une ou plusieurs composantes.

6.1 Cas monovariante stationnaire

Considérons une variable $Z(x)$ de moyenne m et de structure:

$$\gamma(h) = b^1 \gamma^1(h) + b^2 \gamma^2(h) + b^3 \gamma^3(h)$$

où les composantes structurales $\gamma^k(h)$ décrivent par exemple des échelles k différentes. Ce modèle peut s'interpréter comme une décomposition de la variable en composantes (choisies de moyenne nulle) aux différentes échelles :

$$Z(x) = Z^1(x) + Z^2(x) + \dots + m$$

On a:

$$\text{Cov}[Z^k(x), Z(x+h)] = \text{Cov}[Z^k(x), Z^k(x+h)]$$

ce qui permet d'effectuer le cokrigeage de chaque $Z^k(x_0)$ à partir des seules données $Z_\alpha = Z(x_\alpha)$ sur Z .

Si la moyenne m est connue, ce sera un cokrigeage simple :

$$\left(Z^k(x_0) \right)^{CKS} = \sum_{\alpha} \lambda_{\alpha} (Z_{\alpha} - m)$$

A titre d'exercice, on pourra développer le système (identique au système de krigeage de $Z(x_0)$, en changeant le membre de droite selon la composante Z^k à estimer).

A moyenne inconnue, il convient de faire attention au non-biais. Ainsi, pour l'estimation de $Z^3(x_0)$:

$$\begin{aligned} \left[Z^3(x_0) \right]^{CK} &= \sum \lambda_{\alpha} Z_{\alpha} \\ \varepsilon &= Z_0^3 - \sum \lambda_{\alpha} Z_{\alpha} \\ E(\varepsilon) &= 0 - m \sum \lambda_{\alpha} = 0 \Rightarrow \sum \lambda_{\alpha} = 0 \end{aligned}$$

et pour celle de $[Z^3(x_0) + m]$:

$$\begin{aligned} \left[Z^3(x_0) + m \right]^{CK} &= \sum \lambda_{\alpha} Z_{\alpha} \\ \varepsilon &= Z^3(x_0) + m - \sum \lambda_{\alpha} Z_{\alpha} \\ E(\varepsilon) &= 0 + m \left(1 - \sum \lambda_{\alpha} \right) = 0 \Rightarrow \sum \lambda_{\alpha} = 1 \end{aligned}$$

On a la cohérence :

$$\left[Z_0^3 + m \right]^{CK} = \left(Z_0^3 \right)^{CK} + m^K$$

ainsi que globalement:

$$Z_0^K = (Z_0^1)^{CK} + (Z_0^2)^{CK} + (Z_0^3)^{CK} + m^K$$

due à la linéarité des pondérateurs par rapport au second membre du système, qui change selon la quantité à estimer.

Attention, bien que les facteurs cachés Z^k (dépendant de la décomposition structurale, inconnus et non objectifs) soient sans corrélation spatiale, leurs estimations, faites à partir de données communes, sont corrélées.

6.2 Cas monovariante intrinsèque

Le cas intrinsèque est un peu plus compliqué. Considérons par exemple une variable de structure :

$$\gamma(h) = b^1 \gamma^1(h) + b^2 \gamma^2(h) + b^3 \gamma^3(h)$$

On peut l'interpréter comme la somme de 3 composantes intrinsèques sans corrélation spatiale:

$$Z(x) = Z^1(x) + Z^2(x) + Z^3(x)$$

Supposons que l'on cherche à estimer $Z^3(x_0)$. L'erreur s'écrit:

$$\varepsilon = Z_0^3 - \sum \lambda_\alpha Z_\alpha^k = Z_0^3 - \sum \lambda_\alpha (Z_\alpha^1 + Z_\alpha^2 + Z_\alpha^3) = Z_0^3 - \sum \lambda_\alpha Z_\alpha^3 - \sum \lambda_\alpha Z_\alpha^1 - \sum \lambda_\alpha Z_\alpha^2$$

Pour être une combinaison linéaire autorisée, il faudrait à la fois $\sum \lambda_\alpha$ égal à 1 et à 0, ce qui est impossible. On est donc conduit à ne considérer qu'une seule composante purement intrinsèque (disons Z^3), les autres (Z^1 et Z^2) étant stationnaires et de moyenne nulle. On a alors, par exemple:

$$\begin{aligned} [Z^2(x_0)]^{CK} &= \sum \lambda_\alpha Z_\alpha \\ E(\varepsilon) = 0 &\Rightarrow \sum \lambda_\alpha = 0 \end{aligned}$$

ou

$$\begin{aligned} [Z^3(x_0)]^{CK} &= \sum \lambda_\alpha Z_\alpha \\ E(\varepsilon) = 0 &\Rightarrow \sum \lambda_\alpha = 1 \end{aligned}$$

On notera la similitude entre l'estimation de la composante intrinsèque Z^3 dans ce cadre, et l'estimation de (Z^3+m) dans le cas stationnaire à moyenne inconnue :

$$Z(x) = Z^1(x) + Z^2(x) + Z^3(x) + m$$

6.3 Cas multivariante

Le cas multivariante est similaire :

$$\gamma_{ij}(h) = \sum_k b_{ij}^k \gamma^k(h)$$

Dans le cas stationnaire par exemple, ceci peut s'interpréter comme :

$$\begin{aligned} Z_i(x) &= \sum_k Z_i^k(x) + m_i \\ &= \sum_k \sum_q a_{iq}^k Y_q^k(x) + m_i \end{aligned}$$

L'analyse krigéante consiste à extraire, par cokrigéage, les Z_i^k (non objectifs, donnés par la décomposition structurale) ou les Y_q^k (dépendant en outre de la méthode choisie pour factoriser les composantes Z_i^k d'échelle k).

7 Synthèse et autres remarques

7.1 Modèles multivariés

Nous avons vu:

- le modèle de corrélation intrinsèque, où toutes les structures sont identiques ;
- le modèle linéaire de corégionalisation, où toutes les structures sont des combinaisons linéaires d'un certain nombre de structures de base ;
- les modèles factorisés, essentiellement le cas particulier du modèle à résidu, dans lequel les facteurs sont constitués par la variable autokrigeable et par le résidu (ce modèle, qui préside à l'estimation par krigeage du résidu et au krigeage en dérive externe, peut notamment représenter un modèle d'erreur ou un modèle multi-support).

On peut compléter de tels modèles, ou en imaginer bien d'autres, selon le cas traité, par exemple :

- modèles incluant des retards entre certaines composantes ;
- modèles construits à l'aide de convoluées (régularisées) ou de dérivées ;
- modèles d'indicatrices résultant du seuillage d'une FA ;
- etc.

7.2 Simulation

Les modèles présentés ici se prêtent bien à simulation dans le cas (multi-)gaussien (i.e. toute combinaison linéaire est gaussienne), l'absence de corrélation étant alors équivalente à l'indépendance:

- la simulation de p variables en corrélation intrinsèque peut s'obtenir à partir de simulations indépendantes des facteurs issus d'une factorisation quelconque (en nombre p dans le cas général);
- le modèle linéaire de corégionalisation, pour p variables en fonction de k structures de base, se simule également à partir de simulations indépendantes des facteurs (en nombre kp , en général) ;
- le modèle factorisé de p variables se simule à partir de simulations indépendantes de chaque facteur (au nombre de p en général, soit deux dans le modèle à résidu bivariable), chaque facteur pouvant cependant obéir à une structure gigogne.

7.3 Simplification du cokrigeage

Si une des variables, par exemple $Z_1(x)$, a une structure croisée avec chacune des autres variables identique à sa propre structure, alors elle est autokrigeable : son cokrigeage coïncide avec son krigeage propre, dans le cas isotopique, mais également en hétérotopique si Z_1 est connu sur l'ensemble des points échantillonnés.

Dans le cas de deux variables et dans une telle configuration, il peut être avantageux de remplacer $Z_2(x)$ par le résidu $R(x)$ de sa régression linéaire par rapport à la variable autokrigeable $Z_1(x)$ au même point, résidu qui est en effet sans corrélation spatiale avec $Z_1(x)$. C'est le modèle à résidu, où tout cokrigeage s'obtient par un krigeage des facteurs $Z_1(x)$ et de $R(x)$, du moins si $Z_1(x)$ est connu en tout point expérimental.

Toutes ces remarques s'appliquent en particulier au modèle de corrélation intrinsèque : le cokrigeage de Z_1 s'identifie à son krigeage, à supposer que Z_1 est connu en tout point expérimental.

Enfin, de façon plus générale, dans les modèles factorisés, le cokrigeage se réduit au krigeage propre de chacun des facteurs, à supposer que les données puissent s'exprimer en valeurs prises individuellement par ces facteurs (ce qui est notamment le cas en isotopique).

7.4 Voisinage de cokrigeage

Dans le cas monovarié, le choix du voisinage de krigeage (entendons par là les données voisines que l'on retiendra, par exemple les N données les plus proches par secteur angulaire) dépend :

- de l'implantation (nombre, répartition, en particulier densité géographique) des données ;
- de la structure de la variable ;
- de la distance limite de l'hypothèse de stationnarité faite (sur la variable ou ses accroissements).

Différents paramètres peuvent aider à ce choix : variance de krigeage (pour la précision), poids de la moyenne en krigeage simple (pour voir l'influence de l'estimation de la moyenne en krigeage ordinaire, notamment), pente de la régression linéaire valeur vraie sur valeur estimée en krigeage ordinaire (pour se rapprocher du non-biais conditionnel où cette pente vaut 1).

Les mêmes considérations s'appliquent également en multivarié, mais la situation est notablement plus complexe.

Notons, pour mémoire, que le cokrigeage d'une variable est assuré d'être au moins aussi précis que son krigeage propre, si le voisinage de cokrigeage inclut le voisinage de krigeage (ce qui ne va pas toujours de soi en pratique, pour des raisons de nombre de données, notamment) et si bien sûr les modèles mono- et multivarié sont cohérents.

En outre, pour assurer la cohérence entre valeurs cokrigées des différentes variables, il est nécessaire de prendre un voisinage identique pour l'estimation de chacune des variables. Dans le cas isotopique, ceci incite à choisir le voisinage d'après la variable nécessitant le plus grand voisinage (alors que des voisinages plus petits peuvent être suffisants pour les autres variables).

Le cas hétérotopique est beaucoup plus difficile que le cas isotopique, car outre l'implantation géographique des valeurs de chaque variable, joue (à travers le modèle) leur agencement mutuel.

Ainsi, dans le cas de l'estimation d'une variable Z_1 à l'aide d'échantillons sur cette variable et d'une variable Z_2 connue de façon dense, un voisinage collocalisé consistera à retenir un voisinage incluant, outre les données de Z_1 , celles de Z_2 au point cible et en ces mêmes points, mais non les valeurs de Z_2 ailleurs. Un tel voisinage se justifie dans un modèle à résidu où la structure croisée entre Z_1 et Z_2 est identique à la structure de Z_2 : le cokrigeage y est en effet collocalisé, et c'est aussi le modèle permettant le krigeage en dérive externe. Mais un tel voisinage peut entraîner une perte substantielle d'information dans d'autres cas.

Si par exemple, dans un cas stationnaire hétérotopique quelconque, la structure croisée entre Z_1 et Z_2 est identique à celle de Z_1 , avec en outre un résidu de $Z_2(x)$ sur $Z_1(x)$ pépétique, le cokrigeage simple de Z_1 en un point non échantillonné ne fait usage de Z_2 que là où Z_1 est inconnu : le

voisinage est dis-loqué, en opposition radicale au voisinage collocalisé précédent, qui conduirait ici à ignorer la seconde variable Z_2 , sauf au point cible.

7.5 Usage du cokrigeage

Le cokrigeage permet d'améliorer la cohérence entre estimations de différentes variables.

Dans le cas isotopique, il peut améliorer l'estimation d'une variable, si sa structure simple contraste avec ses structures croisées aux autres variables.

Dans le cas hétérotopique, le cokrigeage permet en particulier de tenir compte d'une variable auxiliaire mieux connue. Dans le cas d'une variable auxiliaire dense et du modèle à résidu, notamment, le cokrigeage est collocalisé et coïncide avec un krigeage du résidu, on peut également avoir recours au krigeage en dérive externe.

Le cokrigeage permet encore l'extraction de composantes structurales (analyse krigeante).

Enfin il peut être utilisé pour conditionner une simulation gaussienne multivariable.

8 Indicatrices

Dans cette section, nous examinerons les outils de la géostatistique linéaire appliqués aux indicatrices. Nous considérerons successivement :

- le cas d'une indicatrice ;
- le cas d'une indicatrice et d'une autre variable ;
- le cas de plusieurs indicatrices,

et présenterons, dans ce dernier cas, quelques modèles de base et le cokrigage d'indicatrices.

8.1 Indicatrice d'un ensemble

8.1.1 Variable indicatrice d'un ensemble

La présence ou l'absence d'un ensemble A (un faciès géologique par exemple) en un point x peut être codée par la variable en 0 ou 1 dite indicatrice:

$$1_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

Comme un point appartient nécessairement, soit à A, soit à son complémentaire A^c , leurs indicatrices valent respectivement 1 et 0 ou bien 0 et 1. On a donc :

$$1_A(x) + 1_{A^c}(x) = 1$$

et $1_A(x)1_{A^c}(x) = 0$

On notera les propriétés suivantes :

$$1_A(x)1_A(y) \text{ vaut } 1 \text{ si } x \text{ et } y \text{ appartiennent tous deux à } A, 0 \text{ sinon.}$$

$$1_A(x) = [1_A(x)]^2$$

Moyenne de l'indicatrice en différents points:

$$\frac{1}{N} \sum_i 1_A(x_i) = \text{proportion de points } x_i \text{ appartenant à } A.$$

$$\frac{1}{V} \int_V 1_A(x) dx = \text{proportion de points de } V \text{ appartenant à } A.$$

8.1.2 Modèle probabiliste

L'ensemble A étant maintenant considéré comme un ensemble aléatoire (ensemble dont forme et localisation sont aléatoires), sa variable indicatrice est représentée par une Fonction Aléatoire en 0/1. Cette indicatrice vérifie :

$$E[1_A(x)] = P(x \in A), \text{ soit } p_A \text{ indépendante de } x, \text{ si cette probabilité est stationnaire (c'est-à-dire invariante par translation)}$$

$$E[1_{A^c}(x)] = 1 - p_A$$

$$\text{var}[1_A(x)] = p_A(1-p_A) = \text{var}[1_{A^c}(x)] \leq 0.25$$

Covariance non-centrée de (l'indicatrice de) A:

$$E[1_A(x)1_A(x+h)] = P(x \in A, x+h \in A), \text{ soit } K_A(h) \text{ si stationnaire}$$

Covariance centrée :

$$\text{cov}[1_A(x), 1_A(x+h)] = C_A(h) \text{ si stationnaire}$$

Variogramme :

$$\frac{1}{2}E([1_A(x+h) - 1_A(x)]^2) = \frac{1}{2}[P(x \in A, x+h \notin A) + P(x \notin A, x+h \in A)]$$

$$\text{soit : } P(x \in A, x+h \notin A) = \gamma_A(h) = \gamma_{A^c}(h) \leq 0.5$$

dans le cas stationnaire. (Sa variance étant bornée, une indicatrice qui satisfait à l'hypothèse intrinsèque, ayant donc des accroissements stationnaires, est nécessairement stationnaire.)

On notera que les variogrammes d'un ensemble et de son complémentaire sont les mêmes. Comme un ensemble et son complémentaire n'ont pas la même connexité (cf par exemple un ensemble constitué de lentilles disjointes – disconnectées - entourées par l'ensemble complémentaire : celui-ci est connexe, on peut se rendre d'un point à un autre sans en sortir), on en déduit qu'un variogramme ne peut pas caractériser la connexité d'un ensemble.

On a :

$$P(x+h \in A \mid x \in A) = \frac{K_A(h)}{p_A}$$

$$P(x+h \notin A \mid x \in A) = \frac{\gamma_A(h)}{p_A}$$

Le variogramme d'un ensemble aléatoire ne peut être quelconque. On a en particulier nécessairement :

$$|1_A(z) - 1_A(x)| \leq |1_A(z) - 1_A(y)| + |1_A(y) - 1_A(x)|$$

$$\text{d'où: } 0.5 E[1_A(z) - 1_A(x)]^2 \leq 0.5 E[1_A(z) - 1_A(y)]^2 + 0.5 E[1_A(y) - 1_A(x)]^2$$

et donc l'inégalité triangulaire:

$$\gamma_A(h+h') \leq \gamma_A(h) + \gamma_A(h')$$

Il s'ensuit que:

$$\gamma_A(2h) \leq 2\gamma_A(h)$$

d'où un comportement à l'origine:

$$\gamma_A(h) \propto |h|^\alpha \quad \alpha \leq 1$$

α est lié à l'irrégularité des frontières de A. Il vaut au maximum 1 (variogramme linéaire aux petites distances) pour un ensemble aux contours réguliers. Un comportement parabolique (en h^2) aux petites distances est donc interdit.

8.1.3 Krigeage d'une indicatrice

Supposons les données $x_\alpha \in A, \dots, x_\beta \notin A, \dots$ codées par les indicatrices $1_A(x_\alpha)$, égales à 0 ou 1.

La probabilité conditionnelle $P(x_0 \in A | x_\alpha \in A, \dots, x_\beta \notin A, \dots)$ est en général inaccessible. On peut alors avoir recours au krigeage de l'indicatrice, par exemple un krigeage simple:

$$[1_A(x_0)]^K = \sum_{\alpha=1}^n \lambda_\alpha 1_A(x_\alpha) + \left[1 - \sum_{\alpha=1}^n \lambda_\alpha \right] p_A$$

qui a le sens d'une probabilité, mais peut sortir de $[0, 1]$, ce qui impose une correction postérieure.

8.2 Cas d'une indicatrice et d'une autre variable

8.2.1 Cas général

On considère ici un ensemble A représenté par son indicatrice $1_A(x)$, et une variable $Z(x)$. On peut alors calculer et comparer les moyennes $E[Z(x) | x \in A]$ et $E[Z(x) | x \in A^c]$ de $Z(x)$ sur A et sur son complémentaire A^c respectivement.

Voyons l'interprétation que l'on peut faire dans ce cas des outils structuraux croisés.

La covariance croisée non-centrée:

$$E[1_A(x)Z(x+h)] = E[Z(x+h) | x \in A]P(x \in A)$$

fournit, au facteur $P(x \in A)$ près, la moyenne de Z à distance h d'un point x de A. Le rapport entre variogramme croisé et variogramme simple de A donne quant à lui l'écart moyen entre deux points distants de h de part et d'autre de la frontière:

$$\frac{\gamma_{A,Z}(h)}{\gamma_A(h)} = E[Z(x+h) - Z(x) | x+h \in A, x \notin A]$$

(dans l'hypothèse où cette quantité est symétrique en h).

Naturellement, on peut aussi comparer la variabilité spatiale (variogramme) de $Z(x)$ calculée, sur l'ensemble du champ, sur A (les points courants x et x+h appartenant à A), sur son complémentaire, ou à cheval entre A et A^c .

Si l'on s'intéresse au comportement de $Z(x)$ dans A, on peut également considérer la nouvelle variable (restriction de $Z(x)$ à A) $Z(x)1_A(x)$, égale à $Z(x)$ dans A et à 0 en dehors. Ce sera le cas suivant.

8.2.2 Indicatrice et variable simultanément nulles

On s'intéresse ici à un ensemble A, représenté par son indicatrice $1_A(x)$, et à une variable $Z(x)$ nulle en dehors de A. On a donc la relation:

$$Z(x) = Z(x)1_A(x)$$

Le rapport entre covariance croisée non centrée et covariance non centrée de A

$$\frac{E[1_A(x)Z(x+h)]}{E[1_A(x)1_A(x+h)]} = E[Z(x+h) | x \in A, x+h \in A]$$

égale la moyenne de Z sur l'intersection entre l'ensemble A et son translaté par h. Le rapport entre variogramme croisé et variogramme simple de A s'écrit alors (toujours sous hypothèse de symétrie):

$$\frac{\gamma_{A,Z}(h)}{\gamma_A(h)} = E[Z(x+h) | x+h \in A, x \notin A]$$

Son évolution en fonction de h mesure l'effet des bordures de A sur Z. Il est constant en l'absence d'effet de bord. Ceci est notamment le cas si l'on peut considérer Z(x) comme résultant de la restriction à A d'une variable indépendante de A (dont la structure variographique est alors accessible à partir des seules valeurs de Z(x) dans A).

8.2.3 Variables utiles

Par variables utiles, on entend une paire de variables additives (se prêtant donc bien à une estimation linéaire), dont le rapport est une variable d'intérêt non additive, typiquement en mine :

T(x), représentant le minerai,

Q(x), le métal,

de rapport $Z(x) = \frac{Q(x)}{T(x)}$ égale à la teneur en métal.

C'est le cas classique d'un gisement stratiforme ou filonien étudié à 2 D avec :

T(x) puissance (épaisseur de la couche) au point 2D x,

Q(x) accumulation-métal sur l'épaisseur de la couche,

dont le rapport est égale à la teneur de la puissance.

Soit un bloc défini par le support v de son emprise à 2D. Par définition, sa teneur moyenne Z(v), rapport entre métal et minerai, s'écrit :

$$Z(v) = \frac{\int_v Q(x)dx}{\int_v T(x)dx} = \frac{\frac{1}{v} \int_v Q(x)dx}{\frac{1}{v} \int_v T(x)dx}$$

soit le rapport des régularisées de Q et T :

$$Z(v) = \frac{Q(v)}{T(v)}$$

ou encore:
$$Z(v) = \frac{1}{v} \int_v Z(x) \left(\frac{T(x)}{\frac{1}{v} \int_v T(y) dy} \right) dx$$

Ainsi la teneur moyenne n'est pas la moyenne directe des teneurs, mais la moyenne des teneurs pondérée par les puissances.

De la même façon, on pourra déduire une estimation de la teneur (en un point ou sur un bloc) à partir d'une estimation linéaire des variables utiles puissance et accumulation (par exemple, krigeage ou, mieux, cokrigeage).

Du fait de la relation $Q(x)=Z(x)T(x)$, les 3 variables ne sont pas indépendantes et peuvent présenter des relations variées, aussi est-il instructif d'examiner leurs nuages de corrélation. En particulier:

- $Z(x)$ peut être négativement corrélé à $T(x)$, notamment si $T(x)$ et $Q(x)$ sont indépendants;
- $Z(x)$ peut, en moyenne, augmenter avec $T(x)$ (corrélation positive), alors $Q(x)$ également;
- $Z(x)$ peut être non corrélé ou indépendant de $T(x)$, par exemple si $Q(x)$ varie proportionnellement à $T(x)$.

Les relations structurales, elles aussi, peuvent être variées. En particulier l'observation d'une structure croisée entre T et Q proportionnelle à la structure simple de T correspond à un modèle à résidu avec T autokrigeable:

$$Q(x) = aT(x) + b + R(x)$$

Ce peut être notamment le cas lorsque, en moyenne, le métal varie linéairement en fonction du minerai:

$$E[Q(x)|T(x)] = aT(x) + b \text{ (proportionnel à } T(x) \text{ si } b = 0)$$

soit:

$$E[Z(x)|T(x)] = a + b/T(x) \text{ (constant si } b = 0)$$

Comme autre exemple de variables utiles, citons les réserves récupérables à teneur de coupure z , soit (en désignant par $Z(x)$ une teneur 3D par exemple):

- le minerai, donné par l'indicatrice $1_{Z(x) \geq z}$ (égale à 0 ou 1), une régularisée de cette indicatrice (par exemple, la proportion de minerai sur la hauteur de gradins), ou $1_{Z(v) \geq z}$ pour un support de sélection v ;
- le métal, donné par $Z(x)1_{Z(x) \geq z}$ (égale à 0 ou $Z(x)$), une régularisée (par exemple sur la hauteur de gradins), ou $Z(v)1_{Z(v) \geq z}$,

le rapport entre ces deux variables utiles (ou leurs estimations) représentant la teneur récupérable correspondante. On est alors dans le cas d'une indicatrice et d'une variable simultanément nulles.

La considération de plusieurs coupures possibles, ainsi que le changement de support (passage d'échantillons de support quasi-ponctuel x à un bloc de sélection v) relèvent de la géostatistique non-linéaire.

8.3 Indicatrices de plusieurs ensembles

Bien distinguer:

- ensembles aléatoires indépendants, alors

$$P(x \in A, x \in B) = P(x \in A) P(x \in B)$$

- ensembles disjoints $A \cap B = \emptyset$, alors:

$$1_A(x)1_B(x) = 0 \text{ d'où aussi: } P(x \in A, x \in B) = 0$$

- ensembles emboîtés $A_2 \subset A_1$:

$$1_{A_2}(x) \leq 1_{A_1}(x)$$

8.3.1 Variographie croisée

covariance croisée non-centrée :

$$E[1_A(x)1_B(x+h)] = P(x \in A, x+h \in B) \text{ soit } K_{AB}(h) \text{ si stationnaire}$$

covariance croisée :

$$C_{AB}(h) = K_{AB}(h) - p_A p_B$$

variogramme croisé :

$$\gamma_{AB}(h) = \frac{1}{2} E[1_A(x+h) - 1_A(x)][1_B(x+h) - 1_B(x)]$$

8.3.2 Ensembles indépendants

On a alors $C_{AB}(h) = 0$ et $\gamma_{AB}(h) = 0$.

8.3.3 Ensembles disjoints

On a alors (cas symétrique)

$$\gamma_{AB}(h) = \frac{1}{2} E[1_A(x+h) - 1_A(x)][1_B(x+h) - 1_B(x)]$$

$$= -P(x \in A, x+h \in B)$$

$$\frac{-\gamma_{AB}(h)}{p_A} = P(x+h \in B \mid x \in A)$$

$$\frac{-\gamma_{AB}(h)}{\gamma_A(h)} = P(x+h \in B \mid x \in A, x+h \notin A)$$

Ce dernier rapport est constant en l'absence d'effets de bord $A \rightarrow B$ (soit : de A vers B).

Si de plus les ensembles disjoints (disons A, B et C) forment une partition, les indicatrices des ensembles sont liées par la relation de fermeture:

$$1_A(x) + 1_B(x) + 1_C(x) = 1$$

d'où les relations déjà mentionnées à ce propos, par exemple:

$$C_{AB}(h) + C_{AC}(h) = -C_A(h) \text{ négatif en } h = 0$$

et $\gamma_{AB}(h) + \gamma_{AC}(h) = -\gamma_A(h) \leq 0$

8.3.4 Ensembles emboîtés

Soit par exemple:

$$A_2 \subset A_1 \subset A_0$$

Le rapport entre covariance croisée non centrée de deux ensembles et covariance croisée de l'ensemble le plus grand donne :

$$\frac{E[1_{A_1}(x)1_{A_2}(x+h)]}{E[1_{A_1}(x)1_{A_1}(x+h)]} = \frac{P(x \in A_1, x+h \in A_2)}{P(x \in A_1, x+h \in A_1)} = P(x+h \in A_2 | x \in A_1, x+h \in A_1)$$

il est constant en l'absence d'effets de bord de A_2 dans A_1 .

On a, pour le variogramme croisé (cas symétrique) :

$$\begin{aligned} \gamma_{A_1A_2}(h) &= \frac{1}{2} E[1_{A_1}(x+h) - 1_{A_1}(x)][1_{A_2}(x+h) - 1_{A_2}(x)] \\ &= P(x \notin A_1, x+h \in A_2) \end{aligned}$$

$$\frac{\gamma_{A_1A_2}(h)}{1 - p_{A_1}} = P(x+h \in A_2 | x \notin A_1)$$

$$\frac{\gamma_{A_1A_2}(h)}{\gamma_{A_1}(h)} = P(x+h \in A_2 | x \notin A_1, x+h \in A_1)$$

Ce dernier rapport est constant en l'absence d'effets de bord de A_2 dans A_1 .

8.3.5 Remarque : relation entre ensembles disjoints et emboîtés

Le cas d'ensembles emboîtés peut se ramener à celui d'ensemble disjoints, en posant par exemple:

$$\begin{aligned} C &= A_2 \\ B &= A_1 - A_2 \\ A &= A_0 - A_1 - A_2 \end{aligned}$$

En théorie les familles des indicatrices des ensembles disjoints et des ensembles emboîtés sont équivalentes. En particulier, l'absence d'effets de bord de A_2 dans A_1 (ensembles emboîtés) correspond à l'absence d'effets de bord de A vers B (ensembles disjoints).

Inversement, il n'est intéressant de cumuler des ensembles disjoints en ensembles emboîtés que lorsque ces ensembles peuvent être ordonnés (par exemple des faciès géologiques de granulométrie décroissante).

8.3.6 Cokrigage d'indicatrices

Les indicatrices d'ensembles disjoints ou emboîtés ne sont pas indépendantes. Le cokrigage permet en général d'améliorer la cohérence entre l'estimation des différentes indicatrices (mais, de même que le krigeage, le cokrigage d'une indicatrice peut sortir de [0,1]).

8.4 Quelques modèles de base

Trois types de modèles de base sont illustrés en fin de document (figures aimablement communiquées par Hélène Beucher, donnant images et figures variographiques dans les deux axes, issues du logiciel Isatis). Ils correspondent chacun à 3 faciès disjoints, coloriés en gris (G), jaune (J) et rouge (R), en proportion à peu près égales dans les trois modèles, mais différentes d'un faciès à l'autre.

8.4.1 Modèle sans effet de bord

Il s'agit plus précisément d'un modèle sans effet de bord en sortant du gris. Sortant du gris, on peut donc trouver du jaune ou du rouge selon une probabilité indépendante de la distance. D'où par exemple un variogramme croisé rouge-gris proportionnel au variogramme du gris.

Ce modèle a été construit :

- (1) en mettant tout l'espace en rouge ;
 - (2) en générant le faciès jaune;
 - (3) en superposant indépendamment le faciès gris ;
- (on pourrait continuer à superposer séquentiellement d'autres ensembles indépendants).

Du fait de la hiérarchie inhérente à sa construction, ce modèle se prête bien à une description par cumul des faciès en ensembles emboîtés, ici : [rouge] \subset [rouge + jaune] \subset [rouge + jaune + gris].

Considérons donc un modèle sans effet de bord en montant dans les ensembles emboîtés $A_0 \supset A_1 \supset A_2 \dots$

On a alors par exemple :

$$\frac{\gamma_{A_1 A_2}(h)}{\gamma_{A_1}(h)} = P(x+h \in A_2 \mid x \notin A_1, x+h \in A_1) = \frac{p_{A_2}}{p_{A_1}}$$

On peut montrer que les indicatrices se factorisent à partir des résidus d'indicatrices :

$$H_1(x) = \frac{1_{A_1}(x)}{p_{A_1}} - 1$$
$$\dots, H_{i+1}(x) = \frac{1_{A_{i+1}}(x)}{p_{A_{i+1}}} - \frac{1_{A_i}(x)}{p_{A_i}}, \dots$$

Dans le cas isotopique, il suffit donc de kriger ces résidus pour en déduire le cokrigage de toutes les indicatrices.

8.4.2 Modèle mosaïque à valuations indépendantes

Un tel modèle se construit à partir d'une partition stationnaire de l'espace en compartiments (en l'occurrence, sur la figure: une partition de Voronoi, donnée par les zones d'influence de points poissonniens). A chaque compartiment, on attribue un faciès A, B ou C (ici gris, jaune ou rouge), selon une probabilité a priori p_A , p_B ou p_C , indépendamment des autres compartiments. (Sur la figure, on ne distingue plus chacun des compartiments initiaux: en particulier, les grandes plages de gris sont constituées de plusieurs compartiments mis chacun – par le hasard pur – en gris.)

Si on note:

$$\rho(h) = P(x \text{ et } x+h \text{ appartiennent au même compartiment})$$

on a:

$$K_A(h) = P(x \in A, x+h \in A) = p_A \rho(h) + p_A^2 [1 - \rho(h)]$$

$$K_{AB}(h) = P(x \in A, x+h \in B) = p_A p_B [1 - \rho(h)]$$

$$\gamma_A(h) = p_A(1 - p_A)[1 - \rho(h)]$$

$$\gamma_{AB}(h) = -p_A p_B [1 - \rho(h)]$$

Variogramme simples et croisés sont tous identiques. Il n'y a aucun effet de bord entre faciès.

Le modèle mosaïque à valuations indépendantes correspond précisément le cas où le cokrigage n'apporte pas d'amélioration par rapport au krigeage, du moins en isotopique. En effet les indicatrices sont ici en corrélation intrinsèque, et donc cokrigage coïncide avec krigeage :

$$[1_A(x_0)]^{CK} = [1_A(x_0)]^K$$

Autrement dit, pour estimer l'indicatrice $1_A(x_0)$, seuls comptent les $1_A(x_\alpha)$, non les $1_B(x_\alpha)$, ce qui est assez intuitif vu la construction du modèle.

8.4.3 Modèle de type diffusif

Il s'agit ici d'une FA gaussienne seuillée en gris, jaune et rouge : on passe par le jaune pour se rendre du gris au rouge. D'où les effets de bord : en sortant du gris, par exemple, la probabilité de tomber dans du rouge augmente avec la distance. Les faciès sont ici naturellement ordonnés, et un tel modèle est très utilisé en industrie pétrolière pour décrire des faciès de granulométrie croissante ou décroissante (décroissante par exemple des sables – ici en rouge – vers les argiles – gris - en passant par des sables argileux – jaune -). Du fait de cet ordre, il peut être intéressant de considérer les faciès cumulés (faciès emboîtés dans le sens des granulométries croissantes ou bien décroissantes).

Les structures simples et croisées des indicatrices ne sont pas des modèles simples : elles sont fonction de la structure simple de la FA gaussienne et des seuils (voir cours de simulation). En pratique l'ajustement de telles structures d'indicatrices de faciès passe par la recherche de la structure de la variable cachée gaussienne. Le modèle est très utilisé en simulation (plutôt qu'en estimation par cokrigage), les faciès étant alors directement obtenus par seuillage de la gaussienne simulée.

8.4.4 Vers la géostatistique non-linéaire

L'application de coupures à une FA de loi quelconque conduit à des ensembles emboîtés (par exemple, ensembles de points où la FA excède les valeurs 0, 1, 2, 3...), les valeurs prises par une FA discrète ou discrétisée formant par ailleurs une partition de l'espace en ensembles disjoints (par exemple, ensembles de points où la FA prend respectivement les valeurs 0, 1, 2, 3...). Le cokrigeage d'indicatrices, qui s'appelle alors krigeage disjonctif, permet l'estimation de toute fonction de la FA.

Il se simplifie en krigeage d'indicatrices dans le cas du modèle mosaïque à valuations indépendantes (caractérisé par une absence de déstructuration des ensembles emboîtés quand la coupure varie).

Le krigeage disjonctif est essentiellement utilisé dans le cadre de modèles isofactoriels, où il est obtenu par krigeage de facteurs :

- résidus d'indicatrices dans les modèles hiérarchiques sans effet de bord en montant, ou bien en descendant, dans les étages ;
- autres familles de facteurs dans les modèles de type diffusif (polynômes d'Hermite dans le cas gaussien anamorphosé, polynômes de Laguerre pour des lois gamma, etc).

Voir cours de géostatistique non-linéaire.

9 Références

Chilès, J.-P., and Delfiner, P. 1999. *Geostatistics: Modeling spatial uncertainty*. Wiley, New York, 695 p.

Desbarats, A. J. and Dimitrakopoulos, R. 2000. Geostatistical simulation of regionalized pore-size distributions using min/max autocorrelation factors. *Math. Geol.* Vol. 32 No 8, 919-942.

Künsch, H. R., Papritz, A., and Bassi, F. 1997. Generalized cross-covariances and their estimation. *Math. Geol.* Vol. 29 No 6, 779-799.

Switzer, P. and Green, A. A. 1984. Min/max autocorrelation factors for multivariate spatial imaging. Technical report no 6, Department of Statistics, Stanford University, 14 p.

Wackernagel, H. 1993. Cours de géostatistique multivariable. Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau, 80 p.

Wackernagel, H. 1995. *Multivariate geostatistics : an introduction with applications*. Springer, Berlin. 256 p.

10 Index

(renvoi aux sections)

Analyse krigéante 6

Autokrigeabilité 5.4

Bord (effets de) 8.2.2, 8.3.3, 8.3.4, 8.4

Cohérence entre estimations 0.2, 4.2.1, 4.2.2, 4.3.1, 5.1.1, 7.5

Cokrigeage 4

collocalisé 5.5.3, 7.4

disloqué 5.5.4, 7.4

d'indicatrices 8.3.6, 8.4

intrinsèque 4.4

ordinaire 4.3

(poids de) 4, 6

simple 4.2

(simplification du) 5

(variance de) 4.6.1, 7.4

Combinaison linéaire

autorisée 2.1.2, 2.6.2

de variables 2.10, 3.1.3, 5.1, 5.2.1, 5.3

en différents points 2.1.1, 2.4

Corrélation

entre accroissements 2.5.1, 3.1.1

(coefficient de) 1.4, 1.5, 2.9.2, 2.9.3, 3.1.1, 3.1.2

intrinsèque 3.1, 4.4, 5.4.2, 5.5.5, 8.4.2

(nuage de) 1.1

spatiale 0.1.2.1, 1.6, 2.3, 5.2

statistique 0.1.2.1, 1.6

corrélogramme 2.1.1, 5.3.1

covariance 2.1

croisée 2.2, 2.3

impaire 2.3, 2.4.1, 2.5.2

paire 2.5.2, 2.6.1, 3.1.1, 3.2.2.1

- non-centrée 2.1
- stationnaire 2.1.1
- statistique 1.4
- décalage 2.3, 7.1
- dérivée 2.4.1, 7.1
- écran 5.4.1
- erreur de mesure 3.3, 3.4, 4.5.1, 5.5.4
- facteur/factorisation 3.1.3, 3.1.4, 3.2.2.3, 3.2.3, 5.3, 5.4.2, 7.2, 8.4.1, 8.4.4
- fermeture 2.11, 5.2.1, 8.3.3
- gigognes (structures) 0.1.3, 2.10, 3.2.1
- hétérotopie 0.2, 2.8, 5.2.1, 5.5.2, 5.5.3, 5.5.4, 5.5.7, 7.3, 7.4, 7.5
- indicatrices 8
- intrinsèque 0.1.2.3, 1.6, 2.1.2, 2.6
 - voir aussi corrélation
- isotopie 0.2, 5.3, 5.4.2, 5.5.2, 7.3, 7.4, 7.5
- krigeage
 - d'indicatrice 8.1.3, 8.4.2
 - disjonctif 8.4.4
 - en dérive externe 5.5.8, 5.5.9, 7.5
- modèle
 - diffusif 8.4.3, 8.4.4
 - linéaire de corégionalisation 3.2
 - mosaïque 8.4.2, 8.4.4
 - voir aussi : corrélation intrinsèque, intrinsèque, stationnaire, résidu, facteurs
- moyennes liées 4.5, 5.2.2
- orthogonalité 1.5
- redondances 5.1.2
- régression 1.3
 - linéaire 1.5, 2.5.1, 2.9.2, 5.5.1, 7.4
- résidu 5.5, 7.4, 7.5, 8.2.3
 - d'indicatrices 8.4.1, 8.4.4
- régularisation 2.9.1, 7.1, 8.2.3
- seuillage 7.1, 8.4.3, 8.4.4

simulation 0.2, 7.2, 7.5, 8.4.3

stationnarité 2.1.1, 2.4

support 1.6, 2.9, 3.4

voir aussi régularisation

symétrie (de la structure) 2.5.1, 2.6.2, 2.6.3

voir aussi covariance croisée paire/impair

variogramme 2.1.2

d'indicatrice 8.1.2

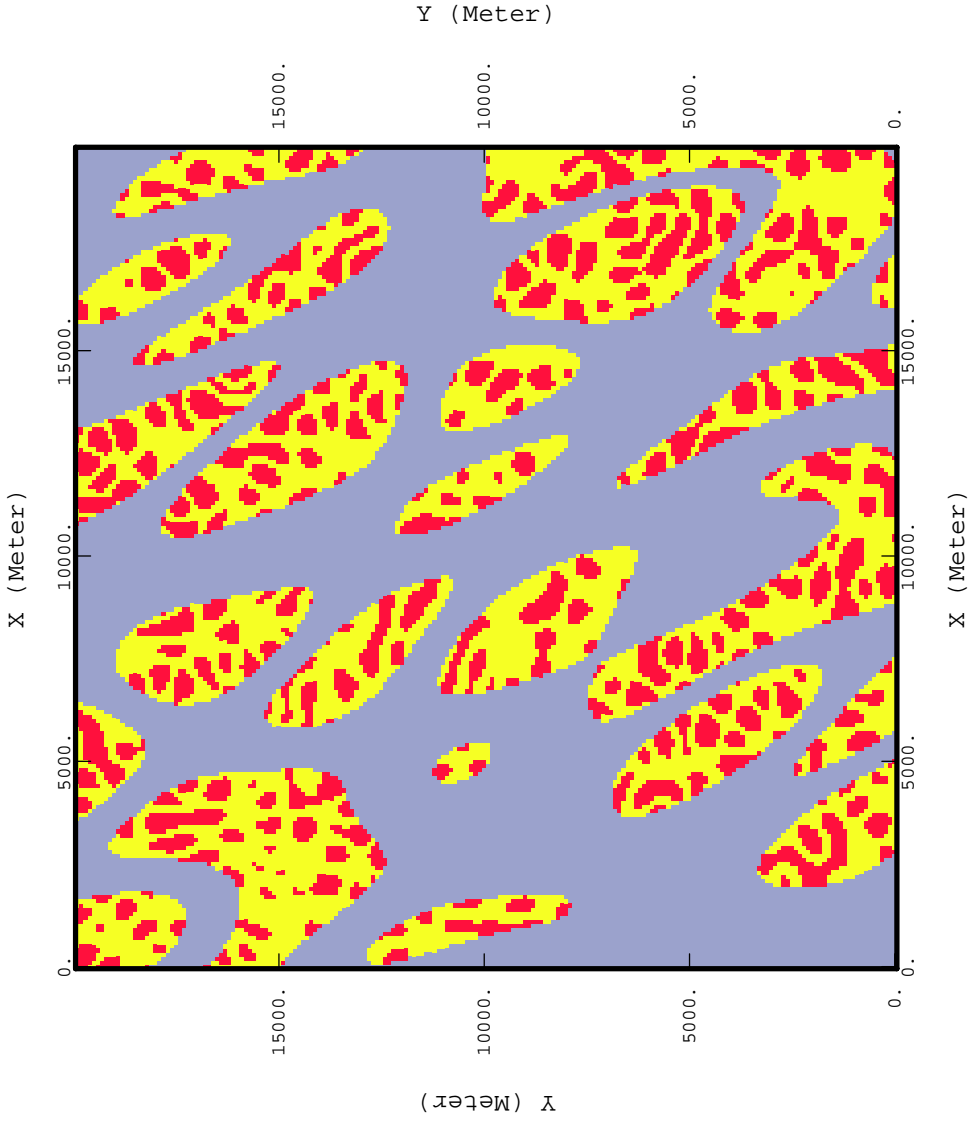
variogramme croisé 2.5

d'indicatrices 8.3.1-4

(pseudo-) 2.7, 3.3

voisinage 7.4

Modèle sans effet de bord (en sortant du gris)



Proportions (%) :

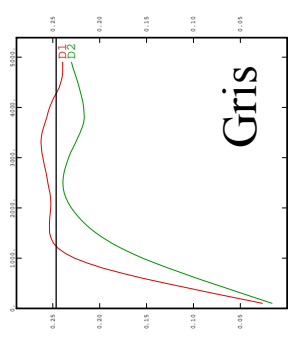
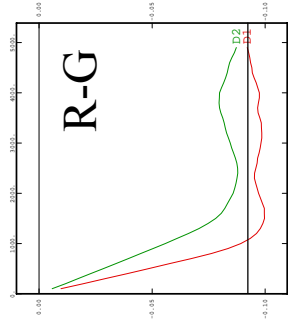
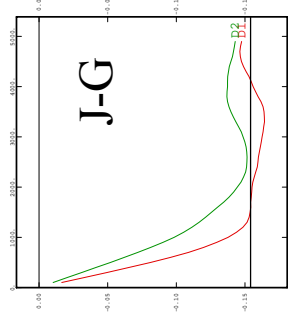
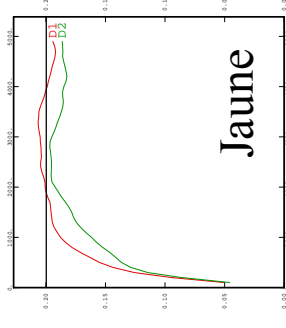
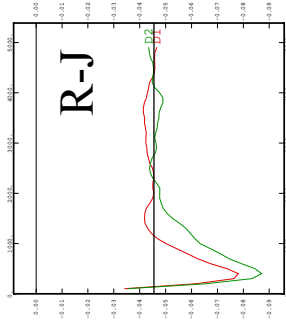
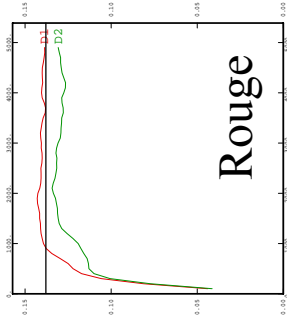
gris - 54.5

jaune - 30.3

rouge - 15.2

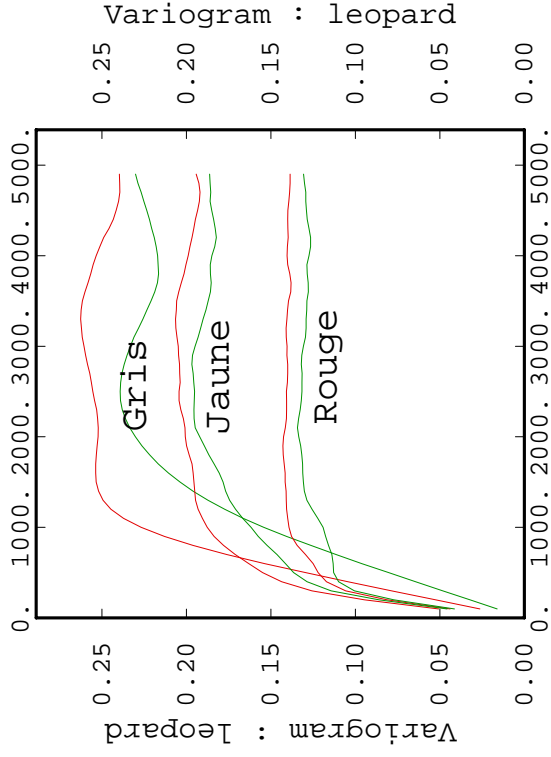
Modèle sans effet de bord (en sortant du gris)

- variogrammes simples et croisés

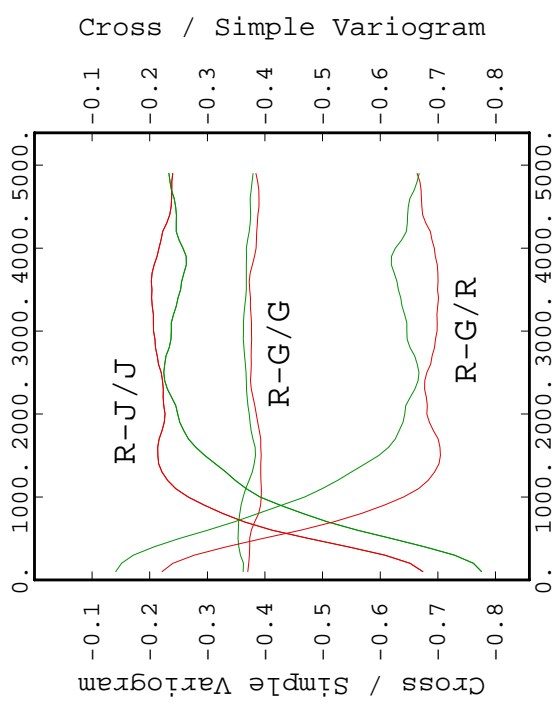


Modèle sans effet de bord (en sortant du gris)

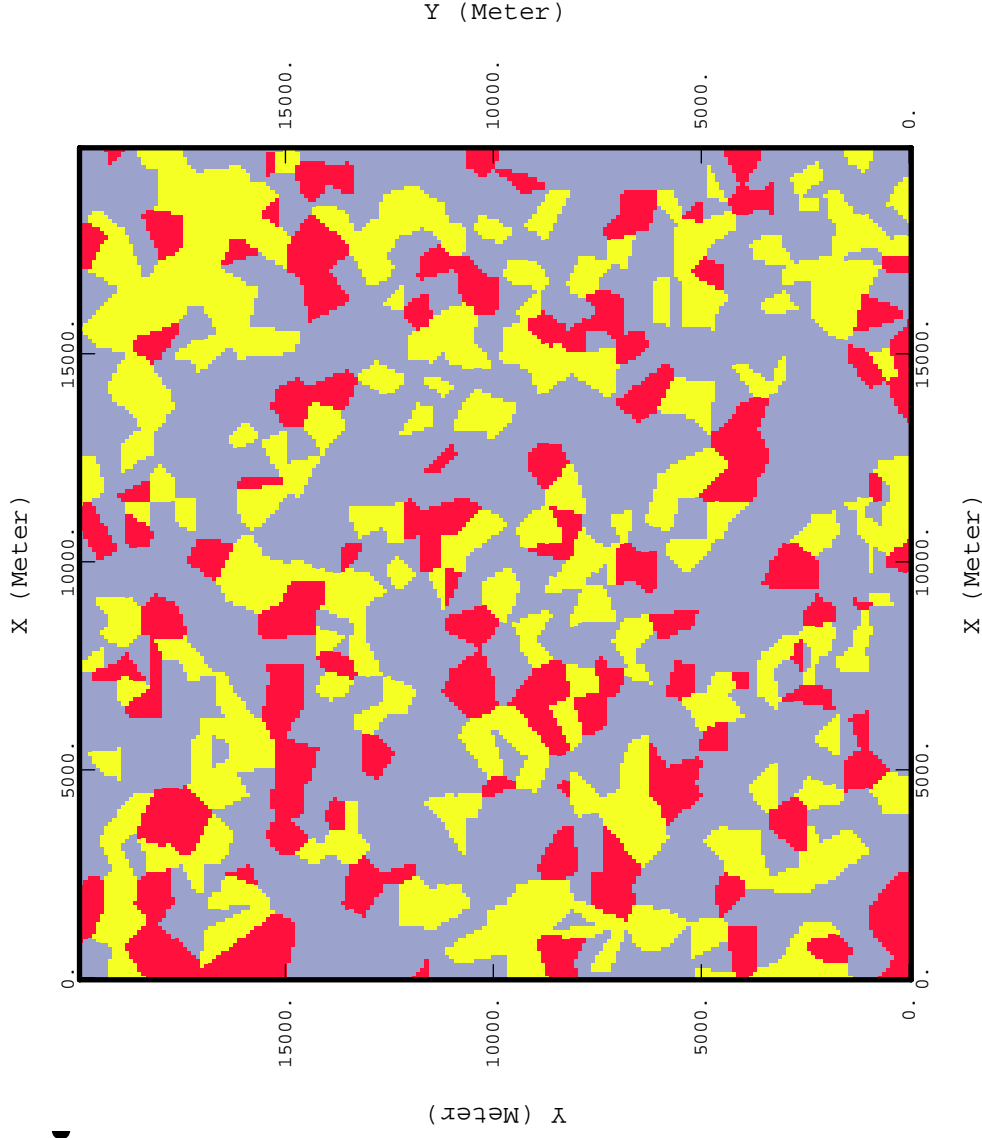
- variogrammes simples



- rapports croisés/simples



Modèle mosaïque à valuations indépendantes



Proportions (%) :

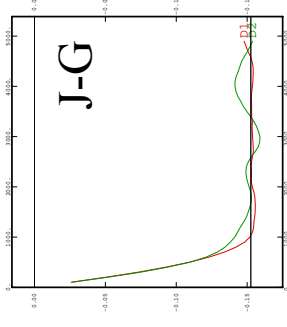
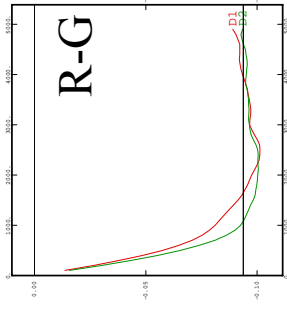
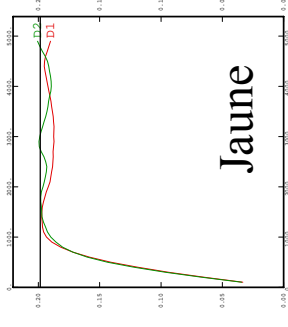
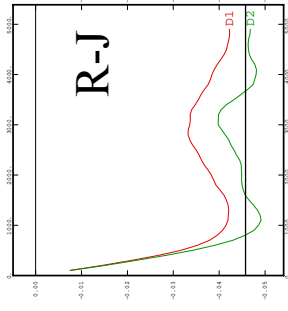
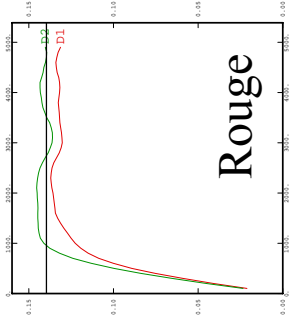
gris - 55.9

jaune - 27.3

rouge - 16.8

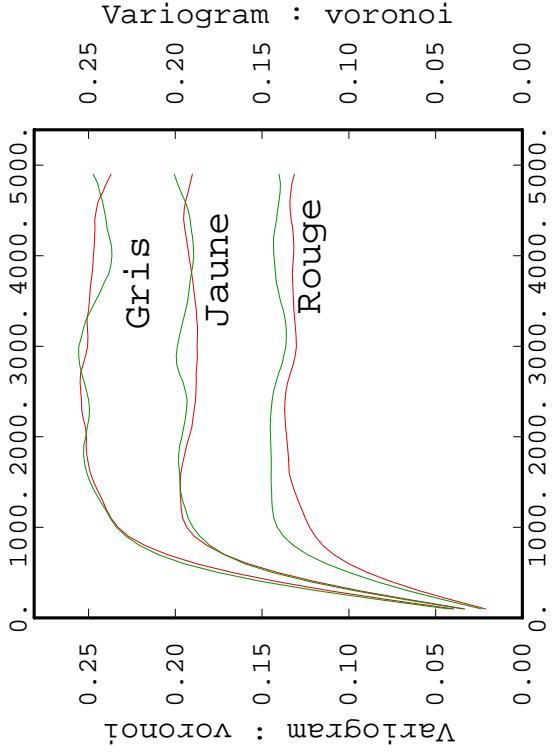
Modèle mosaïque à valuations indépendantes

- variogrammes simples et croisés

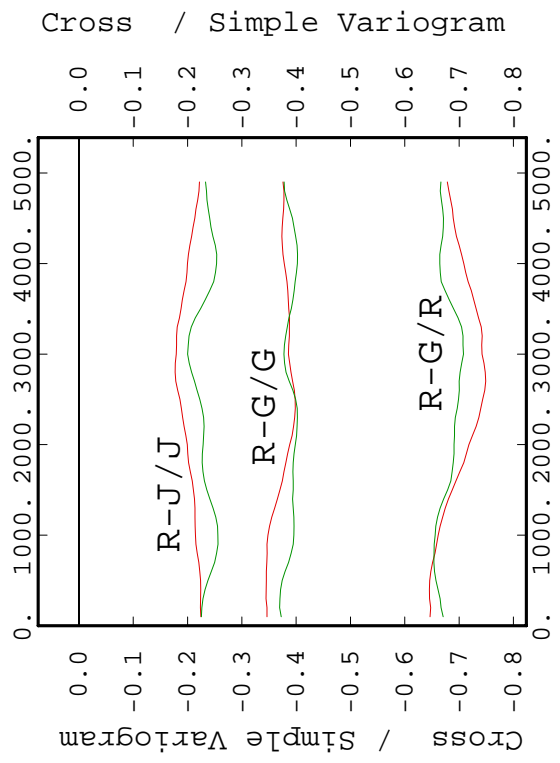


Modèle mosaïque à valuations indépendantes

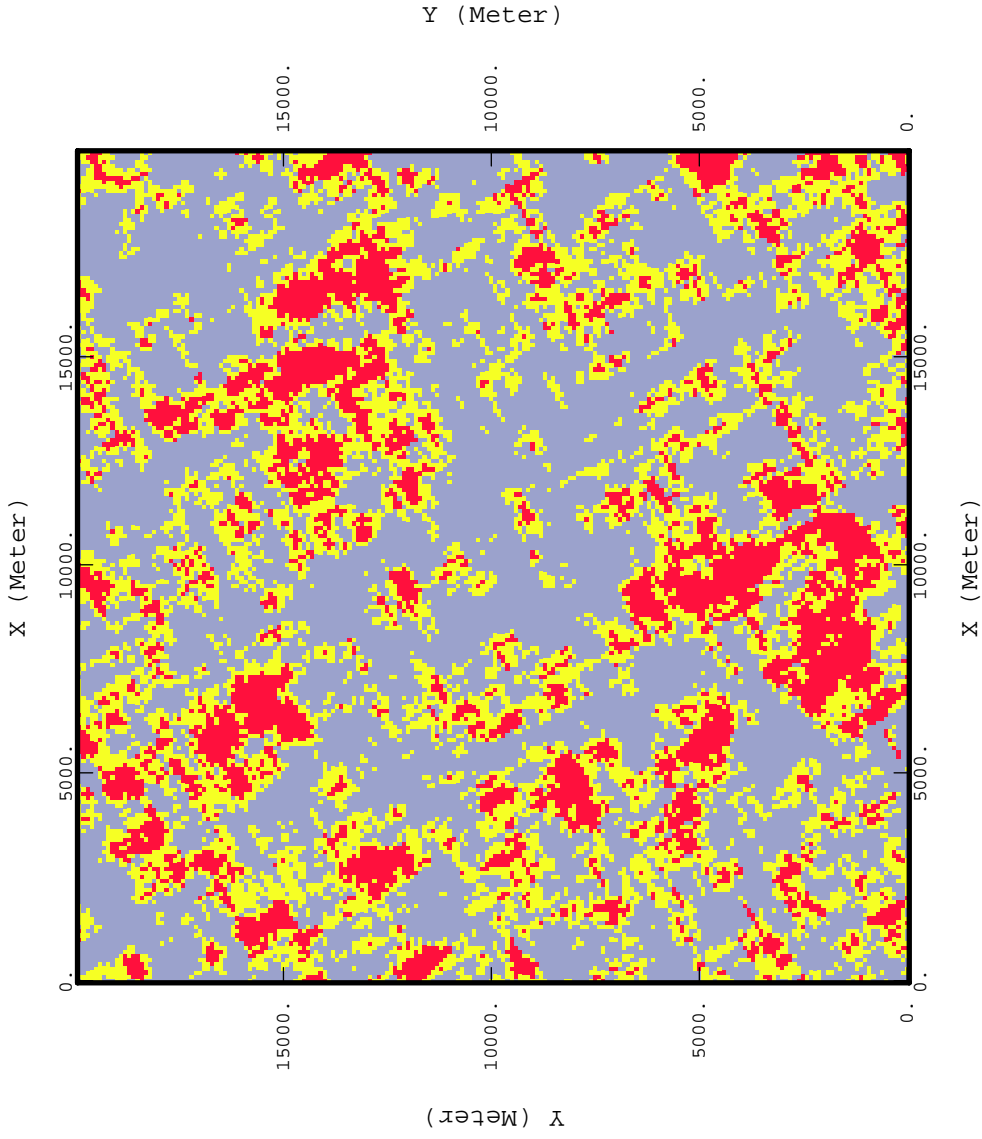
- variogrammes simples



- rapports croisés/simples



Modèle de type diffusif



Proportions (%):

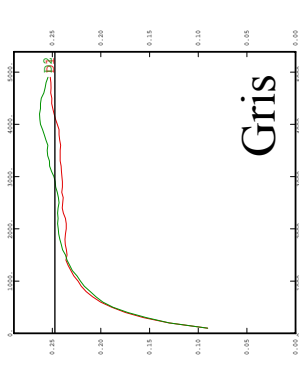
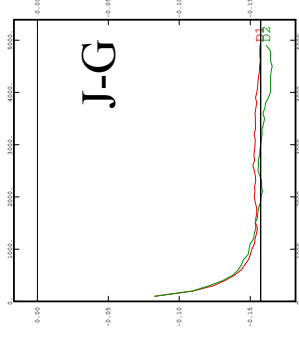
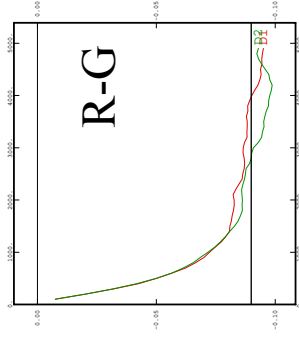
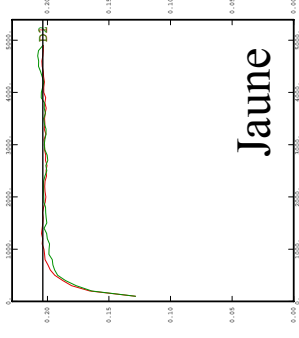
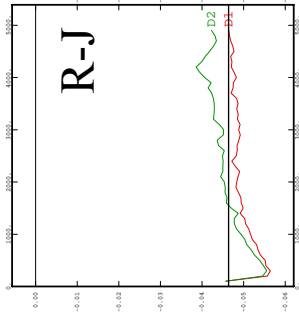
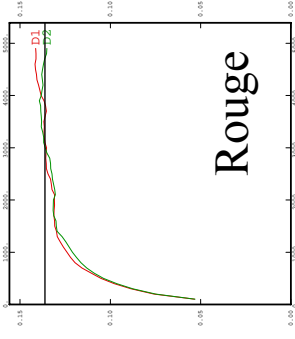
gris - 51.5

jaune - 30.3

rouge - 18.2

Modèle de type diffusif

- variogrammes simples et croisés



Modèle de type diffusif

- variogrammes simples

- rapports croisés/simples

