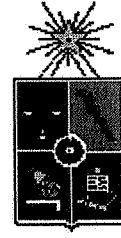




ÉCOLE DES MINES DE PARIS
Centre de Géostatistique



UNIVERSIDAD DE CHILE
Departamento de Ingeniería de Minas

GÉOSTATISTIQUE LINÉAIRE

par Xavier Emery

Janvier 2001

Remerciements

Je tiens à remercier Pierre Chauvet, Chantal de Fouquet, Christian Lantuéjoul (Centre de Géostatistique, Ecole des Mines de Paris¹) et Michel Arnaud (CIRAD de Montpellier²), pour avoir relu et corrigé tout ou partie de ce document.

Je remercie également le CIRAD de Montpellier et CODELCO-CHILE División Andina, qui ont fourni les données des exemples illustratifs et autorisé leur publication.

¹ Ecole des Mines de Paris, Centre de Géostatistique,
35, rue Saint Honoré, 77 305 Fontainebleau Cedex, France

² CIRAD-TERA/ERE, Campus International de Baillarguet
Montferrier-sur-Lez BP5035, 34032 Montpellier Cedex1, France

Table des matières

Avant-propos.....	1
Introduction	13
I. Le concept de variable régionalisée	13
II. Notion de support.....	14
III. Les principes directeurs	15
IV. Les problèmes d'estimation	16
V. Notations	17
Chapitre 1 : géostatistique transitive	19
I. Moyenne globale et abondance.....	19
II. Le covariogramme transitif.....	21
II.1. Définition et propriétés.....	21
II.2. Portée.....	23
II.3. Comportement à l'origine.....	23
II.4. Isotropie.....	23
II.5. Structures gigognes	24
II.6. Covariogramme géométrique	24
III. L'estimation de l'abondance	26
III.1. Echantillonnage régulier	26
III.1.1. Définition de l'estimateur.....	27
III.1.2. Randomisation du réseau d'échantillonnage	27
III.1.3. Commentaires sur la variance d'estimation.....	29
III.1.4. Insuffisance des valeurs expérimentales.....	29
III.1.5. Application à des données agronomiques.....	30
III.2. Echantillonnage aléatoire stratifié	35
III.3. Echantillonnage aléatoire uniforme	36
III.4. Echantillonnage aléatoire préférentiel	38
III.5. Echantillonnage irrégulier quelconque	39
IV. L'estimation locale : le krigeage transitif	40
V. L'analyse variographique	41
V.1. Le covariogramme expérimental	41
V.2. Le covariogramme modélisé	43
VI. Conclusion.....	44
Résumé.....	45

Chapitre 2 : les concepts de base de la géostatistique probabiliste	47
I. Limites de la statistique classique	47
II. Le concept de fonction aléatoire	48
II.1. Définition et choix méthodologique	49
II.2. Caractéristiques d'une fonction aléatoire	51
II.2.1. Loi spatiale	51
II.2.2. Moment du premier ordre (espérance mathématique)	52
II.2.3. Moments du second ordre	52
III. Possibilité de l'inférence statistique : les hypothèses de stationnarité	54
IV. Les combinaisons linéaires autorisées et le calcul de leurs deux premiers moments ..	60
IV.1. Définition des combinaisons linéaires autorisées	60
IV.2. Règles de calcul sous l'hypothèse stationnaire	60
IV.3. Règles de calcul sous l'hypothèse intrinsèque	61
V. L'étude exploratoire des données	62
V.1. Remarques générales	62
V.2. Application à un cas pratique : données de sol	63
V.2.1. Présentation des données	63
V.2.2. Etude de la stationnarité	65
V.3. Application à des données minières	67
V.3.1. Présentation des données	67
V.3.2. Cartes d'implantation	68
V.3.3. Statistiques élémentaires	68
V.3.4. Nuages directionnels des valeurs	70
V.3.5. Nuage de corrélation différée	72
V.3.6. Moyennes et variances locales	72
Résumé	74
 Chapitre 3 : les outils de modélisation de la structure spatiale	 75
I. La covariance stationnaire	75
II. Le variogramme	77
II.1. Définition	77
II.2. Propriétés théoriques	78
II.3. Anisotropie	79
II.4. Palier et portée	79
II.5. Comportement au voisinage de l'origine	81
III. Estimateurs du variogramme et de la covariance	82
III.1. Les estimateurs classiques	82
III.1.1. Estimateur du variogramme	82
III.1.2. Estimateur de la covariance	84
III.2. Propriétés des estimateurs	84
III.2.1. Biais et robustesse	84
III.2.2. Nécessité d'un contrôle des hypothèses	87

IV. Principaux modèles de variogramme.....	87
IV.1. Modèles avec palier ou modèles de transition	88
IV.1.1. Comportement discontinu à l'origine.....	88
IV.1.2. Comportement linéaire à l'origine	88
IV.1.3. Comportement parabolique à l'origine.....	90
IV.1.4. Autres modèles de transition	91
IV.2. Modèles à effet de trou	94
IV.3. Modèles sans palier	97
V. Analyse variographique : ajustement d'un modèle à un variogramme expérimental... 99	
V.1. Le modèle linéaire de régionalisation.....	99
V.2. L'effet de pépite et sa genèse	101
V.3. Le traitement des anisotropies	103
V.4. Recommandations pour le calcul du variogramme expérimental et l'ajustement à un modèle.....	107
V.4.1. Représentativité des données	107
V.4.2. La nuée variographique	107
V.4.3. Paramètres de calcul du variogramme expérimental.....	107
V.4.4. Modélisation.....	109
V.5. Test de la qualité du modèle : la validation croisée.....	111
V.6. Application à trois jeux de données.....	113
V.6.1. Un exemple quasi-stationnaire : les données agronomiques	113
V.6.2. Un exemple non stationnaire : les données de sol.....	121
V.6.3. Un exemple dans l'espace à trois dimensions : les données minières.....	128
Résumé.....	138
Chapitre 4 : changement de support et estimation globale	139
I. Les outils du changement de support	140
I.1. Changement de support (régularisation).....	140
I.2. Variance d'extension et variance d'estimation.....	144
I.3. Variance de dispersion.....	146
I.3.1. Première définition.....	146
I.3.2. Définition générale de la variance de dispersion.....	149
I.3.3. Relations entre variances d'extension et de dispersion.....	150
I.3.4. Relation d'additivité (formule de Krige).....	151
II. Application à l'estimation globale	152
II.1. Remarques préliminaires	152
II.1.1. Contrainte d'homogénéité.....	152
II.1.2. Choix de l'estimateur.....	152
II.1.3. Hypothèses et objectif.....	153
II.1.4. Quel sens objectif donner au biais et à la variance d'estimation globale ?... 153	
II.2. Estimation globale en support ponctuel	154
II.2.1. Echantillonnage aléatoire uniforme.....	154
II.2.2. Echantillonnage aléatoire stratifié	156
II.2.3. Echantillonnage régulier	157
II.2.4. Comparaison des différentes stratégies d'échantillonnage	159
II.3. Support non ponctuel	160

III. Conclusions sur l'estimation globale.....	161
III.1. Facteurs conditionnant la précision de l'estimation	161
III.2. Comparaison avec la statistique descriptive.....	161
III.3. Les problèmes de délimitation du champ.....	162
III.4. Récapitulatif sur l'estimation globale par les méthodes statistiques et géostatistiques.....	163
III.4.1. Rappel des notations	163
III.4.2. Support ponctuel	163
III.4.3. Support non ponctuel.....	164
IV. Application aux données agronomiques	165
V. Complément : géostatistique transitive ou géostatistique intrinsèque ?	167
Résumé.....	170
 Chapitre 5 : l'estimation locale : le krigeage.....	 171
I. Introduction.....	171
II. Les quatre étapes du krigeage.....	172
III. Choix du voisinage de krigeage.....	174
III.1. Voisinage unique	174
III.2. Voisinage glissant.....	175
IV. Le krigeage ponctuel.....	177
IV.1. Le krigeage à moyenne connue (krigeage simple).....	177
IV.2. Le krigeage à moyenne inconnue (krigeage ordinaire)	181
IV.3. Le krigeage intrinsèque ordinaire.....	185
IV.4. Remarques sur le système de krigeage	187
IV.5. Propriétés du krigeage	189
IV.6. Application aux données agronomiques	193
IV.6.1. Choix de la méthode d'estimation	193
IV.6.2. Comparaison entre les valeurs vraies et estimées	194
IV.7. Application aux données de sol.....	199
IV.8. Application aux données minières.....	202
V. Les effets des paramètres du modèle	205
V.1. Type de modèle	205
V.2. Portée	207
V.3. Palier	208
V.4. Effet de pépité.....	210
V.5. Effet de trou.....	211
V.6. Anisotropie	212
V.7. Type de krigeage (simple / ordinaire).....	213
VI. Le krigeage par bloc	214
VI.1. Principe	214
VI.2. Exemple sur les données minières.....	216
VII. Complément : le krigeage aléatoire.....	222
Résumé.....	225

Chapitre 6 : les modèles non stationnaires	227
I. Introduction.....	227
II. Le krigeage universel	228
II.1. La dichotomie du phénomène.....	228
II.1.1. Présentation du modèle.....	228
II.1.2. Propriétés de la dérive.....	229
II.1.3. Commentaires sur la dichotomie	230
II.2. Krigeage universel ponctuel	230
II.2.1. Résidu stationnaire d'ordre deux.....	230
II.2.2. Résidu intrinsèque	232
II.2.3. Propriétés du krigeage universel ponctuel	234
II.3. Krigeage universel non ponctuel	236
II.4. Evaluation de la dérive.....	237
II.4.1. Méthode des moindres carrés	237
II.4.2. Evaluation par moyenne mobile.....	238
II.4.3. Les splines de lissage.....	239
II.5. L'analyse variographique	240
II.5.1. Analyse directe sur les données brutes	240
II.5.2. Analyse variographique du résidu estimé	241
III. Le krigeage intrinsèque d'ordre k.....	243
III.1. Combinaisons linéaires autorisées d'ordre k.....	243
III.2. Fonctions aléatoires intrinsèques d'ordre k	245
III.3. Fonction de covariance généralisée	246
III.3.1. Propriétés théoriques	246
III.3.2. Exemples de covariances généralisées isotropes	247
III.4. Le krigeage intrinsèque d'ordre k ponctuel.....	250
III.5. Propriétés du krigeage intrinsèque d'ordre k	251
III.6. L'analyse variographique.....	252
III.6.1. Détermination de l'ordre k.....	252
III.6.2. Inférence statistique de la covariance généralisée.....	253
III.6.3. Remarques sur la procédure d'ajustement.....	254
III.7. Exemple sur les données de sol.....	255
Résumé.....	260
Chapitre 7 : géostatistique multivariable	261
I. Outils structuraux	262
I.1. Fonctions de covariance simples et croisées.....	262
I.1.1. Hypothèse de stationnarité conjointe	262
I.1.2. Propriétés mathématiques.....	263
I.2. Variogrammes simples et croisés.....	265
I.2.1. L'hypothèse intrinsèque conjointe.....	265
I.2.2. Propriétés théoriques	265
I.2.3. Lien avec la fonction de covariance croisée.....	266
I.2.4. Pseudo-variogramme croisé	267

II. L'analyse variographique multivariable.....	268
II.1. Calcul des covariances et variogrammes simples et croisés expérimentaux.....	268
II.2. Le modèle linéaire de corégionalisation.....	269
II.2.1. Cas stationnaire du second ordre.....	269
II.2.2. Cas intrinsèque strict.....	270
II.3. Ajustement pratique	271
II.3.1. Cas bivariable.....	271
II.3.2. Cas multivariable.....	272
II.4. Illustration sur les données de sol.....	273
III. Le cokrigage	278
III.1. Cokrigage ordinaire stationnaire.....	278
III.1.1. Cas général.....	278
III.1.2. Cas homotopique.....	281
III.2. Cokrigage simple (moyennes connues).....	282
III.3. Cokrigage intrinsèque ordinaire.....	283
III.4. Propriétés du cokrigage ponctuel.....	283
III.5. L'alternative krigage – cokrigage	284
III.6. Application aux données de sol.....	285
III.6.1. Un couple de variables corrélées : % argile + limon / % terre fine.....	285
III.6.2. Un couple de variables peu corrélées : % argile + limon / pF4.2	289
III.7. Cas de simplification du système de cokrigage : "autokrigeabilité"	292
IV. L'analyse krigéante.....	294
IV.1. Décomposition des variables en facteurs.....	294
IV.1.1. Principe.....	294
IV.1.2. Définition des facteurs	296
IV.2. Estimation des facteurs.....	298
IV.3. Estimation des composantes spatiales intermédiaires	299
IV.4. Conclusion sur l'analyse krigéante.....	300
IV.4.1. Atouts et limitations	300
IV.4.2. Exemple sur les données de sol	301
V. Modèles non stationnaires.....	304
V.1. Cokrigage universel.....	304
V.1.1. Hypothèses.....	304
V.1.2. Cokrigage universel ponctuel.....	304
V.1.3. Propriétés du cokrigage universel.....	305
V.2. Cokrigage intrinsèque	306
V.3. Dérive externe.....	307
V.3.1. Un exemple introductif.....	307
V.3.2. Krigage avec une dérive externe	308
V.3.3. Krigage avec dérivées externes multiples.....	309
Résumé.....	310
ANNEXES.....	311

Annexe A : rappels des notions usuelles de probabilité.....	313
I. Espace probabilisé.....	313
II. Variable aléatoire.....	313
III. Fonction de répartition et densité de probabilité.....	314
IV. Espérance et moments d'une variable aléatoire.....	315
V. Variables aléatoires vectorielles.....	316
VI. Loi de probabilité conditionnelle.....	319
VII. Fonction aléatoire.....	320
Annexe B : l'ergodicité.....	323
I. Définition.....	323
II. Moment d'ordre 1 (espérance mathématique).....	324
II.1. Espérance et moyenne spatiale.....	324
II.2. La portée intégrale.....	326
II.3. Calcul empirique de la portée intégrale.....	331
III. Le variogramme.....	332
Annexe C : compléments sur le krigeage universel	
l'évaluation optimale de la dérive et ses implications théoriques	335
I. Hypothèses et objectif.....	335
II. Evaluation de la dérive dans son ensemble.....	336
II.1. Résidu stationnaire d'ordre deux.....	336
II.2. Résidu intrinsèque strict.....	337
III. Evaluation optimale des coefficients de la dérive.....	338
IV. Le variogramme du résidu estimé optimal.....	340
V. L'invariance tensorielle.....	341
VI. Le théorème d'additivité.....	342
VI.1. Résidu stationnaire d'ordre deux.....	342
VI.2. Résidu intrinsèque.....	343
Annexe D : compléments sur les fonctions aléatoires intrinsèques d'ordre k.....	345
I. Formalisation théorique des fonctions aléatoires intrinsèques d'ordre k.....	345
I.1. Notations.....	345
I.2. Première définition.....	346
I.3. Seconde définition.....	346
I.4. Troisième définition.....	347
I.5. Dérive d'une fonction aléatoire intrinsèque d'ordre k.....	348
II. Cas particuliers de représentations.....	349
II.1. Représentations internes.....	349
II.2. Représentations localement stationnaires.....	351

III. Les covariances généralisées.....	353
III.1. Caractérisation des covariances généralisées.....	353
III.2. Propriétés spectrales	354
III.3. Changement d'ordre	355
IV. Fonctions aléatoires intrinsèques générales	355
V. Le krigeage comme méthode d'interpolation.....	356
V.1. Présentation duale du krigeage.....	356
V.2. Equivalence entre splines et krigeage.....	358
Annexe E : introduction aux problèmes de filtrage.....	359
I. Filtrage d'erreurs	359
I.1. Cas d'une erreur non systématique	359
I.2. Cas particulier d'un bruit blanc.....	361
I.3. Cas d'une erreur systématique	363
II. Filtrage d'une dérive temporelle.....	363
Annexe F : une méthode d'analyse des données :	
l'analyse en composantes principales.....	365
I. Introduction.....	365
II. Construction de facteurs synthétiques orthogonaux	366
II.1. Principe.....	366
II.2. Définition des facteurs	366
II.3. Hiérarchisation des facteurs	367
III. Le nuage des échantillons	368
III.1. Deux critères de représentation équivalents.....	368
III.2. Représentation du nuage des échantillons.....	369
III.3. Réduction des variables de départ (standardisation)	370
IV. Le nuage des variables.....	371
IV.1. Corrélation des variables avec les facteurs.....	371
IV.2. Représentation du nuage des variables	372
V. Eléments supplémentaires	373
V.1. Echantillons supplémentaires	373
V.2. Variables supplémentaires.....	373
VI. Conclusion.....	373
Annexe G : l'analyse krigeante.....	375
I. Cas d'une seule variable : le modèle linéaire de régionalisation	375
I.1. Décomposition de la variable.....	375
I.2. Estimation des composantes spatiales : krigeage avec filtrage de composantes ...	376
I.2.1. Krigeage des composantes stationnaires et intrinsèque	376
I.2.2. Krigeage de la moyenne locale.....	378

II. Décomposition de plusieurs fonctions aléatoires en corrélation intrinsèque	379
II.1. Décomposition des variables	379
II.1.1. Cas stationnaire du second ordre.....	379
II.1.2. Cas intrinsèque	381
II.2. Estimation des facteurs spatiaux.....	383
III. Le modèle linéaire de corégionalisation	384
III.1. Décomposition des variables.....	384
III.2. Estimation des facteurs spatiaux.....	387
IV. Différence entre l'analyse krigéante et l'analyse en composantes principales	388
Annexe H : compléments sur les modèles variographiques multivariés	389
I. Le modèle bilinéaire de corégionalisation.....	389
I.1. Le modèle linéaire complexe de corégionalisation.....	389
I.2. Le modèle bilinéaire de corégionalisation.....	390
I.3. Exemples.....	391
II. Fonction aléatoire et ses dérivées partielles	393
III. Fonction aléatoire et son laplacien	395
INDEX ET BIBLIOGRAPHIE	397
Index des mots clés	399
Bibliographie.....	403

Introduction

Le néologisme “géostatistique” a été inventé par Georges Matheron en 1962.

Le préfixe “*géo*” fait référence au domaine des sciences de la terre, qui est historiquement celui où la géostatistique s’est développée. Les champs d’application actuels concernent des domaines très variés, comme l’évaluation de ressources naturelles (minières, pétrolières, forestières, etc.), la pollution, l’agronomie, la topographie, la météorologie, l’océanographie, la géophysique ou l’analyse d’images, pour n’en citer que quelques-uns.

La désinence “*statistique*” se réfère à l’utilisation de méthodes probabilistes. L’originalité de la géostatistique par rapport à la statistique classique est la prise en compte de la dépendance entre les observations lorsque celles-ci sont situées dans l’espace. Toutefois, les méthodes géostatistiques ne sont pas exclusivement probabilistes : il existe une branche, connue sous le nom de “géostatistique transitive”, qui ne fait pas appel au concept des processus aléatoires (cf. chapitre 1).

I. Le concept de variable régionalisée

La géostatistique se définit comme l’étude des **phénomènes régionalisés**, c’est-à-dire des phénomènes qui s’étendent dans l’espace et y présentent une “organisation” ou “structure”. Par “espace”, nous entendons en général l’espace géographique, mais il peut aussi s’agir de l’axe temporel ou d’espaces plus abstraits.

L’objet sur lequel nous allons travailler ne sera pas le phénomène régionalisé lui-même, qui est une réalité physique, mais une description mathématique de cette réalité, à savoir une fonction numérique appelée **variable régionalisée** ou encore *régionalisation*, censée représenter et mesurer correctement ce phénomène.

Par exemple :

- en géostatistique minière, des variables telles que teneur, puissance, accumulation, densité, granulométrie, décrivent un phénomène de minéralisation ;
- la concentration dans l’atmosphère d’un élément polluant, l’altitude topographique d’un point, la température au niveau du sol, la densité de poissons dans une zone océanique, la conductivité mesurée sur un échantillon de terre, sont également des variables régionalisées.

D'un point de vue mathématique, une variable régionalisée est une fonction déterministe, traditionnellement notée z . En général, cette fonction présente deux aspects complémentaires : d'une part, elle a une "structure" spatiale (zones de fortes valeurs / zones de faibles valeurs), mais d'autre part elle varie irrégulièrement et échappe à toute représentation simple. Le propre des méthodes géostatistiques sera de définir des outils synthétiques qui permettront de résumer les principales caractéristiques de la régionalisation.

Un phénomène régionalisé n'ayant jamais une étendue infinie, nous n'étudierons la variable régionalisée qu'à l'intérieur d'un domaine borné $\mathcal{D}$ appelé **champ** de la variable. Ce champ $\mathcal{D}$ peut représenter une zone naturelle en dehors de laquelle z n'est pas définie ; il peut aussi s'agir d'un domaine particulier où la régionalisation présente un intérêt, par exemple les endroits où elle est non nulle, ou supérieure à une limite de détection.

II. Notion de support

Une variable régionalisée peut être définie non seulement en un point, mais également sur une surface ou un volume. La surface ou le volume de base sur lequel on considère la variable régionalisée est appelé **support**. En général, le support des échantillons mesurés est (quasi-) ponctuel, alors que le support qui intéresse le praticien est souvent plus volumineux (citons à titre d'exemple les unités sélectives d'exploitation en évaluation minière). Cette notion est importante en raison de la dépendance entre la taille du support et la distribution statistique des observations : les gros supports présentent moins de valeurs extrêmes et davantage de valeurs intermédiaires que les supports ponctuels. Ainsi, la distribution des valeurs, et en particulier sa variance statistique, dépend du support sur lequel est définie la variable régionalisée. Nous reviendrons ultérieurement sur cet *effet de support*.

Lorsque l'on change de support (par exemple en passant des valeurs quasi-ponctuelles aux valeurs moyennes sur un "bloc"), on obtient une nouvelle variable régionalisée, liée à la variable initiale mais qui a des caractéristiques structurales différentes. Le changement de support s'inscrit dans la théorie plus générale de la **régularisation**. Cette opération consiste à calculer une convoluée de la variable régionalisée par une fonction de pondération, dont le poids total est en général unité ; la variable résultante est spatialement plus régulière que celle de départ, d'où son nom de *régularisée*.

Dans les problèmes de changement de support, il est souhaitable que la variable régionalisée étudiée soit **additive**, c'est-à-dire que sa valeur sur l'union de plusieurs domaines soit égale à la somme ou la moyenne de ses valeurs sur chaque domaine. Cette condition est nécessaire pour que le calcul de la valeur moyenne sur un support plus grand que celui des mesures, garde un sens physique. Dans le cas contraire, on devra se restreindre à étudier des supports égaux à ceux des échantillons : un changement de support sera, sinon interdit, du moins fortement déconseillé.

III. Les principes directeurs

D'ordinaire, la variable régionalisée n'est jamais accessible exhaustivement et n'est connue qu'en certains sites d'échantillonnage. Pour résoudre le problème posé, on ne dispose en général que des valeurs échantillonnées et leur position, parfois d'informations supplémentaires — qualitatives ou quantitatives — sur le phénomène. Le premier principe qui doit régir l'étude géostatistique est le **respect des données**, c'est-à-dire la conformité des manipulations et des hypothèses sous-jacentes aux observations. Ce principe n'empêche pas d'avoir un sens critique quant au mode d'échantillonnage (implantation et représentativité des observations) et à l'exactitude des mesures (valeurs aberrantes).

Pour exploiter les informations disponibles au-delà d'un simple report des données, il sera nécessaire de construire des *modèles*. Mais il faut garder à l'esprit qu'un modèle ne décrit jamais le phénomène dans sa totalité, sans simplification, ni distorsion. La contrainte du **seuil de réalisme** consiste à trouver un modèle qui fournit une description adéquate du phénomène, ni trop simplifiée, ni trop déformée. Soulignons dès à présent l'importance de l'**échelle de travail** : les caractéristiques structurales d'une même régionalisation (régularité, homogénéité spatiale...) dépendent en grande partie de l'échelle d'observation, de sorte qu'un modèle peut parfaitement être adapté à une certaine échelle et se révéler inacceptable à des échelles plus grandes ou plus petites.

La modélisation et les hypothèses qu'elle nécessite reviennent à ajouter de l'information à celle contenue dans les données expérimentales. Il convient de réduire au maximum la complexité du modèle, en évitant d'introduire des hypothèses ou des paramètres "arbitraires", qui ne peuvent être contrôlés expérimentalement : un modèle trop complexe risque de dépasser le seuil de réalisme. Il faut au contraire rechercher le modèle le moins exigeant permettant de résoudre le problème posé (**principe d'économie**).

Enfin, le modèle n'est qu'un intermédiaire de calcul pour résoudre un problème concret. Les résultats d'une étude géostatistique doivent être reformulés en termes *objectifs* ; cette phase de reformulation est appelée **reconstruction opératoire**. Précisons ce que l'on entend par "**objectivité**". Un concept ou un énoncé est objectif s'il peut s'exprimer en termes de *grandeurs régionales*, c'est-à-dire de quantités qui seraient disponibles si la réalité était connue exhaustivement¹. Le résultat final doit être vérifiable expérimentalement et susceptible d'être démenti à partir d'une connaissance exhaustive de la variable régionalisée, même s'il s'agit là d'une hypothèse illusoire (dans la pratique, on ne dispose que d'une connaissance partielle, voire extrêmement limitée, de la réalité).

¹ Par exemple, une espérance mathématique est une entité abstraite, non objective, tandis que la moyenne dans l'espace d'une variable régionalisée est objective. On pourra, à titre d'exercice, s'interroger sur le sens objectif du concept d'*intervalle de confiance*, censé mesurer la "probabilité" qu'une valeur inconnue, mais unique, se situe dans un intervalle donné.

IV. Les problèmes d'estimation

L'estimation² constitue un objectif important de la géostatistique ; elle consiste à évaluer, de la manière la plus précise et la plus pertinente possible, une grandeur qui n'a pas été mesurée, à partir de tout ou partie des valeurs échantillonnées dans le champ. On peut donner comme exemple connu d'estimateur la moyenne arithmétique, qui donne des poids égaux à tous les points d'observation. Une spécificité des méthodes géostatistiques sera non seulement de construire des estimateurs, mais aussi de fournir une mesure de la précision de l'estimation à l'aide d'outils probabilistes.

On distingue deux types d'estimation : les estimations *globales* et les estimations *locales*.

L'**estimation globale** concerne le champ $\mathcal{D}$ entier, que l'on désire caractériser par une valeur unique (à savoir, la moyenne ou la somme de la variable régionalisée étudiée). Il est rare qu'une estimation globale soit suffisante ; il est souvent nécessaire de la compléter par des estimations locales. Par exemple, dans une étude de pollution, il ne suffit pas d'évaluer la pollution moyenne sur l'ensemble de la zone, mais il faut aussi distinguer les secteurs fortement pollués de ceux qui le sont moins.

Les **estimations locales** s'intéressent au contraire aux différents secteurs de la zone d'étude. En général, elles cherchent à évaluer soit la valeur en un site non échantillonné, soit la valeur moyenne sur un bloc (surface ou volume), grâce à une combinaison linéaire pondérée des valeurs mesurées aux sites échantillonnés. Elles doivent tenir compte de la distance entre le secteur à estimer et les sites d'observation : les sites proches auront intuitivement plus de poids que les sites éloignés. La détermination des poids devra également dépendre des caractéristiques structurales de la variable régionalisée, en particulier de son degré de régularité, et de la disposition spatiale des échantillons eux-mêmes : en effet, des observations groupées ont souvent des valeurs similaires et contiennent donc une information redondante. Il conviendra de prendre en compte ces effets au moment de construire l'estimation.

Les problèmes d'estimation ne sont pas les seuls problèmes auxquels répond la géostatistique. L'analyse de la **dépendance spatiale** des données quantifiera les "corrélations" ou redondances d'information entre les valeurs mesurées en différents sites, et déterminera la taille de la "zone d'influence" d'une observation. La plus ou moins grande **continuité et régularité spatiale** de la variable régionalisée sera révélée par des outils faciles à interpréter. L'étude selon différentes directions de l'espace permettra de détecter des **anisotropies**, indiquant que la régionalisation est davantage structurée dans certaines directions que dans d'autres. Il sera éventuellement possible d'**interpréter** le phénomène régionalisé, par exemple en mettant en évidence les divers processus ayant engendré le phénomène étudié à différentes échelles de temps ou d'espace. Finalement, un sujet non traité dans le présent document est la construction de modèles numériques ou "**simulations**" qui reproduisent la structuration spatiale mise en évidence sur les données et permettent d'étudier l'incertitude portant sur une grandeur inconnue.

² La terminologie d'*estimation* n'a pas la même connotation qu'en statistique classique, où l'on estime les *paramètres* d'un modèle ou d'une loi de probabilité. Ici, on estime des grandeurs physiques objectives, c'est-à-dire des grandeurs qui existent indépendamment de l'état de l'information disponible : "estimation" a le sens classique de "prédiction".

V. Notations

Dans la suite, on notera :

- z la variable régionalisée étudiée
- d la dimension de l'espace de travail ; en général, $d = 2$ ou 3 , c'est-à-dire que l'on se place dans l'espace géographique $\mathbb{R}^2$ ou $\mathbb{R}^3$
- $\mathcal{D}$ le champ de la variable régionalisée, qui est un domaine borné de l'espace $\mathbb{R}^d$
- $|D|$ la mesure (longueur, aire, volume...) du domaine D inclus dans $\mathcal{D}$
- $\mathbf{x} \in \mathcal{D}$ un vecteur de localisation dans l'espace, qui varie continûment dans $\mathcal{D}$
- $(x_1, \dots, x_d)$ les coordonnées dans $\mathbb{R}^d$ du vecteur $\mathbf{x}$
- $\{ z(\mathbf{x}) \mid \mathbf{x} \in \mathcal{D} \}$ l'ensemble des valeurs, en général inconnues, prises par la variable z ; le plus souvent, $z(\mathbf{x}) \in \mathbb{R}$. Lorsque l'on considèrera le cadre multivariable, on pourra prendre $z(\mathbf{x}) \in \mathbb{R}^N$ ou, ce qui revient au même, travailler avec N variables régionalisées réelles.
- n le nombre de points de données
- $\{ \mathbf{x}_\alpha, \alpha = 1 \dots n \}$ les sites d'observation dans $\mathcal{D}$.

D'une manière générale, les indices grecs α, β sont réservés pour désigner les points de données. Les matrices et vecteurs seront notés en caractères gras. Par ailleurs, lorsque l'on abordera les modèles probabilistes, les lettres minuscules désigneront les quantités déterministes et les lettres majuscules les quantités aléatoires. Enfin, les estimateurs utilisés en géostatistique seront repérés par une astérisque placée en exposant ; par exemple, $z^*(\mathbf{x})$ représentera une estimation de z au point $\mathbf{x}$.

Chapitre 1 : géostatistique transitive

La géostatistique transitive est la partie **déterministe** de la géostatistique, celle qui étudie directement la variable régionalisée, sans recourir à une interprétation probabiliste. L'adjectif "transitif" se réfère aux zones de transition aux frontières du champ : souvent, la variable régionalisée présente des *effets de bordure*, c'est-à-dire que ses valeurs sont plus faibles en périphérie du champ et plus fortes à l'intérieur. La géostatistique transitive va permettre de prendre en compte cette "interférence" entre la régionalisation et son champ en les analysant *simultanément*, sans chercher à les dissocier. Le principal intérêt de cette approche est qu'aucune hypothèse n'est requise pour mener à bien les calculs, ce qui ne sera pas le cas des méthodes probabilistes.

I. Moyenne globale et abondance

La géostatistique transitive s'intéresse surtout aux problèmes d'**estimation globale**. Il s'agit de caractériser la variable régionalisée par une valeur unique et représentative du champ entier. Cette valeur doit correspondre à une grandeur *objective*, en ce sens qu'elle serait disponible si la réalité était connue exhaustivement. Par exemple, la moyenne globale est un concept objectif car défini sans ambiguïté par la donnée d'un domaine et de la valeur de la variable en tous les points de ce domaine. Toutefois, lorsque se posent des problèmes de délimitation du champ de la régionalisation, la moyenne n'est pas un concept approprié.

Il arrive parfois que la définition du champ fasse intervenir les valeurs prises par la variable régionalisée : il s'agit par exemple du domaine de l'espace où cette dernière est non nulle. Dans ce cas, les frontières du champ sont incertaines, car la connaissance de la variable régionalisée se limite aux mesures effectuées aux sites d'échantillonnage. Le formalisme de la géostatistique transitive va permettre de s'affranchir du besoin de connaître la géométrie exacte du champ.

La géostatistique transitive se propose d'estimer, non pas la *moyenne* globale de la variable régionalisée $z(\mathbf{x})$ sur le champ $\mathcal{D}$, mais sa quantité totale ou *abondance* :

$$q = \int_{\mathcal{D}} z(\mathbf{x}) d\mathbf{x}.$$

La moyenne se déduit de l'abondance en divisant par la surface ou le volume de $\mathcal{D}$, mais ce ne sont pas des concepts équivalents lorsque le champ $\mathcal{D}$ n'est pas identifié avec précision. D'ailleurs, dans de tels cas, c'est souvent l'abondance et non la moyenne que l'on cherche *in fine* à évaluer.

Exemples :

- z est la concentration en nitrate dans le sol, et q la quantité dans la zone entière ;
- z est la teneur en métal dans un gisement minier, et q la quantité de métal associée ;
- z est la superficie d'une culture sur un support donné, et q la surface totale cultivée.

Par convention, on prolonge la variable régionalisée en dehors de $\mathcal{D}$ par zéro¹, de sorte qu'elle est à présent définie sur tout l'espace, mais n'est non nulle qu'à l'intérieur de $\mathcal{D}$:

$$z(\mathbf{x}) = 0 \text{ quand } \mathbf{x} \notin \mathcal{D}.$$

Par suite, on peut écrire²

$$q = \int_{\mathcal{D}} z(\mathbf{x}) d\mathbf{x} = \int z(\mathbf{x}) d\mathbf{x}.$$

L'abondance apparaît comme un concept plus facile à manipuler que la moyenne globale, car indépendant du champ : contrairement à la moyenne, l'intégrale est invariante lorsque l'on étend le domaine d'intégration et peut donc être calculée sur tout l'espace. Réciproquement, les valeurs nulles peuvent être considérées *hors champ* puisqu'elles n'influencent pas sur la valeur de l'abondance. C'est la raison pour laquelle le champ désignera, dans toute la suite de ce chapitre, l'ensemble des points où la régionalisation est non nulle :

$$\mathcal{D} = \{\mathbf{x} \in \mathbb{R}^d / z(\mathbf{x}) \neq 0\}.$$

Du point de vue mathématique, le problème de l'estimation globale se formule ainsi : estimer l'intégrale d'espace d'une fonction qui s'annule en dehors d'un domaine borné (non nécessairement identifié), connaissant les valeurs prises par cette fonction en certains sites de mesure.

¹ Il est judicieux de choisir, dans la mesure du possible, une variable régionalisée qui s'annule "naturellement" sur les frontières de $\mathcal{D}$ (quitte à redéfinir la variable étudiée).

² Dans toute la suite de ce chapitre, le symbole d'intégration désignera l'intégrale sur tout l'espace $\mathbb{R}^d$. La variable régionalisée n'étant non nulle qu'à l'intérieur de $\mathcal{D}$, il n'y aura jamais de problème de convergence, les intégrales se réduisant à des intégrales sur $\mathcal{D}$ (borné).

II. Le covariogramme transitif

II.1. Définition et propriétés

Considérons une variable régionalisée $z(\mathbf{x})$ de champ $\mathcal{D} \subset \mathbb{R}^d$ (en général, $d = 2$ ou 3). On va associer à $z(\mathbf{x})$, fonction qui varie très irrégulièrement dans $\mathcal{D}$, une fonction plus simple, appelée **covariogramme transitif**, et notée traditionnellement g . L'intérêt de cette fonction est qu'elle exprime de manière synthétique les principales caractéristiques structurales de $z(\mathbf{x})$ et du champ $\mathcal{D}$. Le covariogramme transitif est défini sur $\mathbb{R}^d$ par

$$\forall \mathbf{h} \in \mathbb{R}^d, g(\mathbf{h}) = \int z(\mathbf{x}) z(\mathbf{x} + \mathbf{h}) d\mathbf{x}.$$

L'intégrale porte sur tout l'espace géographique ou, ce qui est équivalent, sur le champ $\mathcal{D}$, puisque z s'annule en dehors de $\mathcal{D}$. Une autre écriture de la formule ci-dessus est sous forme d'une autoconvolution :

$$g = z * \check{z}$$

où $\check{z}(\mathbf{x}) = z(-\mathbf{x})$ est l'application transposée de z et où $*$ désigne le produit de convolution.

La variable régionalisée et son covariogramme transitif ne sont pas définis dans le même espace : la première est associée à l'espace géographique (vecteur $\mathbf{x}$), le second à celui des "séparations" (vecteur $\mathbf{h}$). Le covariogramme a une signification objective, mais il faudrait connaître exhaustivement la variable régionalisée pour le déterminer. Il possède les propriétés mathématiques suivantes :

- il n'est non nul que sur un domaine borné, car z s'annule en dehors de $\mathcal{D}$
- parité : $\forall \mathbf{h} \in \mathbb{R}^d, g(\mathbf{h}) = g(-\mathbf{h})$
- inégalité de Cauchy-Schwarz : $\forall \mathbf{h} \neq \mathbf{0}, |g(\mathbf{h})| < g(\mathbf{0})$
- $\int g(\mathbf{h}) d\mathbf{h} = \left[\int z(\mathbf{x}) d\mathbf{x} \right]^2 = q^2$

$$\begin{aligned} \text{En effet, } \int g(\mathbf{h}) d\mathbf{h} &= \int \int z(\mathbf{x}) z(\mathbf{x} + \mathbf{h}) d\mathbf{x} d\mathbf{h} \\ &= \int z(\mathbf{x}) d\mathbf{x} \int z(\mathbf{x} + \mathbf{h}) d\mathbf{h} && \text{en permutant les intégrales} \\ &= \int z(\mathbf{x}) d\mathbf{x} \int z(\mathbf{y}) d\mathbf{y} && \text{en posant } \mathbf{y} = \mathbf{x} + \mathbf{h} \\ &= \left[\int z(\mathbf{x}) d\mathbf{x} \right]^2 = q^2 \end{aligned}$$

- $g(h)$ est une fonction de *type positif*, c'est-à-dire qu'il vérifie :

$$\forall k \in \mathbb{N}^*, \forall \lambda_1 \dots \lambda_k \in \mathbb{R}, \forall \mathbf{x}_1 \dots \mathbf{x}_k \in \mathbb{R}^d, \sum_{i=1}^k \sum_{j=1}^k \lambda_i \lambda_j g(\mathbf{x}_i - \mathbf{x}_j) \geq 0.$$

$$\begin{aligned} \text{En effet, } \sum_{i=1}^k \sum_{j=1}^k \lambda_i \lambda_j g(\mathbf{x}_i - \mathbf{x}_j) &= \sum_{i=1}^k \sum_{j=1}^k \lambda_i \lambda_j \int z(\mathbf{x}) z(\mathbf{x} + \mathbf{x}_i - \mathbf{x}_j) d\mathbf{x} \\ &= \int \sum_{i=1}^k \sum_{j=1}^k \lambda_i \lambda_j z(\mathbf{x} + \mathbf{x}_j) z(\mathbf{x} + \mathbf{x}_i) d\mathbf{x} \\ &= \int \left[\sum_{i=1}^k \lambda_i z(\mathbf{x} + \mathbf{x}_i) \right]^2 d\mathbf{x} \geq 0 \end{aligned}$$

Cette dernière propriété est très contraignante et il est difficile de contrôler si une fonction donnée la satisfait. Comme les données disponibles sur la variable régionalisée ne permettront de connaître le covariogramme transitif que de manière partielle, on sera contraint de choisir un *modèle* de covariogramme parmi les fonctions de type positif. On étudiera ultérieurement la façon dont on modélise le covariogramme transitif à partir d'un ensemble de données expérimentales.

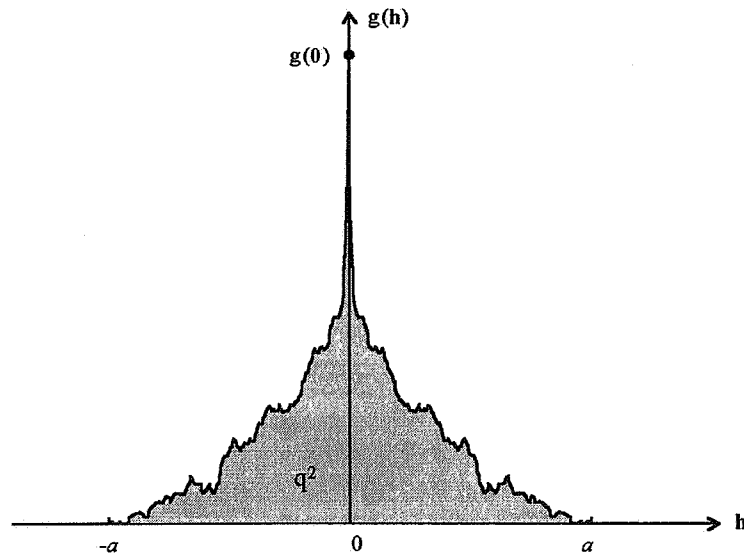


figure 1.1 : allure d'un covariogramme transitif dans R pour une variable régionalisée positive

Nous allons voir que le covariogramme transitif résume les traits structuraux essentiels de la variable régionalisée, et en même temps contient des informations sur la géométrie du champ. Les concepts que nous allons introduire seront repris et étudiés de manière détaillée au chapitre 3 pour d'autres outils structuraux (covariance et variogramme).

II.2. Portée

Dans une direction donnée de l'espace, la *portée* correspond à la distance au-delà de laquelle le covariogramme transitif est identiquement nul ; il s'agit de la plus grande longueur de $\mathcal{D}$ dans cette direction. C'est une quantité géométrique, car elle dépend du champ et non des valeurs de la variable régionalisée. Il convient de signaler que cette propriété disparaîtra en géostatistique probabiliste, lorsque l'on étudiera d'autres outils structuraux (covariance et variogramme) : la portée sera alors une caractéristique de la variable elle-même.

II.3. Comportement à l'origine

Contrairement à la portée, le comportement à l'origine de $g(\mathbf{h})$ dépend de la variable régionalisée elle-même et de ses propriétés de régularité spatiale. Ce lien est mis en évidence par la relation

$$g(\mathbf{0}) - g(\mathbf{h}) = \frac{1}{2} \int [z(\mathbf{x} + \mathbf{h}) - z(\mathbf{x})]^2 d\mathbf{x}.$$

Plus précisément :

- si $z(\mathbf{x})$ est dérivable, $g(\mathbf{h})$ a un comportement parabolique à l'origine ;
- si $z(\mathbf{x})$ est continue (éventuellement par morceaux), $g(\mathbf{h})$ a un comportement linéaire à l'origine ;
- si $z(\mathbf{x})$ est discontinue, $g(\mathbf{h})$ est discontinu à l'origine ; on dit qu'il y a *effet de pépité*. Plutôt qu'une discontinuité, il s'agit généralement d'une décroissance très rapide qui se confond avec une discontinuité (voir par exemple la figure 1.1).

II.4. Isotropie

Par définition, le covariogramme transitif $g(\mathbf{h})$ est *isotrope* si sa valeur ne dépend pas de l'orientation du vecteur $\mathbf{h}$, mais seulement de son module $|\mathbf{h}|$.

Dans le cas contraire, il y a *anisotropie*. L'exemple le plus simple est l'*anisotropie géométrique*, qui s'obtient en déformant un covariogramme isotrope par une transformation linéaire des coordonnées (rotation suivie d'une homothétie) : en étirant ou en raccourcissant les coordonnées dans la direction d'anisotropie, on retrouve le cas isotrope. Toutes les anisotropies ne peuvent cependant pas être traitées aussi facilement.

II.5. Structures gigognes

Un covariogramme gigogne se manifeste par la présence de ruptures de pente, lesquelles indiquent des changements dans la structuration des données en fonction de leur séparation. En pratique, il est représenté par la superposition de plusieurs covariogrammes élémentaires, appelés *structures gigognes*, qui présentent des caractéristiques (portée, comportement à l'origine, anisotropie) différentes.

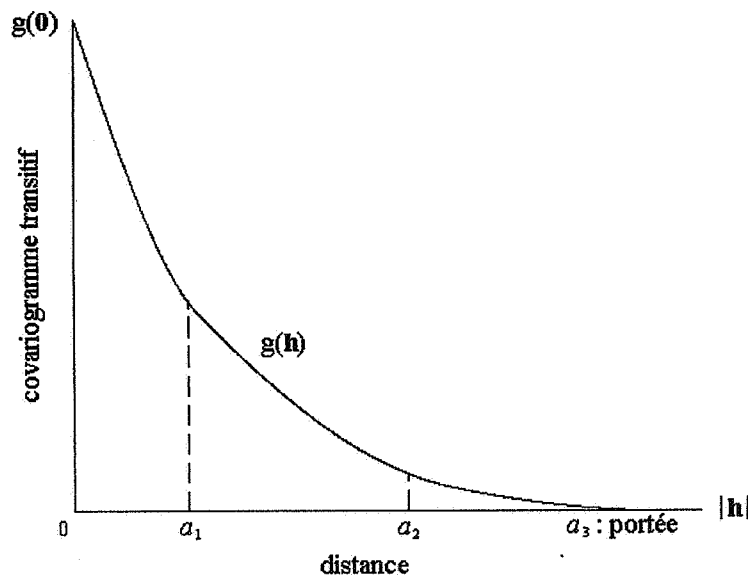


figure 1.2 : covariogramme gigogne composé de trois structures emboîtées

II.6. Covariogramme géométrique

Un cas particulier de covariogramme transif est le **covariogramme géométrique**, noté $K(h)$. Il est défini à partir de la variable indicatrice :

$$1_{\mathcal{D}}(x) = \begin{cases} 1 & \text{si } x \in \mathcal{D} \\ 0 & \text{sinon} \end{cases}$$

par :

$$K(h) = \int 1_{\mathcal{D}}(x) 1_{\mathcal{D}}(x+h) dx = |\mathcal{D} \cap \mathcal{D}_{-h}|.$$

$K(h)$ représente la mesure (longueur, aire, volume...) de $\mathcal{D} \cap \mathcal{D}_{-h}$, où $\mathcal{D}_{-h}$ est le translaté de $\mathcal{D}$ par le vecteur $-h$. En particulier, $K(0)$ est la mesure de $\mathcal{D}$, ce qui implique la relation supplémentaire

$$K(0)^2 = \left[\int 1_{\mathcal{D}}(x) dx \right]^2 = \int K(h) dh.$$

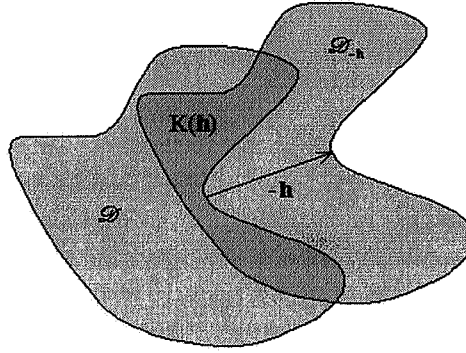


figure 1.3 : covariogramme géométrique d'un domaine $\mathcal{D}$

Le covariogramme géométrique décrit la forme du champ $\mathcal{D}$. Il a toujours un comportement linéaire à l'origine. On montre que, si le champ est convexe, l'opposé de la dérivée à l'origine dans une direction donnée représente le diamètre apparent de $\mathcal{D}$ dans cette direction. Sous certaines conditions, on peut relier les dérivées à l'origine dans toutes les directions au périmètre de $\mathcal{D}$ (cas bidimensionnel) ou à sa surface (cas tridimensionnel). Bien qu'il existe une dérivée à l'origine dans toutes les directions, le covariogramme géométrique n'est pas dérivable en $\mathbf{h} = \mathbf{0}$.

Par exemple, dans $\mathbb{R}^3$, le covariogramme géométrique d'un champ sphérique, de diamètre a , est isotrope et s'écrit :

$$K(\mathbf{h}) = \begin{cases} \frac{\pi a^3}{6} \left(1 - \frac{3|\mathbf{h}|}{2a} + \frac{1}{2} \frac{|\mathbf{h}|^3}{a^3} \right) & \text{pour } |\mathbf{h}| \leq a \\ 0 & \text{pour } |\mathbf{h}| \geq a \end{cases}$$

Il a un comportement linéaire à l'origine, de pente égale à $-\pi a^2/4$. Cette fonction, que l'on retrouvera dans les chapitres suivants, est appelée *modèle sphérique de portée a*.

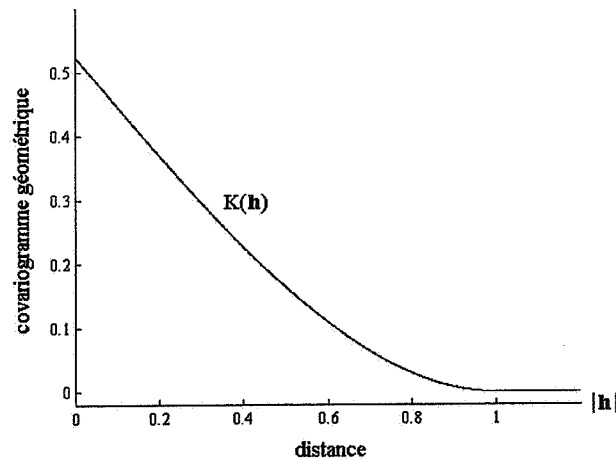


figure 1.4 : covariogramme géométrique de la sphère de diamètre unité

Le covariogramme géométrique présente un intérêt lorsque l'on cherche à estimer la superficie ou le volume du champ, qui n'est autre que l'abondance de la variable indicatrice $1_{\mathcal{D}}(\mathbf{x})$.

III. L'estimation de l'abondance

Nous allons examiner, dans la suite de ce paragraphe, comment estimer l'abondance selon la configuration géométrique des sites échantillonnés : régulière, aléatoire (stratifiée³, uniforme ou préférentielle), ou encore irrégulière quelconque.

Pour alléger les notations, on se limite au cas courant où $\mathcal{D} \subset \mathbb{R}^2$; la généralisation à des espaces de dimension supérieure est immédiate.

III.1. Echantillonnage régulier

On suppose que les observations sont réparties sur une grille régulière, que l'on prendra rectangulaire pour plus de commodité. La cellule rectangulaire élémentaire, notée $[a,b]$, est définie par les vecteurs orthogonaux $\mathbf{a}$ et $\mathbf{b}$. Ses côtés ont pour longueur $|\mathbf{a}|$ et $|\mathbf{b}|$. On choisit arbitrairement un site d'observation $\mathbf{x}_0$ comme origine du réseau d'échantillonnage. Les sites échantillonnés seront repérés par deux indices entiers η et v , de la manière suivante :

$$\mathbf{x}_{\eta,v} = \mathbf{x}_0 + \eta \mathbf{a} + v \mathbf{b} \quad \text{avec } \eta, v \in \mathbb{Z}.$$

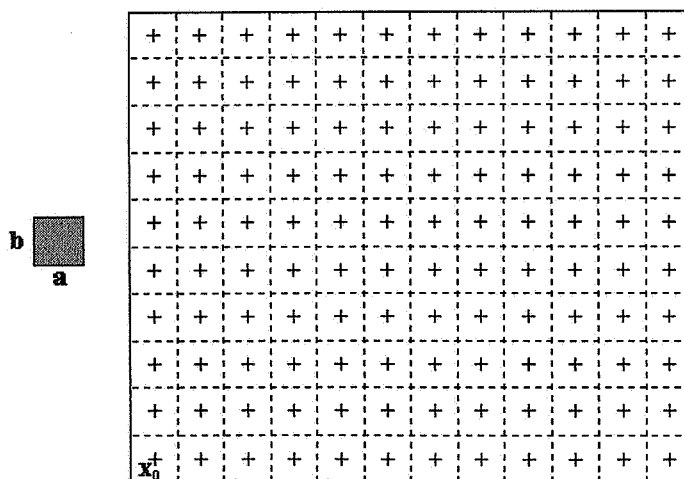


figure 1.5 : échantillonnage régulier (maille rectangulaire)

³ L'espace est partitionné selon une grille régulière et un site est tiré au hasard dans chaque cellule de la grille, indépendamment des autres sites.

III.1.1. Définition de l'estimateur

A partir des mesures $\{ z(\mathbf{x}_0 + \eta \mathbf{a} + \nu \mathbf{b}), \eta, \nu \in \mathbb{Z} \}$, l'abondance peut être estimée par la combinaison linéaire suivante⁴ :

$$q^*(\mathbf{x}_0) = |\mathbf{a}||\mathbf{b}| \sum_{\eta, \nu \in \mathbb{Z}} z(\mathbf{x}_0 + \eta \mathbf{a} + \nu \mathbf{b}).$$

La simplicité de cet estimateur, où chaque observation reçoit le même poids, semble intuitive : quelle observation privilégier lorsque le réseau d'échantillonnage est régulier ? Par ailleurs, comme $z(\mathbf{x})$ s'annule en dehors de $\mathcal{D}$ (borné), la somme ne comporte qu'un nombre fini de termes non nuls. En pratique, il suffit donc de s'assurer que le réseau de prélèvements "déborde" le champ pour pouvoir calculer l'estimateur $q^*(\mathbf{x}_0)$. Nous allons voir que l'approche transitive permet non seulement d'estimer l'abondance, mais aussi d'apprécier l'amplitude de l'erreur commise.

III.1.2. Randomisation du réseau d'échantillonnage

Faute d'information sur la variable régionalisée et sur la géométrie de son champ, il n'y a aucune raison de commencer l'échantillonnage en un endroit plutôt qu'un autre : l'ensemble du réseau d'échantillonnage $\{ \mathbf{x}_{\eta, \nu}, \eta, \nu \in \mathbb{Z} \}$ a été positionné d'une manière quelconque dans l'espace. Cela revient à dire que l'origine $\mathbf{x}_0$ de ce réseau a été implantée "au hasard" ; sa position peut donc être considérée comme une variable aléatoire⁵ $\mathbf{X}_0$, uniformément distribuée dans la surface d'une cellule élémentaire $[a, b]$ de la grille.

On peut alors considérer

$$q^*(\mathbf{X}_0) = |\mathbf{a}||\mathbf{b}| \sum_{\eta, \nu \in \mathbb{Z}} z(\mathbf{X}_0 + \eta \mathbf{a} + \nu \mathbf{b})$$

comme une variable aléatoire, et en calculer l'espérance :

$$\begin{aligned} E[q^*(\mathbf{X}_0)] &= \int_{[a, b]} q^*(\mathbf{x}) \frac{d\mathbf{x}}{|\mathbf{a}||\mathbf{b}|} \\ &= \sum_{\eta, \nu \in \mathbb{Z}} \int_{[a, b]} z(\mathbf{x} + \eta \mathbf{a} + \nu \mathbf{b}) d\mathbf{x} \\ &= \int z(\mathbf{x}) d\mathbf{x} = q \end{aligned}$$

⁴ L'astérisque en exposant indique qu'il s'agit d'une estimation.

⁵ Pour bien distinguer la variable aléatoire de la variable déterministe, la première sera notée avec une majuscule ($\mathbf{X}_0$) et la seconde avec une minuscule ($\mathbf{x}_0$).

L'estimateur de l'abondance est donc **sans biais** :

$$E[q^*(X_0) - q] = 0.$$

Cette égalité signifie que, si l'on répétait de nombreuses fois l'expérience, en déplaçant l'origine x_0 de la grille d'échantillonnage et en calculant à chaque fois l'estimateur $q^*(x_0)$, l'erreur moyenne commise tendrait vers zéro.

L'absence de biais n'implique pas que l'estimateur soit précis. Pour connaître la qualité de l'estimation, il faut mesurer la dispersion de l'erreur autour de zéro, par exemple en calculant sa variance. On montre que l'expression de cette **variance d'estimation**⁶ fait intervenir le covariogramme transitif $g(h)$:

$$\text{var}[q^*(X_0) - q] = \text{var}[q^*(X_0)] = |a||b| \sum_{\kappa, \ell \in \mathbb{Z}} g(\kappa a + \ell b) - \int g(h) dh.$$

En désignant par $\int_{[a,b]}$ l'intégrale sur la cellule $[a,b]$, on peut écrire :

$$\begin{aligned} E\{[q^*(X_0)]^2\} &= |a|^2 |b|^2 E\left\{\left[\sum_{\eta, v \in \mathbb{Z}} z(X_0 + \eta a + v b)\right]^2\right\} \\ &= |a|^2 |b|^2 E\left\{\sum_{\kappa, \ell \in \mathbb{Z}} \sum_{\eta, v \in \mathbb{Z}} z(X_0 + \eta a + v b) z(X_0 + \eta a + v b + \kappa a + \ell b)\right\} \\ &= |a|^2 |b|^2 \sum_{\kappa, \ell \in \mathbb{Z}} \sum_{\eta, v \in \mathbb{Z}} \int_{[a,b]} z(x + \eta a + v b) z(x + \eta a + v b + \kappa a + \ell b) \frac{dx}{|a||b|} \\ &= |a||b| \sum_{\kappa, \ell \in \mathbb{Z}} \int z(x) z(x + \kappa a + \ell b) dx \\ &= |a||b| \sum_{\kappa, \ell \in \mathbb{Z}} g(\kappa a + \ell b) \end{aligned}$$

$$\{E[q^*(X_0)]\}^2 = q^2 = \int g(h) dh \quad (\text{cf. les propriétés du covariogramme transitif}).$$

$$\text{D'où } \text{var}[q^*(X_0)] = E\{[q^*(X_0)]^2\} - \{E[q^*(X_0)]\}^2 = |a||b| \sum_{\kappa, \ell \in \mathbb{Z}} g(\kappa a + \ell b) - \int g(h) dh.$$

Remarque

Il faut bien voir que ces résultats ne font appel à aucune interprétation probabiliste de la variable régionalisée, qui reste une fonction déterministe, mais connue seulement de manière fragmentaire. C'est la raison pour laquelle la géostatistique transitive se range parmi les méthodes d'estimation non stochastiques. **Ce n'est pas la variable régionalisée qui est probabilisée, mais le réseau d'échantillonnage.**

⁶ Ici, la *variance d'estimation* désigne indifféremment la variance de l'erreur d'estimation et celle de l'estimateur lui-même, ces deux quantités ne différant que d'une constante déterministe (à savoir l'abondance q) qui est invisible dans une variance.

III.1.3. Commentaires sur la variance d'estimation

Cette variance quantifie la **précision** de l'estimation. On voit qu'elle ne fait intervenir que le covariogramme transitif $g(\mathbf{h})$ et les dimensions de la maille d'échantillonnage $|\mathbf{a}|$ et $|\mathbf{b}|$. Elle s'interprète comme la différence entre la valeur exacte de l'intégrale de la fonction $g(\mathbf{h})$ et son approximation par une somme discrète sur la grille de prélèvement. Cette différence est d'autant plus faible que

- $|\mathbf{a}|$ et $|\mathbf{b}|$ sont petits, c'est-à-dire l'échantillonnage dense ;
- $g(\mathbf{h})$ est une fonction régulière ; ainsi, pour un échantillonnage fixé, la régularité spatiale de la variable régionalisée détermine la précision de l'estimation.

III.1.4. Insuffisance des valeurs expérimentales

Pour calculer le covariogramme transitif $g(\mathbf{h})$, la connaissance exhaustive de la variable régionalisée est nécessaire. Or cette connaissance est, par hypothèse, fragmentaire : $z(\mathbf{x})$ n'est connue qu'en un nombre fini de sites. Un calcul exact de la variance d'estimation est donc illusoire. On ne peut proposer en fait qu'une *estimation* de cette variance, dont la véritable valeur reste inaccessible en pratique.

A partir des valeurs expérimentales, on peut estimer $g(\mathbf{h})$ aux points $\{\kappa \mathbf{a} + \ell \mathbf{b} ; \kappa, \ell \in \mathbb{Z}\}$ à l'aide de la formule suivante :

$$\forall \kappa, \ell \in \mathbb{Z}, \hat{g}(\kappa \mathbf{a} + \ell \mathbf{b}) = |\mathbf{a}| |\mathbf{b}| \sum_{\eta, \nu} z(\mathbf{x}_0 + \eta \mathbf{a} + \nu \mathbf{b}) z(\mathbf{x}_0 + \eta \mathbf{a} + \nu \mathbf{b} + \kappa \mathbf{a} + \ell \mathbf{b}).$$

Cet estimateur porte le nom de *covariogramme expérimental*. En remplaçant g par $\hat{g}$, on obtient une évaluation du premier terme de l'expression de la variance d'estimation. En ce qui concerne l'intégrale du covariogramme, elle est égale à q^2 et peut être estimée par q^{*2} , lui aussi accessible à l'aide des données expérimentales.

Si à présent on reporte $\hat{g}$ et q^{*2} dans la formule de la variance d'estimation, cette dernière devient identiquement nulle. Ceci signifie que, sans hypothèses supplémentaires, un même ensemble de données ne peut fournir à la fois une estimation et la précision de cette estimation. Pour pouvoir aller plus loin, il faut ajouter de l'information et faire appel à un **modèle** de covariogramme transitif, i.e. à une fonction qui possède les bonnes propriétés mathématiques (fonction paire, maximale à l'origine, de type positif⁷) et s'approche "au mieux" du covariogramme expérimental. C'est ce modèle, défini pour tous les vecteurs $\mathbf{h}$ et pas seulement aux points $\{\kappa \mathbf{a} + \ell \mathbf{b} ; \kappa, \ell \in \mathbb{Z}\}$, qui sera utilisé dans le calcul de la variance d'estimation au lieu du vrai covariogramme.

⁷ Le fait de choisir comme modèle une fonction de type positif est essentiel, puisque cela garantit la positivité de la variance d'estimation globale. Si l'on utilise une fonction $g(\mathbf{h})$ quelconque, on pourrait obtenir des variances d'estimation négatives, ce qui est inacceptable.

III.1.5. Application à des données agronomiques

Nous allons illustrer les concepts précédents sur un jeu de données réelles. Bien que cet exemple soit emprunté au domaine de l'agronomie, les commentaires et explications qui suivent peuvent facilement se transposer à d'autres domaines d'application, en particulier au domaine minier : ce qui est une surface cultivée pourrait très bien être une teneur en métal dans un gisement.

Les données que l'on considère concernent la répartition des surfaces de canne à sucre au nord de Basse-Terre (Guadeloupe). Elles sont extraites d'une cartographie de la sole cannière à l'échelle 1/5000 réalisée par le CIRAD dans le cadre du projet *Connaissance et suivi de la sole cannière par télédétection* (G. Lainé, P. Truong, F. Baleux, 1996). La zone étudiée mesure 19.7 km (nord-sud) sur 21.8 km (est-ouest) ; elle a été divisée en cellules de 1 hectare (100 mètres sur 100 mètres) et reconnue exhaustivement. Ainsi, dans chaque cellule, on connaît la surface réelle de canne à sucre (comprise entre 0 et 1 ha). La variable régionalisée à laquelle on s'intéresse est :

$z(\mathbf{x})$ = surface en canne à sucre dans la cellule contenant $\mathbf{x}$.

La carte 1.5 localise les valeurs de z sur l'ensemble de la zone.

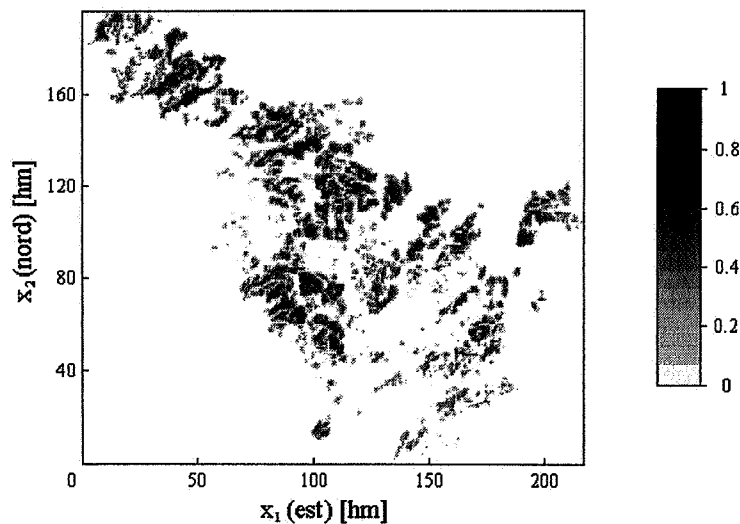


figure 1.6 : surfaces de canne à sucre, Basse-Terre, 1995 (source : CIRAD)

On note que le champ $\mathcal{D}$, qui correspond à l'ensemble des parcelles cultivées (sombres sur la carte), ne couvre pas la totalité de la zone rectangulaire étudiée (en réalité, à peine 20% de la zone). Il apparaît comme relativement morcelé et présente une forte anisotropie (il est plus étendu dans la direction nord-ouest / sud-est que dans la direction orthogonale). Toutefois, au moment où l'on cherche à estimer les superficies agricoles, la géométrie exacte de $\mathcal{D}$ est inconnue, et, faute d'information supplémentaire, les échantillons seront répartis dans la totalité de la zone rectangulaire.

La figure 1.7 donne l'histogramme des valeurs réelles, sur l'ensemble de la zone d'abord, puis sur le seul champ $\mathcal{D}$.

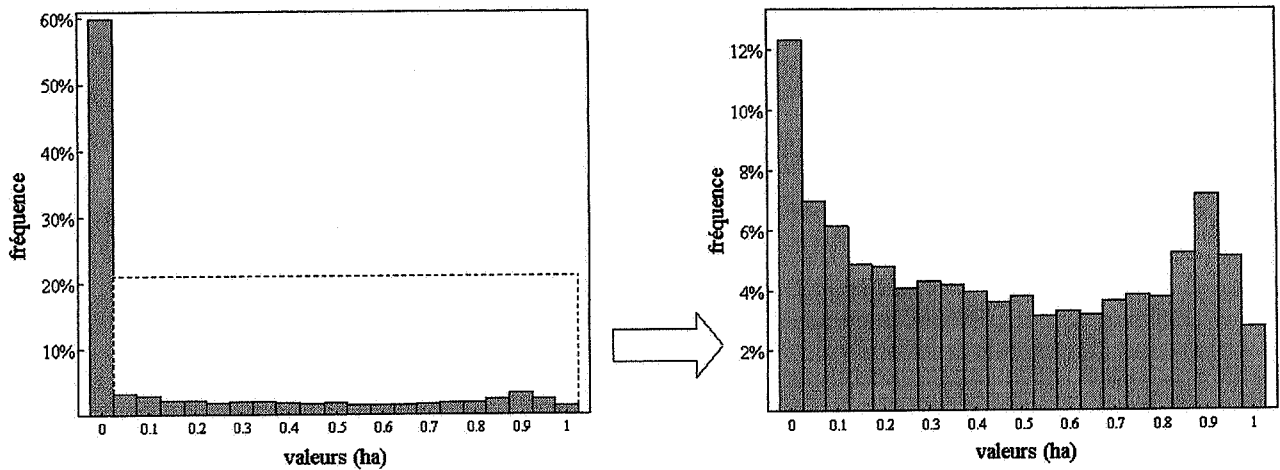


figure 1.7 : histogramme des valeurs réelles
à gauche : toutes les valeurs ; à droite, les valeurs non nulles

Le premier histogramme présente une classe très fréquente correspondant aux 80% de valeurs nulles (hors champ). Les autres classes de valeurs ont des effectifs plus réduits mais plus homogènes (second histogramme).

Ici, la réalité est entièrement connue, ce qui est exceptionnel. On est donc en mesure de calculer, pour tout vecteur $\mathbf{h}$ de $\mathbb{R}^2$, le vrai covariogramme transitif $g(\mathbf{h})$. La figure 1.8 donne les valeurs de $g(\mathbf{h})$ dans les différentes directions du plan ("carte covariographique").

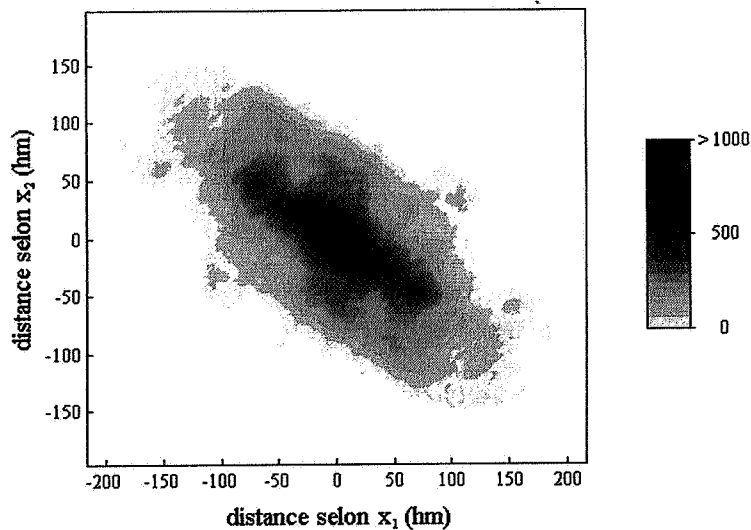


figure 1.8 : covariogramme transitif réel
l'origine est située au centre de l'image

Le covariogramme transitif présente une anisotropie (quasi-géométrique), qui n'est autre que celle du champ $\mathcal{Z}$. On vérifie qu'il est nul aux distances supérieures au diamètre du champ, symétrique par rapport à l'origine $\mathbf{h} = \mathbf{0}$, et maximal en cette origine où il vaut 2402.5 ha². Pour avoir une idée plus précise de $g(\mathbf{h})$, notamment sur son comportement au voisinage de l'origine, on peut le visualiser le long de plusieurs directions du plan.

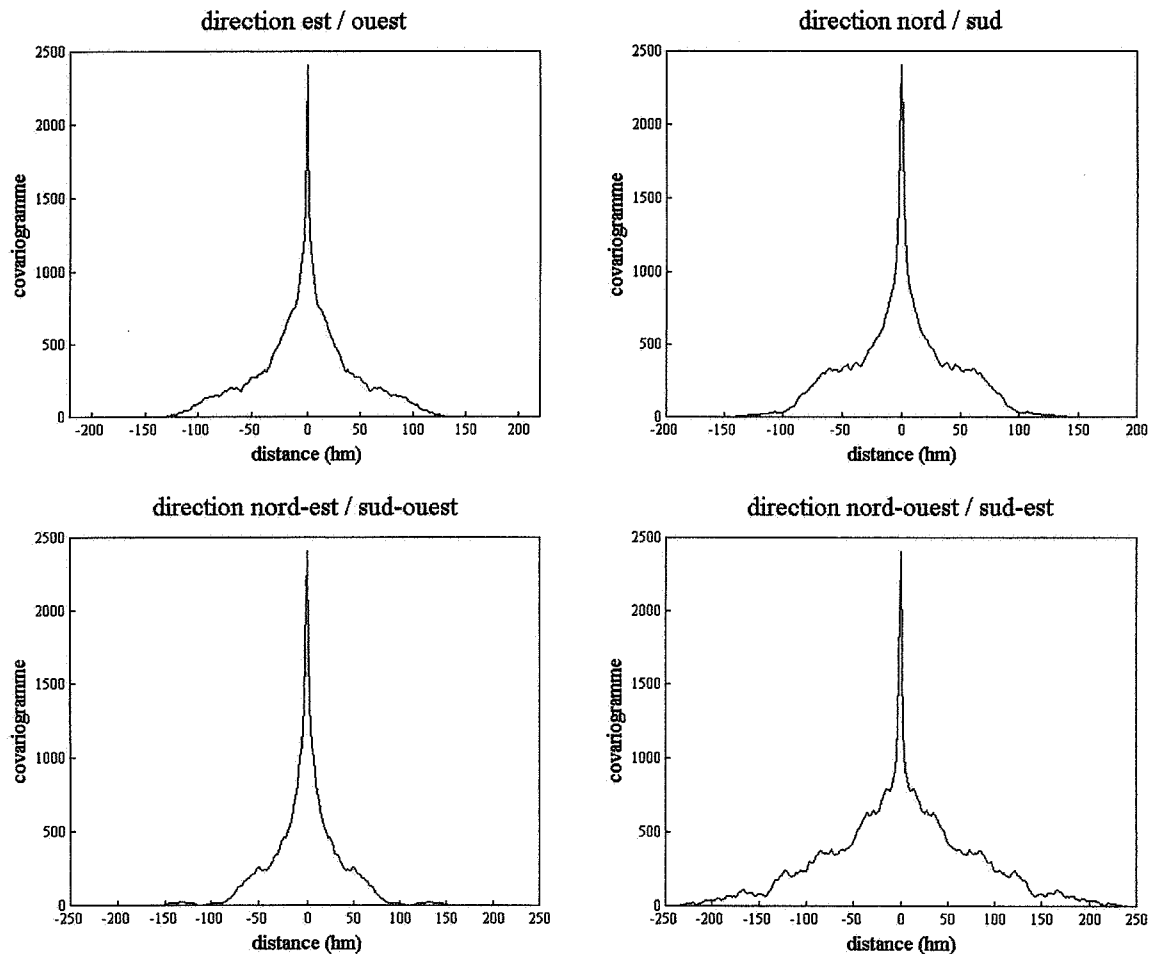


figure 1.9 : covariogramme transitif le long de quatre directions du plan

L'anisotropie apparaît clairement sur ces covariogrammes directionnels, notamment ceux relatifs aux directions nord-est / sud-ouest et nord-ouest / sud-est, qui sont approximativement les directions principales d'anisotropie (c'est-à-dire les directions le long desquelles la portée est respectivement minimale et maximale). Au voisinage de l'origine $\mathbf{h} = \mathbf{0}$, le covariogramme a un comportement linéaire et décroît très rapidement : si l'on adoptait une échelle de travail plus grande, le covariogramme apparaîtrait comme discontinu à l'origine. Par suite, la variable régionalisée est continue, mais présente des transitions rapides, ce qui reflète le fait que les parcelles cultivées apparaissent morcelées dans la zone d'étude.

L'objectif d'une étude pratique est d'évaluer l'abondance, c'est-à-dire la superficie de canne à sucre sur l'ensemble de la zone, à partir d'un échantillonnage fragmentaire (la réalité n'est pas supposée entièrement connue au moment de l'estimation). On pourra comparer l'estimation obtenue avec la valeur réelle de la superficie en canne à sucre, qui vaut 3 416.3 ha. L'objectif d'une telle estimation est la prévision du volume des récoltes en canne à sucre dans la région considérée.

Estimation de l'abondance

On échantillonne selon une grille régulière au taux de 1% (taux habituel en agronomie), c'est-à-dire que l'on échantillonne 1 cellule sur 10 dans les directions nord/sud et est/ouest. A partir de cet échantillon, on estime la superficie en canne à sucre par la somme des superficies mesurées multipliée par le volume de la maille (100 ha).

Si l'on essaie les 100 échantillons possibles, en faisant varier l'origine du réseau de prélèvements, on obtient 100 estimations différentes, dont les valeurs sont distribuées selon l'histogramme ci-dessous.

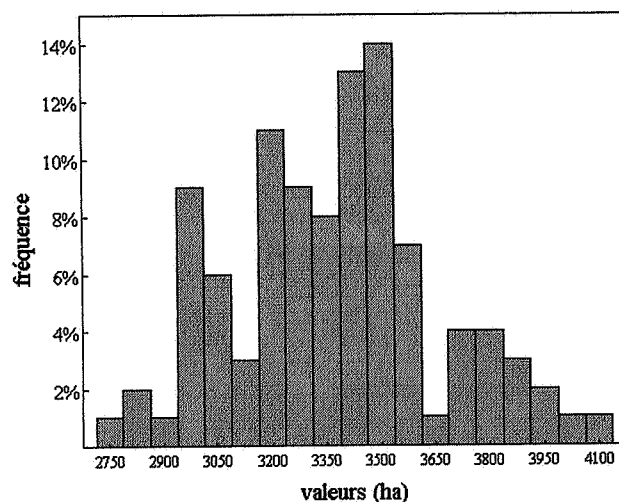


figure 1.10 : histogramme des 100 estimations possibles

L'estimation minimale est de 2 811.4 ha, et l'estimation maximale vaut 4 157.1 ha. On peut vérifier que la moyenne des 100 estimations possibles redonne bien la valeur exacte de la superficie en canne à sucre, à savoir très précisément 3 416.3 ha.

La variance des 100 estimations, qui est aussi la variance des 100 erreurs d'estimation, est égale à 80 119 ha². La connaissance complète de $g(\mathbf{h})$ permet de retrouver exactement cette valeur, grâce à la formule

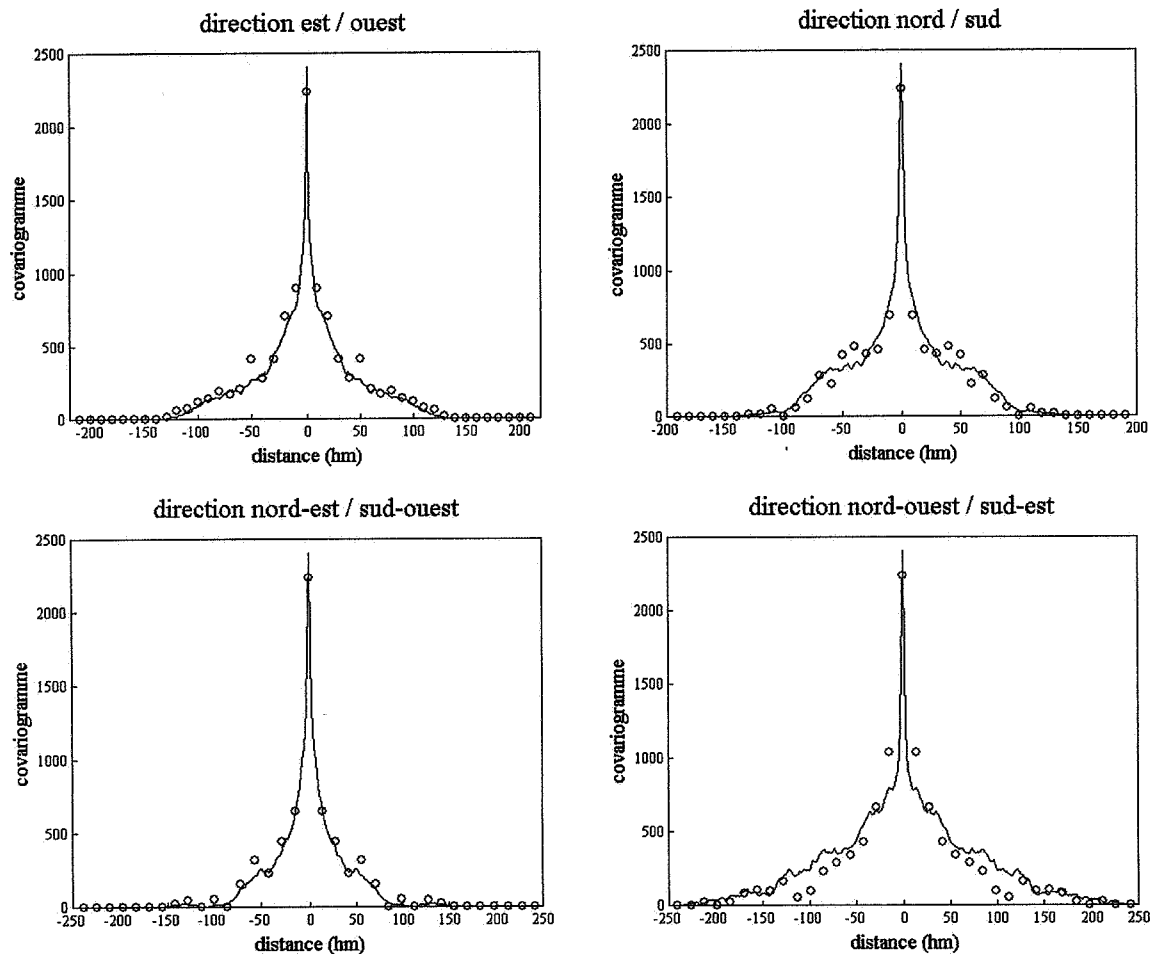
$$\text{var}[q^* - q] = |\mathbf{a}| |\mathbf{b}| \sum_{\kappa, \ell} g(\kappa \mathbf{a} + \ell \mathbf{b}) - \int g(\mathbf{h}) d\mathbf{h}$$

où $\mathbf{a}$ et $\mathbf{b}$ désignent les côtés de la maille d'échantillonnage (ici $|\mathbf{a}| = |\mathbf{b}| = 10$ hm). La variance d'estimation correspond à un écart-type σ^* de 283 ha. Cet écart-type donne l'ordre de grandeur de l'erreur que l'on commet en estimant q par q^* , mais il ne permet pas en théorie de calculer un intervalle de confiance sur l'estimation, car on ne connaît pas la distribution de l'erreur d'estimation⁸.

⁸ En pratique, on a souvent tendance à associer à l'écart-type d'estimation un intervalle de confiance basé sur la distribution gaussienne. Il se trouve ici que cette présomption est relativement bien vérifiée, même si la loi des erreurs d'estimation est loin d'être gaussienne : par exemple, 95% des estimations se situent dans l'intervalle $[q - 2\sigma^*, q + 2\sigma^*]$, et 99% dans l'intervalle $[q - 2.5\sigma^*, q + 2.5\sigma^*]$. Mais la validité d'une telle hypothèse n'est pas garantie en général.

Comparaison entre le covariogramme expérimental et le covariogramme vrai

Sur un échantillonnage particulier, on a calculé le covariogramme expérimental pour des distances multiples de la maille de prélèvement. La figure 1.11 permet de comparer, dans les quatre directions principales du plan, les valeurs du covariogramme expérimental (repérées par des ronds) à celles du covariogramme vrai (ligne continue).



*figure 1.11 : covariogramme expérimental (ronds)
et covariogramme vrai (ligne continue)
le long des quatre directions principales du plan*

Sur cet exemple, on voit qu'il est possible d'estimer de manière très satisfaisante le véritable covariogramme transitive. La variance d'estimation que l'on obtiendrait avec un covariogramme modélisé serait une bonne approximation de la variance réelle.

III.2. Echantillonnage aléatoire stratifié

L'échantillonnage aléatoire stratifié consiste à découper la zone d'étude selon une grille régulière et à tirer au hasard pur la position d'un site dans chaque cellule de cette grille, indépendamment des autres sites. Comme précédemment, on peut repérer chaque site par deux indices entiers η et v , selon la cellule à laquelle il appartient.

On prend comme estimateur de l'abondance :

$$q^*(\mathbf{x}_0, \{\mathbf{x}_{\eta,v}\}) = |\mathbf{a}||\mathbf{b}| \sum_{\eta,v} z(\mathbf{x}_{\eta,v})$$

où $\{\mathbf{x}_{\eta,v} ; \eta, v \in \mathbb{Z}\}$ sont les points échantillonnés dans l'espace $\mathbb{R}^d$;
 $|\mathbf{a}||\mathbf{b}|$ est la surface de la cellule élémentaire (notée $[\mathbf{a}, \mathbf{b}]$) ;
 $\mathbf{x}_0$ est une origine de la grille de stratification choisie arbitrairement.

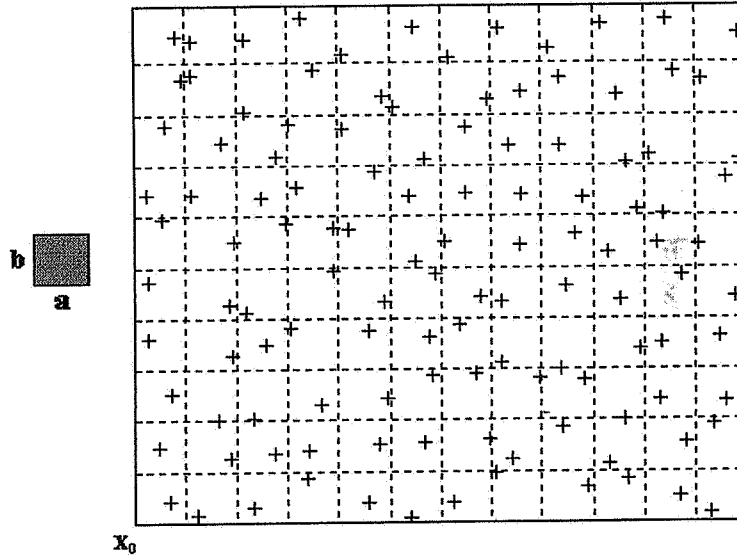


figure 1.12 : échantillonnage aléatoire stratifié

Le raisonnement est analogue à celui du paragraphe précédent : en randomisant la position des échantillons et de l'origine, l'estimateur devient une variable aléatoire dont on peut calculer l'espérance et la variance. Cette démarche se fait en deux étapes. Chaque site $\mathbf{x}_{\eta,v}$ ayant été "tiré au hasard" dans une cellule d'une grille régulière, on randomise dans un premier temps sa position au sein de la cellule à laquelle il appartient, selon une loi de probabilité uniforme, d'où l'obtention de variables aléatoires mutuellement indépendantes $\mathbf{X}_{\eta,v}$ ($\eta, v \in \mathbb{Z}$). Puis on considère l'implantation de l'origine de la grille $\mathbf{x}_0$ comme une autre variable aléatoire $\mathbf{X}_0$ (indépendante des $\mathbf{X}_{\eta,v}$) uniformément distribuée sur une cellule $[\mathbf{a}, \mathbf{b}]$ de la grille.

L'estimateur $q^*(\mathbf{X}_0, \{\mathbf{X}_{\eta,v}\})$ est toujours **sans biais** :

$$E[q^*(\mathbf{X}_0, \{\mathbf{X}_{\eta,v}\})] = q.$$

On montre que la **variance d'estimation** s'écrit :

$$\text{var}[q^*(X_0, \{X_{\eta_v}\}) - q] = |a||b| (g(0) - \bar{g}_{[a,b]})$$

où $\bar{g}_{[a,b]}$ désigne la valeur moyenne de $g(x-y)$ lorsque x et y décrivent séparément $[a,b]$:

$$\bar{g}_{[a,b]} = \frac{1}{|a|^2 |b|^2} \int_{[a,b]} \int_{[a,b]} g(x-y) dx dy.$$

La variance d'estimation est bien positive, car d'après l'inégalité de Cauchy-Schwarz, le covariogramme transitif est maximal à l'origine, d'où $\bar{g}_{[a,b]} < g(0)$.

Cette variance ne dépend que du comportement de $g(h)$ dans un voisinage de l'origine défini par la cellule $[a,b]$. Elle est d'autant plus élevée que la décroissance de $g(h)$ dans $[a,b]$ est rapide ($\bar{g}_{[a,b]}$ petit), donc que la variable régionalisée est peu régulière.

III.3. Echantillonnage aléatoire uniforme

On suppose que les sites d'observation sont implantés au hasard pur, indépendamment les uns des autres, à l'intérieur d'un domaine Σ contenant le champ $\mathcal{D}$ (lequel n'est pas forcément connu).

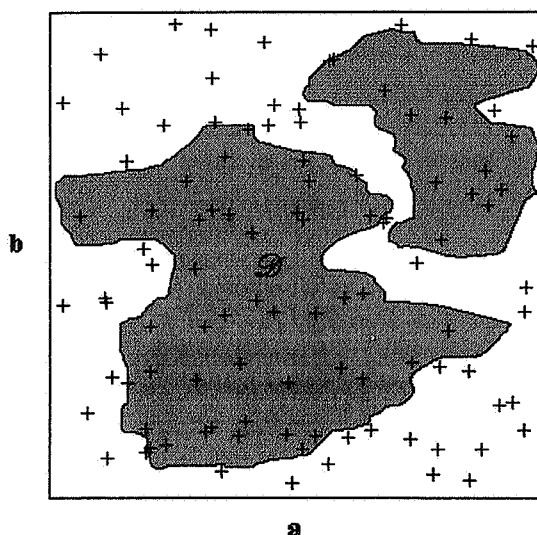


figure 1.13 : échantillonnage aléatoire pur sur $\Sigma = [a,b]$

Un tel ensemble de points peut être vu comme la réalisation d'un processus aléatoire uniforme. Ce processus est caractérisé par le paramètre $\theta = n/|\Sigma|$, qui donne la densité des points, c'est-à-dire le nombre moyen de points par unité de surface ou de volume.

Dans ce cas, on propose l'estimateur suivant de l'abondance :

$$q^*(\mathbf{x}_1, \dots, \mathbf{x}_n) = \frac{1}{\theta} \sum_{\alpha=1}^n z(\mathbf{x}_\alpha).$$

Les positions des sites $\mathbf{x}_\alpha$ ayant été implantées au hasard à l'intérieur du domaine Σ , elles peuvent être randomisées et deviennent des variables aléatoires $\mathbf{X}_\alpha$ indépendantes et uniformément distribuées à l'intérieur de Σ avec une densité

$$\frac{\theta}{n} = \frac{1}{|\Sigma|}.$$

L'estimateur devient aussi une variable aléatoire ; son espérance est égale à l'abondance réelle⁹ :

$$E[q^*(\mathbf{X}_1, \dots, \mathbf{X}_n)] = \frac{1}{\theta} \sum_{\alpha=1}^n \int_{\Sigma} z(\mathbf{x}_\alpha) \frac{\theta d\mathbf{x}_\alpha}{n} = \frac{1}{n} [n \times \int_{\Sigma} z(\mathbf{x}) d\mathbf{x}] = \int_{\mathcal{D}} z(\mathbf{x}) d\mathbf{x} = q.$$

L'estimateur est encore **sans biais**. Quant à la variance d'estimation, elle s'exprime à l'aide du covariogramme transitif g :

$$\begin{aligned} \text{var}[q^*(\mathbf{X}_1, \dots, \mathbf{X}_n)] &= \frac{1}{\theta^2} \sum_{\alpha=1}^n \text{var}[z(\mathbf{X}_\alpha)] \quad \text{en supposant les } \mathbf{X}_\alpha \text{ indépendantes} \\ &= \frac{n}{\theta^2} \left\{ \int_{\Sigma} z^2(\mathbf{x}) \frac{\theta d\mathbf{x}}{n} - \left[\int_{\Sigma} z(\mathbf{x}) \frac{\theta d\mathbf{x}}{n} \right]^2 \right\} \\ &= \frac{1}{\theta} \underbrace{\int_{\Sigma} z^2(\mathbf{x}) d\mathbf{x}}_{g(0)} - \frac{1}{n} \underbrace{\left[\int_{\Sigma} z(\mathbf{x}) d\mathbf{x} \right]^2}_{\int g(\mathbf{h}) d\mathbf{h}} \\ &= \frac{1}{\theta} \left[g(0) - \frac{1}{|\Sigma|} \int g(\mathbf{h}) d\mathbf{h} \right] \end{aligned}$$

On notera que cette variance d'estimation dépend explicitement du domaine Σ dans lequel se répartissent les échantillons. En conséquence, la délimitation plus ou moins judicieuse de Σ se répercutera sur la qualité de l'estimateur.

⁹ La variable régionalisée s'annulant en dehors de $\mathcal{D}$, les intégrales sur Σ s'identifient aux intégrales sur $\mathcal{D}$ ou encore sur tout l'espace.

III.4. Echantillonnage aléatoire préférentiel

Un **échantillonnage** irrégulier est **préférentiel** lorsque le taux d'échantillonnage n'est pas uniforme dans l'espace. Il peut être modélisé à l'aide d'un processus aléatoire à densité variable (ou densité *régionalisée*) : les sites d'observation $\{ \mathbf{x}_\alpha, \alpha = 1 \dots n \}$ sont considérés comme des réalisations indépendantes d'un processus ponctuel de densité $\theta(\mathbf{x})$ sur le domaine échantillonné Σ . Pour obtenir n points en moyenne, on doit poser :

$$\int_{\Sigma} \theta(\mathbf{x}) d\mathbf{x} = n.$$

Pour tenir compte du caractère non uniforme de l'échantillonnage dans l'estimation de l'abondance, on pondère la valeur $z(\mathbf{x}_\alpha)$ observée en chaque point de mesure par l'inverse de la densité de l'échantillonnage $\theta(\mathbf{x}_\alpha)$ en ce même point :

$$q^*(\mathbf{x}_1, \dots, \mathbf{x}_n) = \sum_{\alpha=1}^n \frac{z(\mathbf{x}_\alpha)}{\theta(\mathbf{x}_\alpha)}.$$

En restituant aux sites d'observation $\{ \mathbf{x}_\alpha, \alpha = 1 \dots n \}$ leur caractère aléatoire, q^* devient une variable aléatoire. On montre que, si $\theta(\mathbf{x})$ ne s'annule pas sur Σ , c'est un estimateur **sans biais** :

$$E[q^*] = \sum_{\alpha=1}^n \int_{\Sigma} \frac{z(\mathbf{x}_\alpha)}{\theta(\mathbf{x}_\alpha)} \frac{\theta(\mathbf{x}_\alpha) d\mathbf{x}_\alpha}{n} = n \times \int_{\Sigma} z(\mathbf{x}) \frac{d\mathbf{x}}{n} = \int_{\Sigma} z(\mathbf{x}) d\mathbf{x} = q$$

et que la variance d'estimation vaut :

$$\begin{aligned} \text{var}[q^* - q] &= \sum_{\alpha=1}^n \text{var} \left(\frac{z(\mathbf{X}_\alpha)}{\theta(\mathbf{X}_\alpha)} \right) \\ &= n \times \int_{\Sigma} \frac{z(\mathbf{x})^2 \theta(\mathbf{x}) d\mathbf{x}}{\theta(\mathbf{x})^2 n} - n \times \left(\int_{\Sigma} \frac{z(\mathbf{x}) \theta(\mathbf{x}) d\mathbf{x}}{\theta(\mathbf{x}) n} \right)^2 \\ &= \int \frac{z(\mathbf{x})^2}{\theta(\mathbf{x})} d\mathbf{x} - \frac{1}{n} \int g(\mathbf{h}) d\mathbf{h} \\ &= g_y(0) - \frac{1}{n} \int g(\mathbf{h}) d\mathbf{h} \end{aligned}$$

où g_y désigne le covariogramme transitif de la variable $y(\mathbf{x}) = z(\mathbf{x}) / \sqrt{\theta(\mathbf{x})}$.

On vérifie que pour une densité constante $\theta = n/|\Sigma|$, on retrouve bien les expressions de l'estimateur et de la variance d'estimation données dans le cas d'un échantillonnage aléatoire uniforme.

Un calcul rigoureux de la variance d'estimation requiert une modélisation *conjointe* de g et g_y . A ce titre, on pourra se référer au chapitre 7, qui traite de l'ajustement bivariable pour d'autres outils structuraux (covariances et variogrammes). En particulier, il serait nécessaire de définir un covariogramme "croisé" g_{zy} entre z et y ; les restrictions théoriques peuvent cependant être allégées, car on n'a besoin de connaître que la valeur à l'origine de g_y .

On peut se demander quelle densité minimise la variance d'estimation. Supposons que la variable régionalisée soit $z(\mathbf{x})$ positive et prenons une densité $\theta(\mathbf{x})$ proportionnelle à $z(\mathbf{x})$. Alors, quel que soit le point $\mathbf{X}_\alpha$ échantillonné, $z(\mathbf{X}_\alpha) / \theta(\mathbf{X}_\alpha)$ est constant et donc de variance nulle. On en déduit que, dans ce cas

$$\text{var}[q^* - q] = \sum_{\alpha=1}^n \text{var}[z(\mathbf{X}_\alpha) / \theta(\mathbf{X}_\alpha)] = 0.$$

Il s'agit bien de la valeur minimale possible, puisqu'une variance est toujours positive.

La stratégie d'échantillonnage qui minimise la variance d'estimation consiste donc à tirer les sites avec une densité $\theta(\mathbf{x})$ proportionnelle à $z(\mathbf{x})$. L'estimateur est alors presque sûrement égal à l'abondance, autrement dit l'estimation est exacte. Ce résultat est moins surprenant qu'il n'y paraît, puisqu'un tel échantillonnage nécessite de connaître la densité $\theta(\mathbf{x})$ – et donc $z(\mathbf{x})$ – partout, ce qui n'est pas envisageable en pratique. Cette remarque présente cependant un intérêt lorsque l'on connaît l'allure, ou encore la tendance, de la variable régionalisée (zones de fortes et faibles valeurs...), par exemple au travers d'une variable auxiliaire connue exhaustivement.

III.5. Echantillonnage irrégulier quelconque

La démarche qui vient d'être présentée ne peut s'appliquer que si l'implantation des points de mesure est modélisable de façon appropriée par un processus aléatoire. Dans le cas d'un **échantillonnage quelconque**, il devient difficile de probabiliser l'estimateur, faute de connaître le mode de construction de l'échantillonnage.

Une manière de contourner cette difficulté consiste à diviser le champ en plusieurs sous-zones sur lesquelles la densité de l'échantillonnage peut être considérée comme uniforme, puis à additionner les estimations et les variances d'estimation obtenues dans chaque sous-zone.

Une seconde solution consiste à pondérer les données dans l'estimation de l'abondance, par exemple en attribuant à chaque donnée un poids égal à sa surface ou son volume d'influence dans la zone échantillonnée :

$$q^* = \sum_{\alpha=1}^n |S_\alpha| z(\mathbf{x}_\alpha)$$

où $|S_\alpha|$ est la surface ou le volume d'influence du site $\mathbf{x}_\alpha$.

Les données isolées auront ainsi plus d'importance que les données groupées, dont les valeurs sont en partie redondantes. Pour pouvoir calculer le biais et la variance d'estimation, il faut préalablement convertir l'estimateur en variable aléatoire. Or, si l'échantillonnage est quelconque, on ne peut randomiser la position des échantillons. **Seule une estimation globale est fournie, sans que l'on puisse définir son biais ni sa variance.**

IV. L'estimation locale : le krigeage transitif

Le covariogramme transitif peut servir à l'estimation locale de la variable régionalisée grâce à la technique du **krigeage transitif** que nous allons détailler ci-dessous. En chaque site $\mathbf{x}$ de l'espace, on estime la valeur $z(\mathbf{x})$, en général inconnue, à l'aide d'une combinaison linéaire pondérée des données avoisinantes (y compris celles qui ne sont pas dans le champ), situées en des points $\{\mathbf{x}_\alpha, \alpha = 1 \dots n\}$:

$$z^*(\mathbf{x}) = \sum_{\alpha} \lambda_{\alpha} z(\mathbf{x}_{\alpha}) \quad \text{estimateur de } z(\mathbf{x}).$$

Pour spécifier cet estimateur, un premier critère consiste à rechercher les poids λ_{α} tels que, *si la configuration géométrique des points $\mathbf{x}$ et $\mathbf{x}_{\alpha}$ pouvait être déplacée dans l'espace*, la somme des erreurs quadratiques entre valeurs vraies et estimées serait minimale. Or, cette somme s'écrit :

$$\begin{aligned} \int [z(\mathbf{x}) - z^*(\mathbf{x})]^2 d\mathbf{x} &= \int [z(\mathbf{x}) - \sum_{\alpha} \lambda_{\alpha} z(\mathbf{x}_{\alpha})]^2 d\mathbf{x} \\ &= \int [z(\mathbf{x})^2 - 2 \sum_{\alpha} \lambda_{\alpha} z(\mathbf{x}) z(\mathbf{x}_{\alpha}) + \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} z(\mathbf{x}_{\alpha}) z(\mathbf{x}_{\beta})] d\mathbf{x} \\ &= g(0) - 2 \sum_{\alpha} \lambda_{\alpha} g(\mathbf{x}_{\alpha} - \mathbf{x}) + \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} g(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \end{aligned}$$

Minimiser cette quantité revient à chercher les pondérateurs qui annulent les dérivées partielles par rapport aux inconnues λ_{α} . On obtient le système d'équations suivant :

$$\sum_{\beta} \lambda_{\beta} g(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) = g(\mathbf{x}_{\alpha} - \mathbf{x}) \quad \forall \alpha.$$

On voit que les poids de krigeage dépendent de la structure spatiale de la variable régionalisée et de la géométrie du champ, par l'intermédiaire du covariogramme transitif $g(\mathbf{h})$, ainsi que de la configuration des points de données entre eux et par rapport au point à estimer, par l'intermédiaire des vecteurs $\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}$ et $\mathbf{x}_{\alpha} - \mathbf{x}$. La valeur de l'erreur quadratique moyenne peut servir d'indicateur de la précision de l'estimation.

On peut sophisticationner le critère de détermination des pondérateurs λ_{α} , en imposant la contrainte supplémentaire que l'abondance soit respectée *dans les différents déplacements imaginés*. Cette condition est analogue à la restriction de non biais que l'on verra lorsque l'on étudiera le "krigeage ordinaire" en géostatistique probabiliste.

Il s'agit cette fois de

$$\text{minimiser } \int [z(\mathbf{x}) - z^*(\mathbf{x})]^2 d\mathbf{x} \text{ sous la contrainte } \int z^*(\mathbf{x}) d\mathbf{x} = \int z(\mathbf{x}) d\mathbf{x}.$$

Ce problème peut être résolu en introduisant une inconnue supplémentaire μ appelée *multiplicateur de Lagrange*. On aboutit au système :

$$\begin{cases} \sum_{\beta} \lambda_{\beta} g(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + \mu = g(\mathbf{x}_{\alpha} - \mathbf{x}) & \forall \alpha \\ \sum_{\beta} \lambda_{\beta} & = 1 \end{cases}$$

Ces équations ne sont pas sans lien avec celles que l'on obtiendra dans la version probabiliste de la théorie : elles seront même formellement identiques, à ceci près que le covariogramme transitif sera remplacé par un autre outil structural (la fonction de covariance). Il convient cependant de souligner que le **krigeage transitif est défini sans que soit randomisée la variable régionalisée, et sans introduire d'hypothèse sur cette variable.**

V. L'analyse variographique

Le recours à un modèle de covariogramme est nécessaire pour pouvoir calculer — ou plutôt estimer — les variances d'estimation globale, et aussi pour pouvoir déterminer les poids de krigeage dans l'estimation locale. C'est ce **covariogramme modélisé** qui sera utilisé, au lieu du covariogramme vrai, inaccessible en pratique. L'étape d'estimation et modélisation du covariogramme à partir de l'ensemble des données disponibles est traditionnellement appelée *analyse structurale* ou *analyse variographique*.

V.1. Le covariogramme expérimental

On a tout d'abord besoin de calculer un covariogramme expérimental, qui constitue une première approximation du covariogramme transitif réel. Dans le cas d'un **échantillonnage régulier** (on considère, pour simplifier, un réseau d'origine $\mathbf{x}_0$ et de maille rectangulaire de longueurs $|\mathbf{a}|$ et $|\mathbf{b}|$ dans l'espace $\mathbb{R}^2$), on peut proposer comme covariogramme expérimental la formule suivante :

$$\forall \eta, \nu \in \mathbb{Z}, \hat{g}_1(\eta \mathbf{a} + \nu \mathbf{b}) = |\mathbf{a}| |\mathbf{b}| \sum_{\kappa, \ell} z(\mathbf{x}_0 + \kappa \mathbf{a} + \ell \mathbf{b}) z(\mathbf{x}_0 + \kappa \mathbf{a} + \ell \mathbf{b} + \eta \mathbf{a} + \nu \mathbf{b}).$$

Si l'on randomise l'origine du réseau de prélèvements, $\hat{g}_1$ devient une variable aléatoire, dont on montre qu'elle est sans biais :

$$\forall \eta, \nu \in \mathbb{Z}, E[\hat{g}_1(\eta \mathbf{a} + \nu \mathbf{b})] = g(\eta \mathbf{a} + \nu \mathbf{b}).$$

En effet :

$$\begin{aligned} \forall \eta, \nu \in \mathbb{Z}, E[\hat{g}_1(\eta \mathbf{a} + \nu \mathbf{b})] &= |\mathbf{a}| |\mathbf{b}| \sum_{\kappa, \ell} \int_{[\mathbf{a}, \mathbf{b}]} z(\mathbf{x}_0 + \kappa \mathbf{a} + \ell \mathbf{b}) z(\mathbf{x}_0 + \kappa \mathbf{a} + \ell \mathbf{b} + \eta \mathbf{a} + \nu \mathbf{b}) \frac{d\mathbf{x}_0}{|\mathbf{a}| |\mathbf{b}|} \\ &= \int z(\mathbf{x}) z(\mathbf{x} + \eta \mathbf{a} + \nu \mathbf{b}) d\mathbf{x} = g(\eta \mathbf{a} + \nu \mathbf{b}) \end{aligned}$$

Ainsi, pour un réseau d'échantillonnage régulier, l'estimation du covariogramme est particulièrement simple. C'est d'ailleurs essentiellement dans ce contexte que s'est développée la géostatistique transitive.

Dans le cas d'un **échantillonnage aléatoire stratifié**, on peut chercher à se ramener au cas d'une maille régulière, par exemple en assignant à chacun des nœuds de la grille de stratification la valeur de l'échantillon de la cellule correspondante. On conçoit toutefois que cette méthode d'estimation est fort approximative. Pour l'améliorer, on peut construire une grille régulière plus fine que la grille de stratification et attribuer à chaque nœud la valeur de l'échantillon le plus proche. Dans ce cas, la contrainte de stratification est inutile, et le procédé d'estimation peut être employé pour un **échantillonnage irrégulier quelconque**.

Lorsque la maille de la grille devient infiniment petite, on obtient la formule suivante :

$$\hat{g}_2(\mathbf{h}) = \sum_{\alpha=1}^n \sum_{\beta=1}^n z(\mathbf{x}_\alpha) z(\mathbf{x}_\beta) |S_\alpha \cap S_{\beta-\mathbf{h}}|$$

où S_α est la surface ou le volume d'influence du site $\mathbf{x}_\alpha$ au sein du domaine échantillonné ;
 $S_{\beta-\mathbf{h}}$ représente le translaté de la surface S_β par le vecteur $-\mathbf{h}$.

Cet estimateur correspond au covariogramme de la variable régionalisée $\hat{z}(\mathbf{x}) = \sum_{\alpha=1}^n z(\mathbf{x}_\alpha) 1_{S_\alpha}(\mathbf{x})$, qui attribue à chaque site de l'espace la valeur du site échantillonné le plus proche :

$$\hat{g}_2(\mathbf{h}) = \int \hat{z}(\mathbf{x}) \hat{z}(\mathbf{x} + \mathbf{h}) d\mathbf{x} = \sum_{\alpha=1}^n \sum_{\beta=1}^n z(\mathbf{x}_\alpha) z(\mathbf{x}_\beta) \int \underbrace{1_{S_\alpha}(\mathbf{x}) 1_{S_\beta}(\mathbf{x} + \mathbf{h})}_{\substack{1 \text{ si } \mathbf{x} \in S_\alpha \text{ et } \mathbf{x} + \mathbf{h} \in S_\beta \\ 0 \text{ sinon}}} d\mathbf{x} = \sum_{\alpha=1}^n \sum_{\beta=1}^n z(\mathbf{x}_\alpha) z(\mathbf{x}_\beta) |S_\alpha \cap S_{\beta-\mathbf{h}}|.$$

L'estimateur précédent peut, en théorie, être calculé pour n'importe quel vecteur $\mathbf{h}$. Il faut bien voir toutefois que cette situation est artificielle, car elle provient de la "répétition" de l'information disponible aux nœuds d'une grille très fine. On démontre en particulier que le comportement à l'origine de $\hat{g}_2$ est toujours linéaire, alors qu'il n'en est pas forcément de même du covariogramme réel. Le bon sens suggère donc que l'estimateur $\hat{g}_2$ ne soit appliqué qu'à des vecteurs $\mathbf{h}$ pour lesquels il existe des paires de données séparées de $\mathbf{h}$, éventuellement avec des tolérances sur les distances et les angles.

Une dernière approche pour évaluer le covariogramme consiste à poser, en conservant les notations précédentes :

$$\hat{g}_3(\mathbf{h}) = \sum_{\alpha=1}^n z(\mathbf{x}_\alpha) z(\mathbf{x}_\alpha + \mathbf{h}) |S_\alpha|.$$

A maille régulière, on retrouve l'estimateur classique $\hat{g}_1$. A maille irrégulière, chaque donnée est pondérée par sa surface d'influence dans le domaine échantillonné, qui mesure en quelque sorte son "importance". L'estimateur $\hat{g}_3$ n'est toutefois calculable que si $\mathbf{x}_\alpha + \mathbf{h}$ coïncide exactement avec un point de données, ce qui a peu de chance de se produire si l'échantillonnage est fortement irrégulier. En pratique, pour robustifier l'estimateur, on autorise des tolérances sur la longueur et l'angle du vecteur de séparation $\mathbf{h}$. Or, dans ce cas, à cause justement de ces tolérances, il peut arriver que plusieurs données soient associées à un même site $\mathbf{x}_\alpha + \mathbf{h}$, de sorte que l'on considérera leur contribution moyenne, pondérée par leur surface d'influence :

$$\hat{g}_3^+(\mathbf{h}) = \sum_{\alpha=1}^n z(\mathbf{x}_\alpha) \frac{\sum_{\beta \in N_\alpha^+(\mathbf{h})} z(\mathbf{x}_\beta) |S_\beta|}{\sum_{\beta \in N_\alpha^+(\mathbf{h})} |S_\beta|} |S_\alpha|$$

avec $N_\alpha^+(\mathbf{h}) = \{ \beta \text{ tels que } \mathbf{x}_\beta - \mathbf{x}_\alpha \approx \mathbf{h} \}.$

Pour assurer la parité de l'estimateur (on sait que le covariogramme réel est une fonction paire), il est préférable de prendre la moyenne des valeurs obtenues en $\mathbf{h}$ et $-\mathbf{h}$:

$$\hat{g}_3^{++}(\mathbf{h}) = \frac{1}{2}[\hat{g}_3^+(\mathbf{h}) + \hat{g}_3^+(-\mathbf{h})].$$

Le recours aux surfaces ou volumes d'influence $\{S_\alpha, \alpha = 1 \dots n\}$ impose de définir les frontières du champ ou du domaine échantillonné, ou alors de s'assurer que l'échantillonnage déborde du champ afin que les superficies associées aux échantillons de valeurs non nulles soient toutes finies. On peut en partie éviter ces contraintes si l'on remarque que l'estimateur $\hat{g}_3$ ne nécessite pas de connaître la géométrie des surfaces d'influence, mais seulement leur mesure ; par suite, on pourrait les remplacer par d'autres pondérateurs, par exemple ceux obtenus par la *méthode des cellules* présentée au chapitre 2, tels que leur somme soit égale à la superficie estimée du champ ou celle du domaine échantillonné.

V.2. Le covariogramme modélisé

Le covariogramme expérimental obtenu par l'une ou l'autre de ces méthodes est en réalité une suite discrète de points ; il doit être modélisé, c'est-à-dire ajusté par une courbe définie en continu qui respecte au mieux les points expérimentaux et répond aux contraintes théoriques d'un covariogramme, notamment le fait d'être de **type positif** (ce qui assurera en particulier la positivité des variances d'estimation globale). Il est souvent laborieux de démontrer qu'une fonction est de type positif, de sorte que l'on préfère toujours utiliser des fonctions dont on a démontré une fois pour toutes qu'elles sont de type positif. Cela réduit la gamme des fonctions disponibles mais assure la validité du modèle.

Par ailleurs, il est souhaitable que le modèle soit conforme aux informations qualitatives sur le phénomène. En particulier, on sait que le **comportement à l'origine** du covariogramme transitif reflète les propriétés de régularité spatiale de la régionalisation. Des connaissances préalables sur la structuration du phénomène étudié permettront d'ajuster un modèle respectant le comportement à l'origine du vrai covariogramme.

Deux autres paramètres importants dans la modélisation sont la valeur de la **portée** et la possible présence d'une **anisotropie**. Dans le cas d'un échantillonnage régulier, la portée est connue avec une incertitude égale à la taille de la maille d'échantillonnage. Quant à l'anisotropie, elle peut être due à la régionalisation elle-même ou à la géométrie de son champ.

La procédure pour modéliser le covariogramme transitif à partir d'un covariogramme expérimental est identique à celle utilisée en géostatistique probabiliste pour modéliser les fonctions de covariance et les variogrammes, et sera détaillée au chapitre 3.

Chapitre 2 : les concepts de base de la géostatistique probabiliste

Pour décrire et “comprendre” un phénomène régionalisé, il est nécessaire d’élaborer une représentation mathématique ou *modèle*. Une première solution consiste à utiliser un **modèle déterministe**. Cette approche conduit en général à une évaluation précise des valeurs prises par la régionalisation à partir d’un nombre limité d’observations, mais elle requiert en contrepartie une connaissance détaillée de la genèse du phénomène et des lois physiques ou mathématiques qui régissent l’évolution de la variable régionalisée. Parmi les domaines d’application, citons

- la météorologie : prévision climatique à court terme ;
- la géophysique : détermination de l’intensité et l’orientation du champ gravitationnel et du champ magnétique terrestre dans l’espace et le temps ;
- la théorie du signal : reconstitution d’un signal continu à partir d’un échantillonnage discret, à l’aide de propriétés spectrales.

En général cependant, les phénomènes régionalisés étudiés sont extrêmement compliqués et leur compréhension peut être si partielle qu’une modélisation dans un cadre déterministe est impossible ou illusoire. Des exemples typiques sont fournis par l’évaluation minière, l’exploration pétrolière, la prévision météorologique à long terme, la caractérisation d’une zone polluée, ou encore l’estimation des ressources forestières d’une région. On est alors contraint de renoncer à une description déterministe du phénomène et de recourir à un **modèle probabiliste**. Cette démarche se révèle opératoire, puisqu’elle permet de formaliser à la fois notre connaissance et nos incertitudes sur le phénomène régionalisé.

I. Limites de la statistique classique

En statistique classique, on considère que les valeurs échantillonnées sont des réalisations indépendantes d’une même variable aléatoire, c’est-à-dire qu’elles n’ont pas de relation les unes sur les autres et obéissent à la même loi de probabilité. On cherche alors à estimer les paramètres — espérance et variance notamment — de la distribution de probabilité des valeurs, dont la forme est parfois présumée (gaussienne, lognormale, etc.). Cependant, lorsque les données sont localisées dans l’espace géographique, les hypothèses de la statistique classique sont rarement acceptables.

Il arrive dans certains cas que les observations ne soient pas toutes réalisées dans des conditions identiques (par exemple, les mesures sont effectuées sur des supports différents, ou avec un appareillage expérimental différent). Les propriétés statistiques changent alors d'une observation à l'autre, et les méthodes classiques ne peuvent plus s'appliquer. Mais surtout, l'hypothèse d'indépendance des observations, si elle simplifie les calculs statistiques, est en général peu réaliste dans le cadre spatial : intuitivement, les observations proches ont des valeurs voisines alors que celles qui sont éloignées ont peu de lien entre elles.

Ainsi donc, si la modélisation des phénomènes régionalisés dans un cadre déterministe n'est en général pas envisageable à cause de leur extrême complexité, on ne peut pas non plus assimiler les valeurs mesurées à des grandeurs aléatoires *indépendantes*. La géostatistique probabiliste va fournir une représentation intermédiaire de la régionalisation, à la fois aléatoire et structurée, en se basant sur le concept de fonction aléatoire.

II. Le concept de fonction aléatoire

Au début de ce document, nous avons présenté la notion de variable régionalisée. Il s'agit simplement d'une fonction définie dans l'espace géographique et représentant numériquement le phénomène régionalisé étudié.

Une variable régionalisée possède les caractéristiques contraires en apparence :

- 1) *localement*, elle est souvent très irrégulière et ne peut être représentée par une fonction mathématique déterministe ;
- 2) *globalement*, elle présente une certaine organisation ou "structure" dans l'espace.

La figure 2.1 illustre cette situation : alors qu'à une échelle très locale, la régionalisation paraît très erratique, on distingue, à une échelle plus grande, des zones de fortes valeurs et de forte variabilité, et des zones plus "calmes" et de valeurs moindres.

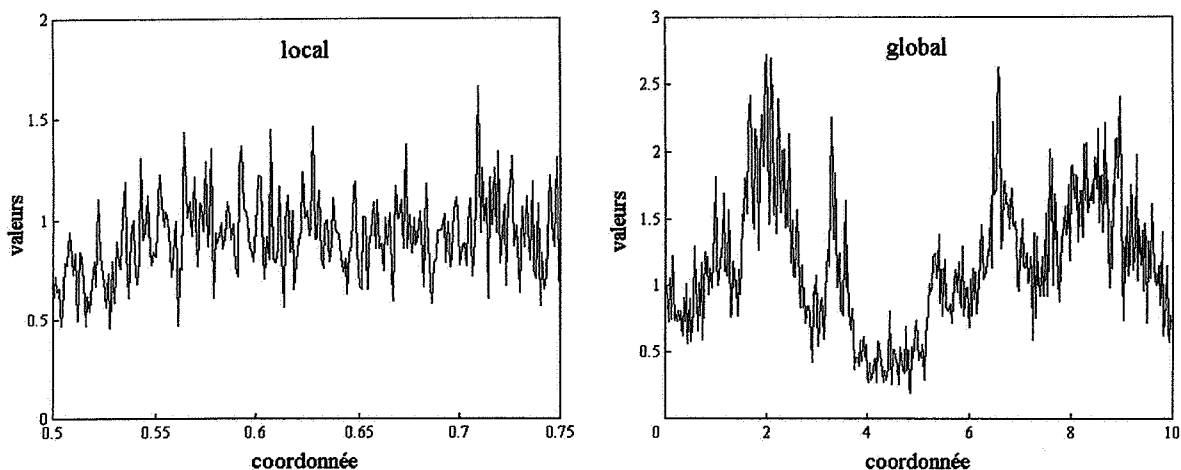


figure 2.1 : aspects local et global d'une variable régionalisée définie dans R

Ce double aspect va pouvoir être pris en compte grâce au formalisme des fonctions aléatoires, sur lequel se fonde la géostatistique probabiliste, appelée aussi **géostatistique intrinsèque** (par opposition à la géostatistique *transitive*, déterministe).

II.1. Définition et choix méthodologique

Les modèles géostatistiques considèrent la valeur $z(\mathbf{x})$ de la variable régionalisée en un point $\mathbf{x}$ du champ $\mathcal{D}$ comme une réalisation d'une variable aléatoire $Z(\mathbf{x})$ ¹. Lorsque $\mathbf{x}$ parcourt $\mathcal{D}$, on obtient une famille de variables aléatoires $\{Z(\mathbf{x}), \mathbf{x} \in \mathcal{D}\}$ qui constitue une **fonction aléatoire** ou *processus stochastique*. L'ensemble $\{z(\mathbf{x}), \mathbf{x} \in \mathcal{D}\}$, qui n'est autre que la variable régionalisée étudiée, est une *réalisation* particulière de la fonction aléatoire.

Contrairement au modèle de la statistique classique, les variables aléatoires $Z(\mathbf{x})$ ainsi définies ne sont pas indépendantes. Au contraire, il existe des “interactions” ou “corrélations” entre elles, lesquelles refléteront la structure spatiale de la régionalisation, en particulier son degré de régularité et continuité, son anisotropie...

Ainsi donc, la géostatistique probabiliste **interprète la variable régionalisée comme une réalisation d'une fonction aléatoire**. Cette approche permet de prendre en compte les aspects erratique et structuré de la régionalisation :

- 1) en chaque point $\mathbf{x}$, $Z(\mathbf{x})$ est une variable aléatoire (d'où l'aspect erratique) ;
- 2) pour tout ensemble de points $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$, les variables aléatoires $\{Z(\mathbf{x}_1), Z(\mathbf{x}_2), \dots, Z(\mathbf{x}_k)\}$ sont reliées par des corrélations qui quantifient la “ressemblance” entre les valeurs qu'elles prennent (d'où l'aspect structuré).

En toute rigueur, une fonction aléatoire est définie à la fois sur l'espace géographique et sur un *espace probabilisé* (c'est la raison pour laquelle on attribue parfois le nom de “modèles topo-probabilistes” aux méthodes géostatistiques). Dire que la variable régionalisée est une réalisation d'une fonction aléatoire s'écrit mathématiquement :

$$\forall \mathbf{x} \in \mathcal{D}, z(\mathbf{x}) = Z(\mathbf{x}, \omega_0)$$

où ω_0 est un événement particulier de l'espace probabilisé². En règle générale, on se contente de noter Z ou $Z(\mathbf{x})$, au lieu de $Z(\mathbf{x}, \omega)$, la fonction aléatoire.

¹ Pour bien distinguer les quantités déterministes des quantités aléatoires, nous noterons les premières avec une *minuscule* et les secondes avec une *Majuscule*.

² Cette remarque prend tout son sens dans la mise en œuvre des techniques de “simulations” ou “modèles numériques”, qui, faute de connaître la variable régionalisée $z(\mathbf{x})$, consistent à générer des réalisations $Z(\mathbf{x}, \omega_i)$ de la fonction aléatoire associée Z , qui constituent des versions possibles et équiprobables de la réalité.

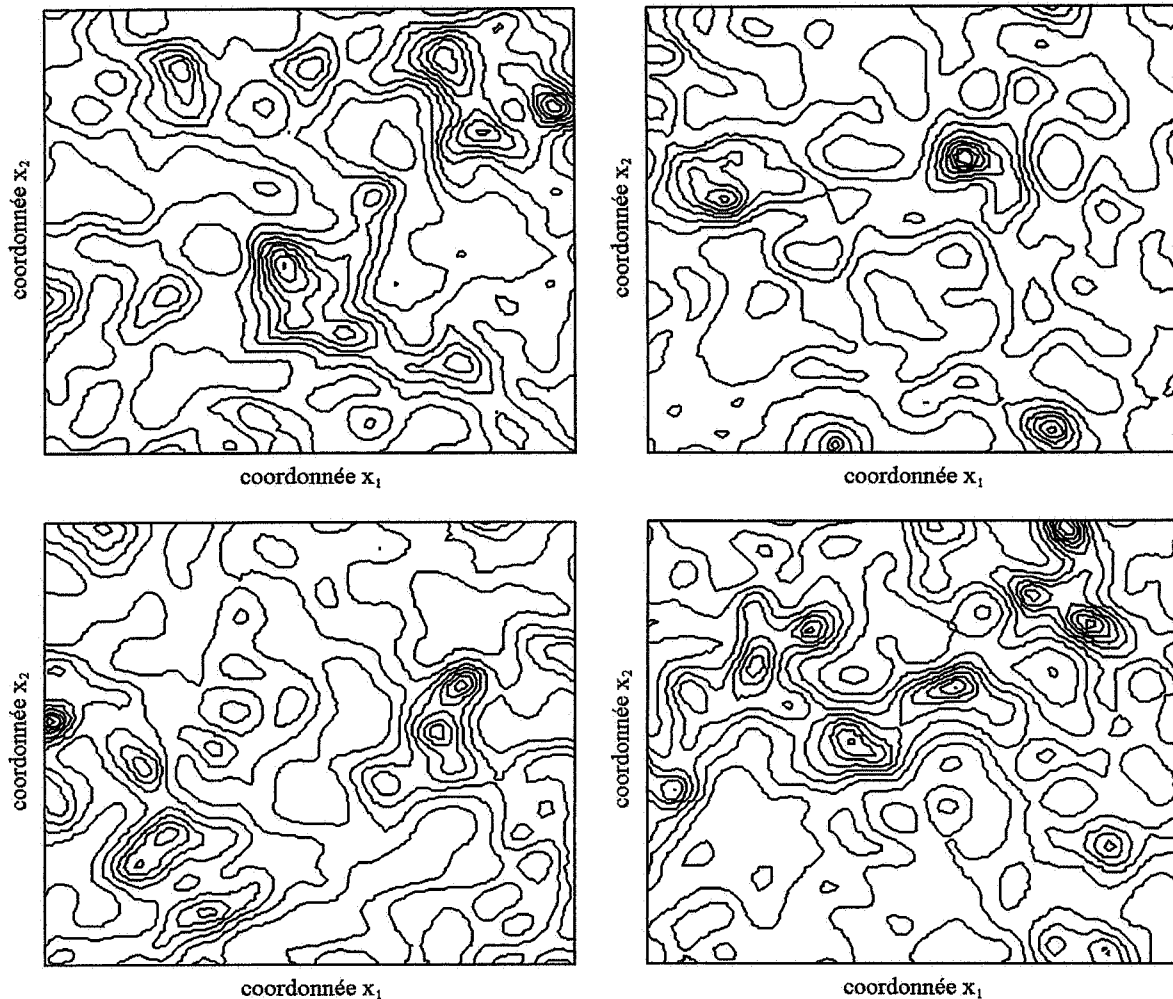


figure 2.2 : quatre réalisations d'une même fonction aléatoire définie dans l'espace à deux dimensions, représentées sous forme de lignes de niveau équiréparties

Le recours au concept de fonction aléatoire est une **décision**, ni vraie ni fausse, mais qui peut se révéler plus ou moins *pertinente*. La fonction aléatoire et l'espace probabilisé sous-jacent sont des objets théoriques qui n'existent pas dans la réalité ; ils ne servent que d'**intermédiaire de calcul** dans la résolution du problème posé. Une fois terminées les manipulations sur la fonction aléatoire, il faut revenir au niveau de la variable régionalisée en remplaçant Z par sa réalisation z . Au besoin, une *reconstruction opératoire* sera nécessaire, c'est-à-dire une formulation des résultats établis dans le modèle probabiliste en termes "concrets" ou objectifs ; par exemple, que signifie un intervalle de confiance ou une variance d'estimation, pour caractériser une erreur d'estimation unique ?

Ajoutons que la détermination de la fonction aléatoire à partir de la variable régionalisée n'est pas une opération univoque : **plusieurs modèles peuvent être acceptables**, dès lors qu'ils sont compatibles avec l'information disponible sur le phénomène régionalisé.

Il convient de signaler qu'un modèle de fonction aléatoire contient plus d'information que les données expérimentales : celles-ci ne sont qu'une collection de valeurs numériques, alors que la fonction aléatoire décrit non seulement la distribution de ces valeurs, mais également leur comportement conjoint (par l'intermédiaire de corrélations, de lois multivariées). En résumé, on introduit dans le modèle de fonction aléatoire une *structure*, et par conséquent on ajoute de l'information à celle contenue dans les données. Pour éviter de prendre une décision non judicieuse, il faut chercher à réduire les hypothèses sur lesquelles le modèle est construit (*principe d'économie*) ; il est inutile, voire dangereux, d'introduire des traits qui ne se manifestent pas dans l'information disponible.

II.2. Caractéristiques d'une fonction aléatoire

II.2.1. Loi spatiale

Considérons une fonction aléatoire $Z(\mathbf{x})$, $\mathbf{x} \in \mathcal{D}$, et une série de points $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$. Le vecteur aléatoire $\{Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)\}$ est caractérisé par une *fonction de répartition* qui dépend de k arguments :

$$F_{\mathbf{x}_1, \dots, \mathbf{x}_k}(z_1, \dots, z_k) = \text{Prob}[Z(\mathbf{x}_1) < z_1, \dots, Z(\mathbf{x}_k) < z_k].$$

L'ensemble des fonctions de répartition, pour tous les entiers k et tous les choix possibles de $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ dans $\mathcal{D}$, constitue la **loi spatiale** de la fonction aléatoire. La donnée de la loi spatiale est une information extrêmement riche, qui fournit de très nombreuses caractéristiques statistiques³ sur Z .

Le nombre d'observations disponibles sur la variable régionalisée étant toujours limité, il est illusoire de vouloir inférer la loi spatiale entière de la fonction aléatoire associée. Des simplifications sont nécessaires pour rendre opératoire le formalisme probabiliste.

En géostatistique linéaire, on n'utilise que les **deux premiers moments**⁴ de la fonction aléatoire, car ils fournissent une description élémentaire de la loi spatiale et sont suffisants pour résoudre la plupart des problèmes rencontrés en pratique ; nous verrons ultérieurement comment réaliser leur inférence à partir des données expérimentales. La contrepartie de cette limitation est que certaines caractéristiques de la fonction aléatoire – et donc de la variable régionalisée – échappent au formalisme de la géostatistique linéaire, notamment la forme de l'histogramme (asymétrie, multimodalité, valeurs limites...), celle des lois multivariées et des lois conditionnelles.

³ Certaines propriétés, comme la continuité ou la dérivabilité des réalisations de Z , la loi des valeurs supérieure et inférieure de Z sur un ensemble, la connexité des ensembles où Z est supérieure à une valeur donnée, échappent à la loi spatiale.

⁴ Les moments d'une fonction aléatoire sont des paramètres descriptifs de sa loi spatiale, assez simples pour être faciles à calculer (cf. annexe A).

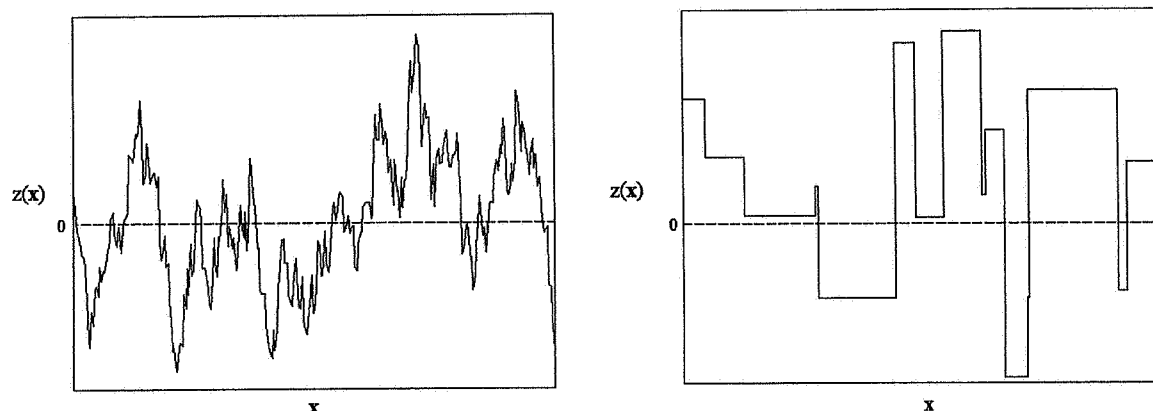


figure 2.3 : deux réalisations de fonctions aléatoires unidimensionnelles ayant mêmes moments jusqu'à l'ordre deux (donc indiscernables en géostatistique linéaire)
à gauche : fonction aléatoire de loi gaussienne, à droite : fonction aléatoire de loi mosaïque

II.2.2. Moment du premier ordre (espérance mathématique)

L'espérance d'une fonction aléatoire Z dépend en toute généralité du point $\mathbf{x}$ considéré ; elle est habituellement notée $m(\mathbf{x})$:

$$E[Z(\mathbf{x})] = m(\mathbf{x}).$$

En un point $\mathbf{x}$ donné, $m(\mathbf{x})$ représente la "moyenne" autour de laquelle se distribuent les valeurs prises par de multiples réalisations indépendantes de la fonction aléatoire. Il s'agit du paramètre descriptif de base du comportement aléatoire de $Z(\mathbf{x})$: on établit en effet que l'espérance est la meilleure approximation d'une variable aléatoire par une constante, au sens des moindres carrés.

II.2.3. Moments du second ordre

En géostatistique linéaire, on considère les moments du second ordre suivant :

- la **variance**, ou variance *a priori* de $Z(\mathbf{x})$. Si elle existe, est définie par :

$$\text{var}[Z(\mathbf{x})] = E \{ [Z(\mathbf{x}) - m(\mathbf{x})]^2 \}.$$

On a également : $\text{var}[Z(\mathbf{x})] = E[Z(\mathbf{x})^2] - m(\mathbf{x})^2$.

La variance est une quantité toujours positive, puisqu'elle est par définition l'espérance d'un carré. Sa racine carrée est appelée **écart-type**. La variance (ou l'écart-type) constitue une mesure de la plus ou moins grande dispersion de $Z(\mathbf{x})$ autour de sa valeur moyenne $m(\mathbf{x})$ et quantifie ainsi son caractère plus ou moins "aléatoire". On montre qu'une variable aléatoire de variance nulle est presque sûrement constante, donc n'est plus aléatoire.

- la **covariance (centrée)** : si les deux variables aléatoires $Z(\mathbf{x}_1)$ et $Z(\mathbf{x}_2)$ possèdent des variances, elles ont aussi une covariance qui dépend de la position des points $\mathbf{x}_1$ et $\mathbf{x}_2$:

$$\begin{aligned}\text{cov}[Z(\mathbf{x}_1), Z(\mathbf{x}_2)] &= E \{ [Z(\mathbf{x}_1) - m(\mathbf{x}_1)][Z(\mathbf{x}_2) - m(\mathbf{x}_2)] \} \\ &= E[Z(\mathbf{x}_1)Z(\mathbf{x}_2)] - m(\mathbf{x}_1)m(\mathbf{x}_2)\end{aligned}$$

La covariance donne une approche élémentaire de la liaison ou "interaction" entre $Z(\mathbf{x}_1)$ et $Z(\mathbf{x}_2)$. C'est une fonction bilinéaire, dont la variance est la forme quadratique associée. En particulier, on a : $\text{cov}[Z(\mathbf{x}_1), Z(\mathbf{x}_1)] = \text{var}[Z(\mathbf{x}_1)]$. L'inégalité de Cauchy-Schwarz relie la covariance entre $Z(\mathbf{x}_1)$ et $Z(\mathbf{x}_2)$ aux variances de $Z(\mathbf{x}_1)$ et $Z(\mathbf{x}_2)$:

$$|\text{cov}[Z(\mathbf{x}_1), Z(\mathbf{x}_2)]| \leq \sqrt{\text{var}[Z(\mathbf{x}_1)] \text{var}[Z(\mathbf{x}_2)]}.$$

- le **(semi-)variogramme** : il est défini comme la moitié de la variance, lorsqu'elle existe, de l'accroissement $Z(\mathbf{x}_1) - Z(\mathbf{x}_2)$ et s'écrit :

$$\gamma(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{2} \text{var}[Z(\mathbf{x}_1) - Z(\mathbf{x}_2)].$$

N.B. : pour alléger l'écriture, on omettra systématiquement le préfixe "semi".

- parfois aussi, on considère le **corrélogramme** ou fonction d'*auto-corrélation*, égal au coefficient de corrélation entre $Z(\mathbf{x}_1)$ et $Z(\mathbf{x}_2)$:

$$\rho(\mathbf{x}_1, \mathbf{x}_2) = \text{corr}[Z(\mathbf{x}_1), Z(\mathbf{x}_2)] = \frac{\text{cov}[Z(\mathbf{x}_1), Z(\mathbf{x}_2)]}{\sqrt{\text{var}[Z(\mathbf{x}_1)] \text{var}[Z(\mathbf{x}_2)]}}.$$

Contrairement à la covariance ou au variogramme, le corrélogramme est adimensionné. L'inégalité de Cauchy-Schwarz implique que ρ prend ses valeurs dans l'intervalle $[-1, 1]$. Un coefficient ρ nul indique que les variables $Z(\mathbf{x}_1)$ et $Z(\mathbf{x}_2)$ sont non corrélées (condition nécessaire pour qu'elles soient indépendantes), un coefficient égal à 1 ou -1 indique qu'elles sont liées linéairement :

$$Z(\mathbf{x}_1) = a Z(\mathbf{x}_2) + b, \text{ avec } \text{signe}(a) = \rho = \pm 1.$$

La loi spatiale de $Z(\mathbf{x})$ et ses moments (espérance, variance, covariance, variogramme, etc.) ne sont pas définis de manière univoque par la donnée de la variable régionalisée, qui n'est qu'une *réalisation* de $Z(\mathbf{x})$. Contrairement au covariogramme transitif, ce ne sont pas des grandeurs régionales mais des objets théoriques ; leur identification ne peut être contrôlée expérimentalement, même si l'on dispose d'une connaissance complète de la réalité. Des hypothèses limitatives sont nécessaires, comme nous allons le voir.

III. Possibilité de l'inférence statistique : les hypothèses de stationnarité

Pour mettre en œuvre le formalisme probabiliste, il est nécessaire de pouvoir déterminer, au moins partiellement, la loi spatiale de la fonction aléatoire $Z(\mathbf{x})$ à partir des données disponibles sur la variable régionalisée $z(\mathbf{x})$. Cette étape est connue sous le nom d'*inférence statistique*.

Deux raisons empêchent de pouvoir réaliser l'inférence statistique en toute généralité : d'une part, on ne dispose que d'une seule réalisation de la fonction aléatoire (à savoir, la variable régionalisée) ; d'autre part, cette réalisation n'est connue que de manière fragmentaire, en certains points d'échantillonnage $\{ \mathbf{x}_\alpha, \alpha = 1 \dots n \}$. Cette seconde restriction n'est toutefois pas aussi problématique que la première : **la question de l'inférence se poserait encore si l'on connaissait exhaustivement la réalité**, car la loi spatiale ou ses moments ne sont pas des grandeurs régionales. Pour pouvoir les évaluer, il faudrait de nombreuses réalisations de la fonction aléatoire, mais celles-ci sont purement théoriques et n'existent pas dans la réalité.

Pour sortir de cette impasse, certaines restrictions sont nécessaires. Elles font appel à la notion de **stationnarité**, qui décrit en quelque sorte une forme d'**homogénéité spatiale** de la régionalisation. L'idée est de permettre l'inférence statistique en remplaçant la répétition sur les réalisations de la fonction aléatoire (inaccessibles) par une répétition dans l'espace : les valeurs que l'on rencontre dans les différentes régions du champ présentent les mêmes caractéristiques et peuvent être considérées comme différentes réalisations du même processus aléatoire⁵. Du point de vue mathématique, les hypothèses de stationnarité consistent à supposer que tout ou partie de la loi spatiale de la fonction aléatoire est invariante par translation, c'est-à-dire que les propriétés probabilistes d'un ensemble de valeurs ne dépendent pas de la position absolue des sites associés, mais seulement de leurs séparations.

Plus précisément, on peut définir plusieurs types de stationnarité.

- **Stationnarité stricte**

La stationnarité stricte correspond à l'invariance par translation de la loi spatiale de la fonction aléatoire ; soit, en termes mathématiques : $\forall k \in \mathbb{N}^*, \forall \mathbf{x}_1, \dots, \mathbf{x}_k \in \mathbb{R}^d, \forall \mathbf{h} \in \mathbb{R}^d$, les vecteurs aléatoires $(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k))$ et $(Z(\mathbf{x}_1 + \mathbf{h}), \dots, Z(\mathbf{x}_k + \mathbf{h}))$ ont la même loi de probabilité conjointe. Cette propriété exprime l'idée que toutes les caractéristiques de la régionalisation sont invariantes dans l'espace. Elle est très contraignante car elle suppose une identité de toutes les lois de probabilité dans l'espace. On peut toutefois la restreindre aux deux premiers moments de la fonction aléatoire, qui sont les outils manipulés par la géostatistique linéaire.

⁵ En fait, ces différentes réalisations ne sont pas indépendantes, de sorte que l'inférence n'est possible que moyennant une hypothèse supplémentaire d'**ergodicité**, qui est souvent admise. L'ergodicité consiste à dire que le domaine $\mathcal{D}$ est suffisamment étendu pour pouvoir confondre la *moyenne spatiale* de z sur $\mathcal{D}$ avec l'*espérance* de Z (cf. annexe B).

• **Stationnarité d'ordre deux (ou du second ordre)**

Une fonction aléatoire $Z(\mathbf{x})$ est stationnaire du second ordre quand ses deux premiers moments (espérance et fonction de covariance) existent et sont invariants par translation :

$\begin{cases} E[Z(\mathbf{x})] = m & \text{indépendant de } \mathbf{x} \\ \text{cov}[Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})] = C(\mathbf{h}) & \text{ne dépend que de } \mathbf{h} \end{cases}$

Cette dernière relation s'écrit encore $\text{cov}[Z(\mathbf{x}_1), Z(\mathbf{x}_2)] = C(\mathbf{x}_1 - \mathbf{x}_2)$.

Remarques

a) La stationnarité du second ordre implique l'existence non seulement de l'espérance et de la covariance, mais aussi des autres moments d'ordre deux : variance *a priori*, variogramme et corrélogramme. Ces moments ne dépendent pas de la position absolue des points où ils sont calculés, mais seulement de leur séparation :

$$\begin{aligned} \blacksquare \quad \forall \mathbf{x}, \text{var}[Z(\mathbf{x})] &= \text{cov}[Z(\mathbf{x}), Z(\mathbf{x})] = C(\mathbf{0}) \text{ constante} \\ \blacksquare \quad \forall \mathbf{x}, \mathbf{x} + \mathbf{h}, \gamma(\mathbf{x} + \mathbf{h}, \mathbf{x}) &= \frac{1}{2} \text{var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] \\ &= \frac{1}{2} \left\{ \underbrace{\text{var}[Z(\mathbf{x})]}_{C(\mathbf{0})} + \underbrace{\text{var}[Z(\mathbf{x} + \mathbf{h})]}_{C(\mathbf{0})} - 2 \underbrace{\text{cov}[Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})]}_{C(\mathbf{h})} \right\} \\ &= C(\mathbf{0}) - C(\mathbf{h}) \quad \text{indépendant de } \mathbf{x} \end{aligned}$$

de sorte que l'on peut écrire : $\gamma(\mathbf{x} + \mathbf{h}, \mathbf{x}) = \gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$.

$$\blacksquare \quad \forall \mathbf{x}, \mathbf{x} + \mathbf{h}, \rho(\mathbf{x} + \mathbf{h}, \mathbf{x}) = \rho(\mathbf{h}) = \frac{C(\mathbf{h})}{C(\mathbf{0})} = 1 - \frac{\gamma(\mathbf{h})}{C(\mathbf{0})}.$$

b) Sous l'hypothèse de stationnarité du second ordre, il est équivalent de travailler avec la fonction de covariance ou avec le variogramme, l'un se déduisant de l'autre par la relation $\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$.

c) Pour peu que les deux premiers moments existent, la stationnarité stricte entraîne la stationnarité du second ordre. En règle générale, la réciproque est fautive, sauf dans le cas gaussien. En effet, si une fonction aléatoire de loi spatiale gaussienne (c'est-à-dire telle que toute combinaison linéaire de valeurs suit une distribution gaussienne) est stationnaire d'ordre deux, elle est aussi strictement stationnaire, car la loi gaussienne est entièrement caractérisée par ses deux premiers moments.

d) L'hypothèse de stationnarité d'ordre deux peut s'interpréter de la manière suivante : la variable régionalisée $z(\mathbf{x})$, réalisation particulière de $Z(\mathbf{x})$, fluctue autour de la même valeur constante dans toutes les régions de l'espace ("traduction" de $E[Z(\mathbf{x})] = \text{constante}$), et ces fluctuations présentent partout la même dispersion ("traduction" de $\text{var}[Z(\mathbf{x})] = \text{constante}$). La vérification de ces propriétés empiriques, avec un certain degré d'approximation, peut permettre de juger de l'opportunité de représenter la variable régionalisée comme une réalisation d'un processus stationnaire du second ordre.

Nous proposons ci-dessous une illustration dans le cas unidimensionnel.

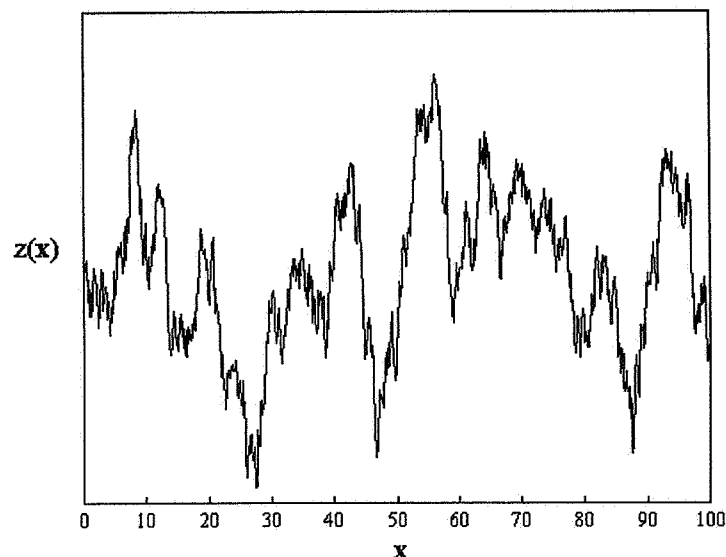


figure 2.4 : une régionalisation raisonnablement stationnaire

Nous verrons comment, au cours d'une étude pratique, accepter ou rejeter l'hypothèse de stationnarité, à partir d'un ensemble de mesures expérimentales. On ne pourra pas utiliser des tests statistiques en raison de l'unicité de la réalisation, mais une étude exploratoire des données permettra d'infirmer ou de corroborer telle ou telle hypothèse.

- e) Enfin, ajoutons que l'échelle de travail a une importance capitale : les caractéristiques structurales d'une même régionalisation et la possibilité de la représenter par un processus stationnaire dépendent de l'échelle d'observation. Sur la figure précédente, l'hypothèse de stationnarité serait moins pertinente si on limitait par exemple l'échelle d'observation aux valeurs d'abscisse comprise entre 47 et 57, pour lesquelles la variable régionalisée croît de manière systématique.

Dans certaines situations, l'hypothèse de stationnarité du second ordre est trop restrictive. Elle implique en particulier l'existence d'une variance *a priori*. Or, un phénomène physique peut présenter une capacité infinie de dispersion, et n'être modélisable par une fonction aléatoire de variance finie. Cette affirmation peut sembler surprenante car, lorsqu'on prélève des échantillons à l'intérieur d'un domaine géographique $\mathcal{D}$, on peut toujours calculer une variance expérimentale. En réalité, cette dernière est une fonction du support de la mesure, en général ponctuel, et du domaine $\mathcal{D}$ dans lequel se répartissent les données (voir, au chapitre 4, la notion de variance de dispersion). La variance *a priori* est infinie lorsque la variance expérimentale croît indéfiniment avec la surface ou le volume de $\mathcal{D}$; elle est finie lorsqu'il existe une limite à la variance expérimentale quand $\mathcal{D}$ devient infini. On est donc conduit à introduire une hypothèse plus faible que celle de stationnarité d'ordre deux, afin de permettre l'occurrence de variances *a priori* infinies.

• **Hypothèse intrinsèque**

Une fonction aléatoire $Z(\mathbf{x})$ est dite intrinsèque quand ses *accroissements* sont stationnaires d'ordre deux, c'est-à-dire :

- 1) $E[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = m(\mathbf{h})$ indépendant de $\mathbf{x}$
- 2) $\text{cov}[Z(\mathbf{x}_1 + \mathbf{h}_1) - Z(\mathbf{x}_1), Z(\mathbf{x}_2 + \mathbf{h}_2) - Z(\mathbf{x}_2)] = C(\mathbf{x}_1 - \mathbf{x}_2; \mathbf{h}_1, \mathbf{h}_2)$
ne dépend que de $\mathbf{x}_1 - \mathbf{x}_2$, $\mathbf{h}_1$ et $\mathbf{h}_2$

On démontre⁶ que la seconde propriété, qui fait intervenir deux points $\mathbf{x}_1$ et $\mathbf{x}_2$ et deux vecteurs d'accroissements $\mathbf{h}_1$ et $\mathbf{h}_2$, est équivalente à la propriété plus simple concernant le variogramme γ :

$$2') \quad \gamma(\mathbf{h}) = \frac{1}{2} \text{var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] \text{ existe et ne dépend que de } \mathbf{h}, \text{ pas de } \mathbf{x}.$$

La fonction $m(\mathbf{h})$ définie par la propriété 1) est appelée la **dérive** de la fonction aléatoire intrinsèque. La linéarité de l'espérance implique que $m(\mathbf{h})$ est nécessairement une fonction linéaire de $\mathbf{h}$:

$$\begin{aligned} m(\mathbf{h}) + m(\mathbf{h}') &= E[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] + E[Z(\mathbf{x} + \mathbf{h} + \mathbf{h}') - Z(\mathbf{x} + \mathbf{h})] \\ &= E[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x}) + Z(\mathbf{x} + \mathbf{h} + \mathbf{h}') - Z(\mathbf{x} + \mathbf{h})] \\ &= E[Z(\mathbf{x} + \mathbf{h} + \mathbf{h}') - Z(\mathbf{x})] = m(\mathbf{h} + \mathbf{h}') \end{aligned}$$

d'où on déduit que $m(\mathbf{h})$ est de la forme $m(\mathbf{h}) = \langle \mathbf{u} | \mathbf{h} \rangle$ où $\mathbf{u}$ est un vecteur fixe et $\langle | \rangle$ désigne le produit scalaire usuel sur $\mathbb{R}^d$.

Dans toute la suite, on ne considérera que les fonctions aléatoires intrinsèques **sans dérive**, c'est-à-dire telles que $m(\mathbf{h}) = 0$. Le cas des fonctions intrinsèques avec dérive est plus complexe à manipuler et sera traité dans un cadre plus général (cf. chapitre 6 et annexe D, les fonctions aléatoires intrinsèques d'ordre k). L'hypothèse intrinsèque simplifiée s'écrit finalement :

- 1) $E[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = 0$
- 2) $\text{var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = 2 \gamma(\mathbf{h})$ ne dépend que de $\mathbf{h}$

L'absence de dérive permet par ailleurs de confondre la variance des accroissements avec l'espérance de leur carré :

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

⁶ On établit que, si γ existe et ne dépend que de $\mathbf{h}$, alors :

$$C(\mathbf{x}_1 - \mathbf{x}_2; \mathbf{h}_1, \mathbf{h}_2) = \gamma(\mathbf{h}_1 + (\mathbf{x}_1 - \mathbf{x}_2)) + \gamma(\mathbf{h}_2 - (\mathbf{x}_1 - \mathbf{x}_2)) - \gamma(\mathbf{h}_1 - \mathbf{h}_2 + (\mathbf{x}_1 - \mathbf{x}_2)) - \gamma(\mathbf{x}_1 - \mathbf{x}_2)$$

ne dépend que de $\mathbf{x}_1 - \mathbf{x}_2$, $\mathbf{h}_1$ et $\mathbf{h}_2$. La propriété 2') entraîne donc la propriété 2). La réciproque est évidente : il suffit de prendre $\mathbf{x}_1 = \mathbf{x}_2 = \mathbf{x}$ et $\mathbf{h}_1 = \mathbf{h}_2 = \mathbf{h}$, de sorte que $C(\mathbf{x}_1 - \mathbf{x}_2; \mathbf{h}_1, \mathbf{h}_2) = C(\mathbf{0}; \mathbf{h}, \mathbf{h}) = 2 \gamma(\mathbf{h})$ ne dépend que de $\mathbf{h}$.

Remarque importante

La stationnarité du second ordre implique l'hypothèse intrinsèque. La réciproque n'est pas nécessairement vraie : l'hypothèse intrinsèque élargit l'hypothèse stationnaire d'ordre deux, laquelle ne porte plus que sur les *accroissements* de la fonction aléatoire. La covariance d'une fonction intrinsèque peut ne pas exister ; elle n'existe que si le variogramme est borné, auquel cas on a la relation

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}).$$

On qualifie de *strictement intrinsèque* une fonction aléatoire intrinsèque mais non stationnaire d'ordre deux.

• Hypothèses quasi-stationnaire et quasi-intrinsèque

Les écritures précédentes ne font pas intervenir explicitement l'**échelle de travail**, qui est pourtant un paramètre essentiel dans la pratique : un modèle acceptable à une certaine échelle peut ne plus l'être à des échelles plus grandes ou plus petites. Dans la plupart des applications (notamment les problèmes d'estimation), il n'est pas utile que les hypothèses stationnaires ou intrinsèques soient valides à l'échelle du champ entier, mais seulement pour des distances inférieures à une distance limite b . Ceci nous conduit à définir la notion de *stationnarité locale*, ou *quasi-stationnarité*.

Une fonction aléatoire est dite **quasi-stationnaire** si elle possède une espérance $m(\mathbf{x})$ variant lentement à l'échelle d'observation (i.e. telle que $m(\mathbf{x} + \mathbf{h}) \approx m(\mathbf{x})$ si $|\mathbf{h}| < b$), et une covariance $C(\mathbf{x} + \mathbf{h}, \mathbf{x})$ *localement* stationnaire, c'est-à-dire ne dépendant que de $\mathbf{h}$ pour des distances $|\mathbf{h}| < b$. L'inférence des fonctions $m(\mathbf{x})$ et $C(\mathbf{x} + \mathbf{h}, \mathbf{x})$ nécessite en pratique de disposer de suffisamment de données dans le voisinage du point $\mathbf{x}$ considéré. L'hypothèse de quasi-stationnarité est donc un compromis entre la taille de la zone considérée comme stationnaire et le nombre de données disponibles. Lorsque l'inférence n'est pas possible à l'échelle locale, on doit se contenter d'une covariance moyenne ("pseudo-stationnaire") calculée sur l'ensemble du champ $\mathcal{D}$:

$$\bar{C}(\mathbf{h}) = \frac{1}{|\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}|} \int_{\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}} C(\mathbf{x} + \mathbf{h}, \mathbf{x}) d\mathbf{x} \quad \text{ne dépend que de } \mathbf{h}.$$

La connaissance de $\bar{C}(\mathbf{h})$ est généralement suffisante pour résoudre les problèmes d'estimation locale ou globale (cf. chapitres 4 et 5) ; seules les variances d'estimation locale seront affectées par l'utilisation de $\bar{C}$ au lieu de C .

De la même manière, on dit qu'une fonction aléatoire est **quasi-intrinsèque** lorsque son variogramme $\gamma(\mathbf{x} + \mathbf{h}, \mathbf{x})$ existe et ne dépend que de $\mathbf{h}$ tant que $|\mathbf{h}| < b$.

Exemple : effet proportionnel

Il arrive parfois que les variogrammes locaux ne diffèrent que d'un facteur multiplicatif, et que l'on puisse écrire, pour $|h| < b$, $\gamma(x+h, x) = \varphi[m^*(x)] \gamma_0(h)$, où γ_0 est un variogramme de référence invariant sur l'ensemble du champ, φ une fonction donnée et $m^*(x)$ la moyenne expérimentale locale en x . On dit alors qu'il y a *effet proportionnel*, direct si φ est une fonction croissante, inverse si φ est décroissante. La variabilité entre deux point x et $x+h$, mesurée par $\gamma(x+h, x)$, est directement liée à la moyenne locale $m^*(x)$ (voir à titre d'illustration la figure 2.1 droite). Un tel phénomène se manifeste souvent lorsque la variable étudiée présente une distribution très dissymétrique, du type lognormal. Il peut être interprété soit comme une quasi-stationnarité, soit – de manière plus théorique – comme le *conditionnement* du variogramme à la moyenne locale $m^*(x)$ (la fonction aléatoire étudiée $Z(x)$ est alors supposée globalement stationnaire d'ordre deux, mais les données disponibles modifient sa loi spatiale *a priori*).

L'effet proportionnel peut être détecté – et la fonction φ modélisée – à partir d'un nuage de corrélation entre les moyennes et variances expérimentales locales, calculées sur des sous-zones de même diamètre b . On montre en effet (cf. chapitre 4, à propos des variances de dispersion) que, sous l'hypothèse d'un effet proportionnel, la variance expérimentale locale $s^{*2}(x)$ est proportionnelle à $\varphi[m^*(x)]$, cette dernière quantité étant pratiquement constante à l'échelle locale. La courbe de régression entre moyennes et variances expérimentales locales s'identifie alors à la fonction φ .

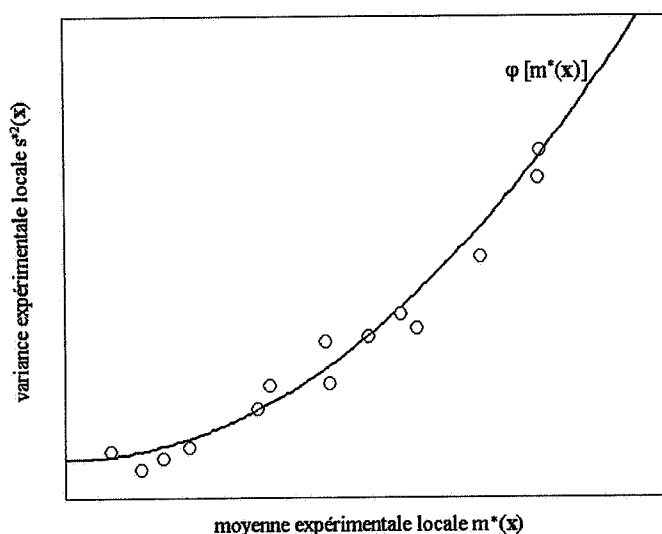


figure 2.5 : ajustement d'un effet proportionnel
les ronds indiquent les points expérimentaux ($m^*(x), s^{*2}(x)$)
la courbe de régression φ (ligne continue) est ici parabolique.

On peut élaborer des modèles quasi-intrinsèques plus complexes, en superposant plusieurs variogrammes présentant des effets proportionnels différents :

$$\gamma(x+h, x) = \varphi_0[m^*(x)] \gamma_0(h) + \varphi_1[m^*(x)] \gamma_1(h) + \dots \quad \text{pour } |h| < b.$$

Par exemple, γ_0 représente la composante discontinue du variogramme (*effet de pépité*, cf. chapitre 3) et γ_1 la composante continue. Dans ce cas, les fonctions φ_0 et φ_1 sont déterminées en examinant les relations de proportionnalité entre les composantes des différents variogrammes locaux que l'on peut former. Le recours à un tel modèle nécessite de disposer de suffisamment de données dans chaque zone de diamètre b pour y construire un estimateur correct de $m^*(x)$ et $\gamma(x+h, x)$.

IV. Les combinaisons linéaires autorisées et le calcul de leurs deux premiers moments

IV.1. Définition des combinaisons linéaires autorisées

La décision de ne considérer que les deux premiers moments de la fonction aléatoire Z étudiée limite le champ d'application du modèle. En particulier, faute de connaître la loi spatiale entière, il ne sera pas possible de déterminer les caractéristiques probabilistes de n'importe quelle expression. En fait, on ne pourra calculer que les deux premiers moments (espérance et variance) des combinaisons linéaires de Z . Ces dernières serviront à construire des estimateurs de grandeurs inconnues, les espérances seront utilisées pour vérifier l'absence de *biais* de ces estimateurs, les variances pour mesurer leur *précision*. On pourrait imaginer recourir à d'autres critères (par exemple, mesurer la précision à l'aide d'intervalles de confiance), mais cela nécessiterait de spécifier la loi spatiale de la fonction aléatoire au-delà des deux premiers moments ; cette exigence sort du cadre de la géostatistique linéaire.

On dira qu'une **combinaison linéaire** est **autorisée** si elle possède une espérance et une variance finies. Dans le cadre stationnaire d'ordre deux, cette condition est toujours satisfaite et toutes les combinaisons linéaires sont autorisées. En revanche, dans le cadre intrinsèque, seuls les *accroissements* de la fonction aléatoire ont une espérance et une variance, de sorte que seules les combinaisons linéaires d'accroissements sont autorisées. Or, on montre que ces dernières s'identifient à la famille des combinaisons linéaires de poids total nul. Ainsi donc, **sous l'hypothèse intrinsèque, on ne pourra calculer que les deux premiers moments des combinaisons linéaires de poids total nul** ; en contrepartie, on saura décrire des phénomènes plus complexes, et traiter des fonctions aléatoires sans variance *a priori*.

Autre implication de l'usage de combinaisons linéaires : du point de vue physique, il est souhaitable de ne manipuler que des variables régionalisées **additives**, c'est-à-dire telles que toute combinaison linéaire (en particulier, toute moyenne) a la même signification que la variable ponctuelle. Cette règle est essentielle lorsque l'on désire effectuer un *changement de support* : définir la moyenne d'une variable régionalisée non additive, comme par exemple le pH, sur un support plus grand que celui de la mesure n'a pas de sens.

IV.2. Règles de calcul sous l'hypothèse stationnaire

On considère une fonction aléatoire $Z(\mathbf{x})$ stationnaire d'ordre deux, d'espérance m et de fonction de covariance $C(\mathbf{h})$, et une combinaison linéaire attribuant les poids λ_i aux valeurs $Z(\mathbf{x}_i)$. Le calcul des deux premiers moments de cette combinaison linéaire est immédiat, grâce aux propriétés de linéarité de l'espérance et de la covariance :

- **espérance** : $E[\sum_i \lambda_i Z(\mathbf{x}_i)] = \sum_i \lambda_i E[Z(\mathbf{x}_i)] = m \sum_i \lambda_i$
- **variance** : $\text{var}[\sum_i \lambda_i Z(\mathbf{x}_i)] = \sum_i \sum_j \lambda_i \lambda_j \text{cov}[Z(\mathbf{x}_i), Z(\mathbf{x}_j)] = \sum_i \sum_j \lambda_i \lambda_j C(\mathbf{x}_i - \mathbf{x}_j)$.

IV.3. Règles de calcul sous l'hypothèse intrinsèque

L'objectif de ce paragraphe est d'explicitier les expressions des deux premiers moments d'une combinaison linéaire autorisée lorsque la fonction aléatoire $Z(\mathbf{x})$ étudiée est intrinsèque sans dérive. En ce qui concerne l'espérance, on sait d'une part que toute combinaison linéaire autorisée est une combinaison linéaire d'accroissements, et d'autre part qu'en l'absence de dérive, les accroissements ont une espérance nulle. On en déduit que **l'espérance de toute combinaison linéaire autorisée est nulle** :

$$E\left[\sum_i \lambda_i Z(\mathbf{x}_i)\right] = 0 \quad \text{lorsque} \quad \sum_i \lambda_i = 0.$$

L'explicitation de la variance est plus délicate.

On fixe un point $\mathbf{x}_0$ et on définit une fonction aléatoire $Y(\mathbf{x})$ par $Y(\mathbf{x}) = Z(\mathbf{x}) - Z(\mathbf{x}_0)$. $Y(\mathbf{x})$ est une combinaison linéaire autorisée, donc admet une fonction de covariance C_Y (en général non stationnaire) :

$$C_Y(\mathbf{x}, \mathbf{y}) = \text{cov}[Y(\mathbf{x}), Y(\mathbf{y})].$$

Pour expliciter $C_Y(\mathbf{x}, \mathbf{y})$, on utilise le variogramme γ de $Z(\mathbf{x})$:

$$\begin{aligned} \gamma(\mathbf{x} - \mathbf{y}) &= \frac{1}{2} \text{var}[Z(\mathbf{x}) - Z(\mathbf{y})] = \frac{1}{2} \text{var}[(Z(\mathbf{x}) - Z(\mathbf{x}_0)) - (Z(\mathbf{y}) - Z(\mathbf{x}_0))] \\ &= \frac{1}{2} \text{var}[Y(\mathbf{x}) - Y(\mathbf{y})] = \frac{1}{2} [C_Y(\mathbf{x}, \mathbf{x}) + C_Y(\mathbf{y}, \mathbf{y}) - 2C_Y(\mathbf{x}, \mathbf{y})] \end{aligned}$$

$$\text{c'est-à-dire : } C_Y(\mathbf{x}, \mathbf{y}) = \frac{1}{2} C_Y(\mathbf{x}, \mathbf{x}) + \frac{1}{2} C_Y(\mathbf{y}, \mathbf{y}) - \gamma(\mathbf{x} - \mathbf{y}).$$

On cherche à calculer la variance d'une combinaison linéaire autorisée, soit

$$\text{var}\left[\sum_i \lambda_i Z(\mathbf{x}_i)\right] \quad \text{sous la condition d'autorisation} \quad \sum_i \lambda_i = 0.$$

La somme des poids étant nulle, on a : $\sum_i \lambda_i Z(\mathbf{x}_i) = \sum_i \lambda_i [Z(\mathbf{x}_i) - Z(\mathbf{x}_0)] = \sum_i \lambda_i Y(\mathbf{x}_i)$, et on peut développer l'expression de la variance à l'aide de C_Y :

$$\begin{aligned} \text{var}\left[\sum_i \lambda_i Z(\mathbf{x}_i)\right] &= \text{var}\left[\sum_i \lambda_i Y(\mathbf{x}_i)\right] = \sum_i \sum_j \lambda_i \lambda_j C_Y(\mathbf{x}_i, \mathbf{x}_j) \\ &= \frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j C_Y(\mathbf{x}_i, \mathbf{x}_i) + \frac{1}{2} \sum_i \sum_j \lambda_i \lambda_j C_Y(\mathbf{x}_j, \mathbf{x}_j) - \sum_i \sum_j \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) \\ &= \frac{1}{2} \sum_i \lambda_i C_Y(\mathbf{x}_i, \mathbf{x}_i) \underbrace{\left(\sum_j \lambda_j\right)}_0 + \frac{1}{2} \sum_j \lambda_j C_Y(\mathbf{x}_j, \mathbf{x}_j) \underbrace{\left(\sum_i \lambda_i\right)}_0 - \sum_i \sum_j \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) \end{aligned}$$

d'où :

Règle de calcul :	$\text{var}\left[\sum_i \lambda_i Z(\mathbf{x}_i)\right] = - \sum_i \sum_j \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) \quad \text{lorsque} \quad \sum_i \lambda_i = 0.$
--------------------------	--

Sous l'hypothèse intrinsèque, la variance d'une combinaison linéaire autorisée se calcule comme s'il existait une covariance C , mais en remplaçant C par $-\gamma$, où γ est le variogramme.

V. L'étude exploratoire des données

V.1. Remarques générales

L'analyse géostatistique met en œuvre, dans un premier temps, les méthodes exploratoires qui permettent d'analyser la distribution et l'homogénéité des données dans l'espace, et de rechercher les observations "atypiques". A la lumière de l'étude exploratoire, on définira éventuellement une division de l'ensemble des données en plusieurs sous-groupes dont l'étude se réalisera séparément, et l'on jugera la pertinence d'une hypothèse de stationnarité pour représenter la régionalisation.

Lorsque les données sont nombreuses, leur description par des **histogrammes** est utile pour détecter des valeurs aberrantes ou suspectes. Il faut prendre garde au fait qu'une donnée suspecte n'est pas forcément fausse et ne doit jamais être éliminée sans raison. Elle peut très bien refléter le comportement réel de la régionalisation ; l'éliminer empêcherait de prévoir l'occurrence de telles valeurs singulières dans des zones non échantillonnées. La visualisation de l'histogramme des données est aussi un premier moyen de vérifier leur homogénéité dans l'espace. Un **découpage du champ** en plusieurs sous-zones sera éventuellement nécessaire ; par exemple, un histogramme multimodal peut conduire à l'identification, parmi les observations, de plusieurs "populations" susceptibles d'être géographiquement séparées. Une telle séparation est parfois imposée par des considérations physiques, qui empêchent de mélanger toutes les données : présence d'un obstacle naturel (faille, rivière...), partition d'une zone minéralisée selon la caractérisation minéralogique ou la texture de la roche, etc. Dans de tels cas, un problème qui se pose fréquemment est celui de la délimitation des sous-zones "homogènes" dans le champ, car il est rare que leurs frontières soient connues ou identifiables avec précision.

Des outils simples permettent de mettre en relief les observations notablement différentes de leurs voisines, en particulier les **cartes d'implantation** (où les valeurs des données sont par exemple repérées par un code de couleur) et les **nuages de "corrélation différée"**⁷.

Les **nuages de corrélation** entre les données et les coordonnées géographiques mettent en évidence le comportement de la variable le long des directions principales de l'espace. Leur étude permet de se donner une idée de la validité d'une hypothèse de stationnarité, aussi bien à échelle locale que globale, ou au contraire de repérer l'existence d'une dérive ("tendance" prononcée dans l'évolution des valeurs). Ils sont aussi un outil efficace pour repérer les observations les plus atypiques : celles correspondant aux points qui s'éloignent du reste de la nuée.

⁷ Pour un vecteur $\mathbf{h}$ donné, le nuage de corrélation différée est le nuage des points $(z(\mathbf{x}_a), z(\mathbf{x}_a + \mathbf{h}))$ tels que $\mathbf{x}_a$ et $\mathbf{x}_a + \mathbf{h}$ coïncident avec des sites de données. Les couples de données fort dissemblables correspondront aux points du nuage éloignés de la première bissectrice.

Enfin, dans le cas multivariable, les **nuages de corrélation entre variables** ainsi que les techniques d'**analyse de données**, comme l'analyse en composantes principales (annexe F), complètent utilement l'approche géostatistique, en donnant une représentation synthétique des données et en étudiant les liaisons entre les variables.

V.2. Application à un cas pratique : données de sol

V.2.1. Présentation des données

Tout au long de ce document, nous suivrons le cheminement d'une étude géostatistique où les différents concepts et hypothèses seront analysés et commentés. Les données utilisées sont extraites d'une étude réalisée au Cirad de Montpellier : "Bilan hydrique et énergétique à l'échelle d'une parcelle", par J. Imbernon, M. Arnaud, G. Bourgeon, M. Brouwers, M. Fortier, J.P. Freteaud, D. Koffi Koffi (novembre 1983). Il s'agit de l'analyse de 110 prélèvements de terre sur trois horizons (profondeurs : 0-20 cm, 40-60 cm et 80-100 cm) répartis sur une grille régulière. Diverses variables ont été mesurées, mais nous ne retiendrons pour l'exemple que le pourcentage d'argile et limon sur l'horizon 40-60 cm. Lorsque nous étudierons le cas multivarié, nous prendrons aussi en compte le pourcentage de terre fine (granulométrie < 2 mm) et la variable pF4.2 (humidité au point de flétrissement permanent, qui représente le taux d'humidité dans le sol que les plantes ne peuvent absorber) mesurée sur la terre fine provenant du même horizon.

La figure 2.6 localise, sous forme de carte à représentation proportionnelle⁸, l'implantation des prélèvements et montre que ceux-ci, s'ils ne remplissent pas la totalité de la surface d'un rectangle, sont néanmoins régulièrement répartis selon une maille carrée de six mètres de côté.

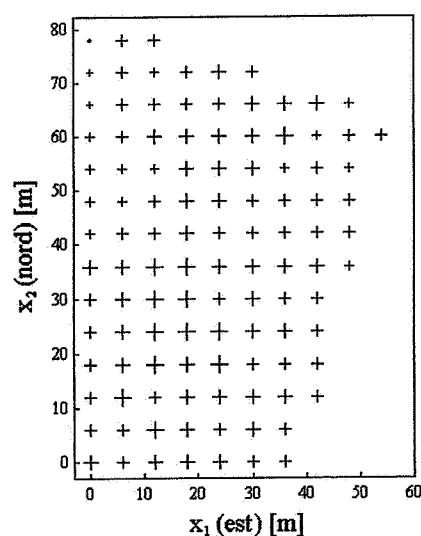


figure 2.6 : localisation des observations (source : CIRAD)

⁸ Chaque échantillon est représenté par une croix dont la taille est proportionnelle à la valeur observée.

Au premier coup d'œil, on peut repérer quelques valeurs extrêmes, notamment des valeurs faibles (croix de petite taille) au nord-ouest de la zone, mais on ne décèle pas la structure spatiale. Il faut aller au-delà de ce simple report des données.

Le tableau ci-dessous donne les statistiques élémentaires de la variable % argile + limon. Il est souvent utile d'inclure les valeurs des extrema, de la médiane et des quartiles (c'est-à-dire des valeurs qui divisent la distribution expérimentale en quatre parties d'effectifs égaux) pour améliorer la connaissance de la distribution des valeurs. Contrairement à la moyenne, les quartiles et la médiane sont des paramètres robustes, peu sensibles aux valeurs très fortes ou très faibles. Ainsi, une moyenne élevée et une médiane faible indiquent une dissymétrie dans la répartition des valeurs.

minimum	25.2
maximum	60.7
moyenne	48.44
écart-type	6.19
variance	38.27
premier quartile	44.6
médiane	48.3
troisième quartile	52.6

tableau 2.1 : statistiques élémentaires de la variable % argile + limon

L'histogramme expérimental représente graphiquement la distribution des valeurs prises par les données (figure 2.7) ; il donne une image de la loi de probabilité monovariée de la fonction aléatoire associée à la régionalisation. Ici, il est légèrement asymétrique.

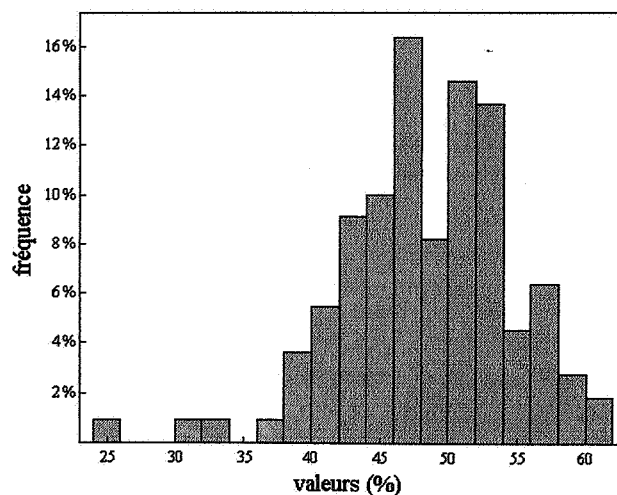


figure 2.7 : histogramme en fréquence du pourcentage d'argile et limon
 $n = 110$ données, médiane = 48.3%, moyenne = 48.44%

L'histogramme permet également d'identifier les données ayant des valeurs extrêmes et / ou suspectes. On peut ainsi observer quelques valeurs plutôt faibles inférieures à 35% ; il s'avère qu'elles proviennent de la zone nord-ouest du champ, plus caillouteuse que le reste de la parcelle, et ne sont donc pas aberrantes.

V.2.2. Etude de la stationnarité

Un outil simple pour étudier la stationnarité consiste à visualiser les nuages des valeurs prises par la variable le long de plusieurs directions de l'espace ; ici, les données étant réparties sur une grille régulière, on choisira comme directions les axes de cette grille, qui coïncident avec les axes de coordonnées. Ces **nuages "directionnels"** permettent d'apprécier l'évolution des valeurs dans l'espace, leur tendance et leur variabilité, et d'accréditer ou non une hypothèse de stationnarité. On peut compléter cette analyse en visualisant les valeurs moyenne et médiane des nuages de points, et en vérifiant graphiquement si ces valeurs restent relativement stables le long des directions de coordonnées.

En l'occurrence, la figure 2.8 montre que les valeurs moyennes et médianes commencent par croître linéairement, puis décroissent le long de la direction est-ouest (coordonnée x_1) ; le nuage correspondant a une forme de dôme. Le long de la direction nord-sud (coordonnée x_2), le nuage, ainsi que ses moyenne et médiane, affichent une certaine "dérive", en ce sens que les valeurs ont tendance à être plus faibles au nord de la parcelle et plus élevées au sud. Il faut savoir que la parcelle est bordée, sur son côté sud-est, par une rivière, ce qui explique que le terrain soit plus limoneux dans cette partie du champ. On remarque encore que, d'une façon générale, la médiane et la moyenne sont voisines, sauf sur la dernière ligne ($x_2 = 78\text{m}$), mais celle-ci ne compte que trois observations.

L'étude des valeurs moyennes correspond à celle du moment d'ordre 1 (espérance) de la fonction aléatoire sous-jacente. Les moments d'ordre deux vont se refléter en partie au niveau de la dispersion des nuages. Bien que ces résultats doivent être interprétés avec prudence, vu le faible nombre de valeurs le long de chaque ligne ou colonne, on remarque que les nuages sont plus dispersés dans la région nord-ouest (x_1 faible, x_2 élevé) que dans la région sud-est. Une hypothèse de stationnarité d'ordre deux risque d'être incompatible avec ces constatations, car ni le moment d'ordre 1 (moyenne) ni ceux d'ordre 2 (variance, pour le moins, si celle-ci existe) ne semblent invariants dans l'espace ; on pourra avoir recours à l'hypothèse intrinsèque qui est moins exigeante.

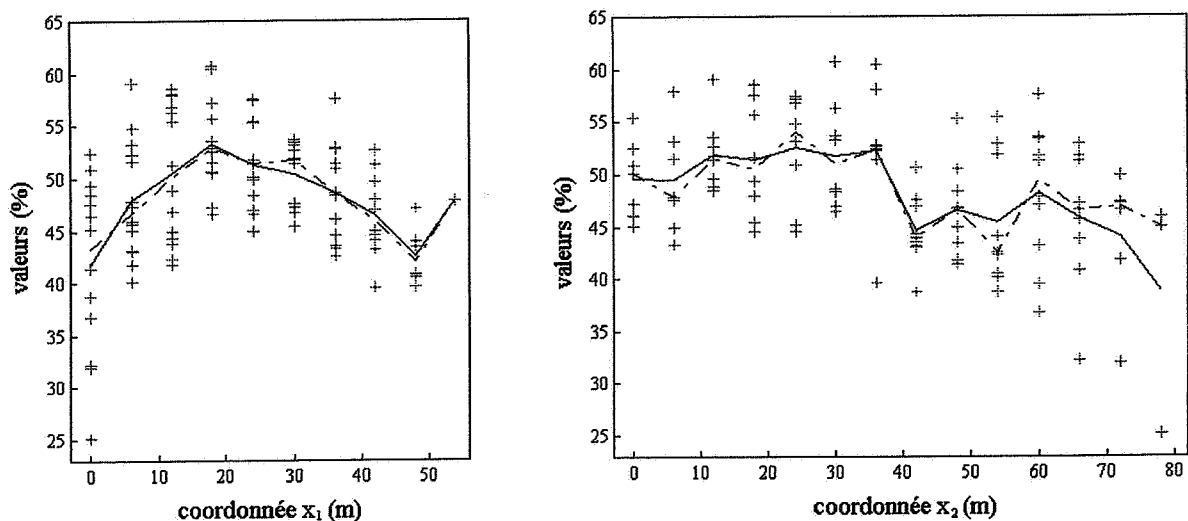


figure 2.8 : nuages directionnels des valeurs
ligne continue = moyenne, pointillés = médiane

Les nuages directionnels permettent également de repérer les observations “originales” que l’histogramme ne met pas en relief (par exemple, une valeur faible entourée de valeurs fortes : même si séparément, chaque valeur est plausible, leur juxtaposition dans l’espace peut présenter les caractères d’une “anomalie”). Ainsi, sur la figure 2.8, les trois valeurs les plus faibles, situées au nord-ouest de la parcelle ($x_1 = 0$ et $x_2 = 66, 72$ et 78), se détachent du reste des nuages directionnels. Ceci signifie une certaine hétérogénéité spatiale dans les valeurs de la régionalisation.

Une autre manière de détecter les observations différentes de leurs voisines est d’examiner un ou plusieurs **nuages de corrélation différée**, i.e. les nuages des points $(z(x_\alpha), z(x_\alpha + h))$ où h est un vecteur donné, x_α et $x_\alpha + h$ coïncident avec des sites d’observation. Lorsque les échantillons sont régulièrement répartis, comme c’est le cas ici, on prend un vecteur h multiple de la maille de la grille d’échantillonnage ; dans le cas contraire, on est contraint d’introduire des tolérances sur la longueur et l’orientation de h , faute de quoi le nuage se réduirait à très peu de points.

Les couples de données les plus dissemblables correspondent aux points qui apparaissent très éloignés de la première bissectrice. Ainsi, sur la figure 2.9, certains points se détachent, que l’on peut identifier sur la carte: dans la zone nord-ouest, les trois valeurs faibles de la première colonne s’opposent aux trois valeurs fortes de la deuxième colonne : $z(0,78) = 25.2$ et $z(6,78) = 46$, $z(0,72) = 31.9$ et $z(6,72) = 47.3$, $z(0,66) = 32.2$ et $z(6,66) = 45.7$. On peut également repérer quelques valeurs localement très différentes, par exemple : $z(36,60) = 57.6$ et $z(42,60) = 39.5$.

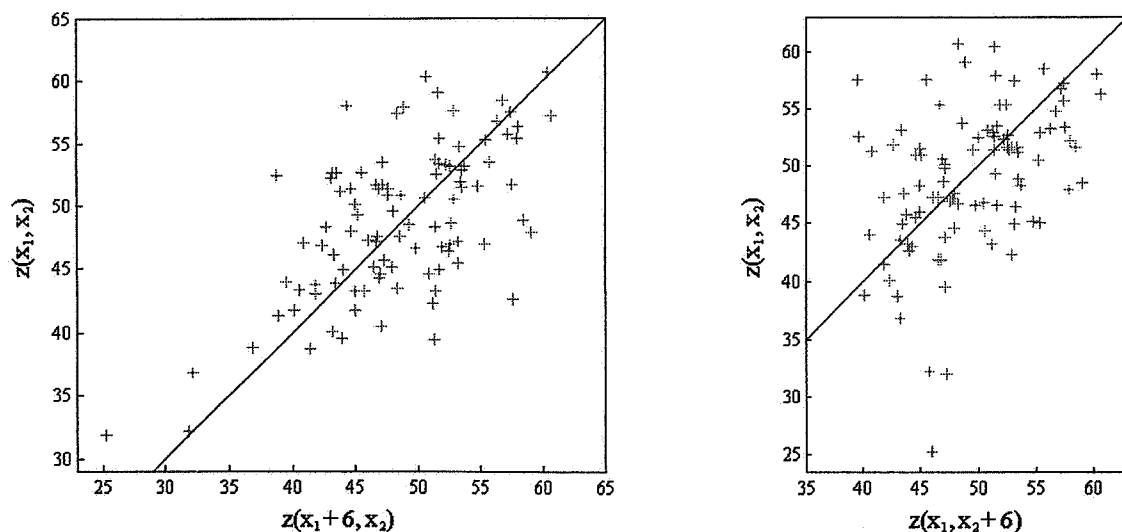


figure 2.9 : nuages des points $(z(x), z(x+h))$, avec $|h| = 6m$ et h orienté selon la direction nord-sud (à gauche) et est-ouest (à droite) sur chaque graphique, la ligne continue représente la première bissectrice

En conclusion de cette étude exploratoire, il ressort qu’une hypothèse de stationnarité, voire de quasi-stationnarité, risque d’être trop exigeante, principalement à cause de la présence d’une dérive assez marquée dans les données et d’une dispersion non homogène des valeurs au sein de la parcelle. Une hypothèse de type intrinsèque a davantage de chance d’être mieux adaptée à un tel jeu de données.

V.3. Application à des données minières

V.3.1. Présentation des données

On s'intéresse à présent à un ensemble de données issu d'un gisement cuprifère. La variable étudiée (teneur en cuivre) a été mesurée sur des échantillons situés le long de sondages, ou "carottes de sondage".

Pour mener à bien l'étude géostatistique, il est indispensable que les mesures aient le même support, ce qui n'est pas le cas des carottes de sondage. On sait, d'après l'effet de support, que les propriétés statistiques, en particulier la variabilité, dépendent du support considéré. Une "régularisation" a donc été menée : elle consiste à replacer les échantillons le long des sondages, puis à découper chaque sondage en morceaux de même longueur, ou **composites**, dont la teneur est égale à la moyenne des mesures initiales pondérées par les fractions de composite associées (figure 2.10). Cette opération suppose que les teneurs sont uniformes sur les supports des échantillons initiaux, pour pouvoir reconstituer le profil des teneurs de chaque sondage.

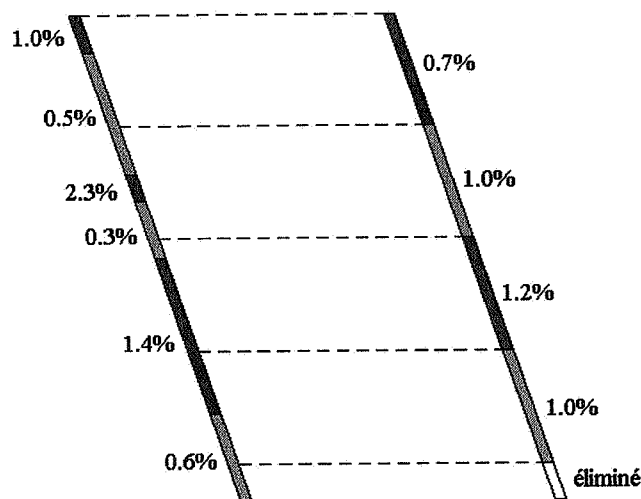
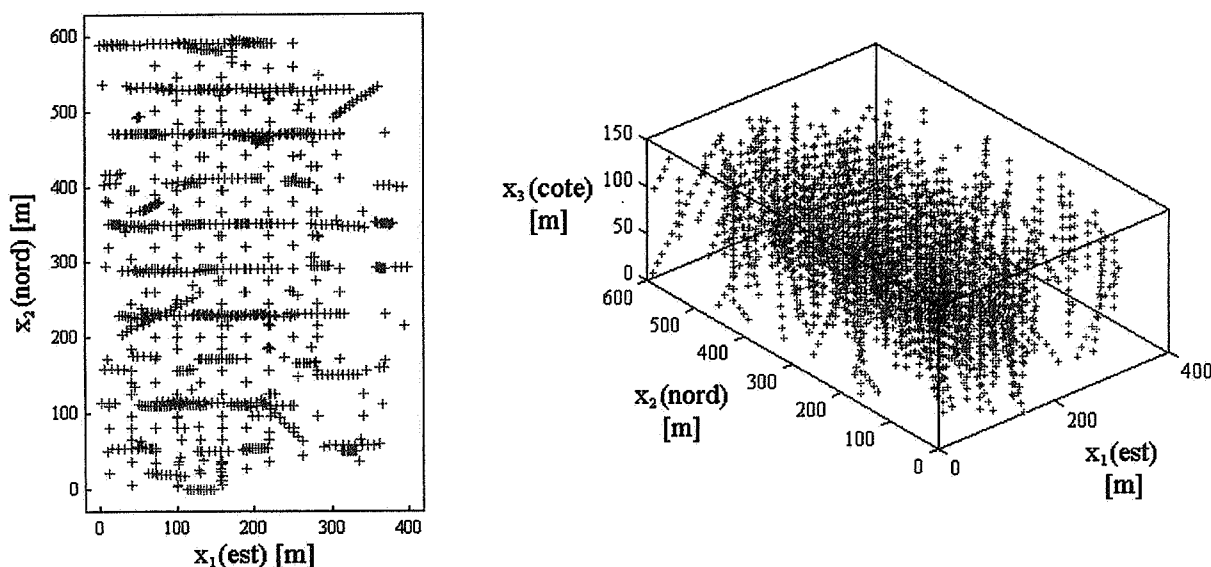


figure 2.10 : principe de la création de composites
à gauche : un sondage avec ses mesures initiales, à droite : les mesures compositées

Dans le cas présent, les échantillons ont été composités tous les 12m, qui correspond à la hauteur du banc dans la fosse (mine à ciel ouvert) ; on obtient ainsi 2376 mesures compositées. On admet que celles-ci peuvent être analysées conjointement, faisant abstraction de leur différence d'orientation dans l'espace (un composite vertical n'a pas exactement les mêmes caractéristiques qu'un composite incliné).

V.3.2. Cartes d'implantation

La carte d'implantation (figure 2.11) montre que la zone échantillonnée couvre une superficie horizontale de 400m × 600m. Le long de la direction verticale, les données sont réparties entre 0 et 130m de profondeur.



*figure 2.11 : cartes d'implantation des échantillons composés
projection horizontale (à gauche) et vue en perspective (à droite)
(source : CODELCO-Chile)*

V.3.3. Statistiques élémentaires

Pour le calcul des histogramme et statistiques élémentaires (figure 2.13), une variable de pondération (en anglais, *declustering*) a été prise en compte afin de corriger les effets dus aux irrégularités d'échantillonnage. L'idée est d'assigner aux données spatialement groupées, en partie redondantes, un poids plus faible qu'aux données isolées, qui sont représentatives d'une portion plus étendue du champ.

La variable de pondération a été déterminée par la *méthode des cellules*. Cette technique consiste à diviser la zone échantillonnée en cellules rectangulaires identiques, et à attribuer à chaque échantillon un poids inversement proportionnel au nombre de données présentes dans la cellule à laquelle il appartient. La mise en œuvre requiert de définir au préalable la taille de la cellule élémentaire. Usuellement, on teste plusieurs tailles et on étudie leur impact sur la valeur de la moyenne (figure 2.12) ; dans cet exemple, on choisit d'utiliser des cellules à base carrée, et dont la hauteur est égale à 40% de la longueur, de façon à compenser la plus grande densité d'échantillonnage dans la direction verticale.

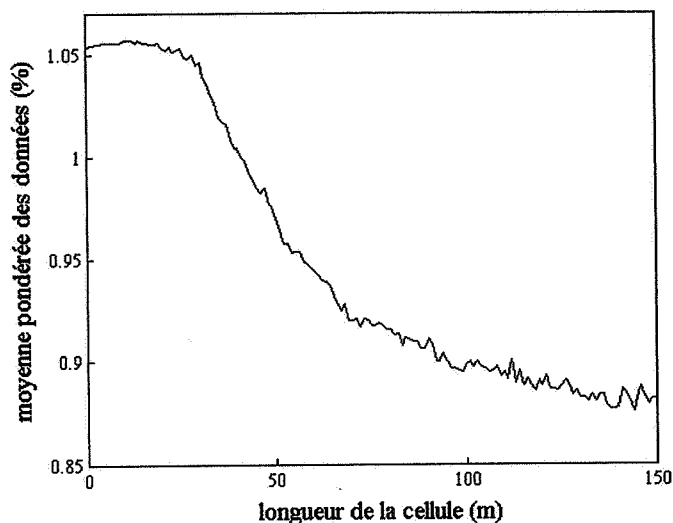


figure 2.12 : influence de la longueur des cellules sur la moyenne

Lorsque les cellules sont infiniment petites, chaque donnée se trouve dans une cellule différente et reçoit donc le même poids ; on obtient par conséquent la moyenne arithmétique des données (1.05%). Quand la longueur augmente, la moyenne pondérée diminue, ce qui indique que les poids calculés privilégient les valeurs faibles au détriment des valeurs fortes ; c'est un signe que l'échantillonnage est plus dense dans les zones de fortes valeurs, qui sont les plus intéressantes économiquement.

Dans ce cas de figure, il est d'usage de prendre la taille qui fournit la valeur minimale de la moyenne, pour justement "compenser" le caractère préférentiel de l'échantillonnage. Cela revient à se placer dans la configuration la plus défavorable. Une telle décision est discutable, et des choix plus "neutres" sont envisageables. Ici, on choisit, de manière conventionnelle, une cellule de taille 100m × 100m × 40m, qui fournit une valeur moyenne d'environ 0.9%.

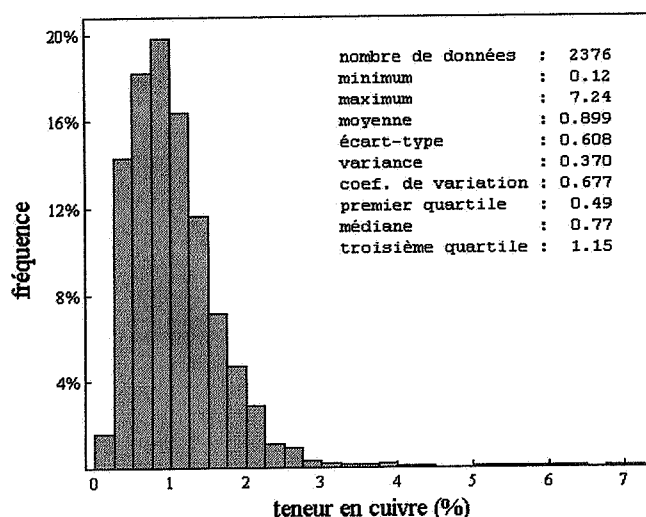


figure 2.13 : histogramme et statistiques élémentaires, tenant compte d'une pondération non uniforme des données

La distribution des valeurs est asymétrique, avec beaucoup de valeurs faibles et quelques unes très fortes ($> 5\%$). La valeur minimale vaut 0.12%, tandis que celle maximale est égale à 7.24%, sans être pour autant aberrante. On peut comparer la distribution expérimentale des valeurs avec une distribution lognormale, à l'aide d'un graphique quantiles contre quantiles (figure 2.14) : un histogramme parfaitement lognormal conduirait à des points alignés ; ici, on constate un léger écart à la lognormalité, notamment en ce qui concerne les premiers quantiles.

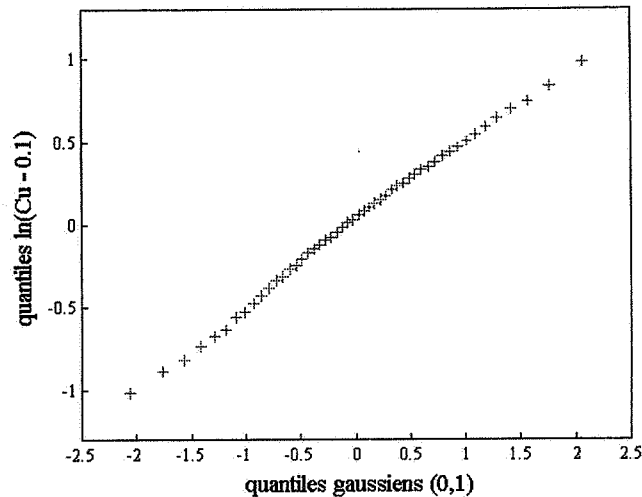


figure 2.14 : graphique quantiles contre quantiles

V.3.4. Nuages directionnels des valeurs

Pour se faire une idée plus précise de la distribution spatiale des valeurs, il est commode de visualiser les nuages des valeurs le long des directions de coordonnées (figure 2.15) ; ces nuages permettent de détecter les observations les plus atypiques et d'étudier la pertinence d'une hypothèse de stationnarité.

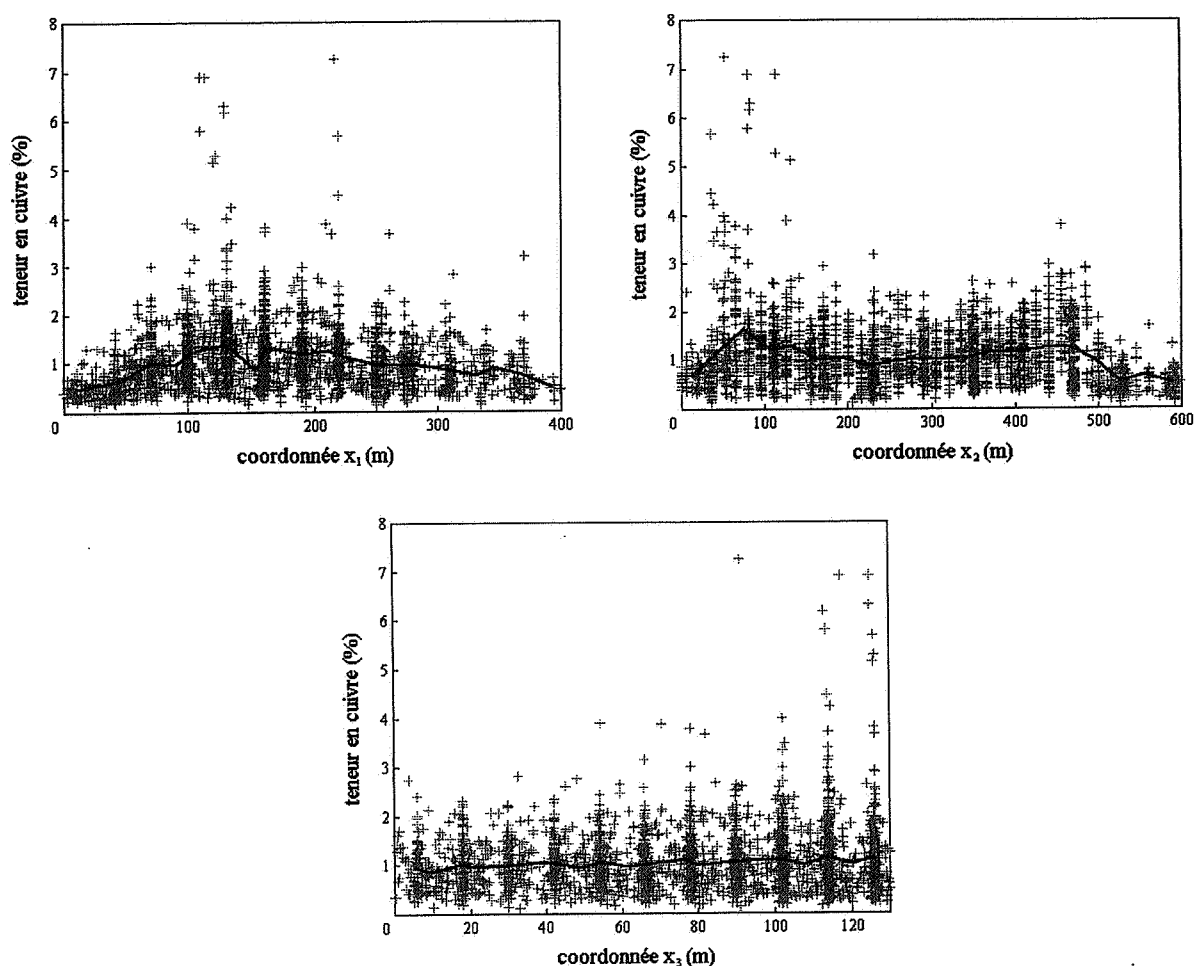


figure 2.15 : nuages directionnels des valeurs
lignes noires = régressions expérimentales

Le long de la direction verticale, la distribution des valeurs est relativement homogène, avec une moyenne remarquablement stable, à l'exception des faibles profondeurs (cote z élevée) où se situent quelques valeurs très fortes. Le long des directions horizontales, les valeurs moyennes sont variables, quoique localement constantes. Elles montrent que la périphérie du champ a tendance à présenter des valeurs plus faibles, et que la zone de plus fortes valeurs est localisée entre les abscisses 100 et 200m, et entre les ordonnées 50 et 100m.

Les deux premiers nuages mettent également en évidence ce que l'on appelle un **effet proportionnel** : la dispersion des valeurs est nettement plus importante dans les zones de fortes valeurs que dans celles de faibles valeurs. Cet effet est fréquent lorsque l'on étudie des variables à distribution asymétrique, comme c'est le cas ici.

Ces remarques mettent en cause la validité d'une hypothèse de stationnarité, mais l'on pourra admettre une hypothèse de **quasi-stationnarité**, avec une moyenne et une variance (dispersion) localement constantes et variant lentement dans le champ. En fait, on montre que l'effet proportionnel n'est pas incompatible avec une hypothèse de stationnarité globale, et qu'il est même la règle lorsque l'on a affaire à une fonction aléatoire stationnaire lognormale.

V.3.5. Nuage de corrélation différée

On peut compléter l'étude de la répartition spatiale des valeurs à l'aide d'un nuage de corrélation différée. En l'occurrence, on visualise les couples de valeurs $\{z(x), z(x+h)\}$, avec $|h| \in [18m, 22m]$, sans se soucier de l'orientation de ce vecteur (figure 2.16) ; l'usage d'une tolérance sur la distance est due au fait que l'échantillonnage est irrégulier.

Les points les plus éloignés de la bissectrice ont été mis en relief et les paires de données correspondantes ont été replacées sur la carte d'implantation. Elles proviennent toutes de la zone de fortes valeurs, qui est aussi celle où la variabilité est la plus forte. C'est une autre illustration de l'effet proportionnel.

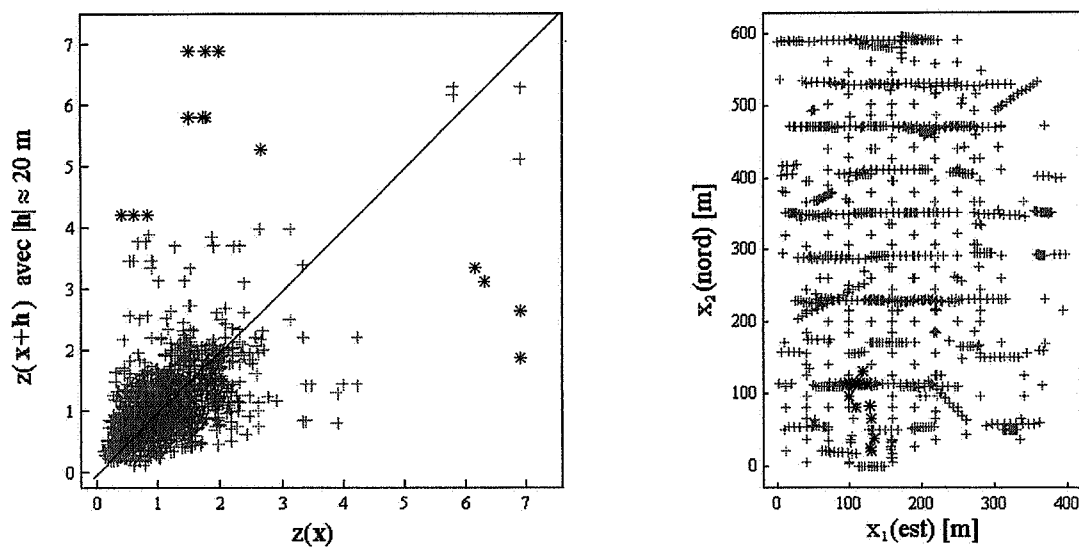
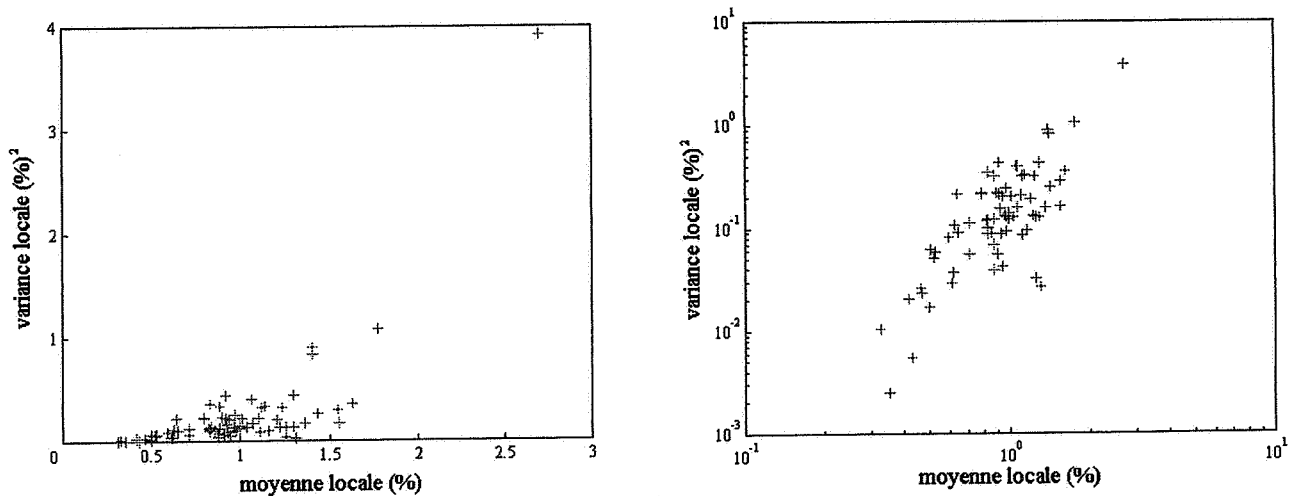


figure 2.16 : nuage de corrélation différée (à gauche) et carte d'implantation (à droite)

V.3.6. Moyennes et variances locales

Un dernier outil qui permet de mettre en évidence et modéliser un effet proportionnel consiste à diviser le champ en cellules identiques, à calculer les moyenne et variance des données dans chaque cellule, et à visualiser le résultat sous forme d'un nuage de points. En l'occurrence, la zone d'étude a été divisée en 72 cellules de taille $100m \times 100m \times 40m$, sur chacune desquelles sont calculées les deux premiers moments (moyenne et variance) des données, en tenant par ailleurs compte de la variable de pondération déterminée antérieurement.



*figure 2.17 : nuage de corrélation entre les moyenne et variance locales
à gauche : échelle arithmétique, à droite : échelle logarithmique*

On note une forte dépendance entre moyennes et variances locales ; le second nuage suggère même une relation linéaire entre leur logarithme. La droite de régression de ce nuage peut servir à modéliser l'effet proportionnel (figure 2.18) :

$$s^{*2}(\mathbf{x}) = 0.152[m^{*}(\mathbf{x})]^{2.29}.$$

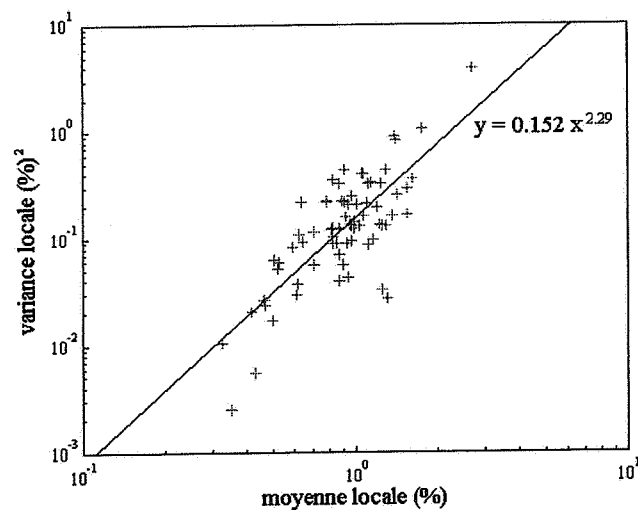


figure 2.18 : modèle de régression entre moyenne et variance locales

Résumé

La géostatistique probabiliste interprète la variable régionalisée étudiée comme une réalisation d'une **fonction aléatoire**. Ce choix méthodologique permet de prendre en compte à la fois l'aspect *erratique* et l'aspect *structuré* du phénomène régionalisé.

En géostatistique linéaire, on n'utilise pas la loi spatiale complète de la fonction aléatoire, mais seulement ses deux premiers moments (espérance et covariance ou variogramme), suffisants pour décrire de manière satisfaisante le comportement et les relations conjointes entre les valeurs prises par la variable régionalisée.

Pour que ces moments puissent être inférés à partir des données expérimentales, des hypothèses limitatives sont nécessaires ; elles font appel aux notions de **stationnarité** et d'**ergodicité** de la fonction aléatoire. Plusieurs hypothèses de stationnarité sont possibles : stationnarité d'ordre deux ou hypothèse intrinsèque, moins exigeante, puisqu'il s'agit de la stationnarité des *accroissements* de la fonction aléatoire. Si l'hypothèse d'ergodicité est souvent admise, il faut en revanche vérifier quelle hypothèse de stationnarité est valide (ou plus exactement judicieuse) à l'aide d'une étude exploratoire des données. Ajoutons pour terminer que la pertinence d'une hypothèse de stationnarité peut dépendre de l'*échelle de travail* envisagée.

Chapitre 3 : les outils de modélisation de la structure spatiale

Les valeurs prises par une variable régionalisée ne sont pas indépendantes, en ce sens qu'une valeur observée en un site apporte de l'information sur les valeurs des sites voisins. L'ensemble des corrélations, ou dépendances, qui existent dans la distribution spatiale des valeurs constitue la "structure" du phénomène régionalisé. Cette notion, intuitive mais pourtant assez vague, est formalisée de manière précise dans l'interprétation probabiliste de la variable régionalisée : la structure est décrite par la *loi spatiale* de la fonction aléatoire associée, qui modélise le comportement conjoint de tout ensemble de valeurs grâce à leur loi multivariable.

En géostatistique linéaire, la description de la loi spatiale est limitée à ses deux premiers moments. Le moment d'ordre 1 (espérance) ne fait intervenir dans sa définition qu'un seul point et n'apporte pas réellement d'information structurale. En revanche, les moments d'ordre 2 (covariance et variogramme notamment) sont définis à l'aide de deux points, c'est-à-dire du plus petit ensemble que l'on puisse considérer pour décrire "l'interaction" entre des valeurs. Ils apportent une description élémentaire et opératoire de la structure spatiale du phénomène régionalisé étudié, et sont pour cette raison aussi appelés *outils structuraux*.

I. La covariance stationnaire

On considère la variable régionalisée étudiée $z(\mathbf{x})$ comme une réalisation d'une fonction aléatoire $Z(\mathbf{x})$ stationnaire d'ordre deux. Dans ce cas, la fonction de covariance C existe et possède les propriétés théoriques suivantes :

- $C(\mathbf{h}) = \text{cov}[Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})]$ ne dépend que de $\mathbf{h}$;
- $C(\mathbf{0}) = \text{var}[Z(\mathbf{x})] = \sigma^2$ constante (variance *a priori*) ;
- parité : $C(\mathbf{h}) = C(-\mathbf{h})$;
- inégalité de Cauchy-Schwarz : $|C(\mathbf{h})| \leq C(\mathbf{0})$ avec, en général, $|C(\mathbf{h})| < C(\mathbf{0})$ pour $\mathbf{h} \neq \mathbf{0}$ (dans le cas contraire, la covariance est nécessairement périodique¹) ;

¹ En effet, en utilisant la bilinéarité de la covariance et l'inégalité de Cauchy-Schwarz, on établit :

$$\forall \mathbf{h}, \mathbf{k} \in \mathbb{R}^d, |C(\mathbf{h} + \mathbf{k}) - C(\mathbf{h})|^2 \leq \text{var}[Z(\mathbf{x} + \mathbf{h} + \mathbf{k}) - Z(\mathbf{x} + \mathbf{h})] \text{var}[Z(\mathbf{x})] = 2[C(\mathbf{0}) - C(\mathbf{k})] \text{var}[Z(\mathbf{x})]$$

de sorte que s'il existe $\mathbf{k}$ tel que $C(\mathbf{0}) = C(\mathbf{k})$, alors $C(\mathbf{h} + \mathbf{k}) = C(\mathbf{h})$ pour tout vecteur $\mathbf{h}$, i.e. $C(\mathbf{h})$ est périodique.

- La covariance mesure en quelque sorte la parenté entre les valeurs prises par la variable régionalisée, en fonction de leur séparation. Les propriétés de régularité spatiale de la régionalisation vont se refléter au niveau du **comportement à l'origine** de la covariance. Plus celui-ci est régulier, plus la fonction aléatoire et la variable régionalisée sont elles-mêmes régulières dans l'espace. Au contraire, plus la décroissance de la covariance à partir de $\mathbf{h} = \mathbf{0}$ est accentuée, plus "l'influence" d'un site sur ses voisins se détériore rapidement, et plus la régionalisation est déstructurée et erratique.
- $C(\mathbf{h})$ est une **fonction de type positif** :

$$\forall k \in \mathbb{N}^*, \forall \lambda_1 \dots \lambda_k \in \mathbb{R}, \forall \mathbf{x}_1 \dots \mathbf{x}_k, \sum_{i=1}^k \sum_{j=1}^k \lambda_i \lambda_j C(\mathbf{x}_i - \mathbf{x}_j) \geq 0.$$

Cette condition exprime la positivité de la variance de la combinaison linéaire $\sum_{i=1}^k \lambda_i Z(\mathbf{x}_i)$.

Signalons qu'une fonction de type positif dans un espace l'est encore dans un espace de dimension inférieure, mais pas nécessairement dans un espace de dimension supérieure.

Ces propriétés sont analogues à celles du covariogramme transitif défini au premier chapitre de ce document. Toutefois, contrairement au covariogramme transitif, la covariance peut prendre des valeurs négatives même si la variable associée est toujours positive, et elle n'admet pas nécessairement de portée finie. En revanche, pour une fonction aléatoire ergodique, la covariance tend vers 0 à l'infini, et on peut toujours lui associer une **portée** finie (distance au-delà de laquelle elle est nulle ou quasi-nulle). Un contre-exemple de covariance ne tendant pas vers 0 à l'infini est une covariance périodique, par exemple $C(\mathbf{h}) = \cos(\mathbf{h})$ dans l'espace à une dimension : la périodicité empêche les valeurs prises en deux endroits distants d'être asymptotiquement non corrélées.

Désormais, nous nous intéresserons essentiellement aux fonctions aléatoires ergodiques, ce qui implique que $C(\mathbf{h})$ tend vers 0 à l'infini, est intégrable (cf. annexe B) et admet un maximum absolu en $\mathbf{h} = \mathbf{0}$.

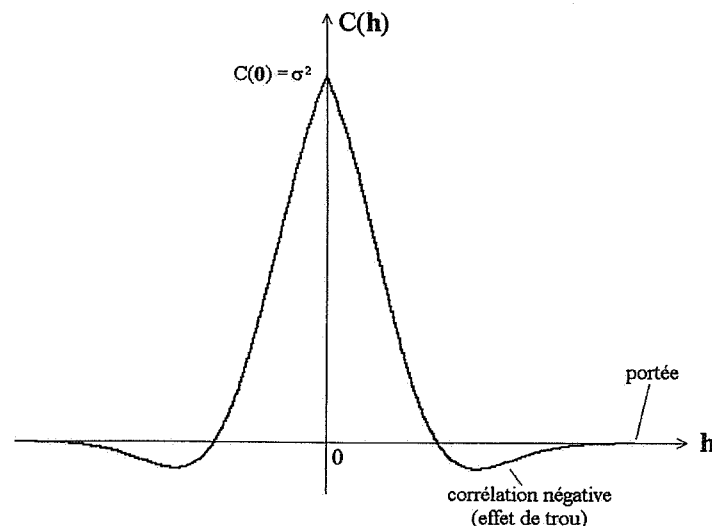


figure 3.1 : exemple de fonction de covariance non monotone (à effet de trou) dans un espace unidimensionnel

Résultats complémentaires

- Le degré maximal d'irrégularité analytique d'une fonction de covariance est toujours localisé en $\mathbf{h} = \mathbf{0}$; ainsi, une covariance continue à l'origine l'est également partout ailleurs.
- Il y a identité entre les fonctions de covariance et les fonctions de type positif. Plus précisément, pour toute fonction $C(\mathbf{h})$ de type positif, on sait construire une fonction aléatoire de loi spatiale gaussienne admettant $C(\mathbf{h})$ comme fonction de covariance. La loi gaussienne est donc compatible avec n'importe quelle fonction de covariance (ce résultat est incertain pour des lois spatiales quelconques).
- Toute somme de fonctions de covariance est encore une covariance. En effet, considérons p fonctions aléatoires $Z_i(\mathbf{x})$, stationnaires d'ordre deux, indépendantes et de covariances respectives $C_i(\mathbf{h})$; alors la fonction aléatoire $Z(\mathbf{x}) = Z_1(\mathbf{x}) + \dots + Z_p(\mathbf{x})$ admet la covariance $C(\mathbf{h}) = C_1(\mathbf{h}) + \dots + C_p(\mathbf{h})$. Cette propriété est utilisée pour construire des modèles complexes de covariance à partir de modèles élémentaires (voir *infra*, le modèle linéaire de régionalisation).
- Tout produit de fonctions de covariance est une covariance. En effet, soient p fonctions aléatoires $Z_i(\mathbf{x})$, stationnaires d'ordre deux, indépendantes et de covariances respectives $C_i(\mathbf{h})$; alors la fonction aléatoire $Z(\mathbf{x}) = Z_1(\mathbf{x}) \times \dots \times Z_p(\mathbf{x})$ admet la covariance $C(\mathbf{h}) = C_1(\mathbf{h}) \times \dots \times C_p(\mathbf{h})$. Cette propriété est toutefois d'un intérêt pratique moindre que la propriété précédente.
- **Théorème de Bochner** : une fonction $C(\mathbf{h})$ continue est de type positif, donc une fonction de covariance, si et seulement si elle est la transformée de Fourier d'une mesure positive sommable χ :

$$C(\mathbf{h}) = \int e^{-i \langle \mathbf{t} | \mathbf{h} \rangle} \chi(d\mathbf{t}), \text{ avec } \int \chi(d\mathbf{t}) = C(\mathbf{0}) \text{ et } i = \sqrt{-1}.$$

Lorsque $C(\mathbf{h})$ est intégrable (portée intégrale finie), la mesure χ est absolument continue et admet une densité S qui est une fonction réelle, paire et positive, et est appelée *fonction de densité spectrale* ou plus simplement *spectre* :

$$C(\mathbf{h}) = \int e^{-i \langle \mathbf{t} | \mathbf{h} \rangle} S(\mathbf{t}) d\mathbf{t}.$$

Ce théorème est peu utilisé en géostatistique, sauf en théorie de la simulation.

II. Le variogramme

II.1. Définition

On suppose à présent que $Z(\mathbf{x})$ est une fonction aléatoire intrinsèque sans dérive, c'est-à-dire :

$$\forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{D}, \begin{cases} E[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = 0 \\ \text{var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = 2\gamma(\mathbf{h}) \end{cases}$$

Compte tenu de la première relation, la seconde égalité s'écrit encore :

$$\gamma(\mathbf{h}) = \frac{1}{2} \text{var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = \frac{1}{2} E \{ [Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]^2 \}.$$

Si, de surcroît, $Z(\mathbf{x})$ est stationnaire d'ordre deux, sa fonction de covariance existe et est liée au variogramme par la relation :

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}).$$

Cette relation montre l'équivalence entre le variogramme et la fonction de covariance dans le cas stationnaire. En règle générale, on préférera travailler avec le variogramme qui permet de prendre en compte des processus plus généraux que la fonction de covariance, car il peut être défini dans le cadre strictement intrinsèque.

On rappelle en effet que l'existence de l'espérance et de la variance des accroissements n'implique pas l'existence des deux premiers moments de la fonction aléatoire $Z(\mathbf{x})$: une fonction aléatoire intrinsèque peut avoir une variance *a priori* infinie bien que la variance de ses accroissements soit finie. Il n'est pas obligatoire qu'une fonction aléatoire intrinsèque ait une espérance constante ou une variance constante. Un contre-exemple classique est celui du mouvement brownien, qui admet une variance infinie (*cf. infra*, paragraphe IV.3).

II.2. Propriétés théoriques

- Le variogramme est pair : $\gamma(\mathbf{h}) = \gamma(-\mathbf{h})$.
- Il s'annule à l'origine : $\gamma(\mathbf{0}) = 0$.
- Il est positif ou nul : $\gamma(\mathbf{h}) \geq 0$.
- A l'infini, il croît moins vite qu'une parabole : $\gamma(\mathbf{h}) / |\mathbf{h}|^2$ reste borné quand $|\mathbf{h}| \rightarrow +\infty$.
Pour une fonction aléatoire intrinsèque sans dérive, on a même $\lim_{|\mathbf{h}| \rightarrow +\infty} \gamma(\mathbf{h}) / |\mathbf{h}|^2 = 0$.
- Rappelons la règle de calcul, fort utile, établie au chapitre précédent :
pour tout système de points $\{\mathbf{x}_i\}_i$ et pour toute combinaison linéaire *autorisée*, c'est-à-dire telle que $\sum_i \lambda_i = 0$, on a :

$$\text{var} \left[\sum_i \lambda_i Z(\mathbf{x}_i) \right] = - \sum_i \sum_j \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j).$$

- $\gamma(\mathbf{h})$ est une fonction de type *négatif conditionnel*,

$$\text{i.e. } \forall k \in \mathbb{N}^*, \forall \lambda_1, \dots, \lambda_k \in \mathbb{R} \text{ tels que } \sum_{i=1}^k \lambda_i = 0, \forall \mathbf{x}_1, \dots, \mathbf{x}_k, \sum_{i=1}^k \sum_{j=1}^k \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) \leq 0.$$

Cette propriété, qui résulte de la règle de calcul précédente, est une condition nécessaire et suffisante pour que γ soit le variogramme d'une fonction aléatoire intrinsèque. L'adjectif *conditionnel* se réfère au fait que l'inégalité n'est vraie que pour une classe restreinte de pondérateurs $(\lambda_i)_i$, à savoir ceux de somme totale nulle.

Autres propriétés

- Pour $t > 0$, la fonction qui à $\mathbf{h} \in \mathbb{R}^d$ associe $e^{-t\gamma(\mathbf{h})}$ est de type positif.
- Toute somme de variogrammes est encore un variogramme.
- Un produit de variogrammes n'est pas nécessairement un variogramme.

Remarque

Le rôle du variogramme est de permettre le calcul des variances des combinaisons linéaires de poids total nul. On pourrait définir le variogramme comme une fonction γ telle que, si $\sum_i \lambda_i = 0$,

$$\text{var}[\sum_i \lambda_i Z(\mathbf{x}_i)] = -\sum_i \sum_j \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j)$$

pour tout système de points $\{\mathbf{x}_i\}_i$. Mais une telle définition n'est pas biunivoque et fournit toute une classe de fonctions qui ne diffèrent que d'une constante additive. S'il est inutile d'introduire une telle indétermination, cette seconde définition du variogramme prendra tout son sens lorsque l'on affaiblira l'hypothèse intrinsèque et que l'on étudiera les fonctions de covariances généralisées (cf. chapitre 6).

II.3. Anisotropie

Le variogramme $\gamma(\mathbf{h})$ est isotrope s'il est identique dans toutes les directions, c'est-à-dire s'il ne dépend pas de l'orientation du vecteur $\mathbf{h}$, mais seulement de son module $|\mathbf{h}|$. Dans le cas contraire, il y a anisotropie ; une telle propriété caractérise un phénomène qui s'étend davantage dans certaines directions. Les modèles théoriques proposés seront isotropes. Il faudra donc en cas d'anisotropie se ramener au cas isotrope. Nous reviendrons en détail sur les différents types d'anisotropie ultérieurement.

II.4. Palier et portée

Le variogramme $\gamma(\mathbf{h})$ mesure la dissemblance, ou encore la variabilité, entre deux valeurs séparées de $\mathbf{h}$. Souvent, $\gamma(\mathbf{h})$ croît à partir de $|\mathbf{h}| = 0$ (signe que la variabilité augmente avec la séparation), puis atteint, à partir d'une distance a , une valeur limite $\gamma(\infty)$ appelée **palier**. Dans ce cas, on montre que la fonction aléatoire est stationnaire d'ordre deux, et que sa variance *a priori* est égale au palier, car :

$$\begin{aligned} \gamma(\mathbf{h}) &= C(\mathbf{0}) - C(\mathbf{h}) \\ \downarrow_{|\mathbf{h}| \rightarrow \infty} & \quad \downarrow_{|\mathbf{h}| \rightarrow \infty} \\ \gamma(\infty) &= C(\mathbf{0}) = \sigma^2 : \text{variance } a \text{ priori.} \end{aligned}$$

Dans ce modèle, deux valeurs $Z(\mathbf{x})$ et $Z(\mathbf{x} + \mathbf{h})$ sont corrélées si la longueur du vecteur de séparation $\mathbf{h}$ est inférieure à la distance a , qui est appelée **portée**. La portée correspond à la notion de *zone d'influence* d'une valeur : au delà de $|\mathbf{h}| = a$, le variogramme est constant et égal à son palier $\gamma(\infty)$, et les variables $Z(\mathbf{x})$ et $Z(\mathbf{x} + \mathbf{h})$ sont indépendantes (en toute rigueur, non corrélées). De tels variogrammes sont appelés modèles de transition².

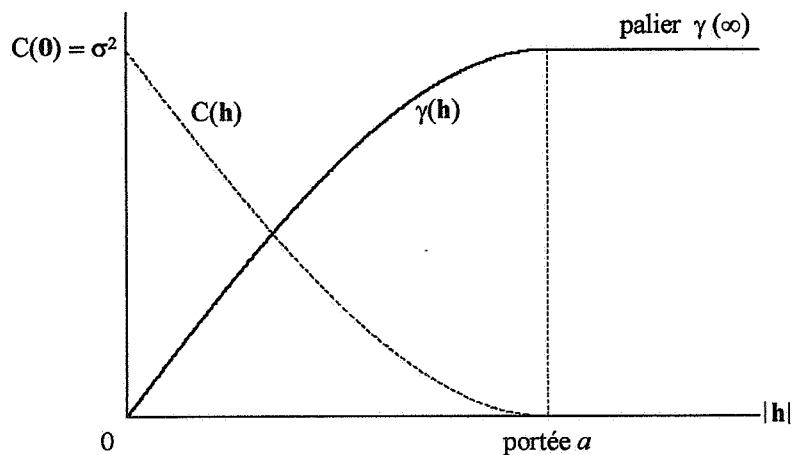


figure 3.2 : modèle de transition

Parfois, le variogramme ne présente ni palier ni portée, et croît indéfiniment lorsque la distance augmente. Une telle situation n'est pas compatible avec l'hypothèse stationnaire : la variance *a priori* est infinie, et la fonction de covariance n'existe pas. La fonction aléatoire correspondante est strictement intrinsèque.

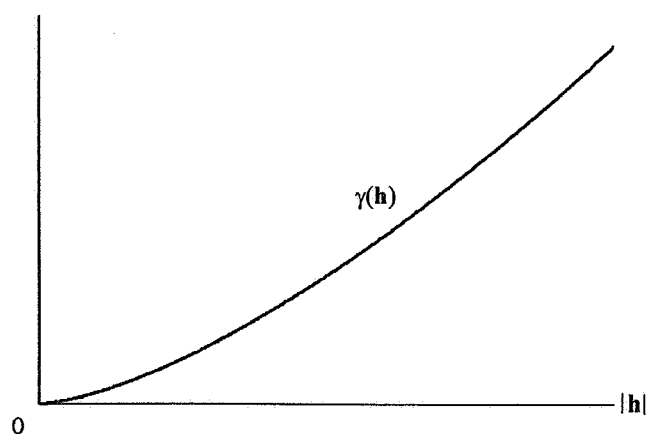


figure 3.3 : variogramme sans palier

² Les fonctions de covariance associées s'identifient à la classe des covariogrammes transitifs, d'où le nom de "modèles de transition".

II.5. Comportement au voisinage de l'origine

Les propriétés analytiques de la fonction aléatoire sont décrites par le comportement de son variogramme au voisinage de l'origine. Cela provient de la relation :

$$\gamma(\mathbf{h}) = \frac{1}{2} E \{ [Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]^2 \}.$$

En règle générale, le comportement à l'origine du variogramme est d'autant plus régulier que la fonction aléatoire – et par suite sa réalisation, la variable régionalisée – est continue et régulière dans l'espace. Plus précisément, on distingue trois grands types de comportements à l'origine :

- **parabolique** : $\gamma(\mathbf{h}) \underset{|\mathbf{h}| \rightarrow 0}{\sim} \omega |\mathbf{h}|^2$, avec ω dépendant éventuellement de l'orientation de $\mathbf{h}$.

Le variogramme est dérivable à l'origine (on montre alors qu'il est dérivable en tout $\mathbf{h}$). La fonction aléatoire est dérivable en moyenne quadratique et la variable régionalisée associée est elle-même très régulière. Ajoutons qu'un comportement à l'origine en $|\mathbf{h}|^\theta$ avec $\theta > 2$ est théoriquement impossible.

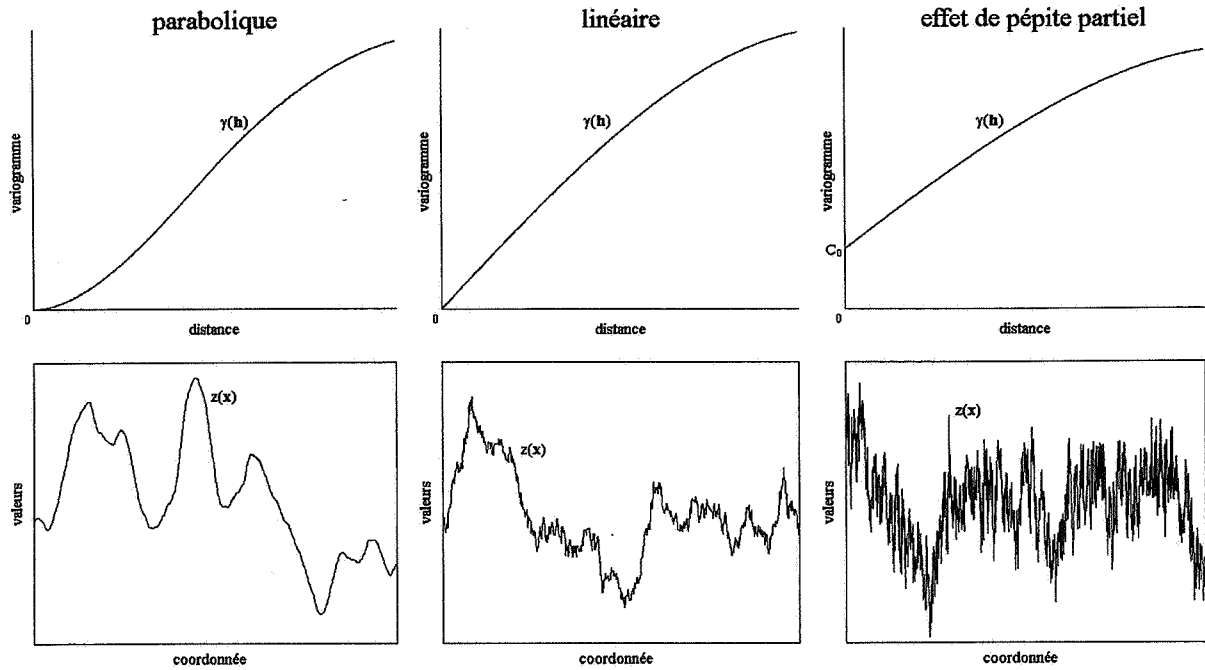
- **linéaire** : $\gamma(\mathbf{h}) \underset{|\mathbf{h}| \rightarrow 0}{\sim} \omega |\mathbf{h}|$.

$\gamma(\mathbf{h})$ est continu en $\mathbf{h} = \mathbf{0}$ mais non dérivable. La fonction aléatoire $Z(\mathbf{x})$ est continue en moyenne quadratique et la variable régionalisée $z(\mathbf{x})$, réalisation particulière de $Z(\mathbf{x})$, est en général continue par morceaux mais non dérivable.

- **discontinu (effet de pépité)** : $\gamma(\mathbf{h}) \xrightarrow{|\mathbf{h}| \rightarrow 0} C_0 > 0$, bien que $\gamma(\mathbf{0}) = 0$.

La dissemblance entre deux points infiniment proches n'est pas négligeable : les valeurs varient à une échelle très petite et leur continuité est imperceptible. Ce phénomène est appelé *effet de pépité*, par référence aux teneurs dans les gisements aurifères, qui changent brusquement lorsqu'il y a des pépites d'or. Il intègre plusieurs facteurs comme les erreurs de mesure, l'absence naturelle de structuration spatiale, ou encore la variabilité à échelle "microscopique" qui n'est pas détectable en raison de la distance entre les échantillons.

Le cas limite est l'**effet de pépité pur** : $\gamma(\mathbf{0}) = 0$ et $\gamma(\mathbf{h}) = C_0$ si $|\mathbf{h}| > 0$. Le variogramme atteint immédiatement son palier : la dissemblance moyenne des valeurs est constante pour toutes les distances $|\mathbf{h}|$. Il y a indépendance (ou plus exactement non corrélation) entre deux points distincts et la géostatistique retrouve tous les résultats de la statistique classique. L'effet de pépité pur correspond à une régionalisation "chaotique", pour laquelle il y a une absence totale de structuration spatiale des valeurs. C'est un cas peu fréquent, à moins que les erreurs de mesure soient très grandes par rapport à l'amplitude des écarts entre valeurs, ou que la distance entre les points d'observation soit supérieure à la portée réelle, auquel cas la structuration spatiale n'est pas décelable et la portée apparente est nulle.



*figure 3.4 : comportement à l'origine du variogramme (au-dessus)
et exemples de variables régionalisées associées (en-dessous)*

III. Estimateurs du variogramme et de la covariance

III.1. Les estimateurs classiques

III.1.1. Estimateur du variogramme

On se pose à présent le problème de l'inférence du variogramme à partir d'un ensemble de données expérimentales $\{z(\mathbf{x}_1), \dots, z(\mathbf{x}_n) / \mathbf{x}_\alpha \in \mathcal{D} \subset \mathbb{R}^d\}$. La variable régionalisée $z(\mathbf{x})$ est considérée comme une réalisation d'une fonction aléatoire $Z(\mathbf{x})$, que l'on suppose **intrinsèque** et sans dérive. On peut alors définir un estimateur du variogramme de la manière suivante :

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} [z(\mathbf{x}_\alpha) - z(\mathbf{x}_\beta)]^2$$

où $N(\mathbf{h}) = \{(\alpha, \beta) \text{ tel que } \mathbf{x}_\alpha - \mathbf{x}_\beta = \mathbf{h}\}$;

$|N(\mathbf{h})|$ est le nombre de paires distinctes de l'ensemble $N(\mathbf{h})$.

Ceci revient à remplacer l'espérance dans l'expression du variogramme

$$\gamma(\mathbf{h}) = \frac{1}{2} E \{ [Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]^2 \}$$

par la moyenne arithmétique sur les paires d'échantillons séparés de $\mathbf{h}$.

L'estimateur ainsi défini porte le nom de **variogramme expérimental**. Il ne s'agit pas à proprement parler d'une fonction, mais d'une suite discrète de points, car il n'est calculable que pour les valeurs de $\mathbf{h}$ telles que $N(\mathbf{h})$ est non vide.

Quand les données sont irrégulièrement espacées dans R^d , le nombre de paires $|N(\mathbf{h})|$ intervenant dans le calcul de $\hat{\gamma}(\mathbf{h})$, pour un vecteur $\mathbf{h}$ donné, est souvent égal à 0 ou 1... Le variogramme expérimental a alors un aspect très erratique et est impossible à interpréter. Pour le rendre plus robuste, on autorise des **tolérances de calcul**, sur les distances et sur les directions :

$$\hat{\gamma}^+(\mathbf{h}) = \frac{1}{2 |N^+(\mathbf{h})|} \sum_{N^+(\mathbf{h})} [z(\mathbf{x}_\alpha) - z(\mathbf{x}_\beta)]^2$$

où $N^+(\mathbf{h}) = \{(\alpha, \beta) \text{ tel que } \mathbf{x}_\alpha - \mathbf{x}_\beta \in T(\mathbf{h})\} = \bigcup_{\mathbf{h}' \in T(\mathbf{h})} N(\mathbf{h}')$;

$T(\mathbf{h})$ est une région de tolérance autour de $\mathbf{h}$, de la forme $[\mathbf{h} - \Delta\mathbf{h}, \mathbf{h} + \Delta\mathbf{h}]$ dans le cas unidimensionnel. Dans le cas multidimensionnel, il existe des tolérances à la fois sur la longueur de $\mathbf{h}$ et sur son orientation, comme l'illustre le schéma ci-dessous.

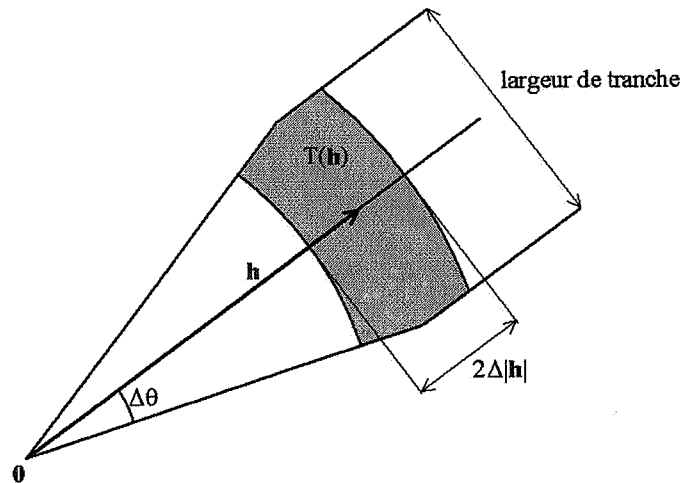


figure 3.5 : région de tolérance $T(\mathbf{h})$ autour du vecteur $\mathbf{h}$ (cas bidimensionnel)

La "largeur de tranche" a pour but de limiter l'écartement du cône de tolérance à une étendue maximale. Dans l'espace à trois dimensions, on introduit deux largeurs de tranche : l'une horizontale et l'autre verticale.

III.1.2. Estimateur de la covariance

Dans le cadre stationnaire, la *méthode des moments*, qui consiste à remplacer les moments de la loi spatiale par les moments empiriques correspondants (ainsi, l'espérance mathématique par la moyenne arithmétique des échantillons), donne l'estimateur de la covariance suivant :

$$\hat{C}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} [z(\mathbf{x}_\alpha) - \bar{z}][z(\mathbf{x}_\beta) - \bar{z}]$$

où $\bar{z} = \frac{1}{n} \sum_{\alpha=1}^n z(\mathbf{x}_\alpha)$ est un estimateur de l'espérance de Z , n étant le nombre d'échantillons.

De la même manière que précédemment, on introduit des tolérances sur les distances et les angles lorsque l'échantillonnage est irrégulier.

On peut remarquer que $\hat{\gamma}(\mathbf{h}) \neq \hat{C}(\mathbf{0}) - \hat{C}(\mathbf{h})$.

III.2. Propriétés des estimateurs

On peut randomiser les estimateurs précédents en remplaçant, dans leur expression, les données expérimentales $\{z(\mathbf{x}_\alpha), \alpha = 1 \dots n\}$ par leur équivalent probabiliste $\{Z(\mathbf{x}_\alpha), \alpha = 1 \dots n\}$. Ceci permet d'étudier la qualité des estimateurs à l'aide d'outils statistiques, comme leur espérance et leur variance, qui vont mesurer respectivement leur biais et leur robustesse.

III.2.1. Biais et robustesse

Si Z est une fonction aléatoire intrinsèque sans dérive, alors $\hat{\gamma}(\mathbf{h})$ est **sans biais** :

$$E[\hat{\gamma}(\mathbf{h})] = \gamma(\mathbf{h}).$$

Un indicateur de la **robustesse** de $\hat{\gamma}(\mathbf{h})$ est sa variance relative

$$\text{var}[\hat{\gamma}(\mathbf{h})] / [\gamma(\mathbf{h})]^2.$$

Plus cette quantité est élevée, plus le variogramme expérimental est susceptible de différer de son espérance $\gamma(\mathbf{h})$, et plus l'inférence statistique est difficile. La variance relative ne peut toutefois être explicitée que dans des cas particuliers, car son calcul requiert la connaissance de la loi spatiale de $Z(\mathbf{x})$ au-delà de ses deux premiers moments.

Plusieurs facteurs peuvent être responsables d'un manque de robustesse du variogramme expérimental, en particulier :

- la distance considérée : la variance relative de $\hat{\gamma}(\mathbf{h})$ peut prendre des valeurs démesurées aux "grandes" distances (pour fixer les idées, celles supérieures à la moitié du diamètre du champ) ;
- les irrégularités d'échantillonnage qui ont parfois tendance à déstructurer le variogramme expérimental, conduisant à un effet de pépite exagéré ;
- une forte dissymétrie dans la distribution des données, qui présente quelques valeurs très élevées ; ces valeurs peuvent avoir un impact considérable dans le calcul du variogramme expérimental, car elles interviennent dans un carré.

Dans ce cas, on peut effectuer une transformation de la variable (par exemple le passage au logarithme lorsque la variable étudiée est positive), de façon à réduire les écarts entre valeurs très élevées et valeurs faibles, et ainsi limiter l'instabilité numérique du variogramme expérimental. Il faut toutefois garder à l'esprit que les résultats obtenus après transformation ne sont pas directement transposables à la variable brute : le variogramme de la variable transformée n'est pas le même que celui de la variable initiale. Sous certaines hypothèses concernant la loi bivariable de la fonction aléatoire, il existe une formule de retour ; par exemple, dans le cas où la transformée logarithmique $Y = \ln(Z)$ est stationnaire et bigaussienne, i.e. si toute paire $(Y(\mathbf{x} + \mathbf{h}), Y(\mathbf{x}))$ suit une loi normale à deux dimensions, on a :

$$\gamma_Z(\mathbf{h}) = m_Z^2 \exp(\sigma_Y^2) \{1 - \exp[-\gamma_Y(\mathbf{h})]\}$$

où $m_Z = E(Z)$ et $\sigma_Y^2 = \text{var}(Y) = \ln[1 + \text{var}(Z) / m_Z^2]$.

La transformation logarithmique amplifie les écarts entre valeurs faibles, et pose problème lorsque la variable peut s'annuler. Pour éviter ces inconvénients, on préfère parfois la "log-translation", qui à Z fait correspondre la variable $Y = \ln(1+Z/\mu)$, où la constante μ représente une valeur de référence pour Z (par exemple la moyenne arithmétique des données sur cette variable). La log-translation atténue l'importance des valeurs très élevées, sans pour autant accentuer l'écart entre valeurs faibles. Si Y est bigaussienne, on établit la relation suivante entre le variogramme de Z et celui de Y :

$$\gamma_Z(\mathbf{h}) = [(m_Z + \mu)^2 + \text{var}(Z)] \{1 - \exp[-\sigma^2 \gamma_Y(\mathbf{h}) / \text{var}(Y)]\}$$

avec $m_Z = E(Z)$ et $\sigma^2 = \ln[1 + \text{var}(Z) / (\mu + m_Z)^2]$.

Des méthodes alternatives ont été proposées pour pallier les problèmes de robustesse du variogramme expérimental :

- changer le mode du calcul du variogramme expérimental : on rencontre, dans certaines publications, des estimateurs qui, au lieu de calculer la *moyenne* des carrés des écarts, considèrent leur *médiane* ou leurs quantiles ; d'autres sont basés sur la moyenne ou la médiane de la *racine carrée* des écarts. L'intérêt de tels estimateurs est d'atténuer l'influence des valeurs extrêmes dans le calcul, mais leur emploi ne se justifie que moyennant des hypothèses sur la loi spatiale de la fonction aléatoire, en particulier une hypothèse de normalité qui est rarement vérifiée (dans le cas contraire, on risque fort d'obtenir un estimateur biaisé). Une approche plus satisfaisante, qui ne requiert aucune hypothèse supplémentaire, consiste à pondérer chaque paire de données dans le calcul, de façon à corriger les effets dus aux irrégularités d'échantillonnage et / ou à la dissymétrie de la distribution des valeurs.
- changer l'outil structural : certains auteurs recommandent l'étude de variogrammes "relatifs" censés améliorer la description des traits structuraux de la régionalisation (portée, anisotropie...) ; pour chaque distance, la valeur du variogramme est normalisée par le carré d'une moyenne expérimentale. Bien que leurs estimateurs soient robustes, ces outils ne peuvent remplacer le variogramme classique dans le formalisme géostatistique et doivent être utilisés dans un but purement descriptif. Une dernière possibilité est de recourir à l'estimateur de la covariance ou de la covariance non centrée (cf. *infra*), plus stable que celui du variogramme classique.

Dans le cadre stationnaire d'ordre deux, l'estimateur de la covariance est plus **robuste** que celui du variogramme, mais est **biaisé**³ :

$$E[\hat{C}(\mathbf{h})] \neq C(\mathbf{h}) .$$

Le biais est d'autant plus important que le nombre d'échantillons est petit. Il provient du recours, dans le calcul de la covariance expérimentale, à une estimation de l'espérance de Z par la moyenne arithmétique des échantillons. Le variogramme expérimental ne possède pas cet inconvénient, car il évite d'estimer l'espérance de Z .

Une alternative plus performante peut être le recours à la **covariance non centrée**, notée C_{nc} , qui, sous l'hypothèse stationnaire d'ordre deux, ne dépend que de la séparation entre échantillons :

$$C_{nc}(\mathbf{h}) = E[Z(\mathbf{x} + \mathbf{h}) Z(\mathbf{x})] = C(\mathbf{h}) + m^2, \text{ où } m = E[Z(\mathbf{x})] .$$

La covariance non centrée est liée au variogramme par la relation :

$$\gamma(\mathbf{h}) = C_{nc}(\mathbf{0}) - C_{nc}(\mathbf{h})$$

et constitue par conséquent un outil structural équivalent à la covariance centrée ou au variogramme (sous réserve de stationnarité). Elle peut être estimée sans biais par :

$$\hat{C}_{nc}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} z(\mathbf{x}_\alpha) z(\mathbf{x}_\beta)$$

où $N(\mathbf{h}) = \{ (\alpha, \beta) \text{ tel que } \mathbf{x}_\alpha - \mathbf{x}_\beta = \mathbf{h} \}$. Comme toujours, dans le cas d'un échantillonnage irrégulier, on introduit des tolérances dans la définition de $N(\mathbf{h})$.

Dans le cadre stationnaire, $\hat{C}_{nc}(\mathbf{0}) - \hat{C}_{nc}(\mathbf{h})$ est un estimateur non biaisé du variogramme $\gamma(\mathbf{h})$. Cet estimateur atténue l'influence des valeurs fortes par rapport au variogramme expérimental classique, puisque les valeurs des données n'interviennent plus dans un carré. Il est particulièrement intéressant dans le cas de données à distribution très dissymétrique, présentant quelques valeurs très fortes qui rendent instable le variogramme expérimental classique. Il est cependant indispensable de vérifier au préalable que l'hypothèse de stationnarité est acceptable (par exemple, en s'assurant que le variogramme expérimental, aussi erratique soit-il, possède un palier).

³ Par exemple, pour $\mathbf{h} = \mathbf{0}$, $\hat{C}(\mathbf{0})$ est l'estimateur classique de la variance *a priori* $C(\mathbf{0}) = \sigma^2$:

$$\hat{C}(\mathbf{0}) = \frac{1}{n} \sum_{\alpha=1}^n [z(\mathbf{x}_\alpha) - \bar{z}]^2 \quad (\text{variance statistique des échantillons})$$

dont l'espérance est, sous l'hypothèse simplificatrice que les données sont non corrélées :

$$E[\hat{C}(\mathbf{0})] = \frac{n-1}{n} \sigma^2 = \left(1 - \frac{1}{n}\right) C(\mathbf{0}) .$$

Sous l'hypothèse de non corrélation des données, on a un biais égal à $C(\mathbf{0})/n$ pour l'estimateur de la variance *a priori* $C(\mathbf{0})$. En réalité, comme le phénomène étudié manifeste une structure, les données sont corrélées et le biais est encore plus important, ce qui peut devenir problématique.

III.2.2. Nécessité d'un contrôle des hypothèses

L'estimateur classique de la covariance, qui peut toujours être calculé sur un ensemble de données, a tendance à rendre plausible l'hypothèse stationnaire, même lorsqu'elle n'est pas acceptable ; la covariance expérimentale est alors totalement artificielle et ne reflète pas les traits structuraux du phénomène régionalisé. Il convient donc de contrôler au préalable la validité de l'hypothèse stationnaire, au moyen d'une étude exploratoire des données.

Le variogramme expérimental peut au contraire être utilisé hors du cadre stationnaire (du moins tant que l'on a affaire à une fonction aléatoire intrinsèque sans dérive) et constitue un outil plus fiable que la covariance expérimentale. En présence d'une dérive, il est biaisé, mais on montre que le biais est faible pour les petites distances (*cf.* chapitre 6).

IV. Principaux modèles de variogramme

Nous avons vu qu'une fonction est un variogramme si et seulement si elle est de type négatif conditionnel. C'est une propriété très contraignante et difficile à contrôler, même lorsque l'on travaille dans un espace unidimensionnel. C'est pourquoi, on choisit souvent un modèle de variogramme parmi les fonctions dont on sait qu'elles sont de type négatif conditionnel. Nous allons donner ci-dessous quelques exemples de fonctions ayant cette propriété, et qui sont les modèles les plus couramment utilisés en pratique. Sauf mention contraire, les modèles de variogramme sont valides dans un espace de dimension quelconque.

Les deux principales caractéristiques d'un variogramme sont :

- son comportement à l'origine, qui traduit le degré de régularité de la régionalisation ;
- la présence ou l'absence d'un palier, i.e. $\gamma(\mathbf{h}) = \gamma(\infty) = \text{constante}$ quand $|\mathbf{h}| > a$. Rappelons que l'existence d'un palier est synonyme de stationnarité d'ordre deux ; il existe alors une fonction de covariance qui se déduit du variogramme par la relation

$$C(\mathbf{h}) = \gamma(\infty) - \gamma(\mathbf{h}).$$

On ne s'intéresse ici qu'aux modèles théoriques **isotropes**, c'est-à-dire qui ne dépendent que de $r = |\mathbf{h}|$. Ils peuvent être classés en trois catégories :

- les modèles avec palier et avec un comportement à l'origine
 - discontinu (effet de pépité pur)
 - linéaire (modèles sphérique, exponentiel...)
 - parabolique (modèles gaussien, cubique...)
- les modèles avec palier et effet de trou : sinus cardinal, modèle J de Bessel...
- les modèles sans palier : modèles puissance, linéaire et logarithmique.

IV.1. Modèles avec palier ou modèles de transition

A ces variogrammes correspondent des modèles de covariance, donnés par la relation

$$\gamma(r) = C(0) - C(r).$$

IV.1.1. Comportement discontinu à l'origine

modèle pépitique de palier C

$$\gamma(r) = \begin{cases} 0 & \text{pour } r = 0 \\ C & \text{pour } r > 0 \end{cases}$$

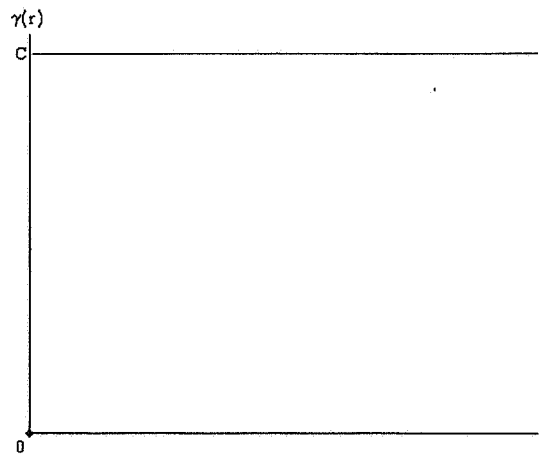


figure 3.6 : modèle pépitique pur

Ce modèle traduit une absence de structuration spatiale, due à des erreurs de mesure, ou à la présence d'une microstructure indétectable expérimentalement.

IV.1.2. Comportement linéaire à l'origine

modèle sphérique de portée a et de palier C (valide dans $\mathbb{R}^d$, $d \leq 3$)

$$\gamma(r) = \begin{cases} C \left(\frac{3}{2} \frac{r}{a} - \frac{1}{2} \frac{r^3}{a^3} \right) & \text{pour } 0 \leq r \leq a \\ C & \text{pour } r \geq a \end{cases}$$

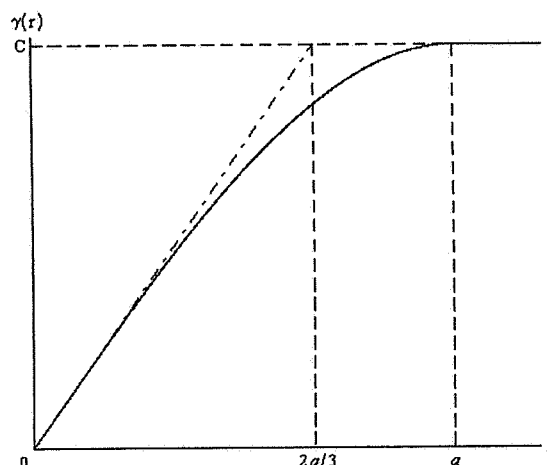


figure 3.7 : modèle sphérique

L'effet de pépite peut être vu comme un cas particulier d'un schéma sphérique de portée infiniment petite. Du point de vue physique, il existe néanmoins une différence fondamentale entre les modèles pépétique et sphérique : le premier représente une régionalisation discontinue à l'échelle d'observation, pour laquelle les valeurs changent brusquement d'un point à l'autre, tandis que le second décrit une régionalisation continue.

modèle exponentiel de paramètre a et de palier C

$$\gamma(r) = C \left(1 - \exp \left(-\frac{r}{a} \right) \right).$$

Contrairement au modèle sphérique qui atteint son palier pour $r = a$, le modèle exponentiel n'atteint son palier qu'asymptotiquement. On peut toutefois définir une portée pratique égale à $3a$ pour laquelle le variogramme atteint 95% de la valeur de son palier.

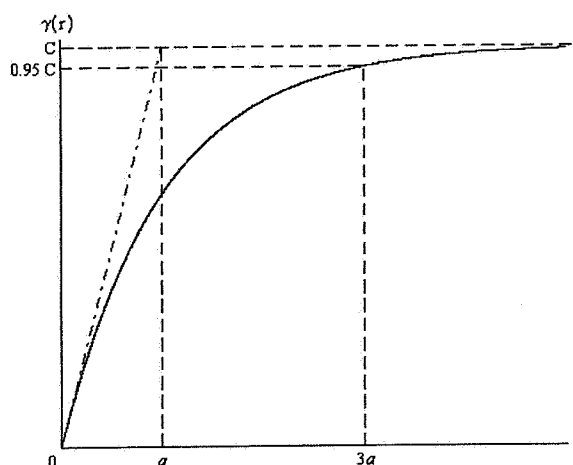


figure 3.8 : modèle exponentiel

IV.1.3. Comportement parabolique à l'origine

modèle cubique de portée a et de palier C (valide dans $\mathbb{R}^d$, $d \leq 3$)

$$\gamma(r) = \begin{cases} C \left(7 \frac{r^2}{a^2} - \frac{35}{4} \frac{r^3}{a^3} + \frac{7}{2} \frac{r^5}{a^5} - \frac{3}{4} \frac{r^7}{a^7} \right) & \text{pour } 0 \leq r \leq a \\ C & \text{pour } r \geq a \end{cases}$$

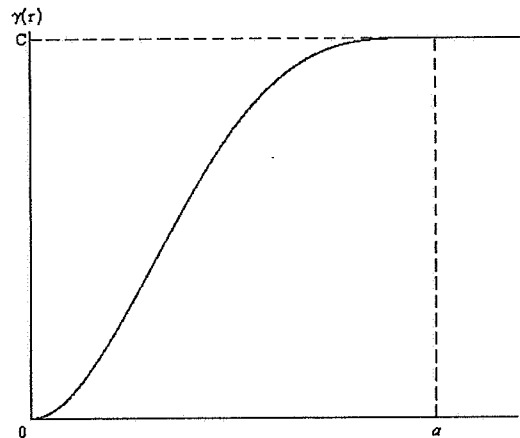


figure 3.9 : modèle cubique

modèle gaussien de paramètre a et de palier C

$$\gamma(r) = C \left(1 - \exp \left(- \frac{r^2}{a^2} \right) \right).$$

Le palier est atteint asymptotiquement et la portée pratique peut être prise égale à $a\sqrt{3}$.

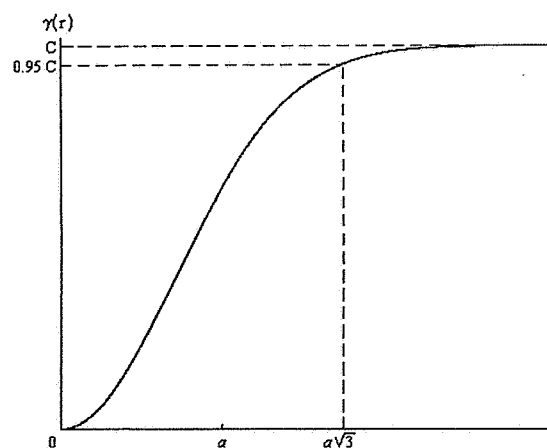


figure 3.10 : modèle gaussien

Même si ce variogramme semble proche du variogramme cubique (à un facteur d'échelle près sur la portée), les fonctions aléatoires associées sont assez différentes : le modèle cubique caractérise des fonctions aléatoires dérivables en moyenne quadratique, tandis que le modèle gaussien correspond à des fonctions aléatoires *infiniment* différentiables.

IV.1.4. Autres modèles de transition

Nous présentons ci-dessous d'autres modèles, de palier C et portée a , d'usage moins fréquent :

- le **modèle triangulaire** (valide seulement dans l'espace $\mathbb{R}$)

$$\gamma(r) = \begin{cases} C r / a & \text{pour } 0 \leq r \leq a \\ C & \text{pour } r \geq a \end{cases}$$

La covariance associée est égale, à un facteur multiplicatif près, au covariogramme géométrique du segment de longueur a dans l'espace $\mathbb{R}$. Plus généralement, un modèle autorisé dans $\mathbb{R}$ est

$$\gamma(r) = \begin{cases} C r^\alpha / a & \text{pour } 0 \leq r \leq a \\ C & \text{pour } r \geq a \end{cases}$$

avec $0 < \alpha \leq 1$.

- le **modèle quadratique** (valide dans $\mathbb{R}^d$, $d \leq 3$)

$$\gamma(r) = \begin{cases} C \left(2 \frac{r}{a} - \frac{r^2}{a^2} \right) & \text{pour } 0 \leq r \leq a \\ C & \text{pour } r \geq a \end{cases}$$

- le **modèle circulaire** (valide dans $\mathbb{R}^d$, $d \leq 2$)

$$\gamma(r) = \begin{cases} C \left(\frac{2r}{\pi a} \sqrt{1 - \frac{r^2}{a^2}} + \frac{2}{\pi} \arcsin \left(\frac{r}{a} \right) \right) & \text{pour } 0 \leq r \leq a \\ C & \text{pour } r \geq a \end{cases}$$

La covariance associée est égale, à un facteur multiplicatif près, au covariogramme géométrique du cercle de diamètre a dans l'espace $\mathbb{R}^2$.

- le **modèle pentasphérique** (valide dans $\mathbb{R}^d$, $d \leq 5$)

$$\gamma(r) = \begin{cases} C \left(\frac{15r}{8a} - \frac{5r^3}{4a^3} + \frac{3r^5}{8a^5} \right) & \text{pour } 0 \leq r \leq a \\ C & \text{pour } r \geq a \end{cases}$$

La covariance associée est égale, à un facteur multiplicatif près, au covariogramme géométrique de l'hypersphère de diamètre a dans l'espace $\mathbb{R}^5$.

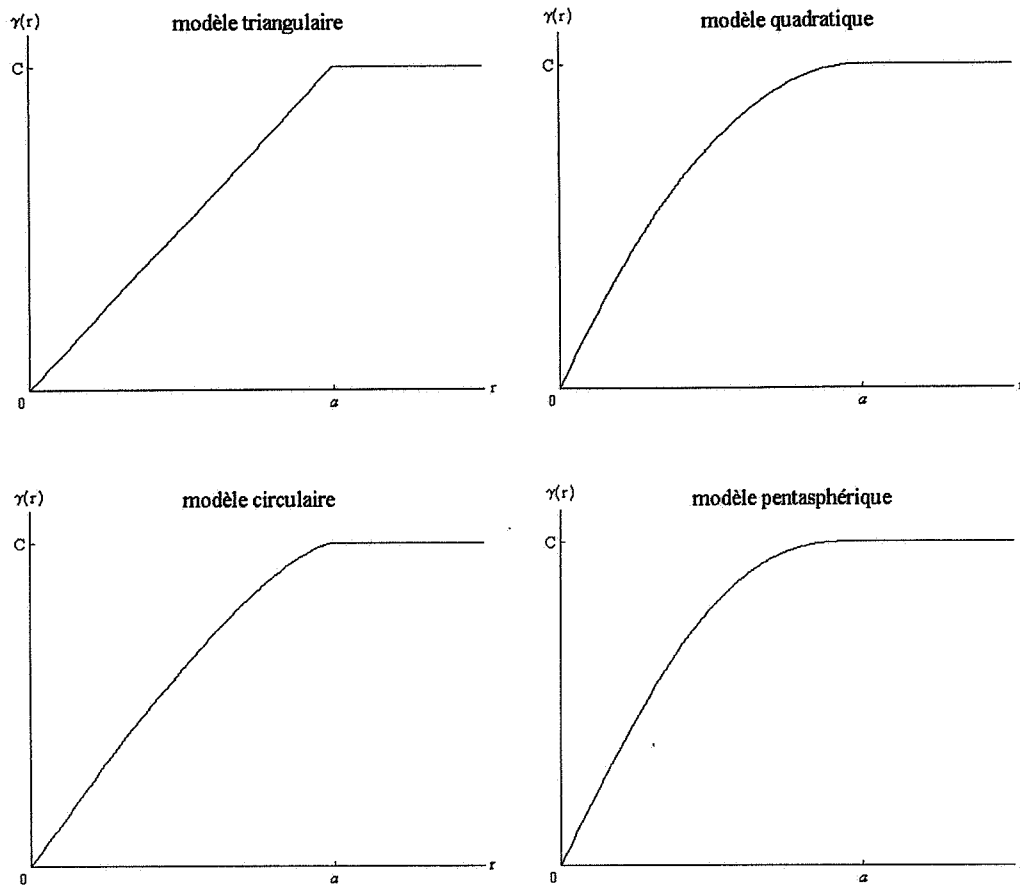


figure 3.11 : modèles de transition triangulaire, quadratique, circulaire et pentasphérique

Les modèles précédents utilisent deux paramètres : le *palier* C et un facteur d'échelle a lié à la portée. D'autres modèles de transition font appel à un troisième paramètre α :

- le **modèle stable** de palier C , facteur d'échelle a et paramètre α :

$$\gamma(r) = C \{ 1 - \exp[-(r/a)^\alpha] \} \quad \text{avec } 0 < \alpha \leq 2 \quad (\text{portée pratique : } a\sqrt[3]{3}).$$

Les modèles exponentiel et gaussien sont en fait des cas particuliers de ce modèle, correspondant aux paramètres $\alpha = 1$ et $\alpha = 2$.

- le **modèle gamma** de palier C , facteur d'échelle a et paramètre α

$$\gamma(r) = C \left\{ 1 - \frac{1}{(1 + r/a)^\alpha} \right\} \quad \text{avec } \alpha > 0 \quad (\text{portée pratique : } a(\sqrt[3]{20} - 1)).$$

Le modèle gamma de paramètre $\alpha = 1$ porte le nom de *modèle hyperbolique*.

- le **modèle de Cauchy** de palier C , facteur d'échelle a et paramètre α

$$\gamma(r) = C \left\{ 1 - \frac{1}{(1 + r^2/a^2)^\alpha} \right\} \quad \text{avec } \alpha > 0 \quad (\text{portée pratique : } a\sqrt[3]{20} - 1).$$

- le modèle de Bessel modifié de palier C, facteur d'échelle a et paramètre α

$$\gamma(r) = C \left\{ 1 - \frac{1}{2^{\alpha-1} \Gamma(\alpha)} (r/a)^{\alpha} K_{\alpha}(r/a) \right\} \quad \text{avec } \alpha > 0$$

où Γ est la fonction d'Euler interpolant la factorielle

K_{α} la fonction de Bessel modifiée de deuxième espèce d'ordre α , définie par :

$$K_{\alpha}(u) = \frac{\pi}{2 \sin(\alpha\pi)} \left\{ \sum_{k=0}^{\infty} \frac{1}{k! \Gamma(-\alpha + k + 1)} \left(\frac{u}{2}\right)^{2k-\alpha} - \sum_{k=0}^{\infty} \frac{1}{k! \Gamma(\alpha + k + 1)} \left(\frac{u}{2}\right)^{2k+\alpha} \right\}$$

(si α est entier, on obtient K_{α} par passage à la limite).

Lorsque α est un demi-entier, le variogramme est un polynôme multiplié par une exponentielle ; le cas $\alpha = 1/2$ correspond au modèle exponentiel.

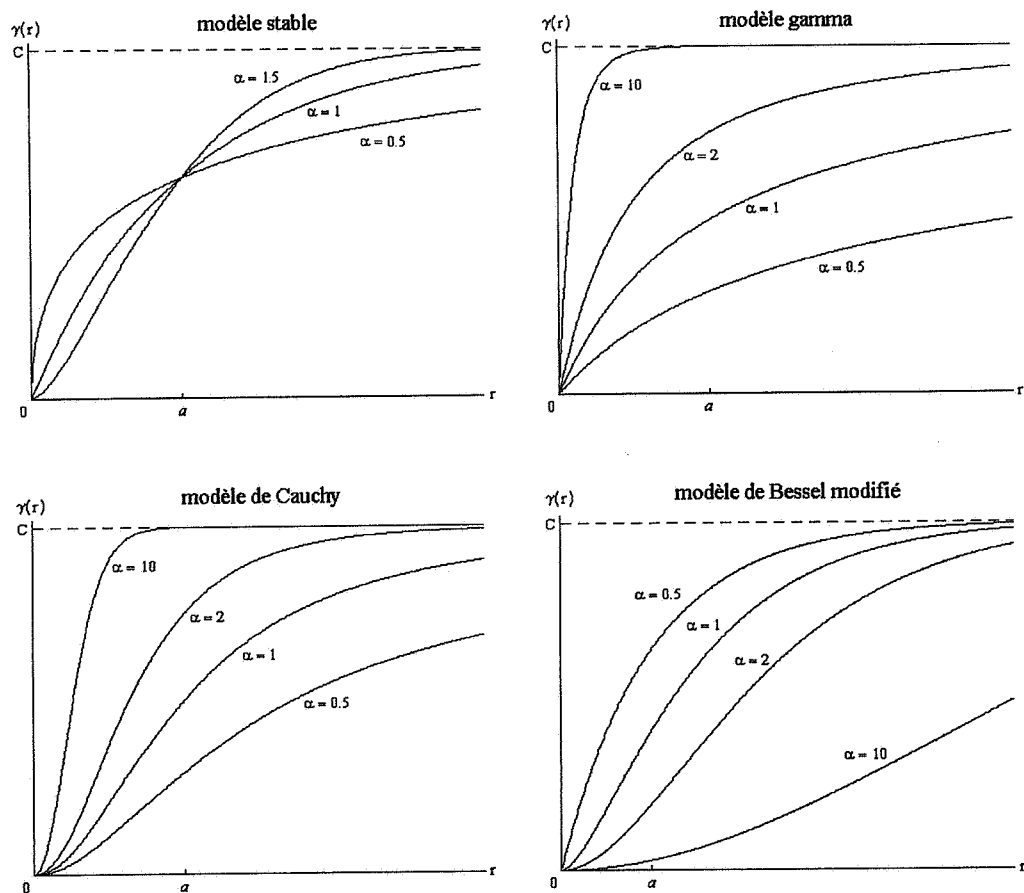


figure 3.12 : autres modèles de transition

Ces modèles couvrent une large gamme de fonctions lorsqu'on fait varier le paramètre α .

D'autres modèles de transition sont les modèles à effet de trou, que nous allons détailler à présent.

IV.2. Modèles à effet de trou

L'effet de trou se manifeste lorsque le variogramme n'est pas monotone, mais présente une ou plusieurs oscillations. Ces oscillations ont en général une interprétation physique, qu'il convient de mettre en évidence (phénomène périodique "amorti" par exemple).

modèle cosinus de période a et amplitude $2C$ (valide seulement dans $\mathbb{R}$)

$$\gamma(r) = C \{1 - \cos(2\pi r / a)\}.$$

Un tel variogramme oscille indéfiniment ; il ne possède ni portée ni portée pratique. La fonction aléatoire associée est une sinusoïde parfaite de période a ; seule sa phase est aléatoire.

modèle cosinus amorti de paramètres a , b , α et palier C (valide dans $\mathbb{R}$)

$$\gamma(r) = C \{1 - \exp(-br^\alpha) \cos(2\pi r / a)\} \text{ avec } 0 < \alpha \leq 2.$$

modèle sinus cardinal de paramètre a et palier C (valide dans $\mathbb{R}^d$, $d \leq 3$)

$$\gamma(r) = C \left\{ 1 - \frac{\sin(r/a)}{r/a} \right\}$$

A titre d'information, la portée pratique est égale à $20.37a$ et la demi pseudo-période à $4.49a$, distance pour laquelle variogramme vaut $1.21 C$. Le quotient entre cette valeur et le palier (i.e. 1.21), qui mesure l'amplitude de l'effet de trou, est le maximum que l'on puisse obtenir par un modèle isotrope à trois dimensions.

modèle J de Bessel de paramètres a et α et palier C (valide dans $\mathbb{R}^d$, $d \leq 2(\alpha+1)$)

$$\gamma(r) = C \left\{ 1 - \left(\frac{r}{2a} \right)^{-\alpha} \Gamma(\alpha+1) J_\alpha \left(\frac{r}{a} \right) \right\}$$

où Γ est la fonction d'Euler interpolant la factorielle

J_α est la fonction de Bessel de première espèce d'ordre α , définie par :

$$J_\alpha(u) = \left(\frac{u}{2} \right)^\alpha \sum_{k=0}^{\infty} \frac{(-1)^k}{k! \Gamma(\alpha+k+1)} \left(\frac{u}{2} \right)^{2k}.$$

Les modèles cosinus et sinus cardinal sont en fait des cas particuliers de ce modèle ; ils correspondent aux paramètres $\alpha = -1/2$ et $\alpha = 1/2$.

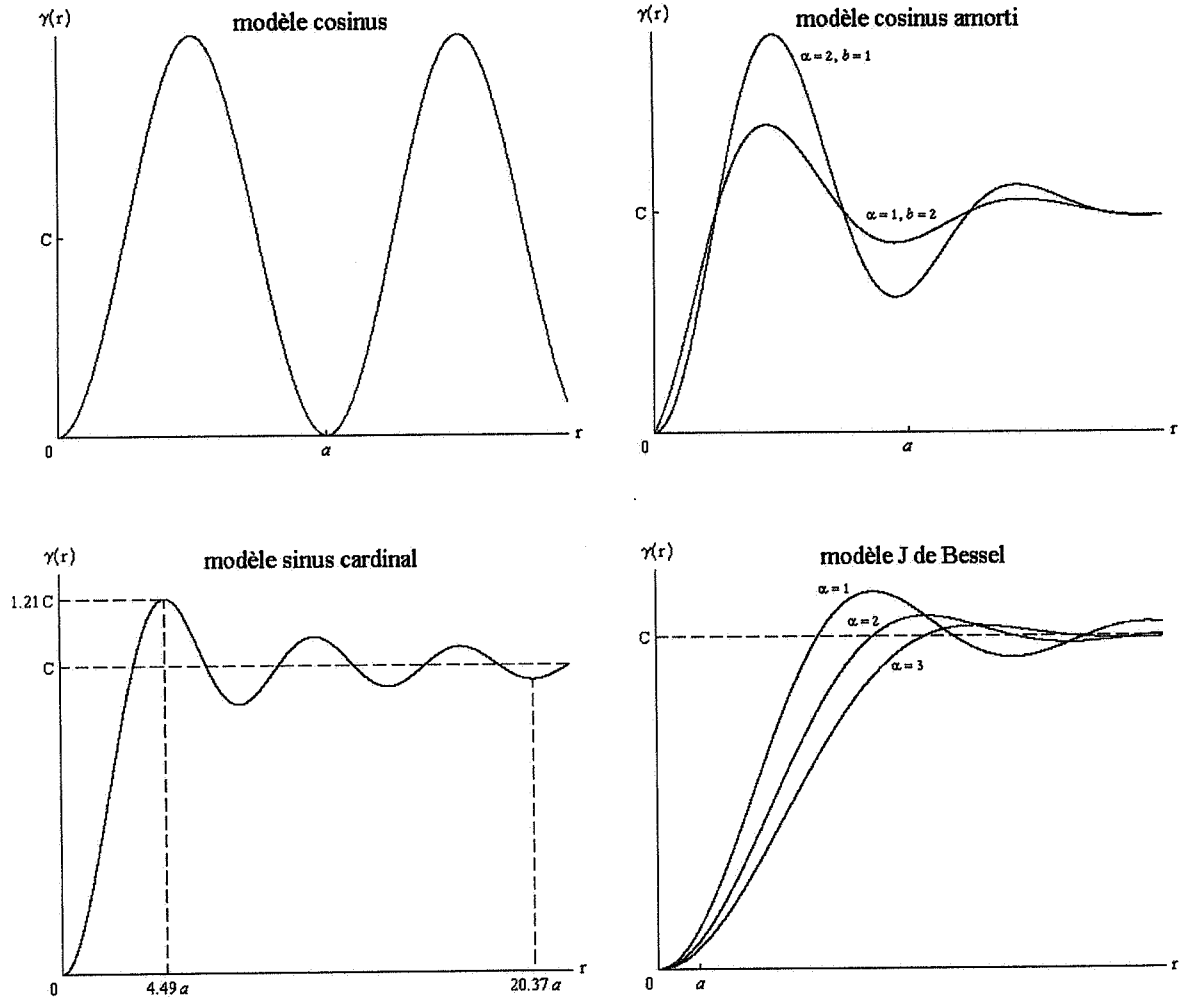


figure 3.13 : modèles à effet de trou périodiques et pseudo-périodiques

Les modèles ci-dessus sont périodiques ou pseudo-périodiques et leur comportement à l'origine est parabolique. Il existe d'autres modèles à effet de trou, non périodiques et dont le comportement à l'origine est linéaire :

modèles polynomiaux tronqués, de portée a et palier C (valides dans $\mathbb{R}^d$, $d \leq 3$)

$$\gamma(r) = \begin{cases} C \left(\frac{5}{2} \frac{r}{a} - \frac{5}{2} \frac{r^3}{a^3} + \frac{r^5}{a^5} \right) & \text{pour } 0 \leq r \leq a \\ C & \text{pour } r \geq a \end{cases}$$

$$\gamma(r) = \begin{cases} C \left(\frac{35}{12} \frac{r}{a} - \frac{35}{8} \frac{r^3}{a^3} + \frac{7}{2} \frac{r^5}{a^5} - \frac{25}{24} \frac{r^7}{a^7} \right) & \text{pour } 0 \leq r \leq a \\ C & \text{pour } r \geq a \end{cases}$$

$$\gamma(r) = \begin{cases} C \left(\frac{7}{2} \frac{r}{a} - 7 \frac{r^3}{a^3} + 7 \frac{r^5}{a^5} - \frac{5}{2} \frac{r^7}{a^7} \right) & \text{pour } 0 \leq r \leq a \\ C & \text{pour } r \geq a \end{cases}$$

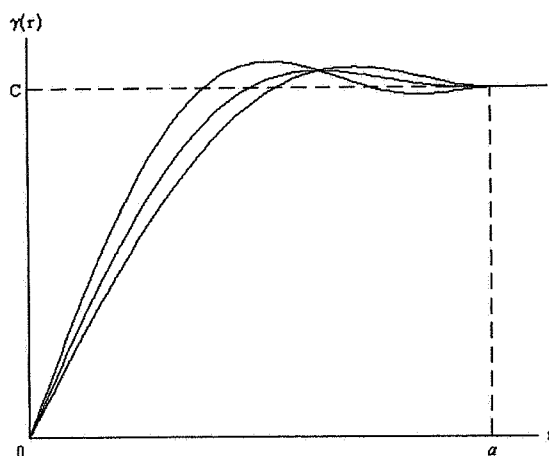


figure 3.14 : modèles polynomiaux tronqués

modèle stable généralisé, de paramètres a , α , β et de palier C (valides dans $\mathbb{R}^d$, $d \leq \alpha / \beta$)

$$\gamma(r) = C \{ 1 - [1 - \beta(r/a)^\alpha] \exp[-(r/a)^\alpha] \} \text{ avec } 0 < \alpha \leq 2 \text{ et } 0 \leq \beta \leq \alpha.$$

Plus les paramètres α et β sont proches de 2 et α , plus l'effet de trou est prononcé ; lorsque $\beta = 0$, on retrouve le modèle stable. Le paramètre α permet de régler le comportement à l'origine (en r^α), le paramètre β l'amplitude de l'effet de trou (lorsque α est fixé), tandis que a est un facteur d'échelle lié à la portée pratique du modèle.

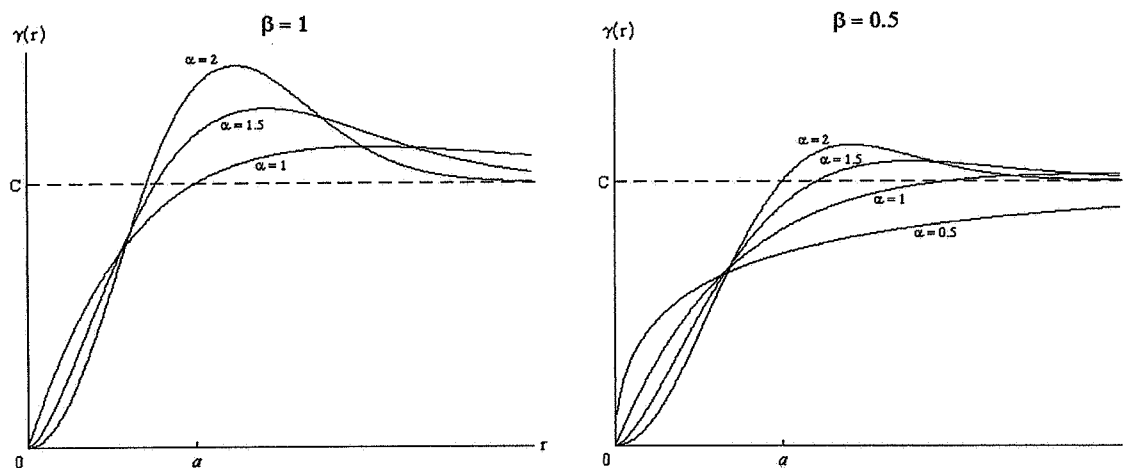


figure 3.15 : modèle stable généralisé

IV.3. Modèles sans palier

Ces modèles débordent du cadre stationnaire d'ordre deux. Ils correspondent à des fonctions aléatoires intrinsèques strictes.

modèle puissance

$$\gamma(r) = \omega r^\theta \quad \text{avec } 0 < \theta < 2.$$

L'exposant θ est lié au degré de continuité de la variable régionalisée. A la limite où θ tend vers 0, le variogramme est purement pépétique. Au contraire, lorsque θ croît et tend vers 2, le comportement à l'origine devient proche d'un comportement parabolique, ce qui indique que la fonction aléatoire $Z(\mathbf{x})$ devient de plus en plus régulière⁴.

A titre d'exemple, le mouvement brownien⁵ fractionnaire de paramètre $\eta \in]0,1[$ est un processus aléatoire dans $\mathbb{R}$ qui a pour variogramme le modèle puissance $\gamma(r) = r^{2\eta}$.

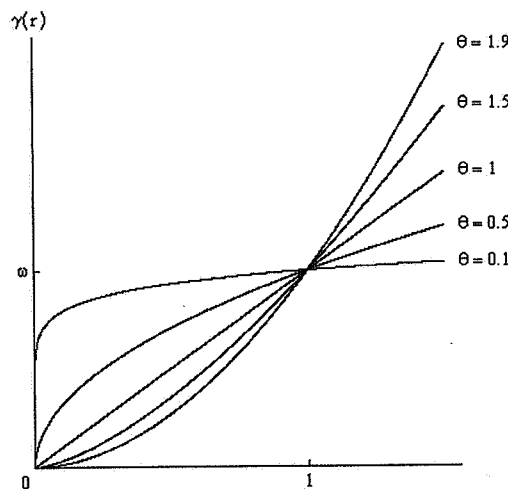


figure 3.16 : modèle puissance

⁴ Si θ est égal à 2, la fonction aléatoire $Z(\mathbf{x})$ se réduit à une dérive linéaire, de la forme $Z(\mathbf{x}) = Z_0 + \langle \mathbf{U} | \mathbf{x} \rangle$, où $\mathbf{U}$ est un vecteur aléatoire de l'espace et Z_0 une variable aléatoire. Cette situation ne présente pas d'intérêt pratique ; en outre, elle pose des problèmes d'ergodicité, puisque le variogramme *expérimental* d'une réalisation $z(\mathbf{x}) = z_0 + \langle \mathbf{u} | \mathbf{x} \rangle$ sera parabolique le long de la direction parallèle à $\mathbf{u}$, mais nul le long des directions orthogonales. C'est la raison pour laquelle la borne $\theta = 2$ n'est pas considérée.

⁵ Le mouvement brownien est un processus aléatoire à accroissements gaussiens stationnaires. Une de ses propriétés est l'invariance par homothétie de ses caractéristiques stochastiques, qui ne dépendent donc pas de l'échelle d'observation. Une autre propriété remarquable est le caractère markovien : si l'on se place le long d'une droite de l'espace sur laquelle on connaît la valeur en un point (pris comme origine), les variations *futures* sont indépendantes des variations *passées*. Pour introduire des corrélations entre ces variations, a été inventé le mouvement brownien fractionnaire, avec un paramètre supplémentaire $\eta \in [0,1]$. Le mouvement brownien classique correspond à $\eta = 1/2$; un paramètre $\eta > 1/2$ indique un phénomène plus régulier, tandis que $\eta < 1/2$ décrit un mouvement très irrégulier, qui a tendance à faire incessamment demi-tour.

modèle linéaire

$$\gamma(r) = \omega r$$

Il s'agit d'un cas particulier du modèle puissance, avec un exposant égal à 1.

modèle logarithmique

$$\gamma(r) = 3\alpha \ln(r)$$

où α est un scalaire positif appelé *dispersion absolue*.

Le modèle logarithmique est très particulier, car il vaut $-\infty$ en $r = 0$, ce qui "contredit" la définition même d'un variogramme. En fait, il ne se réfère pas à une fonction aléatoire, mais à une distribution aléatoire. Or, les régularisées de cette distribution, c'est-à-dire les valeurs moyennes sur des supports non ponctuels, deviennent des fonctions aléatoires. Ceci signifie que le modèle logarithmique n'est utilisable que pour décrire des variables non ponctuelles (physiquement, aucune mesure n'est rigoureusement ponctuelle et s'effectue toujours sur un support plus ou moins volumineux)⁶. On se référera au chapitre 4 pour des détails sur le procédé de régularisation.

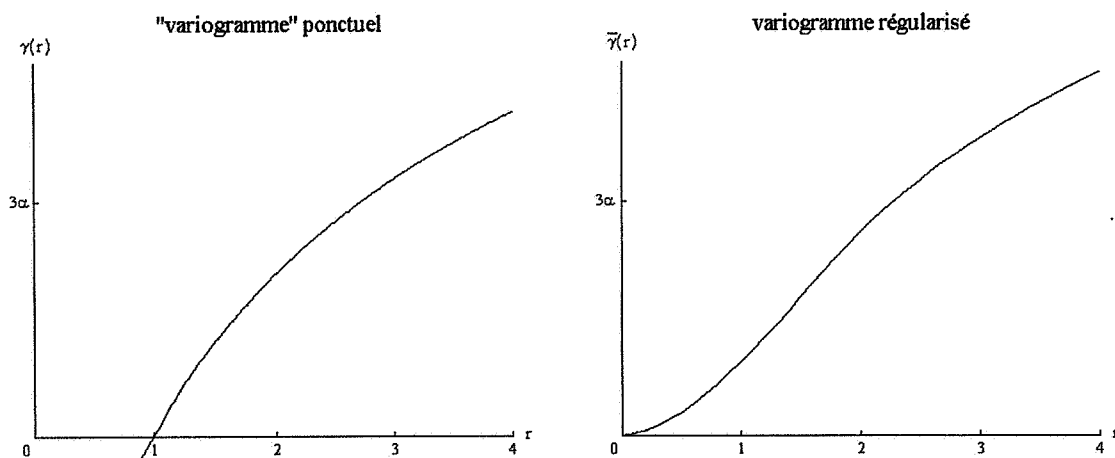


figure 3.17 : modèle logarithmique

Ce modèle, utilisé historiquement pour la description des gisements aurifères sud-africains, est tombé en désuétude.

⁶ Même si elle est plus abstraite, la théorie des distributions aléatoires est sans doute plus adaptée que celles des fonctions aléatoires lorsqu'une description ponctuelle du phénomène régionalisé est abusive.

V. Analyse variographique : ajustement d'un modèle à un variogramme expérimental

Nous allons nous intéresser uniquement au variogramme, mais tout ce qui suit peut être transposé, lorsque l'hypothèse de stationnarité du second ordre est vérifiée, aux fonctions de covariance. Rappelons que le variogramme fournit un cadre d'étude plus général, car il peut être défini hors du contexte stationnaire d'ordre deux.

Le variogramme expérimental ne peut être utilisé directement. D'une part, il n'est défini que pour certaines distances et est donc "incomplet". D'autre part, il n'a aucune raison d'être de type négatif conditionnel et, en l'absence de cette propriété, les variances des combinaisons linéaires de $Z(x)$ ont des chances d'être négatives. L'idée est alors de rechercher un modèle théorique de variogramme qui soit proche du variogramme expérimental. Cette étape de modélisation, appelée analyse structurale ou **analyse variographique**, est la phase essentielle d'une étude géostatistique, une mauvaise modélisation ne pouvant qu'entraîner de mauvais résultats.

Avant de donner des conseils pour l'ajustement d'un modèle, nous allons voir que les modèles de base proposés ci-dessus, en dépit de leur simplicité et de leur faible nombre, sont capables de rendre compte de la plupart des situations pratiques. Ainsi, nous verrons comment ils peuvent être combinés de façon à intégrer différents niveaux d'échelle (schémas gigognes). Nous étudierons aussi comment le variogramme prend en compte la variabilité à courte distance (effet de pépité). Enfin, nous terminerons en montrant les principaux types d'anisotropie que l'on peut rencontrer et dans quelle mesure il est possible de se ramener au cas isotrope.

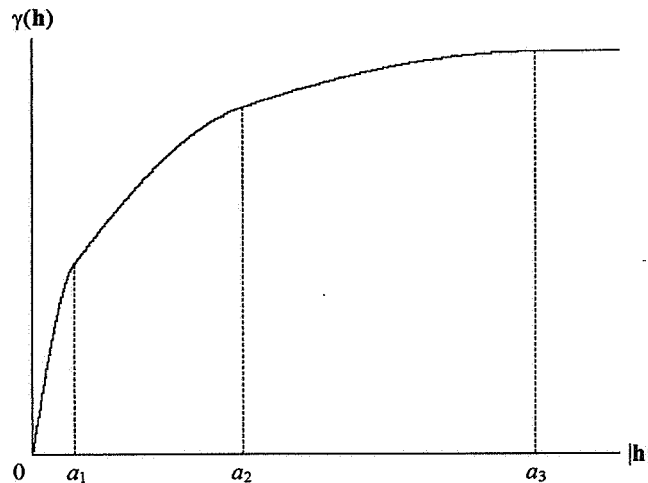
V.1. Le modèle linéaire de régionalisation

Le variogramme expérimental mesure la dissemblance moyenne entre deux observations en fonction de leur séparation. En pratique, il présente souvent des changements de pente, plus ou moins marqués, qui indiquent le passage à une structuration différente des valeurs à partir de certaines distances.

On peut ainsi mettre en évidence différentes échelles de variation dans la régionalisation. Le variogramme théorique peut être vu comme la superposition de plusieurs schémas élémentaires, appelés aussi **structures gigognes**. Chaque échelle d'observation intègre toutes les structures des niveaux inférieurs. Par exemple, si l'on travaille à l'échelle du mètre, le variogramme incorporera la variabilité agissant aux échelles millimétrique, centimétrique,... mais la variabilité kilométrique n'apparaîtra pas. Il y a donc une hiérarchie entre les structures, qui s'emboîtent les unes dans les autres, d'où le qualificatif de *gigognes*.

L'analyse variographique d'un schéma gigogne fait appel à un modèle d'usage courant, le **modèle linéaire de régionalisation**, qui représente le variogramme $\gamma(\mathbf{h})$ par une somme de variogrammes de base :

$$\gamma(\mathbf{h}) = \gamma_1(\mathbf{h}) + \gamma_2(\mathbf{h}) + \dots + \gamma_s(\mathbf{h}) .$$



*figure 3.18 : exemple de schéma gigogne
constitué de trois modèles sphériques emboîtés*

Idéalement, chaque structure de base caractérise la variabilité à une échelle spatiale particulière. La fonction aléatoire peut être vue comme la somme de processus indépendants (ou du moins à accroissements spatialement non corrélés), dont les variogrammes sont les différentes structures reconnues dans $\gamma(\mathbf{h})$ ⁷ :

$$Z(\mathbf{x}) = Z_1(\mathbf{x}) + Z_2(\mathbf{x}) + \dots + Z_s(\mathbf{x}) .$$

Ainsi, le phénomène régionalisé (ou plus exactement, sa représentation mathématique, la fonction aléatoire) apparaît comme la superposition de composantes indépendantes agissant simultanément, mais à des échelles différentes. Nous reviendrons sur le modèle linéaire de régionalisation et son interprétation lors de l'étude des techniques d'analyse krigéante (en géostatistique multivariable, cf. chapitre 7 et annexe G).

Il faut noter qu'en pratique la décomposition précédente n'est pas unique. Les différentes échelles de variabilité d'un phénomène réel sont rarement aussi différenciées. En outre, il est préférable, mais non indispensable, que les composantes d'un schéma gigogne aient une signification physique. Très souvent cependant, l'emploi des schémas gigognes ne représente qu'une commodité d'ajustement. Par ailleurs, il est clair qu'une composante à petite échelle ne pourra être identifiée que si la maille d'échantillonnage est suffisamment fine. De même, une composante à grande échelle n'apparaîtra sur le variogramme que si le diamètre du domaine échantillonné est assez important.

⁷ Une telle interprétation n'est pas possible au niveau du covariogramme transitif (géostatistique transitive), car celui-ci intègre des propriétés géométriques du champ qui interfèrent avec celles de la régionalisation.

V.2. L'effet de pépité et sa genèse

L'effet de pépité correspond à un phénomène qui ne présente aucune "structure" spatiale. Il est symptomatique soit du fait que les mesures sont entachées d'erreurs, soit de la présence d'une *microstructure*, c'est-à-dire d'une structure gigogne atteignant très rapidement sa portée à l'échelle d'observation considérée.

Erreurs de mesure

Si les données $\{Z(\mathbf{x}_\alpha), \alpha = 1 \dots n\}$ sont affectées d'erreurs $\{\varepsilon(\mathbf{x}_\alpha), \alpha = 1 \dots n\}$, non corrélées entre elles ni avec les valeurs de $Z(\mathbf{x})$, de moyenne nulle et de variance C_0 , le variogramme de la variable "bruitée" $Z(\mathbf{x}) + \varepsilon(\mathbf{x})$ diffère de celui de la variable sous-jacente $Z(\mathbf{x})$ par un effet de pépité d'amplitude C_0 :

$$\gamma_{Z+\varepsilon}(\mathbf{h}) = \begin{cases} 0 & \text{si } \mathbf{h} = \mathbf{0} \\ \gamma_Z(\mathbf{h}) + C_0 & \text{sinon} \end{cases}$$

La présence d'un effet de pépité peut aussi être due à des erreurs, non pas sur les valeurs mesurées, mais sur les localisations des sites d'observation. Nous reviendrons sur cette particularité lors de l'étude du *krigeage aléatoire* (fin du chapitre 5).

Présence d'une microstructure

Dans le modèle linéaire de régionalisation, la notion d'échelle est essentielle. Considérons par exemple un schéma gigogne composé de deux structures : la première de portée a centimétrique et de palier C , la seconde de portée kilométrique. Le variogramme $\gamma(\mathbf{h})$ présente alors au voisinage de l'origine une zone de croissance atteignant très rapidement (au bout de quelques centimètres) le palier C . A l'échelle kilométrique, cette croissance se confondra avec une discontinuité à l'origine, c'est-à-dire un effet de pépité, d'amplitude C .

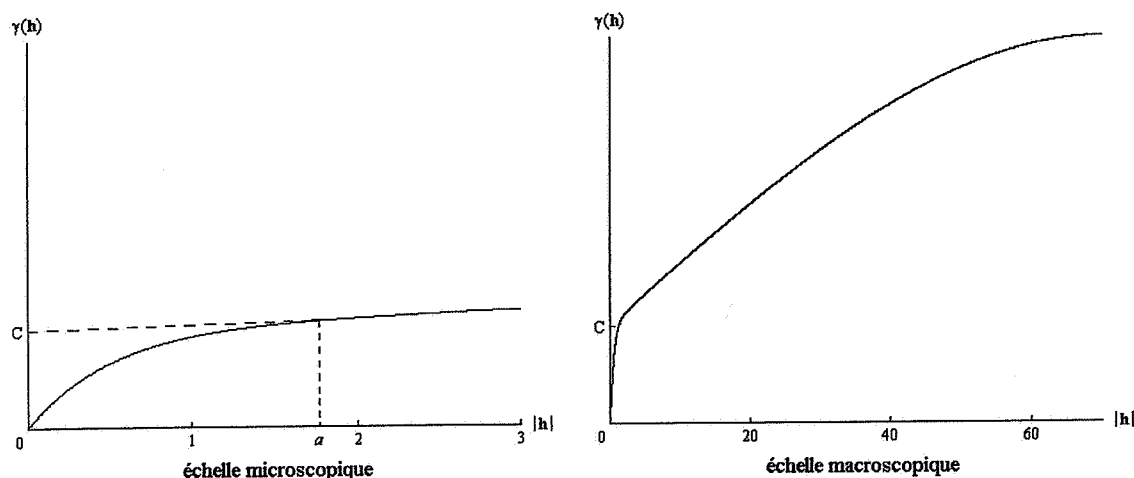


figure 3.19 : genèse d'un effet de pépité

Ainsi donc, un effet de pépite peut être la manifestation d'une structure gigogne dont les dimensions caractéristiques sont très inférieures à l'échelle de travail : seul le palier de cette "microstructure" est mesurable (il est donné par l'amplitude de l'effet de pépite), ses autres traits sont imperceptibles.

Bruit blanc

Le concept d'erreurs de mesure correspond à un processus aléatoire sans corrélation spatiale, mais de nature discrète, qui se superpose au processus sous-jacent $Z(\mathbf{x})$ aux seuls points de données, et n'est pas défini aux autres points. D'autre part, la notion de microstructure équivaut à celle d'un processus défini continûment, de portée très faible mais non nulle. Que se passe-t-il si l'on considère l'effet de pépite comme une composante de portée exactement nulle, correspondant à ce que l'on qualifie de "bruit blanc" en théorie du signal ?

En toute rigueur, il ne peut exister de fonction aléatoire dont le variogramme est un effet de pépite pur (portée exactement nulle) et d'espérance supposée nulle pour simplifier. Dans le cas contraire, la moyenne de cette fonction aléatoire sur tout support non ponctuel, aussi petit soit-il, serait nulle, car elle mélangerait une infinité de valeurs non corrélées et d'espérance nulle, qui finiraient par se neutraliser. En outre, la transformée de Fourier de la covariance pépitique est nulle sur toute la gamme des fréquences.

Un tel processus peut cependant être défini dans la théorie des distributions aléatoires et être associé à une covariance représentée par une mesure de Dirac de poids C_0 à l'origine (cette mesure, notée $C_0 \delta$, peut être vue comme une pseudo-fonction, infinie en 0, nulle partout ailleurs, et telle que son intégrale soit égale à C_0). La transformée de Fourier de cette mesure est constante et égale à C_0 , d'où la dénomination de *bruit blanc*, par référence au spectre de la lumière blanche. Ce modèle diffère du précédent, où l'on faisait tendre la portée d'une microstructure vers 0, sans pour autant que la variance devienne infinie.

La justification de cette définition est qu'aucune mesure n'est rigoureusement ponctuelle, et seules sont accessibles des valeurs moyennes sur des supports petits (cette explication est identique à celle qui a été fournie pour définir le modèle logarithmique de variogramme). Or, la moyenne, ou plus généralement une convoluée de la distribution aléatoire de covariance $C_0 \delta$ par une fonction ϕ , est une fonction aléatoire usuelle, de covariance proportionnelle au covariogramme transitif de ϕ :

$$C_\phi(\mathbf{h}) = C_0 \int \phi(\mathbf{x}) \phi(\mathbf{x} + \mathbf{h}) d\mathbf{x}.$$

Cette formule correspond au procédé de régularisation, qui sera étudié au chapitre 4. Si l'on prend comme fonction ϕ l'indicatrice d'un support v petit divisée par le volume de v (ce qui revient à dire que les valeurs mesurées sont les valeurs moyennes sur v et non des valeurs rigoureusement ponctuelles), on obtient une covariance régularisée proportionnelle au covariogramme géométrique de v :

$$C_v(\mathbf{h}) = \frac{C_0}{|v|^2} \int 1_v(\mathbf{x}) 1_v(\mathbf{x} + \mathbf{h}) d\mathbf{x} = \frac{C_0}{|v|^2} |v \cap v_{-\mathbf{h}}|.$$

En particulier, bien que le support des mesures ne soit pas ponctuel, leur variance n'est pas nulle ; elle vaut $C_v(\mathbf{0}) = C_0 / |v|$ et varie en raison inverse du volume des prélèvements. La portée de $C_v(\mathbf{h})$ est quant à elle égale au diamètre de v .

En conclusion, la présence d'un "bruit blanc" strictement ponctuel d'amplitude C_0 se manifeste, au niveau du variogramme des mesures de support v , par un phénomène de transition de variance $C_0 / |v|$ et de portée égale au diamètre de v . Ce résultat subsiste d'ailleurs si, au lieu de considérer l'effet de pépite par une mesure de Dirac, on l'interprète comme une microstructure de portée très petite vis-à-vis de v ; dans ce cas, la constante C_0 coïncide avec la *portée intégrale* de la microstructure (cf. annexe B). Le modèle de bruit blanc est donc une idéalisation de cette représentation, où la portée de la microstructure devient nulle et son palier infini, de sorte que la portée intégrale reste invariante. Il permet de comprendre pourquoi il peut subsister une constante pépitique lorsque l'on travaille sur des mesures non ponctuelles (par exemple, des valeurs composées), alors que le modèle de portée nulle et variance constante ne permet pas de l'expliquer.

On voit également que, si la composante péritique est importante et occulte la structure sous-jacente de la fonction aléatoire, cela peut signifier que le support des mesures est trop petit. Un support plus volumineux conduirait à une diminution de la constante péritique et une meilleure appréciation de la structuration des valeurs.

V.3. Le traitement des anisotropies

Les anisotropies se manifestent lorsque le comportement du variogramme diffère selon les directions de l'espace. En l'absence d'anisotropie, $\gamma(\mathbf{h})$ ne dépend que du module du vecteur $\mathbf{h}$ et non de son orientation. Comme les modèles théoriques de variogrammes sont définis dans le cas isotrope, nous devons examiner les transformations qui permettent le passage à des modèles anisotropes.

En pratique, les anisotropies s'identifient en comparant les variogrammes expérimentaux calculés le long de plusieurs directions, par exemple, dans le cas bidimensionnel, le long des directions orientées de 0°, 45°, 90° et 135° par rapport à l'axe des abscisses. On complète souvent cet examen en dessinant la "**carte variographique**", c'est-à-dire la carte des isovaleurs du variogramme expérimental en fonction du vecteur $\mathbf{h}$ (donc de son module et de sa direction). Lorsqu'il y a isotropie, les variogrammes expérimentaux dans les différentes directions se chevauchent et la carte variographique dessine des cercles, ou des sphères dans le cas tridimensionnel. Le variogramme dépend seulement de la longueur du vecteur $\mathbf{h}$.

Dans le cas contraire, on est en présence d'une anisotropie. On distingue plusieurs types d'anisotropie, notamment l'**anisotropie géométrique** et l'**anisotropie zonale**.

- **Anisotropie géométrique**

Une anisotropie est dite *géométrique* lorsque la carte variographique dessine des ellipses, ou des ellipsoïdes dans le cas tridimensionnel, concentriques. Les variogrammes directionnels ont la même forme mais ont des portées différentes (cadre stationnaire : variogrammes à paliers) ou des pentes différentes (cadre intrinsèque strict : variogrammes non bornés).

Dans le cas bidimensionnel, l'anisotropie géométrique est complètement définie par son **coefficient d'anisotropie** a/b , rapport entre le petit axe a et le grand axe b de l'une quelconque des ellipses d'isovaleurs, et l'**angle d'anisotropie** θ , formé par le grand axe et l'axe des abscisses. Le premier indique "l'intensité" de l'anisotropie (d'autant plus marquée qu'il s'éloigne de 1), tandis que le second indique son orientation. Dans le cas tridimensionnel, on doit définir deux coefficients d'anisotropie et deux angles.

On peut écrire :

$$\gamma(\mathbf{h}) = \gamma_0(|\mathbf{A} \mathbf{h}|)$$

où γ_0 est un variogramme isotrope et où $\mathbf{A}$ est une matrice (2×2) ou (3×3) qui dépend des angle(s) et coefficient(s) d'anisotropie. Ceci signifie qu'une simple transformation linéaire des coordonnées spatiales (rotation suivie d'une homothétie) suffit à rétablir l'isotropie.

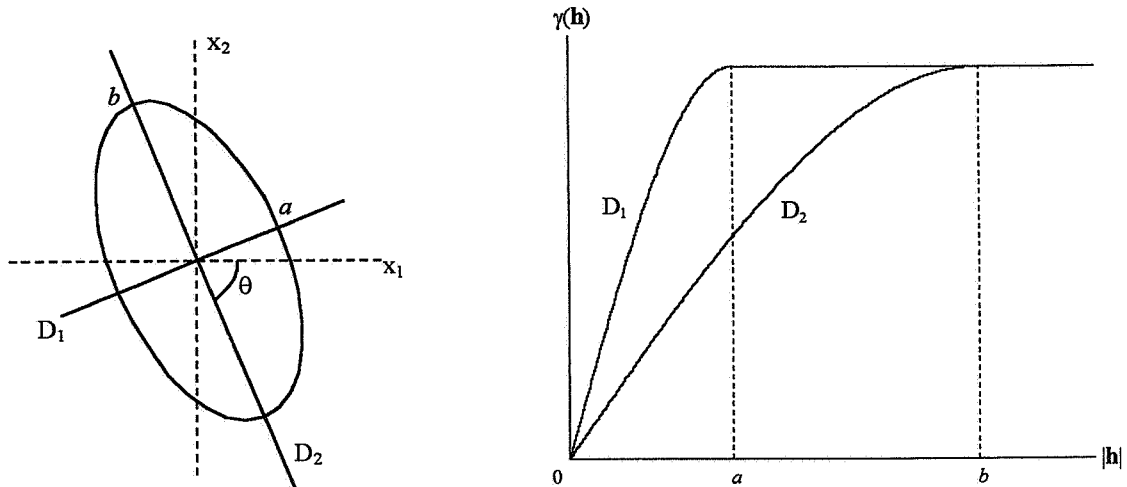


figure 3.20 : exemple (2D) d'anisotropie géométrique de rapport a/b et d'angle θ

à gauche : l'ellipse des portées **à droite :** les variogrammes directionnels dans les directions principales d'anisotropie

On peut obtenir des formes plus complexes d'anisotropie en superposant plusieurs modèles à anisotropies géométriques de rapports et d'angles différents.

- Anisotropie zonale**

Il existe des variogrammes ou covariances dont l'anisotropie ne peut être corrigée par une transformation linéaire des coordonnées, notamment ceux qui présentent un palier variable selon la direction de l'espace.

Au sens strict, l'anisotropie *zonale* est un modèle dans lequel la fonction aléatoire $Z(\mathbf{x})$ ne dépend que d'une coordonnée. Cela arrive lorsqu'on travaille dans l'espace à trois dimensions ; par exemple, dans l'étude d'un phénomène sédimentaire, la variabilité est plus importante le long de la verticale, qui traverse les strates, que dans les strates. En première approximation, $Z(\mathbf{x})$ est constante dans le plan horizontal, de sorte que son variogramme $\gamma(\mathbf{h})$ ne dépend que de la composante verticale de $\mathbf{h}$.

Plus précisément, en notant θ l'angle entre $\mathbf{h}$ et la direction verticale, et $\mathbf{u}$ un vecteur unitaire orienté verticalement, on peut écrire :

$$\gamma(\mathbf{h}) = \gamma(|\mathbf{h}| \cos \theta \mathbf{u}).$$

γ admet le même palier dans toutes les directions de l'espace, sauf dans le plan horizontal pour lequel $\cos \theta = 0$ et $\gamma(\mathbf{h}) = 0$.

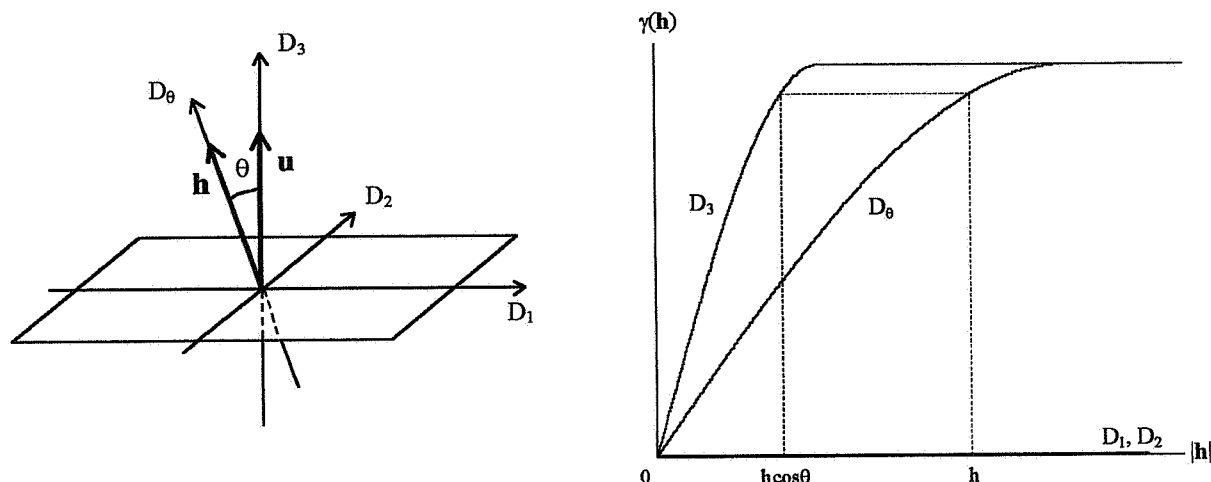


figure 3.21 : exemple (3D) d'anisotropie zonale pure de direction verticale

à droite : les variogrammes directionnels dans les directions horizontales (D_1, D_2), verticale (D_3) et une direction intermédiaire (D_θ)

Le plus souvent, le palier du variogramme n'est pas rigoureusement nul dans le plan horizontal, mais est plus faible que le palier le long de la direction verticale. Ainsi, au sens large, une anisotropie est dite *zonale* quand la valeur du palier varie avec la direction. On a affaire en règle générale à un schéma gigogne dans lequel certaines composantes sont zonales au sens strict, et s'annulent le long d'une droite ou dans un plan de l'espace.

La modélisation de ce type d'anisotropie s'opère de la manière suivante. Supposons par exemple que, dans l'espace à trois dimensions, la valeur du palier le long de l'axe vertical (x_3) soit plus grande que celle dans le plan horizontal (x_1, x_2). On commence par ajuster un modèle isotrope γ_1 dans le plan horizontal (x_1, x_2), puis on ajoute un modèle γ_2 à anisotropie géométrique de grande portée horizontale (en jouant sur son coefficient d'anisotropie). De cette façon, γ_2 sera quasiment nul dans le plan horizontal sur toute la gamme des distances qui nous intéressent, et ne perturbera pas l'ajustement dans ce plan : $\gamma_2(\mathbf{h}) \approx \gamma_2(h_3)$.

Le modèle définitif sera : $\gamma(\mathbf{h}) = \gamma_1(\mathbf{h}) + \gamma_2(h_3)$.

Au lieu d'un modèle γ_1 isotrope, on peut aussi utiliser un modèle à anisotropie géométrique, avec des portées différentes, mais finies, le long des directions (x_1, x_2, x_3). Le modèle de fonction aléatoire sous-jacent est constitué par l'addition de deux processus non corrélés Z_1 et Z_2 de variogrammes respectifs γ_1 et γ_2 :

$$Z(\mathbf{x}) = Z_1(\mathbf{x}) + Z_2(x_3).$$

Ajoutons pour terminer que la superposition de plusieurs anisotropies zonales et / ou géométriques de caractéristiques différentes (angles et coefficients d'anisotropie) permet de faire face à la plupart des situations rencontrées en pratique. Mais il faut savoir que certains types d'anisotropies échappent à ces modèles (cf. page suivante).

- **Anisotropie induite par une factorisation de la covariance**

Un troisième type d'anisotropie peut être facilement modélisé à partir de schémas élémentaires isotropes, mais il échappe au modèle linéaire de régionalisation. Il fait appel à la propriété qu'un produit de covariances est encore une covariance. Par exemple, on peut définir une covariance bidimensionnelle, produit de deux covariances unidimensionnelles distinctes :

$$\forall \mathbf{h} = (h_1, h_2) \in \mathbb{R}^2, C(\mathbf{h}) = C_1(h_1) \times C_2(h_2).$$

Le variogramme associé est :

$$\gamma(\mathbf{h}) = C(0) - C_1(h_1) \times C_2(h_2).$$

Contrairement aux modèles d'anisotropie géométrique et zonale, un tel variogramme présente une forme différente le long des directions de l'espace, bien que le palier soit le même.

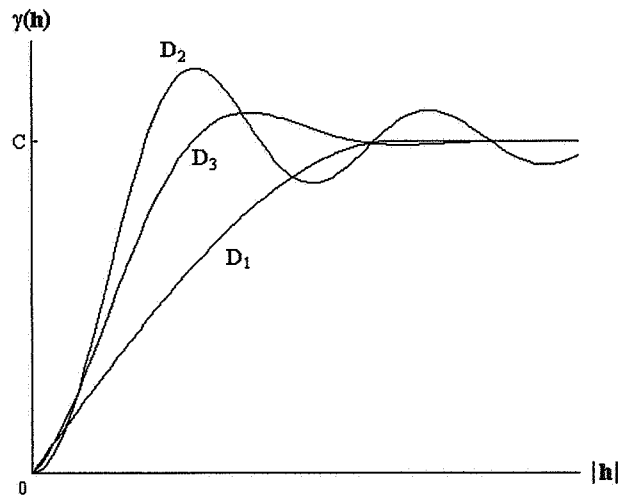


figure 3.22 : exemple d'anisotropie produit, avec un schéma sphérique le long de D_1 et un schéma sinus cardinal le long de D_2 , orthogonale à D_1 ; la direction D_3 correspond à la bissectrice entre D_1 et D_2

Le modèle de fonction aléatoire sous-jacent est :

$$Z(\mathbf{x}) = Z_1(x_1) \times Z_2(x_2)$$

où Z_1 et Z_2 sont deux fonctions aléatoires unidimensionnelles, sans corrélation spatiale mutuelle, et de covariances respectives C_1 et C_2 .

On peut évidemment sophisticationner le modèle, en prenant le produit de plusieurs covariances, mono ou multi-dimensionnelles, isotropes ou à anisotropie géométrique ou zonale.

V.4. Recommandations pour le calcul du variogramme expérimental et l'ajustement à un modèle

V.4.1. Représentativité des données

Si les données sont implantées de manière **préférentielle**, le variogramme expérimental peut ne pas être représentatif de la structuration de la régionalisation sur l'ensemble du champ. Dans ce cas, plusieurs alternatives sont envisageables :

- délimiter des sous-zones reconnues de manière homogène et les étudier séparément ;
- travailler sur une partie des données ;
- pondérer les données dans le calcul du variogramme expérimental ;
- s'aider d'outils peu sensibles aux hétérogénéités d'échantillonnage (par exemple, les variogrammes relatifs) pour améliorer la description de la structure spatiale ;
- si cela est possible, incorporer dans le modèle variographique les informations concernant le mode d'échantillonnage.

V.4.2. La nuée variographique

Pour étudier la stabilité numérique du variogramme expérimental, il est utile de visualiser la “**nuée variographique**”, i.e. le nuage⁸ des carrés des différences $\{ [z(\mathbf{x}_\alpha + \mathbf{h}) - z(\mathbf{x}_\alpha)]^2 / 2 \}$ en fonction du vecteur $\mathbf{h}$, ou simplement de son module $|\mathbf{h}|$. Le variogramme expérimental est obtenu en réalisant, pour chaque distance, la moyenne des points correspondants de la nuée variographique. L'examen de cette nuée permettra de repérer les couples de données responsables des fortes valeurs du variogramme expérimental ; on pourra ainsi mettre en évidence les mesures notablement différentes de leurs voisines, et s'interroger sur leur éventuel caractère “aberrant”.

V.4.3. Paramètres de calcul du variogramme expérimental

Le calcul du variogramme expérimental fait intervenir plusieurs paramètres, qui relèvent du choix de l'utilisateur : directions de calcul, pas et tolérances. Une utilisation maladroite de ces paramètres peut conduire à des artefacts sans relation avec le phénomène régionalisé, et avoir des conséquences fâcheuses dans la modélisation et sur les résultats ultérieurs.

⁸ Au lieu de visualiser la nuée entière des valeurs $\{ [z(\mathbf{x}_\alpha + \mathbf{h}) - z(\mathbf{x}_\alpha)]^2 / 2 \}$, on peut aussi ne repérer que leurs extrema, leur moyenne (laquelle n'est autre que la valeur du variogramme expérimental en $\mathbf{h}$) et leurs quartiles, qui donnent une image synthétique de leur distribution. Si la moyenne sort de l'étendue interquartile (intervalle entre les premier et troisième quartiles), cela peut indiquer que la valeur du variogramme expérimental pour la distance considérée est fortement influencée par les valeurs extrêmes, donc peu stable.

• Choix des directions de calcul

Pour des raisons de simplicité de modélisation et de visualisation, et aussi parce que les modèles théoriques utilisés dérivent de fonctions isotropes, on n'ajuste pas le variogramme entier, mais ses intersections avec des droites de l'espace. En d'autres termes, la modélisation se basera sur un ou plusieurs variogrammes expérimentaux *directionnels*, alors que le variogramme théorique est défini dans tout l'espace.

Les directions de calcul doivent être les plus significatives, et sont évidemment liées à l'anisotropie du phénomène. Leur choix se fait habituellement en visualisant la **carte variographique**. En cas d'anisotropie géométrique ou zonale, on calculera le variogramme expérimental le long des directions reconnues comme les directions principales d'anisotropie ; en cas d'isotropie (i.e. au cas où les variogrammes directionnels se superposent à quelques fluctuations près), on considérera un variogramme "omnidirectionnel" :

$$\bar{\gamma}^+(r) = \frac{1}{2|N^+(r)|} \sum_{N^+(r)} [z(\mathbf{x}_\alpha) - z(\mathbf{x}_\beta)]^2 \quad \text{où } N^+(r) = \{ (\alpha, \beta) \text{ tel que } |\mathbf{x}_\alpha - \mathbf{x}_\beta| \approx r \}.$$

• Pas de calcul

Chaque variogramme directionnel est calculé pour des distances multiples d'une distance élémentaire, ou "pas". Le choix du pas peut avoir une répercussion importante dans le calcul. Un pas trop fin conduira à un variogramme expérimental très erratique, tandis qu'un pas trop élevé ne permettra pas d'apprécier les détails de la structuration spatiale, en particulier le comportement à l'origine du variogramme. Quand l'échantillonnage est à maille régulière⁹, le pas est presque imposé. Dans le cas contraire, on pourra s'aider de la "nuée variographique" pour trouver un pas tel que le nombre de paires intervenant dans le calcul des points du variogramme expérimental soit suffisant (pour fixer les idées, supérieur à la moitié du nombre de données), afin que ces points soient représentatifs.

• Tolérances sur les distances et sur les angles

L'usage de tolérances sur les distances et sur les angles a pour but de lisser et robustifier le variogramme expérimental lorsque l'échantillonnage n'est pas régulier. En général, on utilise une tolérance sur les distances égale à la moitié du pas : dans la direction considérée, le *i*-ème point expérimental est calculé à partir des paires de données dont la séparation est dans l'intervalle $[(i - 0.5) \times pas, (i + 0.5) \times pas]$. De la sorte, toutes les distances sont utilisées une et une seule fois.

On peut aussi prendre une tolérance plus faible. Dans ce cas, les classes de distance qui interviennent dans le calcul du variogramme expérimental ne sont plus contiguës, et des points de la nuée variographique ne sont pas pris en compte ; cette situation apparemment défavorable (on perd de l'information) n'est pas toujours la plus mauvaise, surtout à maille régulière ou quasi-régulière.

⁹ par exemple dans le cas d'une maille carrée de longueur *a*, on choisira naturellement un pas de calcul égal à *a* dans les directions principales de la maille, et à $a\sqrt{2}$ dans les directions diagonales.

Il convient de mettre en garde l'utilisateur contre l'usage de tolérances excessives, aussi bien sur les distances que sur les angles, qui peuvent lisser artificiellement le variogramme expérimental ; bien qu'un variogramme lisse soit en général plus facile à modéliser, cela ne garantit pas qu'il soit plus représentatif des données.

Par ailleurs, dans le cas d'un modèle de transition, les tolérances sur les distances produisent un biais sur l'estimation de la portée : la portée apparente est, en moyenne, égale à la portée réelle augmentée de la moitié de l'intervalle de tolérance. Quant aux tolérances angulaires, elles tendent à estomper une éventuelle anisotropie du variogramme (les directions principales d'anisotropie sont mélangées avec des directions intermédiaires, de sorte que l'anisotropie apparaît moins prononcée qu'elle ne l'est en réalité).

Il faut donc trouver un juste équilibre dans le choix des tolérances. Il est d'usage de compléter la définition des tolérances angulaires par celle d'écartements maximaux : *largeur de tranche* et, dans l'espace à trois dimensions, *hauteur de tranche* ; ce dernier paramètre est important en pratique, car il évite de trop mélanger des directions horizontales avec des directions inclinées, dont le comportement est souvent bien différent.

- **Représentativité des points expérimentaux**

Il convient toujours de prendre garde à la représentativité des points expérimentaux. On pourra afficher, pour chaque point expérimental, le **nombre de paires de données** ayant servi à son calcul ; un nombre de paires faible indique un point peu fiable. Egalement, il est parfois utile de visualiser la nuée variographique, afin d'identifier des données responsables d'une instabilité numérique du variogramme expérimental.

V.4.4. Modélisation

La représentation d'une variable régionalisée par un variogramme est une opération subjective : il n'existe pas de "vrai" modèle sous-jacent. En pratique, il convient de s'assurer que le modèle respecte les principales caractéristiques du variogramme expérimental (comportement à l'origine, existence ou non d'un palier, anisotropie) et qu'il ne soit pas inutilement complexe.

Il n'existe pas de méthode infaillible pour ajuster un schéma théorique à un variogramme expérimental. En général, conformément au modèle linéaire de régionalisation, on superpose plusieurs structures élémentaires, qui, idéalement, modélisent des composantes spatiales d'échelles différentes. On définit tout d'abord la famille de variogrammes que l'on souhaite faire intervenir dans le modèle, puis on ajuste "au mieux" leurs paramètres (palier, portée...). On cherchera à éviter l'usage exclusif de méthodes d'ajustement automatiques. Au contraire, la modélisation d'un variogramme expérimental doit être un **travail interactif** où l'utilisateur doit avoir son mot à dire.

On peut ajouter les remarques suivantes.

- Le choix d'un modèle avec ou sans **palier** se décide en fonction du comportement aux grandes distances des variogrammes expérimentaux directionnels. Ce comportement peut cependant être chaotique, car peu de données interviennent dans le calcul des points correspondants, la distance étant proche de la dimension maximale du champ. En général, le variogramme expérimental est peu fiable au-delà d'une distance égale à la moitié du diamètre du champ dans la direction considérée. L'ajustement du palier éventuel se fait autour des fluctuations du variogramme et / ou de la variance empirique des données.
- Le **comportement du variogramme à l'origine** peut ne pas être accessible si la maille de l'échantillonnage est trop lâche. Ainsi, une microstructure peut être masquée et apparaître comme un effet de pépite si les plus petites distances pour lesquelles le variogramme expérimental est calculable sont supérieures à la portée pratique. On est souvent conduit à adopter un schéma linéaire à l'origine avec une constante pépitique, qui s'obtient en extrapolant le variogramme expérimental jusque l'axe des ordonnées. Il arrive que l'effet de pépite *apparent* soit négatif, et ne peut être modélisé tel quel (un variogramme est toujours positif). Une solution raisonnable consiste à utiliser un modèle à comportement parabolique à l'origine. Mais cette situation doit inciter à s'interroger sur les paramètres de calcul utilisés : l'effet de pépite négatif peut être artificiel, ou du moins exagéré, à cause d'un usage impropre de la tolérance ; cela peut aussi être dû au support des données, s'il n'est pas ponctuel.
- L'ajustement à un modèle ne se fait pas au vu du seul variogramme expérimental, mais doit **intégrer toute l'information disponible sur le phénomène**. Ainsi, une anisotropie ou une périodicité ne doit être modélisée que si elle est évidente sur les variogrammes expérimentaux ou explicable par des considérations physiques. On pourra s'aider des méthodes d'analyse exploratoire des données (nuages de corrélation différée, nuées variographiques...) pour améliorer la connaissance de la structure du phénomène et garantir le réalisme du modèle : comportement à l'origine (reflétant le degré de régularité de la variable régionalisée), stationnarité, anisotropie...
- Enfin, la plupart des applications (problèmes d'estimation notamment) ne requièrent pas la connaissance du variogramme pour toutes les distances. Il convient de bien ajuster le variogramme jusqu'à l'**échelle de travail** envisagée ; au-delà, l'ajustement n'aura plus d'importance. Par exemple, cette échelle représente le diamètre du "voisinage" où l'on prendra l'information pour estimer une valeur en un point non échantillonné (travail en *voisinage glissant*, cf. chapitre 5 sur le krigeage).

En conclusion, l'analyse variographique fait appel à la fois à l'information disponible sur le phénomène étudié, à un certain savoir-faire dans le choix et l'utilisation des outils et à un sens de l'approximation. C'est l'étape essentielle de toute étude géostatistique et certainement celle qui se prête le moins à l'automatisme.

V.5. Test de la qualité du modèle : la validation croisée

On a la possibilité de vérifier l'adéquation entre les données et le modèle structural adopté, et par la même occasion de comparer la qualité des différents modèles envisageables, en utilisant la technique de la **validation croisée**. On suppose que l'on a ajusté le variogramme expérimental à un schéma théorique dont les paramètres ont été déterminés : palier, portée, anisotropie... Le principe de la validation croisée est d'estimer tour à tour (par krigeage, cf. chapitre 5) chaque observation en ne tenant compte que des observations restantes. On peut alors calculer l'erreur d'estimation (différence entre la valeur estimée et la valeur vraie) en chaque point de donnée, et la comparer à l'écart-type d'estimation donné par le krigeage et censé quantifier la précision de l'estimation.

Cette procédure permet de tester le modèle théorique de variogramme et de détecter les données pour lesquelles le modèle est inapproprié, à savoir celles qui auront été mal estimées. Elle est également utile pour définir, de manière adéquate, la recherche des données à utiliser pour l'estimation locale (cf. chapitre 5, travail en voisinage glissant).

La validation croisée est usuellement présentée sous forme de tests graphiques, en particulier :

- le nuage de corrélation entre les valeurs mesurées $\{z(\mathbf{x}_\alpha), \alpha = 1 \dots n\}$ et les valeurs estimées $\{z^*(\mathbf{x}_\alpha), \alpha = 1 \dots n\}$;
- l'histogramme des erreurs standardisées¹⁰

$$\frac{z^*(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha)}{\sigma^*(\mathbf{x}_\alpha)}$$

où $\sigma^*(\mathbf{x}_\alpha)$ est l'écart-type de krigeage pour le point $\mathbf{x}_\alpha$. Une estimation sera jugée bonne si l'erreur standardisée associée est située dans un intervalle $[-\alpha, \alpha]$, où $\alpha = 2.5$ par exemple (borne choisie arbitrairement) ; la donnée correspondante est alors qualifiée de *robuste* ;

- le nuage de corrélation entre les erreurs standardisées et les valeurs estimées¹¹ ;
- la carte d'implantation des données, où sont repérées les valeurs "mal" estimées, c'est-à-dire celles dont l'erreur standardisée associée sort de l'intervalle $[-\alpha, \alpha]$.

Un modèle de variogramme sera d'autant meilleur que l'histogramme des erreurs standardisées est resserré et centré sur 0 et que le nuage de corrélation entre valeurs mesurées et estimées est proche de la première bissectrice.

¹⁰ L'étude des erreurs standardisées, au lieu des erreurs absolues, permet de s'affranchir des contraintes de nature géométrique. Quel que soit le modèle variographique, l'erreur absolue a tendance à être élevée pour une donnée isolée, et faible pour une donnée située à proximité d'autres données. Cet effet est corrigé lorsque l'on considère l'erreur standardisée, car l'écart-type de krigeage quantifie la précision que l'on attend de l'estimation ; par construction, l'erreur standardisée a une dispersion (variance) égale à 1, indépendamment de la configuration géométrique des échantillons.

¹¹ Ce test permet de vérifier si la moyenne des erreurs commises, ou *biais*, dépend des valeurs estimées, auquel cas on dit qu'il existe un *biais conditionnel* (cf. chapitre 5). Une telle situation est à éviter, et l'on recherchera un modèle variographique et un voisinage de krigeage qui minimisent le biais conditionnel.

On peut également essayer de satisfaire au mieux les critères statistiques suivants.

- 1) La moyenne des erreurs et celle des erreurs standardisées mesurent le *biais* de l'estimateur et doivent être proches de zéro. En fait, ce critère est d'un d'intérêt secondaire vis-à-vis de la modélisation du variogramme, car nous verrons que les valeurs obtenues par krigeage sont par construction sans biais, indépendamment du choix du variogramme. Ainsi, quel que soit le modèle utilisé, la moyenne des erreurs tendra toujours vers zéro.
- 2) La moyenne des erreurs quadratiques, qui quantifie la *précision* de l'estimateur, doit être minimale. Au lieu d'une moyenne arithmétique, on peut éventuellement calculer une moyenne pondérée, qui assigne un poids faible aux données groupées et un poids élevé aux données isolées.
- 3) La moyenne des erreurs quadratiques *standardisées* doit être proche de 1 :

$$\varepsilon \equiv \frac{1}{n} \sum_{\alpha=1}^n \left(\frac{z^*(\mathbf{x}_{\alpha}) - z(\mathbf{x}_{\alpha})}{\sigma^*(\mathbf{x}_{\alpha})} \right)^2 \approx 1.$$

Ce critère indique une bonne adéquation du modèle variographique dans l'appréciation de l'amplitude des erreurs commises, c'est-à-dire que le variogramme choisi n'est ni trop optimiste ni trop pessimiste.

Dans le modèle probabiliste, ε est la réalisation d'une variable aléatoire E d'espérance égale à 1 (l'espérance de chaque terme de la somme étant égale à la variance unité des erreurs standardisées). Pour pouvoir interpréter la valeur expérimentale ε de E , il faudrait connaître sa variance, qui est un moment d'ordre 4 de $Z(\mathbf{x})$ et n'est pas accessible. On peut toutefois donner un ordre de grandeur pour la variance de la somme, en supposant pour simplifier que les erreurs standardisées suivent une loi *gaussienne* réduite et sont *indépendantes*. La somme des erreurs quadratiques standardisées suit alors une loi du χ^2 à n degrés de liberté, qui a pour variance $2n$, de sorte que la variance de E est égale à $2/n$ et tend vers 0 lorsque n devient grand. On peut donc s'attendre à ce que la moyenne expérimentale ε soit représentative de l'espérance de E , c'est-à-dire proche de 1.

- 4) Le coefficient de corrélation entre les valeurs estimées et les valeurs mesurées doit être le plus proche possible de 1.
- 5) Le nombre de données non robustes (mal estimées) doit être le plus faible possible. Pour fixer les idées, ce nombre est jugé satisfaisant s'il représente moins de 5% du total des données.

Ces critères permettent de comparer la qualité de différents ajustements possibles. Toutefois, il est bien évident que l'ensemble de ces critères ne pourront être satisfaits simultanément par un seul modèle. Améliorer un critère risque d'en détériorer un autre.

V.6. Application à trois jeux de données

Afin d'illustrer les concepts introduits dans les paragraphes qui précèdent, nous allons mener l'analyse variographique des données agronomiques, pédologiques et minières qui ont été présentées aux chapitres précédents.

V.6.1. Un exemple quasi-stationnaire : les données agronomiques

Nous allons reprendre l'exemple introduit au premier chapitre de ce document, concernant la répartition des parcelles de canne à sucre au nord de Basse-Terre (Guadeloupe). La variable régionalisée étudiée, $z(\mathbf{x})$, est la superficie de canne à sucre dans des cellules de 1 ha, et est connue sur la totalité d'une zone rectangulaire de dimensions 19.7 km (nord-sud) sur 21.8 km (est-ouest). On se souvient en particulier que $z(\mathbf{x})$ présente une très forte proportion de valeurs nulles, et que les parcelles de canne à sucre ne s'étendent que le long de la diagonale nord-ouest / sud-est de la zone rectangulaire. De fait, une étude sur la totalité du rectangle n'aurait pas grand sens, car elle mélangerait des données provenant de zones manifestement hétérogènes.

Il est nécessaire de délimiter aussi précisément que possible le champ $\mathcal{D}$ de la variable régionalisée, sur lequel cette dernière peut être considérée comme homogène. Cette **contrainte d'homogénéité spatiale** ne se posait pas en géostatistique transitive, mais elle est essentielle en géostatistique probabiliste, faute de quoi les hypothèses de stationnarité que l'on serait amené à formuler ne seraient pas acceptables. En réalité, la zone rectangulaire de départ incorpore des parties incultivables (zones urbaines, zones maritimes, marécages, parc naturel et reliefs) où la superficie de canne à sucre est fatalement nulle. La suppression de ces parties permet de définir une zone "potentiellement cultivée" de 16 050 ha, que l'on définit comme le champ $\mathcal{D}$ de la régionalisation, et que l'on considèrera comme homogène. La carte ci-dessous présente ce champ (parties grisées), à la géométrie un peu complexe.

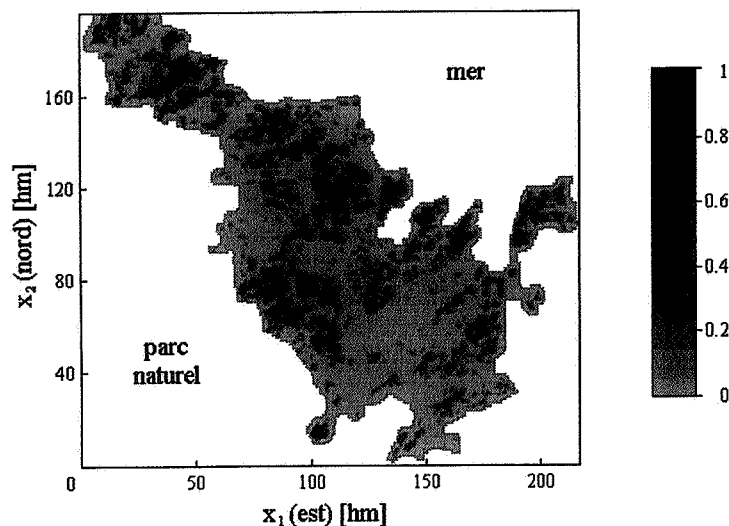


figure 3.23 : carte des parcelles de canne à sucre dans le champ étudié

L'histogramme des valeurs sur le champ est toujours très dissymétrique, avec une très forte proportion de valeurs nulles (figure 3.24).

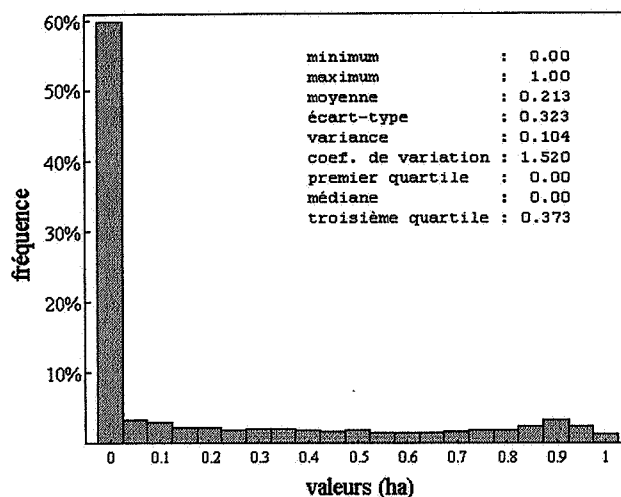


figure 3.24 : histogramme de $z(x)$ dans $\mathcal{D}$

On suppose que, sur le champ $\mathcal{D}$, $z(x)$ est la réalisation d'une fonction aléatoire $Z(x)$ localement stationnaire d'ordre deux. On pourra s'en convaincre en visualisant les profils moyens de $z(x)$ le long des directions nord-sud et est-ouest (figure 3.25), c'est-à-dire les moyennes des valeurs prises par $z(x)$ le long des axes de coordonnées. Ces profils montrent une moyenne *localement* constante et des fluctuations raisonnablement uniformes, compatibles avec une hypothèse quasi-stationnaire. Signalons toutefois que, bien que valable localement, l'hypothèse de stationnarité ne serait pas pertinente à l'échelle du champ¹².

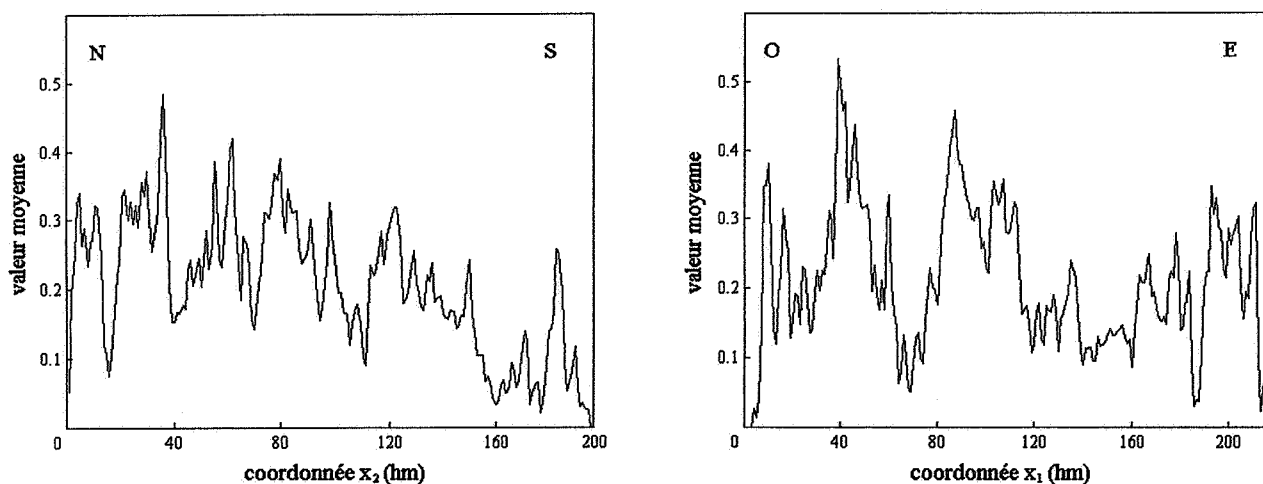


figure 3.25 : profil moyen le long des directions nord-sud (à gauche) et ouest-est (à droite)

¹² En effet, la régionalisation présente un *effet de bord* : la périphérie du domaine est très peu cultivée ; la raison en est que, dans la délimitation du champ, faute de connaître exactement la répartition des parcelles de canne à sucre (puisque c'est le but recherché), on a laissé une "marge de sécurité" en bordure de zone où l'on a moins de chance de trouver des surfaces agricoles.

Etude du variogramme régional

Comme la réalité est entièrement connue, on peut envisager de conserver l'ensemble des données pour l'analyse variographique. Le "variogramme" que l'on obtient, calculé sur l'ensemble des valeurs prises par $z(\mathbf{x})$, est appelé **variogramme régional** ; il est défini par :

$$\gamma_{\text{reg}}(\mathbf{h}) = \frac{1}{2 |\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}|} \int_{\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}} [z(\mathbf{x} + \mathbf{h}) - z(\mathbf{x})]^2 d\mathbf{x}$$

où $\mathcal{D}_{-\mathbf{h}}$ est le translaté du champ $\mathcal{D}$ par le vecteur $-\mathbf{h}$.

On voit que, lorsque la réalité n'est connue que de manière fragmentaire, le variogramme expérimental que l'on calcule à l'aide des données disponibles est une *approximation* du variogramme régional. Au niveau du modèle probabiliste, c'est-à-dire lorsque l'on remplace dans l'expression précédente la variable régionalisée par la fonction aléatoire associée, le variogramme régional devient une variable aléatoire

$$\gamma_{\text{reg}}(\mathbf{h}) = \frac{1}{2 |\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}|} \int_{\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}} [Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]^2 d\mathbf{x}$$

qui est un estimateur sans biais du variogramme réel $\gamma(\mathbf{h})$: $E[\gamma_{\text{reg}}(\mathbf{h})] = \gamma(\mathbf{h})$.

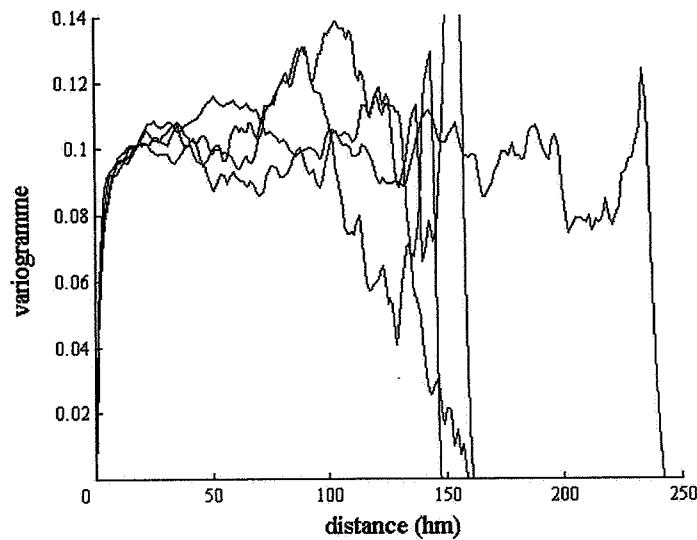
Il faut bien prendre garde à ne pas confondre le variogramme régional $\gamma_{\text{reg}}(\mathbf{h})$ et le "vrai" variogramme $\gamma(\mathbf{h})$. Ce dernier est un objet théorique, défini comme la moitié de l'espérance de l'écart quadratique $[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]^2$: ce n'est pas une *grandeur régionale*, i.e. il n'est pas déterminable de manière univoque, même si l'on connaît la totalité de la régionalisation, comme c'est le cas ici. Au contraire, le variogramme régional est bien une grandeur régionale, mais il ne vérifie pas les propriétés mathématiques du variogramme (en particulier, il n'est pas de type négatif conditionnel). Une **modélisation** est donc **indispensable**, même lorsque l'on dispose du variogramme régional.

L'écart entre le variogramme régional et le variogramme théorique est appelé *fluctuation*, et est caractérisé par la *variance de fluctuation*

$$D_{\text{fluctuation}}^2(\mathbf{h}) = E\{[\gamma_{\text{reg}}(\mathbf{h}) - \gamma(\mathbf{h})]^2\}.$$

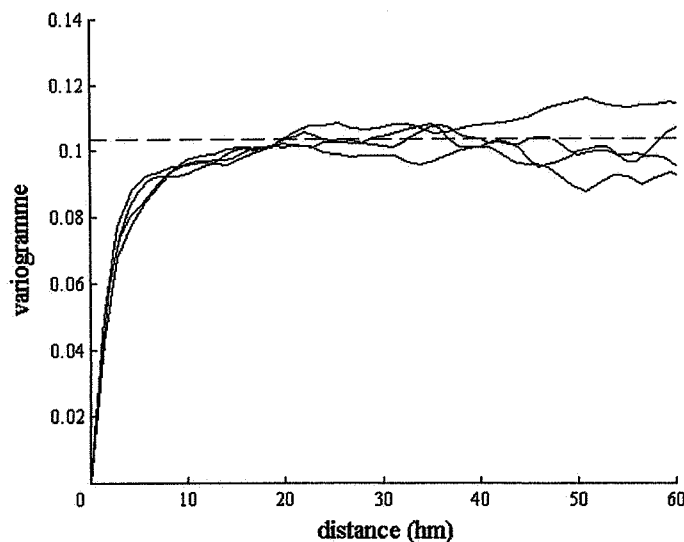
Cette variance, si elle était calculable, mesurerait l'écart que l'on pourrait tolérer entre le variogramme régional et le modèle de variogramme choisi. Le comportement de cette variance de fluctuation n'a été examiné que pour quelques cas particuliers, desquels il ressort que la variance de fluctuation n'est faible (donc le variogramme régional représentatif du variogramme théorique) que lorsque $|\mathbf{h}|$ est petit, et qu'elle s'amplifie considérablement quand $|\mathbf{h}|$ augmente. En pratique, **pour des distances supérieures à la moitié du diamètre du champ, on doit s'attendre à des fluctuations énormes du variogramme régional autour du variogramme vrai, et l'inférence n'est plus possible.**

Illustrons cette remarque en visualisant les valeurs du variogramme régional le long des quatre directions principales du plan pour toutes les valeurs possibles de $|h|$.



*figure 3.26 : variogramme régional dans les directions principales du plan
(nord/sud, est/ouest, nord-est/sud-ouest et nord-ouest/sud-est)*

Comme annoncé, le variogramme régional présente de très fortes fluctuations pour les grandes valeurs de $|h|$. Au-delà de 80 hm, ce qui correspond à peu près à la moitié du diamètre du champ étudié, les fluctuations deviennent démesurées et l'inférence n'est pas envisageable à ces distances. Intéressons-nous de plus près à ce qui se passe aux petites distances.



*figure 3.27 : variogramme régional dans les directions principales du plan
pour des distances comprises entre 0 et 60 hm
(les pointillés représentent la variance statistique des valeurs)*

Une première remarque s'impose : aux distances considérées (0-60 hm), le variogramme régional est **isotrope**. Cette constatation est d'autant plus remarquable que le covariogramme transitif calculé sur les mêmes données montrait une très forte anisotropie (cf. chapitre 1). L'explication tient au fait que ce dernier synthétise les propriétés structurales de la variable régionalisée *et de son champ*, alors que le variogramme est indépendant de la géométrie du champ. On en déduit que l'anisotropie du covariogramme transitif est d'origine purement géométrique (elle est due à la forme étirée du champ, et non à la variable elle-même). Il est utile dans ce cas de calculer le variogramme régional omnidirectionnel, qui servira de base à l'ajustement d'un modèle isotrope.

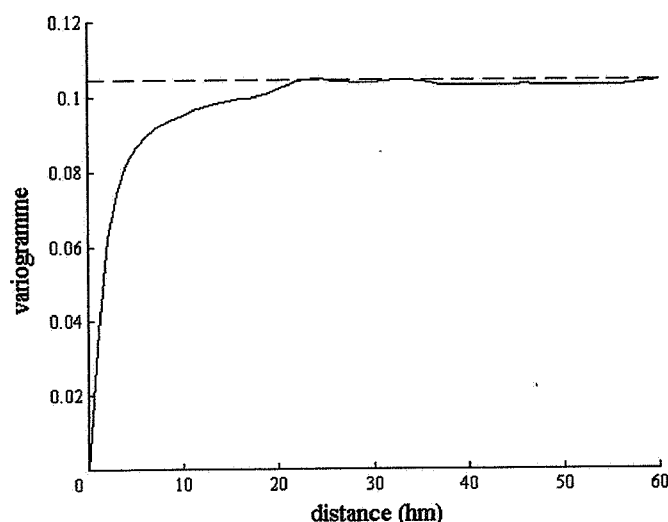


figure 3.28 : variogramme régional omnidirectionnel

Le variogramme régional est continu, mais croît très vite à partir de l'origine (signe que la variable régionalisée se déstructure assez rapidement), et se stabilise autour d'un palier à partir de $|h| \approx 15$ hm. Ce palier est voisin de la variance statistique des données (égale à 0.104), ce qui est conforme à l'hypothèse quasi-stationnaire : $\gamma(\infty) = \sigma^2$. On propose le modèle suivant, composé de deux structures emboîtées : un schéma sphérique de portée 3 hm et de palier 0.04, et un schéma exponentiel de portée pratique 10 hm et de palier 0.06.

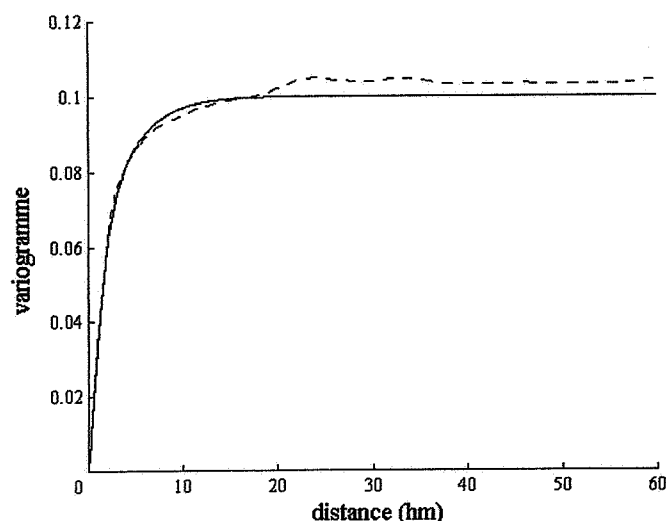
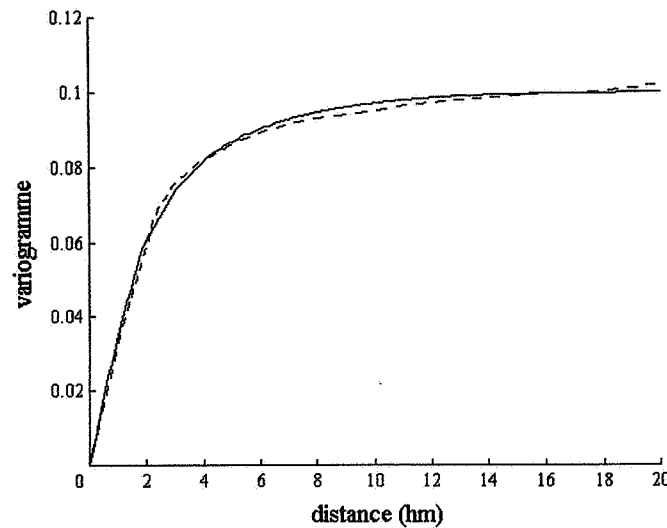


figure 3.29 : variogramme régional omnidirectionnel (pointillés) et son modèle isotrope (ligne continue)

On vérifie notamment que le modèle proposé est en très bon accord avec le variogramme régional aux faibles distances.



*figure 3.30 : variogramme régional omnidirectionnel (pointillés)
et le modèle isotrope de variogramme (ligne continue)*

Etude du variogramme expérimental

En pratique, on ne connaît la réalité que partiellement, et le variogramme régional est inaccessible. Seul un variogramme expérimental est calculable à partir des données disponibles. L'écart entre variogrammes expérimental et régional doit être interprété comme une erreur d'estimation. Dans le modèle de fonction aléatoire, cette erreur est d'espérance nulle, étant donné que le variogramme expérimental et le variogramme régional ont tous deux comme espérance le variogramme théorique. Mais le plus important est la *variance* de l'erreur d'estimation, qui mesure son amplitude potentielle ; cette variance est liée au nombre de points de données disponibles et à leur implantation dans le champ $\mathcal{D}$. On établit qu'elle prend en général des valeurs modérées, c'est-à-dire que **le variogramme expérimental est représentatif du variogramme régional, pourvu que les données soient suffisamment nombreuses et bien réparties dans le champ.**

On propose parfois une règle empirique, selon laquelle un point du variogramme expérimental est fiable si le nombre de couples de données mis en jeu pour son calcul est au moins égal à 30 à 50. Cette règle n'est pas parfaite, car elle ne prend pas en compte le nombre total de données disponibles, qui conditionne le niveau de précision que l'on est en droit d'attendre d'un point expérimental : il est évident que l'on sera plus exigeant si l'on bénéficie de plusieurs milliers d'observations (auquel cas 30 couples ne représentent pas grand chose) que si l'on ne dispose que d'une centaine de données.

Pour reproduire une situation réelle, on a échantillonné de manière régulière le champ au taux de 1%, en sélectionnant une cellule sur 10 dans les directions nord / sud et est / ouest. La figure 3.31 localise les observations et donne leur histogramme.

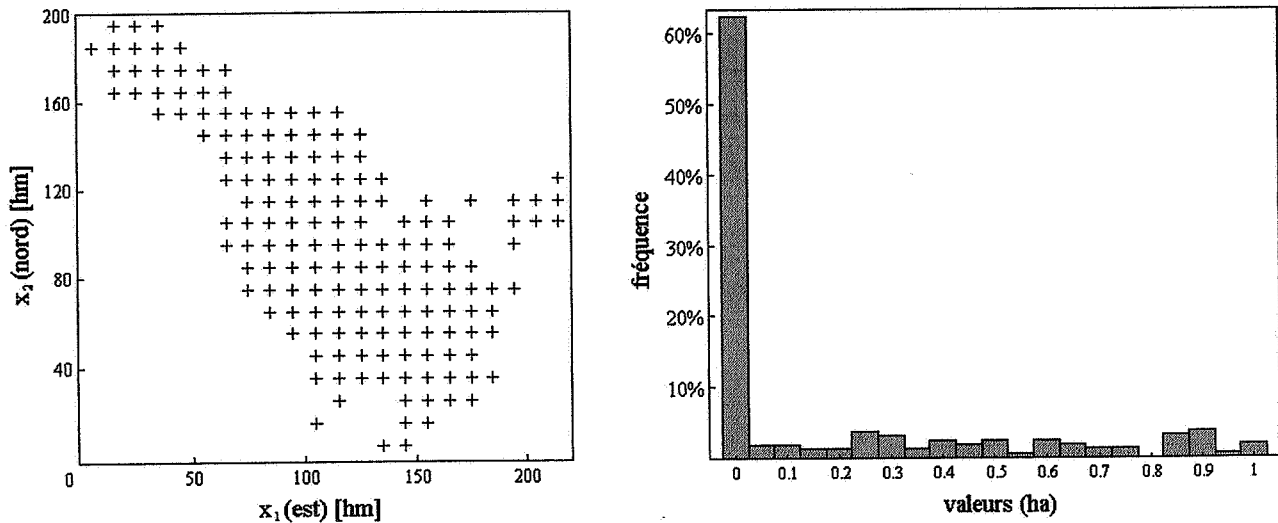


figure 3.31 : carte et histogramme des valeurs échantillonnées

Sur cet échantillon particulier, comptant 162 données, on peut calculer le variogramme expérimental dans les quatre directions principales du plan. On obtient les variogrammes directionnels suivants, où les points expérimentaux sont repérés par des ronds.

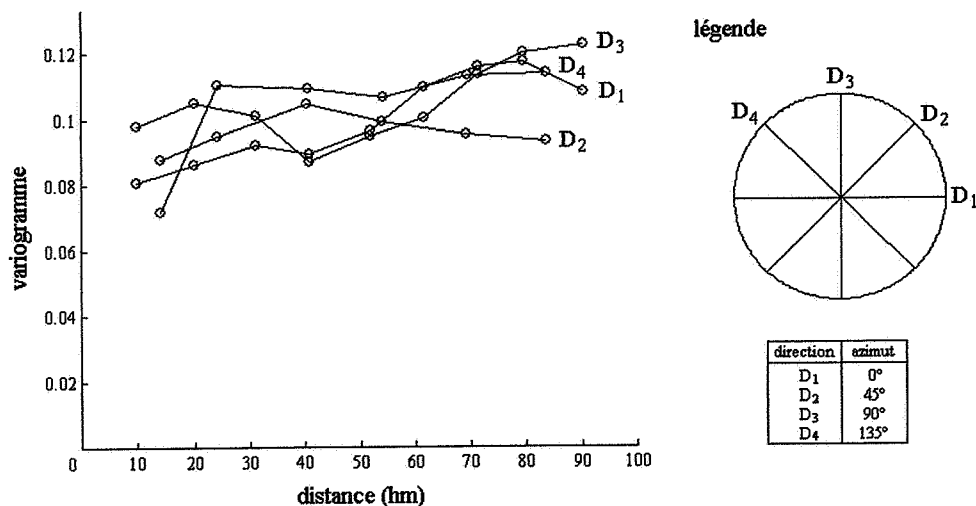


figure 3.32 : variogrammes expérimentaux directionnels

A l'instar du variogramme régional, les variogrammes expérimentaux directionnels se stabilisent autour d'un palier. Les écarts entre les différentes directions doivent être interprétés avec prudence, à cause du faible nombre de données intervenant dans les calculs. Par ailleurs, les chevauchements entre variogrammes directionnels ne permettent pas de distinguer une anisotropie évidente ; dans ce cas, une hypothèse simplificatrice consiste à supposer que la régionalisation est isotrope et à calculer un variogramme omnidirectionnel, sur lequel sera ajusté le modèle définitif (figure 3.33).

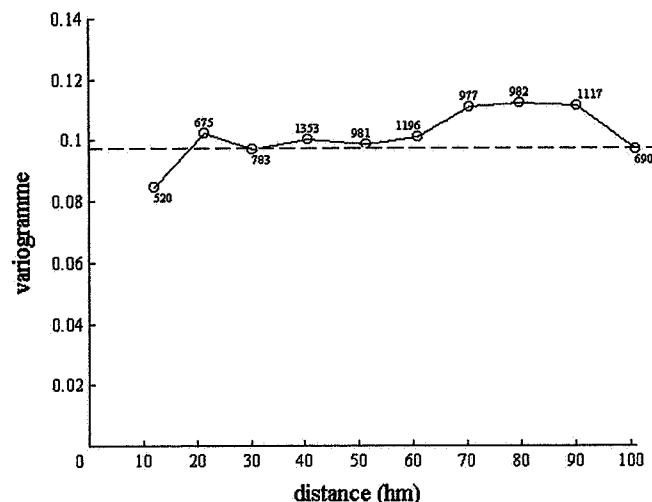


figure 3.33 : variogramme expérimental omnidirectionnel
(les pointillés représentent la variance expérimentale des échantillons)

Les nombres de couples de données intervenant dans le calcul des points expérimentaux sont indiqués sur le schéma ; ils sont tous supérieurs à 500, ce qui indique que ces points expérimentaux peuvent être jugés fiables : le variogramme expérimental est *a priori* une bonne approximation du variogramme régional.

On notera bien que le comportement aux distances inférieures à 10 hm n'est pas accessible, la distance minimale entre deux points de données étant précisément 10 hm. Dans la modélisation, il faut faire appel aux connaissances qualitatives que l'on peut avoir de la variable étudiée pour ajuster un modèle qui soit cohérent au voisinage de l'origine. Ainsi, on sait que la variable étudiée est continue par morceaux (on ne trouve pas de "pépité" de canne à sucre dans la nature) ; cependant, à l'échelle d'observation (à savoir l'hectomètre), ces parcelles sont très morcelées à l'intérieur de la zone, et la régionalisation apparaît presque comme discontinue. A cette échelle, la structure "microscopique" est indétectable et il n'est pas absurde de la modéliser par un effet de pépité. On propose donc comme premier modèle le schéma gigogne suivant : un effet de pépité d'amplitude 0.04 superposé à un schéma sphérique de portée 20 hm et de palier 0.06.

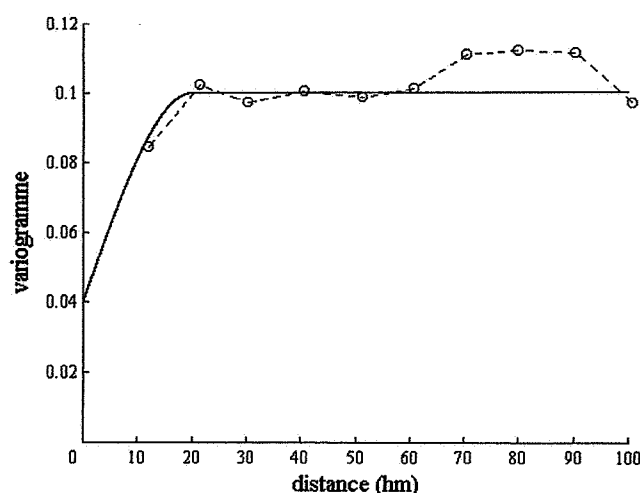
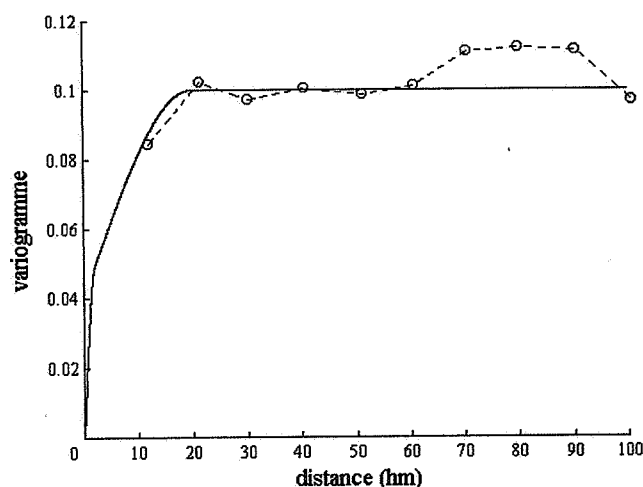


figure 3.34 : variogramme expérimental omnidirectionnel (pointillés)
et modèle isotrope avec effet de pépité (ligne continue)

Ce modèle diffère sensiblement du modèle de variogramme régional : en particulier, la portée du schéma sphérique est plus grande ici (ceci est dû aux aléas de l'échantillonnage et au fait que la portée n'est connue expérimentalement qu'avec une imprécision de 10 hm). De plus, on note que l'effet de pépité possède une amplitude de 0.04 sur un total de 0.1, c'est-à-dire que 40% de la variabilité correspond, dans ce modèle, à une absence de structuration, ce qui est peu favorable.

On peut améliorer l'ajustement du variogramme à l'origine si l'on ajoute l'information supplémentaire (connue par exemple d'après des études d'années antérieures) que la superficie moyenne d'une parcelle de canne à sucre est légèrement supérieure à 1.5 ha, soit un peu plus d'une cellule. En effet, ceci implique qu'une parcelle empiète, en moyenne, sur 4 à 6 cellules ; sa zone d'influence sera par conséquent d'environ 200 m. On peut alors remplacer l'effet de pépité par un schéma sphérique de portée 200 m, qui explique mieux la structuration des données aux courtes distances. Sur le variogramme régional, on avait d'ailleurs trouvé un schéma sphérique de 300 m, ce qui corrobore cette analyse.



*figure 3.35 : variogramme expérimental omnidirectionnel (pointillés)
et modèle isotrope sans effet de pépité (ligne continue)*

A l'issue de l'analyse variographique, la structure des données est décrite et modélisée. Les modèles de variogramme vont permettre de résoudre les problèmes d'estimation (globale et locale) que le praticien est amené à poser.

V.6.2. Un exemple non stationnaire : les données de sol

A titre de second exemple, nous allons mener l'analyse variographique des données de sol présentées au chapitre 2. On peut d'ores et déjà s'attendre à des complications par rapport au cas précédent, l'étude exploratoire de ces données ayant montré qu'une hypothèse de stationnarité risquait de ne pas être adaptée. Nous allons envisager ici une modélisation en admettant la validité de l'hypothèse intrinsèque, mais l'exemple sera repris lors de l'étude des processus non stationnaires (cf. chapitre 6).

Nuée variographique

La figure 3.36 présente la nuée variographique, calculée de manière isotrope, i.e. en fonction de la seule distance $|h|$ et non de la direction. On observe que la dispersion de la nuée a tendance à augmenter avec la distance : ceci signifie que l'on est d'autant plus susceptible de trouver des couples de données dissemblables lorsque leur séparation est grande. Les points les plus élevés de la nuée sont repérés par des ronds ; les couples de données correspondants sont localisés sur une carte.

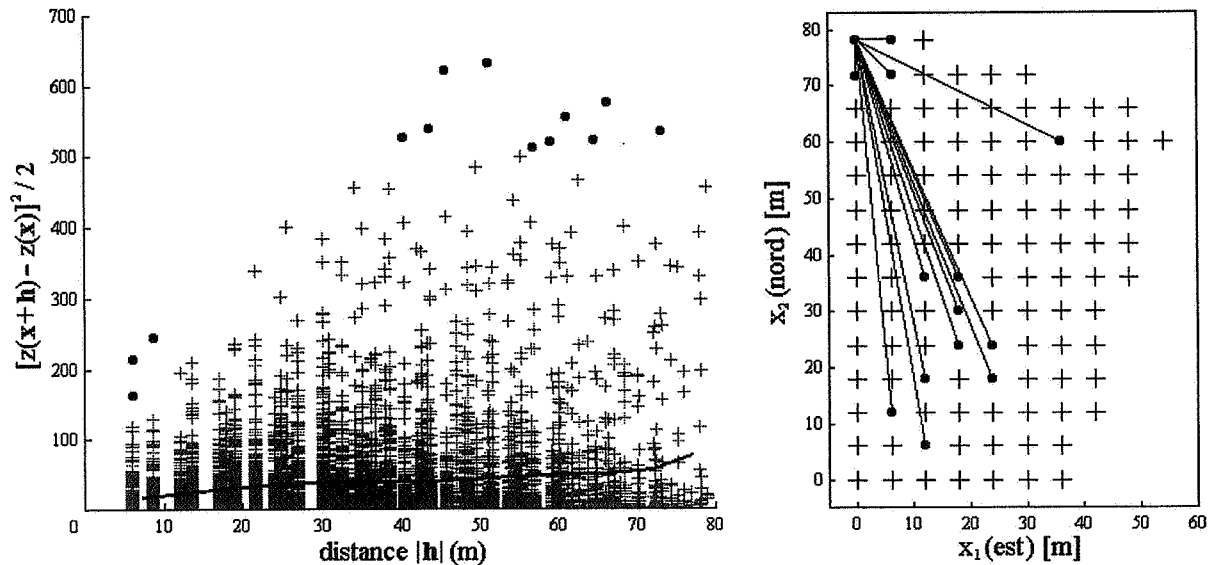


figure 3.36 : nuée variographique de la variable % argile + limon (à gauche) et carte d'implantation correspondante (à droite)

Tous les points mis en évidence sur la nuée variographique sont créés par la donnée située dans le coin en haut à gauche, qui présente une valeur notablement plus faible que les autres. Cette valeur faible est responsable d'écarts importants par rapport aux autres valeurs.

Variogrammes expérimentaux

Les variogrammes expérimentaux selon les directions 0° (D_1), 45° (D_2), 90° (D_3) et 135° (D_4) par rapport à la direction est-ouest ont été calculés pour des distances multiples de la maille d'échantillonnage (à savoir, 6 mètres pour D_1 et D_3 et 8.5 mètres pour D_2 et D_4). La variance empirique des données, égale à 37.9, est signalée par des pointillés.

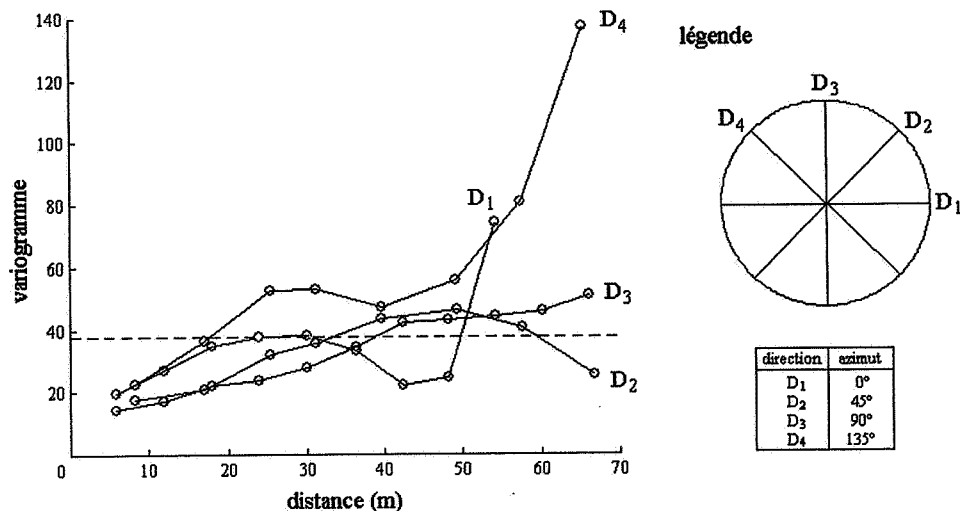


figure 3.37 : variogrammes selon les directions
 0° (D_1), 45° (D_2), 90° (D_3) et 135° (D_4)

Les variogrammes directionnels sont notablement différents, ce qui est symptomatique d'une anisotropie. En gardant cette remarque à l'esprit, on peut néanmoins calculer un variogramme omnidirectionnel ; celui-ci affiche un comportement croissant prononcé (bien supérieur à la variance expérimentale des données) et ne se stabilise pas autour d'un palier, montrant le caractère non stationnaire de la variable.

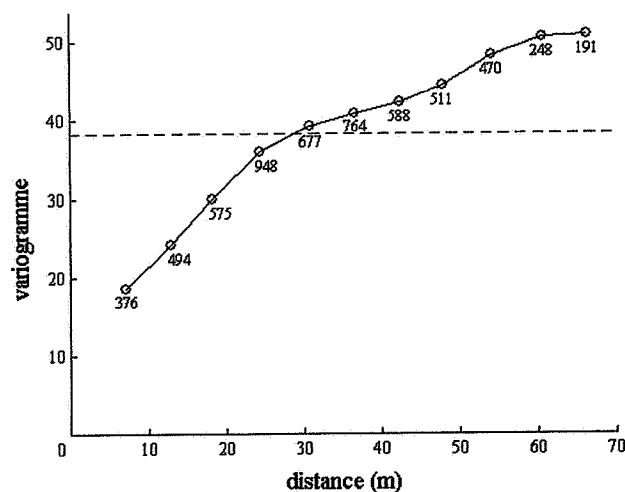
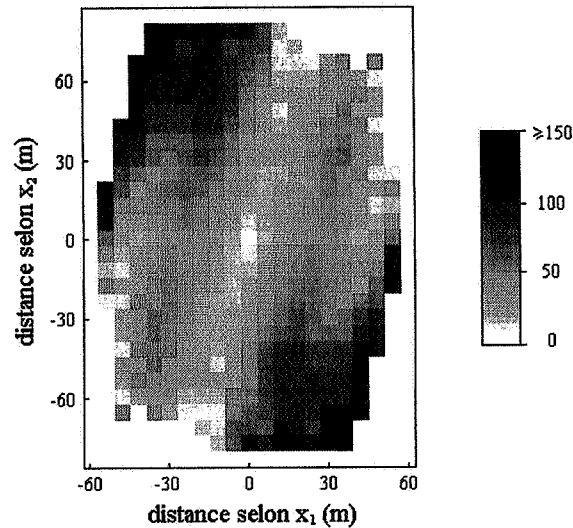


figure 3.38 : variogramme omnidirectionnel
 les nombres de paires de données qui interviennent dans les calculs sont indiqués à côté de chaque point expérimental

Le calcul de la “carte variographique” permet de dessiner une carte bidimensionnelle des valeurs du variogramme expérimental et de mieux visualiser l’anisotropie, notamment son angle et son rapport dans le cas d’une anisotropie géométrique. La figure 3.39 représente la carte variographique en niveaux de gris. Elle met en évidence une forte anisotropie d’angle $\theta \approx 70^\circ$ par rapport à l’axe est-ouest.



*figure 3.39 : étude de l'anisotropie de la variable
% argile + limon (carte variographique)*

Remarques sur le plan d'échantillonnage

Il est important à ce stade de donner quelques éléments concernant l'échantillonnage des valeurs mesurées. En effet, c'est à partir d'un schéma d'échantillonnage de la zone étudiée que l'on obtient les données nécessaires à l'estimation de la variable régionalisée. Il est donc essentiel de bien préparer cette partie du projet géostatistique. A ce titre, il est instructif de revenir sur le plan d'échantillonnage des données de sol, et de le compléter.

Dans cet exemple, les échantillons sont disposés sur une grille régulière. Ce choix est motivé par des considérations pratiques : on recherche une homogénéité de la reconnaissance sur le champ étudié, préférable à une répartition aléatoire des observations qui laisserait des zones mal reconnues. A cela s'ajoute le fait qu'en pratique les échantillonnages réguliers conduisent à des estimations globalement plus précises que les échantillonnages irréguliers. En général, la maille carrée, ou rectangulaire en cas d'anisotropie attendue, est préférée à une maille hexagonale ou triangulaire, plus difficile à mettre en œuvre sur le terrain ; les dimensions du carré ou du rectangle de base permettent alors de déterminer, selon les directions, la plus petite distance pour laquelle on peut évaluer le variogramme expérimental. Ainsi, pour les données de sol, si l'on s'en tient au réseau régulier d'observations, le comportement du variogramme pour des distances inférieures à six mètres (longueur de la maille carrée) est inaccessible.

Pour évaluer le comportement du variogramme à plus petite échelle, on a décidé de compléter le plan d'échantillonnage. Plus précisément, on a placé, en trois endroits tirés au hasard dans la zone, des "croix de sondage" où les variables étudiées ont été échantillonnées sur une maille plus dense (1 mètre). La figure 3.40 montre le plan d'échantillonnage complet pour les données de sol, avec les trois croix de sondage.

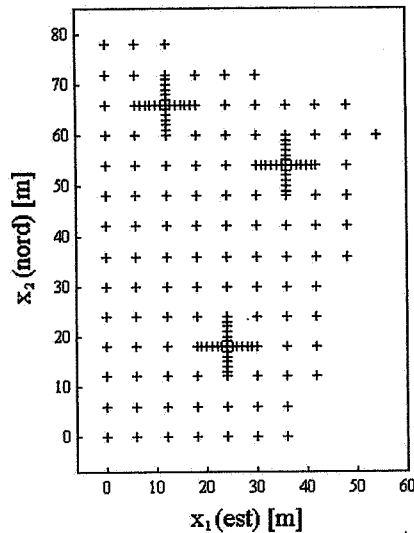


figure 3.40 : plan d'échantillonnage complet
avec trois croix de sondage (170 données au total)

Comportement aux courtes distances des variogrammes expérimentaux

Les croix de sondage étant implantées de manière non préférentielle, il est intéressant de calculer les variogrammes expérimentaux directionnels des données issues de ces croix de sondage. En effet, le pas de calcul étant plus petit (1 mètre au lieu des 6 mètres sur la grille régulière), on est en mesure d'étudier le comportement des variogrammes expérimentaux à l'échelle du mètre.

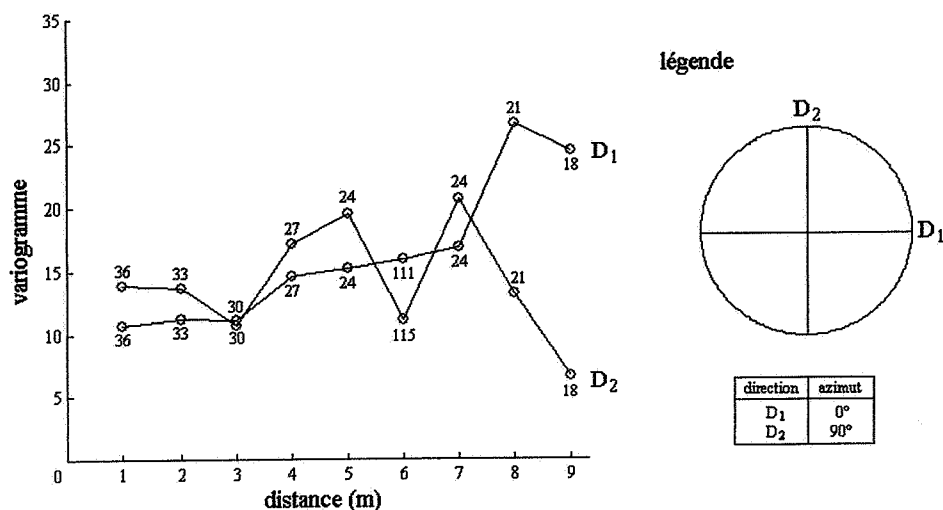


figure 3.41 : variogrammes directionnels le long des croix de sondage

Ces variogrammes ne tendent pas vers zéro aux petites distances. Ils révèlent la présence d'une structure de portée inférieure au mètre et d'amplitude approximativement égale à 10. Faute de connaître sa portée exacte, cette "microstructure" sera modélisée par un effet de pépite.

Ajustement à un modèle

Il convient de chercher un modèle conforme aux constatations précédentes, à savoir anisotrope et non stationnaire. Nous avons réalisé un ajustement à un modèle puissance en $|h|^{1.8}$, à anisotropie géométrique d'angle 70° et de rapport 1/5. Pour améliorer l'ajustement aux petites distances, un effet de pépite de palier 9.4 a été ajouté, ainsi qu'un schéma sphérique de palier 20.9 et portée respectivement égale à 40 m et 20 m le long des axes d'anisotropie. Signalons dès à présent qu'un tel modèle, bien que permis par la théorie, est à la limite de la classe des modèles autorisés par l'hypothèse intrinsèque (le variogramme est presque parabolique aux grandes distances). Le but de cet exemple est avant tout pédagogique ; au cours d'une étude réelle, on aura sans doute intérêt à rechercher des modèles plus généraux que ceux permis par l'hypothèse intrinsèque (cf. chapitre 6).

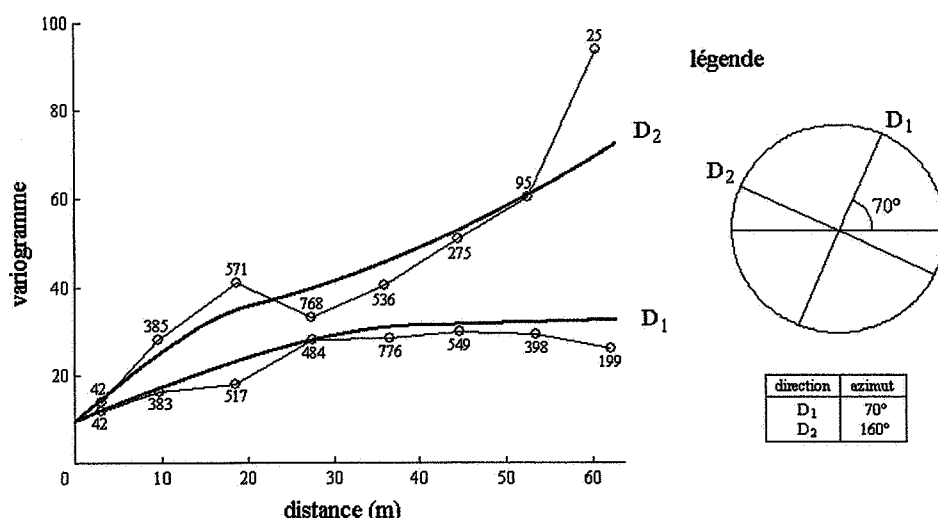


figure 3.42 : ajustement à un modèle anisotrope
 (D₁ : direction nord-nord-est – sud-sud-ouest / D₂ : direction ouest-nord-ouest – est-sud-est)
 lignes brisées : variogrammes directionnels expérimentaux ; lignes épaisses : modèle

Le modèle s'ajuste relativement bien aux valeurs expérimentales. On ne pourra jamais dire si ce modèle est le "vrai". Cette proposition n'a même pas de sens, car, rappelons-le, il n'existe pas de véritable modèle ; même si l'on connaissait exhaustivement la réalité, on ne connaîtrait pas le variogramme qui n'a pas de signification objective.

En revanche, il est possible de soumettre le modèle adopté à une validation croisée, qui va permettre de juger de sa qualité et sa pertinence.

Validation croisée

La figure 3.43 donne les tests graphiques de validation croisée antérieurement mentionnés. L'estimation de chaque donnée est réalisée à partir des échantillons avoisinants, en se limitant à trois échantillons par octant d'espace (cf. chapitre 5, travail en voisinage glissant). Les valeurs mal estimées, au nombre de quatre ici, sont repérées par des ronds et replacées sur la carte d'implantation des données. Trois de ces valeurs proviennent de la zone nord-ouest, qui présente de fortes variations entre échantillons voisins (cf. l'étude exploratoire des données au chapitre 2).

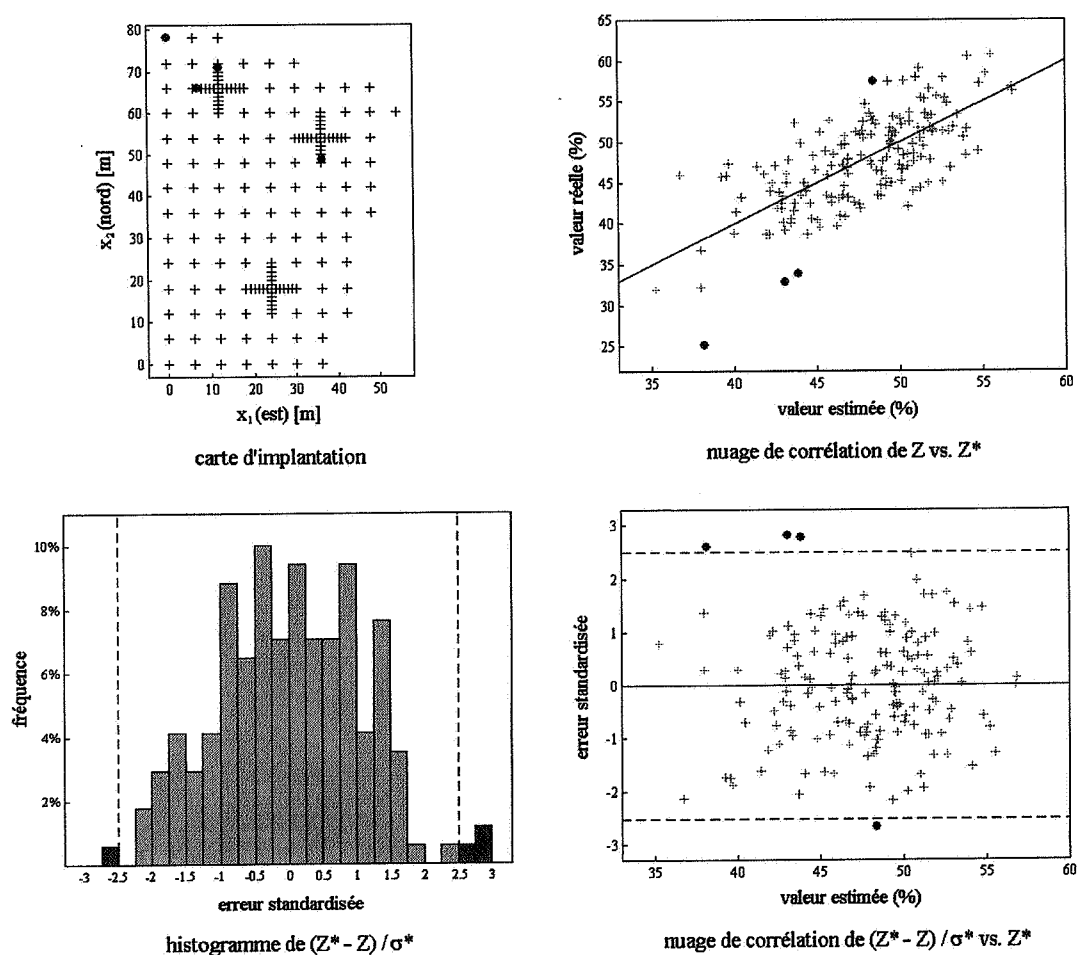


figure 3.43 : tests graphiques pour la validation croisée des données de sol

Les principales statistiques relatives aux erreurs et aux erreurs standardisées sont résumées dans les tableaux 3.1 et 3.2.

	moyenne	variance
erreurs	0.069	17.23
erreurs standardisées	0.016	1.14

tableau 3.1 : statistiques de validation croisée basées sur les 170 données

	moyenne	variance
erreurs	-0.073	14.93
erreurs standardisées	-0.016	1.00

tableau 3.2 : statistiques de validation croisée basées sur les 166 données robustes (erreurs standardisées situées dans l'intervalle $[-2.5 ; 2.5]$)

La moyenne des erreurs est faible et la variance des erreurs standardisées est proche de 1, ce qui est satisfaisant, d'autant que seules quatre données (soit 2.4% du total) ont été mal estimées.

V.6.3. Un exemple dans l'espace à trois dimensions : les données minières

Cette fois-ci, les données disponibles se situent dans l'espace à trois dimensions et sont distribuées irrégulièrement, ce qui rend le choix des paramètres de calcul plus complexe : pour que les résultats expérimentaux soient fiables, il sera indispensable de définir des tolérances sur les distances et sur les angles.

La première tâche consiste à identifier les directions d'anisotropie, si celles-ci existent. L'outil le plus commode est sans doute la visualisation de la carte variographique, ou plus exactement de ses sections principales, horizontale et verticales, passant par l'origine. Le calcul de la carte variographique est généralement effectué aux nœuds d'une grille régulière, dont l'utilisateur doit définir la taille de la maille ; lorsque les données sont irrégulièrement réparties, on procède par essais successifs jusqu'à aboutir à un résultat interprétable. En l'occurrence, on prend une maille de longueur 20m dans le plan horizontal et 12m (hauteur des composites) dans la direction verticale.

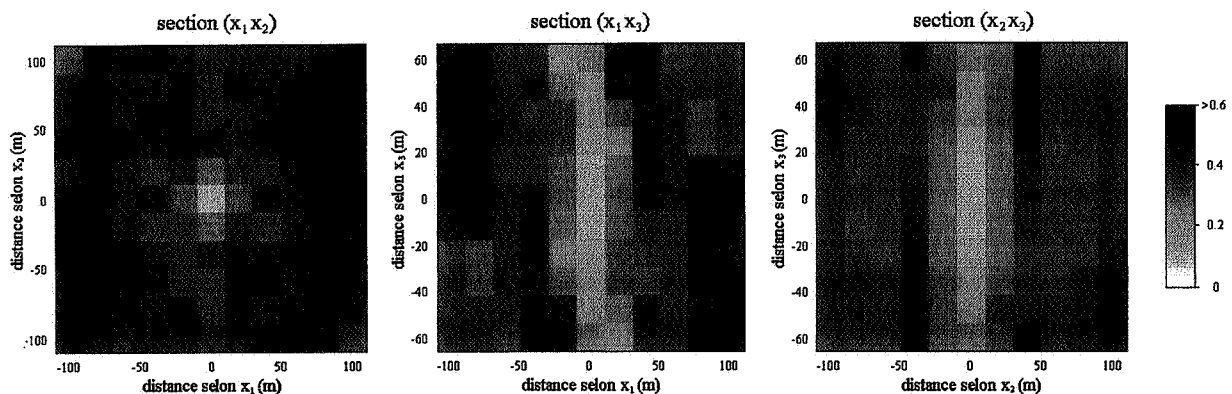


figure 3.44 : carte variographique des données minières
visualisation des trois sections principales

La section horizontale ne révèle aucune anisotropie évidente dans ce plan. En revanche, les sections verticales montrent des “bandes” qui indiquent que la structuration des données n’est pas la même dans la direction verticale que dans le plan horizontal. On est donc amené à calculer deux variogrammes directionnels : l’un “omnihorizontal”, qui mélange toutes les directions horizontales, l’autre vertical.

Variogramme expérimental

	pas	nombre de pas	tolérance sur les distances	tolérance angulaire	hauteur de tranche
horizontal	20m	11	10m	90°	2m
vertical	12m	9	6m	8°	-

tableau 3.3 : paramètres de calcul du variogramme expérimental

A part le premier point le long de chaque direction et les derniers points le long de la direction verticale, les nombres de paires de données intervenant dans les calculs sont toujours supérieurs à 1000, indiquant que les points expérimentaux sont fiables.

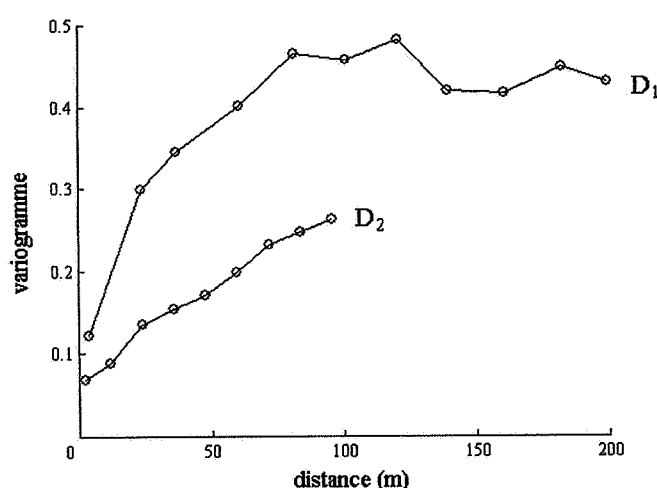


figure 3.45 : variogramme expérimental omnihorizontal (D_1) et vertical (D_2)

Ces variogrammes directionnels mettent en évidence une anisotropie zonale : la valeur du palier est différente selon la direction de calcul. On remarque aussi la présence d'un effet de pépite, d'amplitude inférieure à 0.1, commun aux deux variogrammes directionnels.

Modélisation

La modélisation d'un tel variogramme se fait en deux étapes. On ajuste d'abord, par un modèle isotrope, la direction de moindre palier, ici la direction verticale. Puis, on ajoute une structure gigogne à anisotropie zonale pure, nulle le long de la direction verticale (portée verticale infinie) : ainsi, le modèle n'est pas perturbé dans cette direction et peut être ajusté au variogramme expérimental omnihorizontal. Ici, on obtient un ajustement satisfaisant avec le modèle suivant, composé d'un effet de pépite, un schéma sphérique isotrope et un schéma exponentiel :

$$\forall \mathbf{h} \neq \mathbf{0}, \gamma(\mathbf{h}) = 0.06 + 0.18 \text{sph}(100\text{m}, 100\text{m}) + 0.2 \exp(40\text{m}, \infty)$$

où la première portée correspond au plan horizontal et la seconde à la direction verticale.

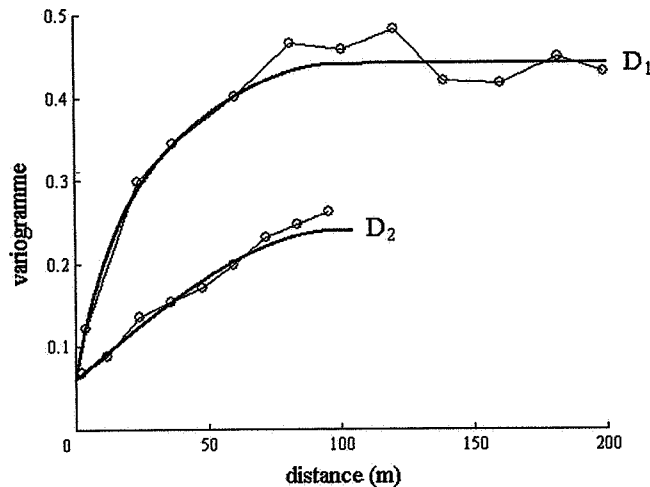


figure 3.46 : *variogrammes expérimental et modélisé le long des directions principales d'anisotropie*

Pour parfaire l'ajustement, on pourrait éventuellement modifier la définition du schéma sphérique, en utilisant des portées horizontale et verticale différentes, mais cette étape est inutile ici.

Le modèle est ensuite mis à l'épreuve à l'aide d'une validation croisée.

Validation croisée

Les tests graphiques de validation croisée sont présentés sur la figure 3.47. L'estimation de chaque donnée n'utilise que les 24 échantillons les plus proches, à raison de 3 par octant d'espace (*cf.* chapitre 5, krigeage en voisinage glissant). Parmi les 2376 données, seules 55 (soit 2.3% du total) ont été "mal estimées", ce qui est fort satisfaisant. La plupart de ces dernières se situent dans les zones de fortes valeurs mises en évidence au cours de l'étude exploratoire (*cf.* chapitre 2) : ce sont aussi les zones de plus forte variabilité, ce qui explique que les estimations sont moins précises.

Le quatrième graphique, qui représente les erreurs standardisées en fonction des valeurs estimées, mérite un commentaire. Le nuage de points correspondant est horizontal, centré autour de la droite d'ordonnée nulle, ce qui indique que l'estimation ne souffre pas de biais conditionnel (la moyenne des erreurs commises est approximativement nulle, indépendamment de la valeur estimée).

On note toutefois que la dispersion du nuage augmente avec la valeur estimée : le nuage a une forme triangulaire (dans le langage statistique, on dit qu'il y a hétéroscédasticité). C'est un symptôme de l'*effet proportionnel* mentionné au chapitre 2 : dans les zones de fortes valeurs, qui correspondent à celles où les valeurs estimées sont fortes, la variabilité est plus importante, donc les erreurs d'estimation ont tendance à présenter une amplitude plus grande.

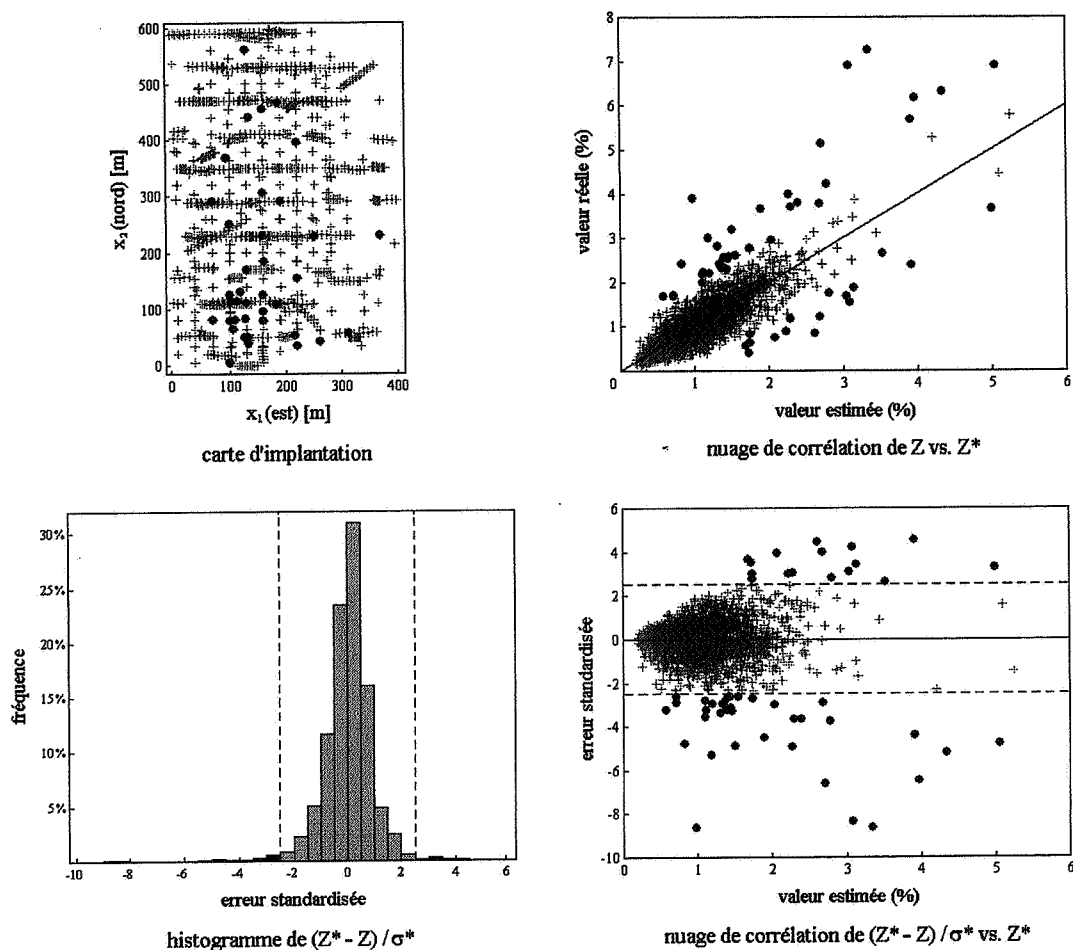


figure 3.47 : tests graphiques de validation croisée pour les données minières

	moyenne	variance
erreurs	0.0033	0.126
erreurs standardisées	0.0062	0.974

tableau 3.4 : statistiques de validation croisée basées sur les 2376 données

	moyenne	variance
erreurs	0.0193	0.071
erreurs standardisées	0.0501	0.572

tableau 3.5 : statistiques de validation croisée basées sur les 2321 données robustes (erreurs standardisées situées dans l'intervalle $[-2.5 ; 2.5]$)

Les estimations sont globalement sans biais (erreurs de moyenne presque nulle). Par ailleurs, les erreurs standardisées ont une variance proche de 1, ce qui montre que le modèle de variogramme permet d'apprécier correctement l'amplitude des erreurs commises. On note également que la variance des erreurs commises diminue de plus de 40% lorsque l'on ôte les données non robustes ; cela indique que l'incertitude se concentre surtout dans les zones correspondantes, qui sont celles de fortes valeurs.

Alternatives pour l'analyse variographique

La variance des données, bien que localement constante, varie d'une zone à l'autre du champ ; à cause de l'effet proportionnel mis en évidence au chapitre 2, on sait qu'elle est plus élevée dans les zones de fortes teneurs que dans celles de faibles teneurs. Cette remarque n'est pas compatible avec un variogramme uniforme sur le champ, car l'existence d'un palier unique implique que le processus admet une variance constante et égale à ce palier.

Nous proposons ci-après quelques alternatives pour compléter la description variographique des données. Les deux premières (variogramme modulé et variogrammes locaux) cherchent à intégrer l'effet proportionnel dans le modèle variographique. Les trois suivantes (variogrammes relatifs, logarithmique, pondéré) ont pour but d'améliorer la description de la structuration des données et la "lisibilité" du variogramme expérimental, au cas où celui-ci serait trop erratique ou exagérément pépétique¹³.

Citons deux possibilités offertes à l'utilisateur pour prendre en compte l'effet proportionnel dans le modèle :

- moduler le palier du variogramme selon une fonction de la moyenne locale, en posant :

$$\forall \mathbf{x} + \mathbf{h} \in V, \gamma(\mathbf{x} + \mathbf{h}, \mathbf{x}) = \varphi[m^*(\mathbf{x})] \gamma_0(\mathbf{h})$$

où V est un bloc centré en $\mathbf{x}$

$m^*(\mathbf{x})$ est la moyenne expérimentale locale en $\mathbf{x}$ (locale = à l'échelle du bloc V)

γ_0 est le variogramme ajusté globalement.

Les caractéristiques structurales sont invariantes sur le champ (elles sont décrites par le variogramme γ_0), seul change le niveau de variabilité locale donné par le palier du variogramme. Il s'agit du modèle le plus simple qui intègre l'effet proportionnel. L'ajustement de la fonction φ se fait en tenant compte de la régression existant entre les moyennes et variances locales (cf. chapitre 2). Nous reviendrons sur ce modèle lors de la mise en œuvre du krigeage (chapitre 5).

- considérer des variogrammes locaux : on partitionne le domaine d'étude en plusieurs sous-zones, dans chacune desquelles on calcule et ajuste un variogramme propre. La comparaison de ces modèles locaux permettra éventuellement de trouver une formule mathématique pour spécifier le variogramme sur tout le champ, par exemple de la forme :

$$\forall \mathbf{x} + \mathbf{h} \in V, \gamma(\mathbf{x} + \mathbf{h}, \mathbf{x}) = \varphi_0[m^*(\mathbf{x})] \gamma_0(\mathbf{h}) + \varphi_1[m^*(\mathbf{x})] \gamma_1(\mathbf{h}) + \dots + \varphi_s[m^*(\mathbf{x})] \gamma_s(\mathbf{h})$$

où $\gamma_0, \gamma_1, \dots, \gamma_s$ sont les structures gigognes utilisées dans la modélisation des variogrammes locaux et V est un bloc centré en $\mathbf{x}$ de support égal à celui des sous-zones locales.

On pourra mettre en évidence un effet proportionnel propre à chaque structure gigogne (ainsi, il arrive que l'effet proportionnel affecte la composante pépétique et la composante continue du variogramme de manière différente). Un inconvénient de cette approche est que le nombre de données servant au calcul de chaque variogramme local est réduit, d'où un manque de robustesse, et qu'il n'est pas possible d'obtenir un modèle de variogramme aux grandes distances (celles supérieures au diamètre de V).

¹³ S'il est associé à un *échantillonnage préférentiel*, l'effet proportionnel tend à déstructurer le variogramme expérimental : aux courtes distances, le variogramme expérimental n'est en effet représentatif que des zones densément échantillonnées, qui sont souvent celles de fortes valeurs et de forte dispersion ; en conséquence, la variabilité est surestimée aux courtes distances, ce qui se traduit par un effet de pépète exagéré, voire prédominant.

A titre d'exemple, la figure 3.48 donne les variogrammes locaux correspondant à trois sous-zones disjointes du champ, de taille $200\text{m} \times 200\text{m} \times 130\text{m}$ (des sous-zones plus petites conduisent à des variogrammes trop erratiques, par manque de données). On observe que, dans les zones 1 et 3, de valeurs moyennes et variances élevées, l'anisotropie est plus marquée et la part de l'effet de pépité moindre que dans la zone 2, de valeur moyenne intermédiaire. Le modèle précédent (variogramme modulé) est par conséquent encore simpliste, et il conviendrait de rechercher un modèle plus sophistiqué, dans lequel les caractères structuraux changent selon la région du champ considérée.

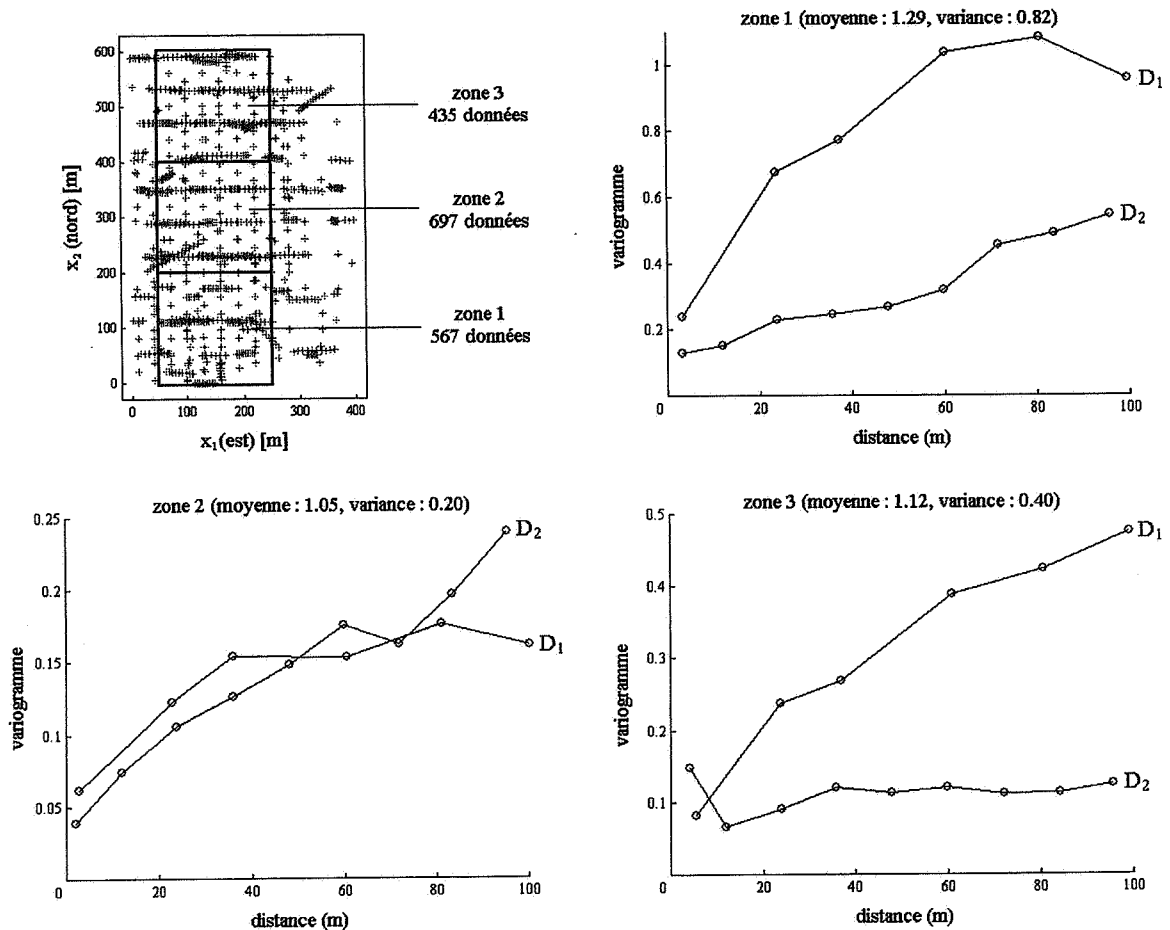


figure 3.48 : carte d'implantation des données (en haut à gauche)
et variogrammes locaux associés à trois sous-zones disjointes
 D_1 : plan horizontal, D_2 : direction verticale

Les trois méthodes qui suivent cherchent à améliorer la description du phénomène, au cas où le variogramme expérimental est peu stable ou n'est pas représentatif de la structuration de la variable régionalisée (cas où l'échantillonnage est préférentiel dans les zones de fortes valeurs par exemple, ce qui peut avoir pour effet de déstructurer le variogramme expérimental) :

- utiliser les variogrammes relatifs : pour améliorer la description de la structure spatiale, sans recourir à une partition du champ, certains auteurs recommandent l'utilisation de "variogrammes relatifs". Les plus connus sont les variogrammes relatifs général et par paires.

▪ *variogramme relatif général*

Cet outil consiste à normer la valeur du variogramme en h par une fonction de la moyenne des données séparées de h :

$$\hat{\gamma}_{RG}(h) = \frac{\hat{\gamma}(h)}{\phi[\hat{m}(h)]}$$

$$\text{avec } \hat{m}(h) = \frac{1}{2 |N(h)|} \sum_{N(h)} [z(x_\alpha) + z(x_\beta)]$$

$N(h) = \{ (\alpha, \beta) \text{ tel que } x_\alpha - x_\beta = h \}$, avec éventuellement des tolérances sur les distances et les angles lorsque l'échantillonnage est irrégulier.

Typiquement, on choisit comme fonction ϕ une parabole : $\phi[\hat{m}(h)] = [\hat{m}(h)]^2$.

Il ne faut pas confondre la moyenne $\hat{m}(h)$, qui se réfère à une séparation particulière, avec la moyenne locale $m^*(x)$ précédemment définie, qui se réfère à l'espace géographique. Ainsi, il n'y a pas de lien direct entre le variogramme relatif général et le variogramme "modulé" vu précédemment.

Si les données sont réparties de manière régulière, $\hat{m}(h)$ sera égale à leur moyenne globale, et le variogramme relatif ne diffèrera du variogramme classique que d'un facteur multiplicatif (la moyenne au carré).

Au contraire, si les données sont irrégulièrement réparties, $\hat{m}(h)$ sera égale, pour les petites valeurs de h , à la valeur moyenne des zones densément échantillonnées ; si celles-ci coïncident avec les zones de fortes valeurs (échantillonnage préférentiel), le variogramme relatif divisera le variogramme classique par une constante élevée. Lorsque la distance augmente, $\hat{m}(h)$ devient plus représentative de la moyenne globale, et la normalisation du variogramme classique est moins importante. On assiste donc à une diminution de l'effet de pépité relatif, ce qui permet de corriger la déstructuration du variogramme expérimental classique due à un échantillonnage préférentiel en présence d'un effet proportionnel.

▪ *variogramme relatif par paires*

La normalisation intervient au niveau de chaque paire de données :

$$\hat{\gamma}_{RP}(h) = \frac{1}{2 |N(h)|} \sum_{N(h)} \frac{[z(x_\alpha) - z(x_\beta)]^2}{[z(x_\alpha) + z(x_\beta)]^2 / 4}$$

avec toujours $N(h) = \{ (\alpha, \beta) \text{ tel que } x_\alpha - x_\beta = h \}$.

Les variogrammes relatifs sont plus robustes que le variogramme classique à la présence de valeurs extrêmes, dont l'influence est réduite à cause de la normalisation effectuée. Des problèmes de définition se posent lorsque la variable $z(x)$ peut s'annuler, auquel cas le dénominateur peut être nul.

Le principal inconvénient de ces variogrammes relatifs est qu'ils ne peuvent être utilisés tels quels dans le formalisme géostatistique : les équations permettant de résoudre les problèmes d'estimation font appel au variogramme ou à la covariance, pas aux variogrammes relatifs. Au mieux, ces derniers fournissent un supplément d'information et améliorent la description de la structuration spatiale des données (portée et anisotropie notamment).

Sur les données de cuivre, on observe que les variogrammes relatifs conservent l'anisotropie zonale constatée sur le variogramme classique. En revanche, le palier est modifié dans les deux cas, et la portée apparente est plus élevée dans le cas du variogramme relatif par paires ; en outre, la part de l'effet de pépite a légèrement diminué par rapport au variogramme classique.

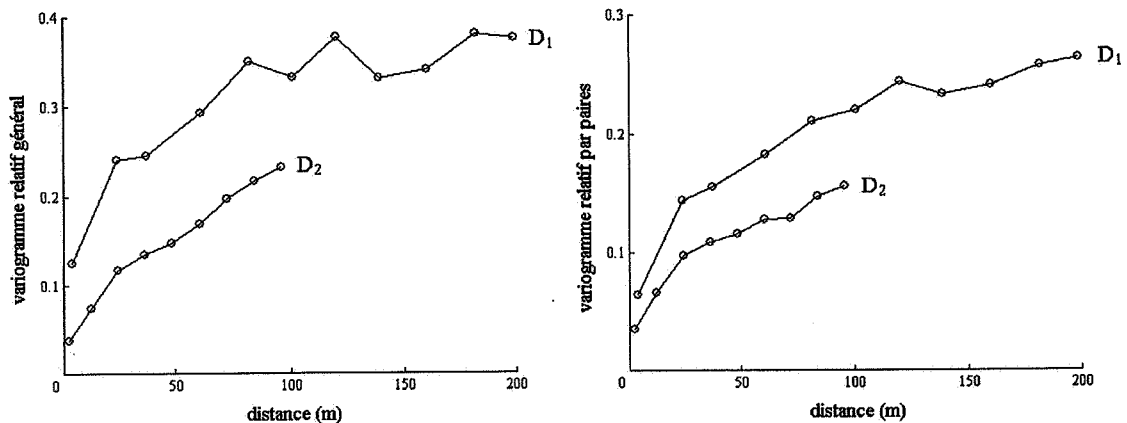


figure 3.49 : variogrammes relatifs général (à gauche) et par paires (à droite) sur les données minières ; D_1 : horizontal, D_2 : vertical

- utiliser le variogramme logarithmique

Au lieu de s'intéresser à la variable $Z(\mathbf{x})$ elle-même, il peut être avantageux d'effectuer l'analyse variographique d'une transformée de cette variable. La transformation la plus couramment utilisée est le passage au logarithme : $Y(\mathbf{x}) = \ln [Z(\mathbf{x})]$ ou, plus généralement, $Y(\mathbf{x}) = \ln [Z(\mathbf{x}) + \beta]$ où β est un scalaire ; cette transformation a comme particularité d'atténuer les différences entre les valeurs fortes et d'estomper, voire faire disparaître, l'effet proportionnel.

Le problème principal concerne le retour à la variable $Z(\mathbf{x})$. Pour pouvoir déduire le variogramme γ_Z de cette dernière à partir du variogramme logarithmique γ_Y , on est obligé de faire des hypothèses sur les lois mono et bi-variables de la fonction aléatoire $Y(\mathbf{x})$, ce qui sort du cadre de la géostatistique linéaire. L'hypothèse la plus courante est celle d'une distribution monovariante gaussienne et d'une loi bivariable bigaussienne. Il existe plusieurs tests permettant de vérifier cette dernière assertion ; en particulier, le variogramme d'ordre 1 ou "madogramme"

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

est, sous l'hypothèse bigaussienne, proportionnel au variogramme usuel $\gamma_Y(\mathbf{h})$, plus précisément :

$$\frac{\sqrt{\gamma_Y(\mathbf{h})}}{\gamma_Y^{(1)}(\mathbf{h})} = \sqrt{\pi} \quad (\text{indépendant de } \mathbf{h}).$$

Ce test de binormalité est satisfait expérimentalement sur les données de cuivre (figure 3.50). En outre, on a vu (cf. chapitre 2) que les valeurs des logarithmes suivent une distribution approximativement normale, de sorte que l'hypothèse gaussienne est aussi vérifiée au niveau de la loi monovariante.

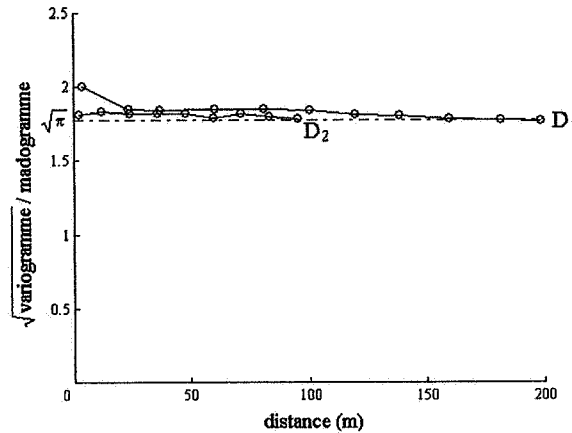


figure 3.50 : test de binormalité du logarithme : racine carré du variogramme divisée par le madogramme en fonction de la séparation, le long des directions horizontale (D_1) et verticale (D_2)

On peut utiliser la formule¹⁴ suivante, valide sous l'hypothèse bigaussienne, liant les variogrammes arithmétique et logarithmique :

$$\gamma_Z(\mathbf{h}) = m_Z^2 \exp(\sigma_Y^2) \{1 - \exp[-\gamma_Y(\mathbf{h})]\}$$

où $m_Z = E(Z)$

$$\sigma_Y^2 = \text{var}(Y) = \ln[1 + \text{var}(Z) / m_Z^2].$$

Par ailleurs, si l'on considère que la moyenne m_Z est localement constante mais varie à l'échelle du champ, et que la variance $\text{var}(Z)$ est proportionnelle à m_Z^2 (ce qui est approximativement vrai, d'après la régression trouvée au chapitre 2), alors σ_Y^2 est constante, ce qui signifie que l'effet proportionnel ne se manifeste pas sur la variable logarithmique. Le variogramme logarithmique ne dépend que de la séparation, tandis que le variogramme arithmétique est proportionnel au carré de la moyenne locale :

$$\gamma_Z(\mathbf{x} + \mathbf{h}, \mathbf{x}) = m_Z^2(\mathbf{x}) \exp(\sigma_Y^2) \{1 - \exp[-\gamma_Y(\mathbf{h})]\}.$$

Cette relation est à rapprocher du modèle précédent, où le palier du variogramme est modulé par une fonction de la moyenne expérimentale locale ; elle peut éventuellement être utilisée comme modèle.

Sur les données minières, les variogrammes logarithmique et arithmétique ressemblent davantage au variogramme relatif par paires qu'au variogramme expérimental classique (la portée horizontale est plus élevée sur les premiers que sur le dernier). Ils sont presque identiques, mais cette similitude est fortuite.

¹⁴ Cette formule est valable dans quelques autres modèles particuliers, où le logarithme n'obéit pas à une loi bigaussienne. Elle peut aussi être étendue au cas où Y est une fonction aléatoire intrinsèque, à condition que m_Z et σ_Y soient considérés comme inconnus ; dans ce cas, bien que γ_Y n'admet ni palier ni portée, il n'est en pas de même de γ_Z , qui reste un variogramme de transition. Par exemple, si γ_Y est un modèle linéaire de pente ω , γ_Z sera un schéma exponentiel de portée pratique $3/\omega$; toutefois, si le diamètre du champ est petit par rapport à cette portée, γ_Z aura l'apparence d'un variogramme linéaire, sans portée apparente.

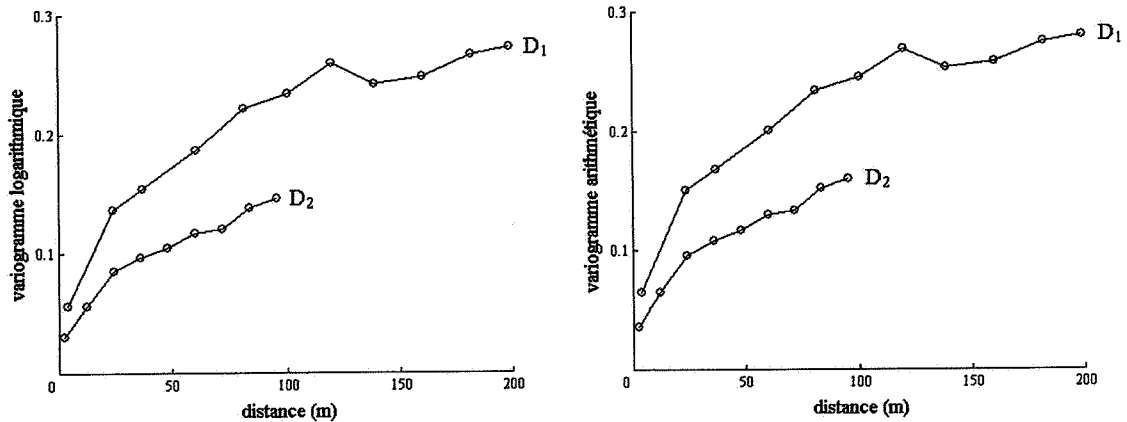


figure 3.51 : variogrammes logarithmique γ_Y (à gauche) et arithmétique γ_Z (à droite) sur les données minières ; D_1 : horizontal, D_2 : vertical

- pondérer les données dans le calcul du variogramme

La pondération des données est une opération fréquente dans le calcul des histogrammes et des statistiques monovariées (moyenne, variance, quantiles...), cf. chapitre 2. Elle est cependant peu répandue pour le calcul des variogrammes, bien que ce soit un outil très peu restrictif.

Supposons que chaque point de donnée $\mathbf{x}_\alpha$ ($\alpha = 1 \dots n$) soit affecté d'un poids ω_α ; on peut attribuer à chaque paire de données un poids proportionnel au produit des poids individuels, ce qui conduit à un estimateur du variogramme de la forme :

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2} \frac{\sum_{N(\mathbf{h})} \omega_\alpha \omega_\beta [z(\mathbf{x}_\alpha) - z(\mathbf{x}_\beta)]^2}{\sum_{N(\mathbf{h})} \omega_\alpha \omega_\beta}$$

avec $N(\mathbf{h}) = \{ (\alpha, \beta) \text{ tel que } \mathbf{x}_\alpha - \mathbf{x}_\beta \approx \mathbf{h} \}$.

Cette formule donne une estimation non biaisée du variogramme théorique ; elle permet de corriger, en partie du moins, les effets dus aux irrégularités et au caractère préférentiel de l'échantillonnage. Contrairement au variogramme logarithmique, aucune hypothèse supplémentaire sur la loi bivariable n'est requise. On retrouve le variogramme expérimental classique en prenant des poids identiques pour toutes les données.

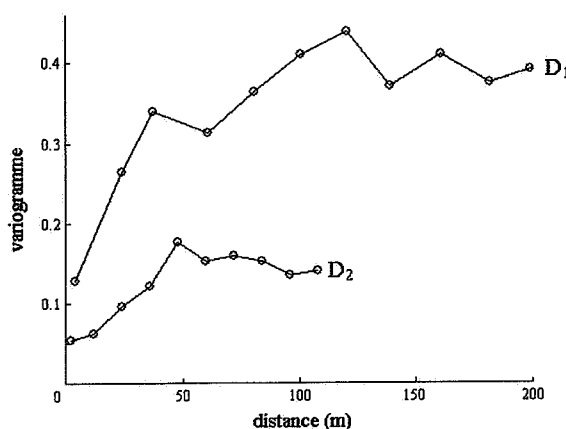


figure 3.52 : variogramme pondéré, le long des directions horizontale (D_1) et verticale (D_2)

Résumé

En géostatistique linéaire, la structure spatiale de la variable régionalisée, interprétée comme une réalisation d'une fonction aléatoire, est modélisée par une fonction de **covariance** ou par un **variogramme**. Ces outils mesurent la corrélation entre les valeurs de la variable en fonction de leur séparation, et décrivent de manière synthétique les caractéristiques du phénomène régionalisé : variabilité à courte distance et régularité spatiale (*comportement à l'origine*), zone d'influence d'une observation (*portée*), structurations d'échelles différentes (*schéma gigogne*), directions préférentielles (*anisotropie*).

Le variogramme est souvent préféré à la covariance, car plus général (il peut être défini hors de l'hypothèse de stationnarité d'ordre deux) et d'inférence plus fiable. Sa connaissance va notamment permettre de résoudre les problèmes d'estimation.

L'**analyse variographique**, c'est-à-dire l'inférence et la modélisation du variogramme à partir des données expérimentales, est l'étape essentielle d'une étude géostatistique. Il convient de porter une attention particulière au choix des paramètres de calcul du variogramme expérimental. En général, la modélisation utilise le *modèle linéaire de régionalisation*, qui consiste à superposer plusieurs modèles théoriques de base. Mais l'ajustement d'un modèle ne doit jamais être automatique ; il fait au contraire appel au sens physique de l'utilisateur, à ses connaissances qualitatives et son expérience du phénomène. Un contrôle de l'adéquation du modèle aux données disponibles est possible grâce à la technique de **validation croisée**.

Chapitre 4 : changement de support et estimation globale

En géostatistique intrinsèque¹, l'estimation globale consiste à évaluer la valeur moyenne de la variable régionalisée sur la totalité du champ et à mesurer la précision de l'estimation, compte tenu du modèle variographique et de la disposition géométrique des échantillons. Un tel problème constitue en général une étape préliminaire dans une étude géostatistique, et a vocation à être complété par des estimations locales, afin de distinguer les zones du champ de fortes valeurs et celles de valeurs plus faibles.

Lorsque les données sont nombreuses et leur répartition suffisamment uniforme sur le champ $\mathcal{D}$, l'expérience montre qu'il n'est guère utile de proposer des estimateurs sophistiqués et que l'on peut se contenter d'estimateurs très simples : pour fixer les idées, ce sera le plus souvent la moyenne arithmétique des observations, parfois une moyenne pondérée si l'on souhaite tenir compte des irrégularités d'échantillonnage. Par ailleurs, des estimateurs comme le krigeage par blocs (cf. chapitre 5) requièrent, pour leur mise en œuvre, une hypothèse de stationnarité, laquelle risque de ne pas être valide à l'échelle du champ. Une alternative consiste à réaliser l'estimation de petits secteurs, puis à calculer la moyenne de ces estimations locales, mais cette approche ne permet pas de connaître la précision de l'estimation globale obtenue.

En géostatistique linéaire, où l'on ne manipule que les deux premiers moments de la fonction aléatoire associée à la variable régionalisée, la précision d'une estimation est mesurée par la variance de l'erreur commise (appelée *variance d'estimation*), qui quantifie sa plus ou moins grande dispersion autour de la quantité à estimer. **L'intérêt de la géostatistique par rapport à la statistique classique est que le calcul de la variance d'estimation prend en compte la structure spatiale de la régionalisation et la configuration géométrique des observations.**

Nous allons au préalable poser les bases de la théorie de la régularisation d'une variable sur un support non ponctuel, et définir des notions qui seront utiles pour calculer les variances d'estimation globale en géostatistique intrinsèque.

¹ On rappelle que la *géostatistique intrinsèque* désigne la partie probabiliste de la géostatistique, celle qui fait appel au concept de fonction aléatoire (par opposition à la géostatistique transitive).

I. Les outils du changement de support

On désigne par $Z(\mathbf{x})$ la fonction aléatoire associée à la variable régionalisée $z(\mathbf{x})$, $\mathcal{D}$ son champ, C sa fonction de covariance et γ son variogramme. Dans un premier temps, on se placera dans le cadre **stationnaire du deuxième ordre** et on utilisera comme outil structural la fonction de covariance C . Puis on généralisera les résultats obtenus au cadre **intrinsèque**, en remplaçant C par $-\gamma$ dans l'expression des variances des combinaisons linéaires autorisées (celles de poids total nul).

I.1. Changement de support (régularisation)

Soit V un domaine borné inclus dans $\mathcal{D}$. On note $Z(V)$ la variable aléatoire obtenue en faisant la moyenne de $Z(\mathbf{x})$ sur V :

$$Z(V) = \frac{1}{|V|} \int_V Z(\mathbf{x}) d\mathbf{x}.$$

$Z(V)$ est aussi appelée *régularisée* de la variable ponctuelle $Z(\mathbf{x})$ sur le support V . Dans le cadre stationnaire d'ordre deux, $Z(V)$ admet la même espérance que $Z(\mathbf{x})$. Déterminons ses moments du second ordre.

- **variance**

La variance de la variable régularisée s'exprime à l'aide de la fonction de covariance ponctuelle :

$$\text{var}[Z(V)] = \bar{C}(V, V) = \frac{1}{|V|^2} \int_V \int_V C(\mathbf{x} - \mathbf{y}) d\mathbf{x} d\mathbf{y}.$$

Cette variance est toujours inférieure à la variance *a priori* de $Z(\mathbf{x})$, égale à $C(0)$, d'après l'inégalité de Cauchy-Schwarz :

$$|C(\mathbf{x} - \mathbf{y})| \leq C(0) \Rightarrow |\bar{C}(V, V)| \leq C(0).$$

C'est une illustration de l'**effet de support** : les valeurs des blocs V sont moins dispersées que les valeurs ponctuelles.

- **covariance**

On définit, pour deux domaines bornés V et V' inclus dans $\mathcal{D}$, la covariance

$$\text{cov}[Z(V), Z(V')] = \bar{C}(V, V') = \frac{1}{|V||V'|} \int_V \int_{V'} C(\mathbf{x} - \mathbf{y}) d\mathbf{x} d\mathbf{y}.$$

Ainsi, la fonction de covariance C_V de la variable régularisée $Z(V)$ s'écrit :

$$C_V(\mathbf{h}) = \text{cov}[Z(V), Z(V_h)] = \bar{C}(V, V_h)$$

où V_h désigne le translaté du support V par le vecteur $\mathbf{h}$.

Tout comme la covariance ponctuelle, C_V ne dépend que du vecteur de séparation $\mathbf{h}$, ce qui montre que $Z(V)$ est stationnaire du second ordre. Sauf de rares exceptions, son **comportement à l'origine** est plus régulier que celui de C (par exemple, parabolique alors que C est linéaire à l'origine), d'où le nom de *régularisation* donné au changement de support. Enfin, sa **portée** est égale à celle de C augmentée du diamètre de V .

En effet, si $Z(\mathbf{x})$ et $Z(\mathbf{x} + \mathbf{h})$ sont non corrélées pour $|\mathbf{h}| > a$ (portée de C), alors $Z(V)$ et $Z(V_h)$ seront non corrélées lorsque tout point de V et tout point de V_h seront distants d'au moins a , donc lorsque les centres de V et V_h seront séparés de $a' = a + \text{diamètre de } V$.

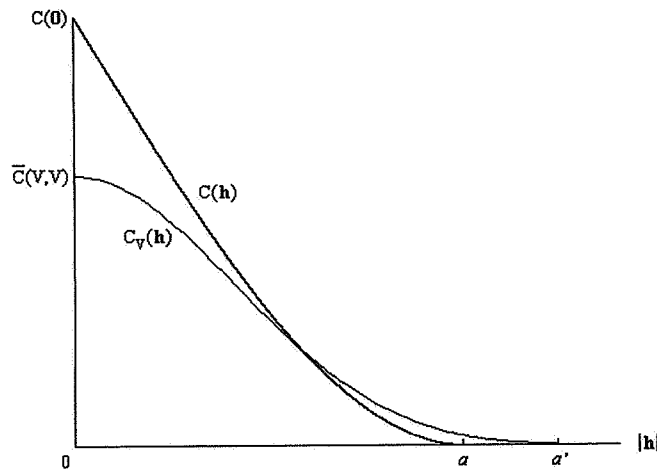


figure 4.1 : covariances ponctuelle (trait épais) et régularisée (trait fin)

- **variogramme**

Il est également possible de déterminer le variogramme γ_V de la régularisée $Z(V)$ à l'aide des (co)variances de blocs :

$$\gamma_V(\mathbf{h}) = \frac{1}{2} \text{var}[Z(V_h) - Z(V)] = \bar{C}(V, V) - \bar{C}(V, V_h).$$

Cette expression peut se réécrire en termes de variogrammes :

$$\gamma_V(\mathbf{h}) = \bar{\gamma}(V, V_h) - \bar{\gamma}(V, V)$$

où $\bar{\gamma}(V, V) = \frac{1}{|V|^2} \int_V \int_V \gamma(\mathbf{x} - \mathbf{y}) d\mathbf{x} d\mathbf{y} = C(0) - \bar{C}(V, V)$

$$\bar{\gamma}(V, V_h) = \frac{1}{|V|^2} \int_V \int_{V_h} \gamma(\mathbf{x} - \mathbf{y}) d\mathbf{x} d\mathbf{y}.$$

Cette dernière formule a l'avantage de rester valide dans le cadre intrinsèque strict, c'est-à-dire lorsque $Z(\mathbf{x})$ n'admet pas de covariance.

Le **comportement à l'origine** de γ_V est plus régulier que celui de γ . En ce qui concerne son **comportement à l'infini**, on peut écrire, lorsque $|\mathbf{h}|$ est grand vis-à-vis des dimensions du bloc V :

$$\bar{\gamma}(V, V_h) = \frac{1}{|V|^2} \int_V \int_{V_h} \gamma(\mathbf{x} - \mathbf{y}) \, d\mathbf{x} \, d\mathbf{y} \approx \frac{1}{|V|^2} \int_V \int_{V_h} \gamma(\mathbf{h}) \, d\mathbf{x} \, d\mathbf{y} = \gamma(\mathbf{h})$$

de sorte que l'on a, asymptotiquement :

$$\gamma_V(\mathbf{h}) \underset{|\mathbf{h}| \rightarrow +\infty}{\sim} \gamma(\mathbf{h}) - \bar{\gamma}(V, V).$$

Aux grandes distances, le variogramme régularisé se déduit du variogramme ponctuel par soustraction du terme constant $\bar{\gamma}(V, V)$ qui ne dépend que du support V et donne en quelque sorte une mesure de l'importance de l'effet de support.

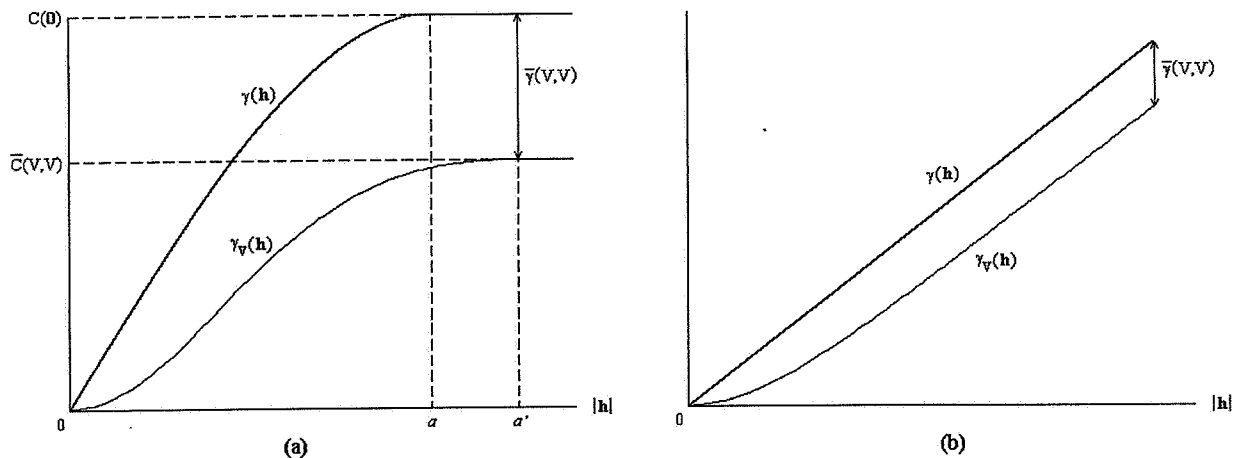


figure 4.2 : variogrammes ponctuel (trait épais) et régularisé (trait fin)
(a) : cadre stationnaire, (b) : cadre intrinsèque strict

En pratique, pour passer du variogramme ponctuel au variogramme régularisé, on conserve le même type de modèle (sphérique, exponentiel...), et l'on corrige le palier et la portée. Cette procédure est approximative ; on sait en particulier que le comportement à l'origine est altéré par un changement de support. Elle ne se justifie que lorsque le diamètre de V est petit vis-à-vis de l'échelle de travail considérée (V est par exemple le support des échantillons, supposé petit). Son avantage est qu'elle permet d'obtenir de manière simple un modèle autorisé pour γ_V (fonction de type négatif conditionnel).

Théorie générale de la régularisation

On peut généraliser les résultats précédents en définissant une variable régularisée par une fonction de pondération p par :

$$Z_p(\mathbf{x}) = \int Z(\mathbf{x} + \mathbf{h}) p(\mathbf{h}) d\mathbf{h}$$

où l'on suppose en général $\int p(\mathbf{h}) d\mathbf{h} = 1$.

La variable régularisée apparaît comme une moyenne mobile pondérée de la variable ponctuelle. On retrouve le changement de support classique, en considérant une fonction de pondération égale à l'indicatrice du bloc centré à l'origine, divisée par la mesure de ce bloc.

On peut écrire Z_p sous la forme d'un produit de convolution :

$$Z_p = Z * \check{p}$$

où $\check{p}$ est l'application transposée de p : $\check{p}(\mathbf{h}) = p(-\mathbf{h})$.

On note P le covariogramme transitif de p , tel qu'il a été défini au chapitre 1 :

$$P = p * \check{p}.$$

Alors, la fonction de covariance C_p et le variogramme γ_p de la régularisée s'expriment en fonction de ceux de la variable ponctuelle sous la forme suivante :

$$C_p = C * P$$

$$\gamma_p = \gamma * P - \text{constante telle que } \gamma_p(\mathbf{0}) = 0.$$

Ces formules permettent de déduire la covariance ou le variogramme des valeurs moyennes à partir de ceux de mesures ponctuelles, ou réciproquement (ce qui est plus délicat).

Remarque : régularisation d'un effet de pépite

La formule du changement de support classique, appliquée à une covariance ponctuelle purement pépitique, donne une covariance régularisée identiquement nulle, ce qui signifie que les valeurs des blocs (aussi petits soient-ils) ne sont plus aléatoires. Cette situation n'est physiquement pas acceptable et demande à être corrigée.

Si l'on assimile l'effet de pépite à une microstructure, de portée très petite mais positive, la covariance régularisée ne sera pas toujours nulle. Les calculs montrent que sa valeur à l'origine (variance) sera égale à la portée intégrale de la microstructure divisée par le volume du bloc considéré. Quant à sa portée, elle sera pratiquement égale au diamètre du bloc ; si les distances mises en jeu sont toutes supérieures à ce diamètre, la covariance régularisée se comportera exactement comme un effet de pépite, mais d'amplitude moindre.

Une seconde description consiste à représenter l'effet de pépite par une mesure de Dirac (cf. chapitre 3). La covariance régularisée est alors un effet de pépite d'amplitude inversement proportionnelle au support considéré, ce qui rejoint la description précédente. Cela revient à considérer que les mesures ne sont jamais exactement ponctuelles, et que la covariance $C(\mathbf{h})$ est déjà une régularisée de la mesure de Dirac.

I.2. Variance d'extension et variance d'estimation

On appelle **variance d'extension** d'un domaine V' à un domaine V , et on note $\sigma_E^2(V, V')$, la variance de la différence² $Z(V') - Z(V)$:

$$\sigma_E^2(V, V') = \text{var}[Z(V') - Z(V)].$$

On peut interpréter cette définition en considérant $Z(V)$ comme une moyenne à estimer et $Z(V')$ comme son estimateur ; $Z(V') - Z(V)$ est alors l'*erreur d'estimation*, d'espérance nulle (sous réserve de stationnarité d'ordre deux ou intrinsèque) et de variance égale à la variance d'extension $\sigma_E^2(V, V')$.

On peut expliciter l'expression de cette variance à l'aide du formalisme précédent :

$$\sigma_E^2(V, V') = \overline{C}(V, V) + \overline{C}(V', V') - 2\overline{C}(V, V').$$

Cette relation montre que la variance d'extension de V à V' dépend :

- de la fonction de covariance C , qui modélise la structure de la variable régionalisée (en particulier sa régularité spatiale)
- de la forme du domaine V à estimer, par l'intermédiaire du terme $\overline{C}(V, V)$
- de la forme du domaine V' qui sert d'estimateur, par l'intermédiaire de $\overline{C}(V', V')$
- de la position relative des domaines V et V' , par l'intermédiaire de $\overline{C}(V, V')$.

On notera que la variance d'extension ne dépend pas des valeurs prises par la variable régionalisée, et que la quantité à estimer et l'estimateur jouent un rôle symétrique : on aura la même précision si l'on estime $Z(V)$ par $Z(V')$ ou réciproquement.

On peut exprimer la variance d'extension à l'aide du variogramme γ de $Z(\mathbf{x})$:

$$\sigma_E^2(V, V') = 2\overline{\gamma}(V, V') - \overline{\gamma}(V, V) - \overline{\gamma}(V', V')$$

où $\overline{\gamma}(V, V')$ est la valeur moyenne de $\gamma(\mathbf{x} - \mathbf{y})$ lorsque $\mathbf{x}$ et $\mathbf{y}$ décrivent respectivement V et V' . Cette formule est moins restrictive que celle utilisant la fonction de covariance C , car elle reste valable dans le cadre strictement intrinsèque, où la covariance n'existe pas et seul le variogramme est défini.

² A titre de curiosité, la variance d'extension d'un point $\mathbf{x}$ à un point $\mathbf{x} + \mathbf{h}$ n'est fonction que de $\mathbf{h}$ et vaut $\sigma_E^2(\mathbf{x}, \mathbf{x} + \mathbf{h}) = \text{var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = 2\gamma(\mathbf{h})$. La variance d'extension apparaît ainsi comme une généralisation du variogramme au cas de variables non ponctuelles.

Le formalisme des variances d'extension est le plus général qui soit pour calculer une variance d'estimation : $Z(V)$ est la quantité à estimer, et $Z(V')$ son estimateur, sans contrainte sur V ni sur V' . Le cas qui nous intéressera dans l'estimation globale est celui où le domaine V est le champ $\mathcal{D}$ de la variable régionalisée et le domaine V' les n sites échantillonnés ; $Z(V)$ est alors la moyenne exhaustive sur $\mathcal{D}$ et $Z(V')$ la moyenne arithmétique des échantillons :

$$Z(V') = Z^* = \frac{1}{n} \sum_{\alpha=1}^n Z(x_{\alpha}).$$

La formule générale de la variance d'extension s'écrit :

$$\begin{aligned} \text{var}[Z^* - Z(\mathcal{D})] &= \frac{1}{n^2} \sum_{\alpha=1}^n \sum_{\beta=1}^n C(x_{\alpha} - x_{\beta}) - 2 \sum_{\alpha=1}^n \frac{1}{n|\mathcal{D}|} \int_{\mathcal{D}} C(x_{\alpha} - x) dx \\ &\quad + \frac{1}{|\mathcal{D}|^2} \int_{\mathcal{D}} \int_{\mathcal{D}} C(x - y) dx dy \end{aligned}$$

ou, en termes de variogramme :

$$\begin{aligned} \text{var}[Z^* - Z(\mathcal{D})] &= -\frac{1}{n^2} \sum_{\alpha=1}^n \sum_{\beta=1}^n \gamma(x_{\alpha} - x_{\beta}) + 2 \sum_{\alpha=1}^n \frac{1}{n|\mathcal{D}|} \int_{\mathcal{D}} \gamma(x_{\alpha} - x) dx \\ &\quad - \frac{1}{|\mathcal{D}|^2} \int_{\mathcal{D}} \int_{\mathcal{D}} \gamma(x - y) dx dy \end{aligned}$$

Cette expression donne ce que l'on appelle la **variance d'estimation**³ de $Z(\mathcal{D})$ par les données $Z(x_{\alpha})$. En règle générale, cette variance est d'autant plus faible que le réseau de prélèvements est dense et réparti de manière homogène dans le champ, et que le variogramme est régulier, donc que la variable régionalisée est continue. En revanche, elle est indépendante des valeurs observées ; on dit aussi que la variance d'estimation n'est pas *conditionnelle*. Ceci permet, lorsqu'on établit un réseau d'échantillonnage, de prévoir à l'avance sa qualité dans l'estimation globale, à condition cependant de connaître le modèle variographique.

On peut généraliser la formule de la variance d'estimation au cas où l'estimateur est une moyenne pondérée des données :

$$Z^* = \sum_{\alpha=1}^n \omega_{\alpha} Z(x_{\alpha})$$

Dans ce cas, on a :

$$\text{var}[Z^* - Z(\mathcal{D})] = -\sum_{\alpha=1}^n \sum_{\beta=1}^n \omega_{\alpha} \omega_{\beta} \gamma(x_{\alpha} - x_{\beta}) + 2 \sum_{\alpha=1}^n \frac{\omega_{\alpha}}{|\mathcal{D}|} \int_{\mathcal{D}} \gamma(x_{\alpha} - x) dx - \frac{1}{|\mathcal{D}|^2} \int_{\mathcal{D}} \int_{\mathcal{D}} \gamma(x - y) dx dy$$

Les pondérateurs $\{\omega_{\alpha}, \alpha = 1 \dots n\}$ peuvent être ceux employés pour corriger l'histogramme des effets dus aux irrégularités de l'échantillonnage et à son caractère préférentiel (cf. chapitre 2).

³ La terminologie de "variance d'estimation" peut parfois prêter à confusion : il s'agit de la variance de l'erreur d'estimation, et non de la variance de l'estimateur.

L'expression de la variance d'estimation est difficile à calculer numériquement pour peu que le nombre d'échantillons soit important ou que la géométrie de $\mathcal{D}$ soit compliquée. Nous allons voir, dans la seconde partie de ce chapitre, comment se simplifie cette formule pour des types d'échantillonnage particuliers. Mais auparavant, nous avons besoin de définir la notion de variance de dispersion.

I.3. Variance de dispersion

I.3.1. Première définition

Soient V un domaine de l'espace géographique et v un domaine plus petit contenu dans V . On cherche à mesurer la *dispersion* dans V des valeurs de la variable $Z(v)$ de support v . On va d'abord examiner le cas particulier où v est réduit à un point, et définir la dispersion de la variable ponctuelle $Z(x)$ lorsque x parcourt le domaine V .

Variance de dispersion d'un point dans un domaine V

On désigne par $z(V)$ la valeur moyenne sur V de la variable régionalisée. La variance ou dispersion statistique de $z(x)$ dans V s'exprime par la moyenne sur V des écarts quadratiques entre $z(x)$ et $z(V)$:

$$s^2(\circ|V) = \frac{1}{|V|} \int_V [z(x) - z(V)]^2 dx.$$

Ainsi définie, la dispersion statistique est une quantité déterministe (son expression ne fait intervenir ni variable ni fonction aléatoire). On cherche à donner une version équivalente de cette notion dans le cadre du modèle de fonction aléatoire. Si l'on remplace la variable régionalisée par la fonction aléatoire associée, la dispersion statistique devient une variable aléatoire et s'écrit :

$$S^2(\circ|V) = \frac{1}{|V|} \int_V [Z(x) - Z(V)]^2 dx.$$

La *variance de dispersion de $Z(x)$ dans V* est définie comme l'espérance de cette variable aléatoire et notée $D^2(\circ|V)$:

$$D^2(\circ|V) = E\{S^2(\circ|V)\} = \frac{1}{|V|} E\left\{\int_V [Z(x) - Z(V)]^2 dx\right\}.$$

Dans le cadre *stationnaire d'ordre deux*, on montre que $D^2(\circ|V)$ peut aussi s'écrire

$$D^2(\circ|V) = C(\mathbf{0}) - \overline{C}(V, V)$$

où $C(\mathbf{0}) = \sigma^2$ est la variance *a priori* de $Z(\mathbf{x})$

$C(V, V)$ est la variance de la variable $Z(V)$, régularisée de Z sur V .

La variance de dispersion d'un point dans le domaine V apparaît comme la différence entre la variance ponctuelle de $Z(\mathbf{x})$, ou variance *a priori*, et celle de $Z(V)$. En particulier, elle ne coïncide pas avec la variance *a priori*, et dépend du domaine V dans lequel se distribuent les échantillons ponctuels (figure 4.3). Ainsi, lorsque l'on calcule la variance empirique de données réparties dans un champ $\mathcal{D}$, ce n'est pas la variance *a priori* de $Z(\mathbf{x})$ que l'on estime, mais la dispersion d'un point dans le champ, ou encore $D^2(\circ|\mathcal{D})$.

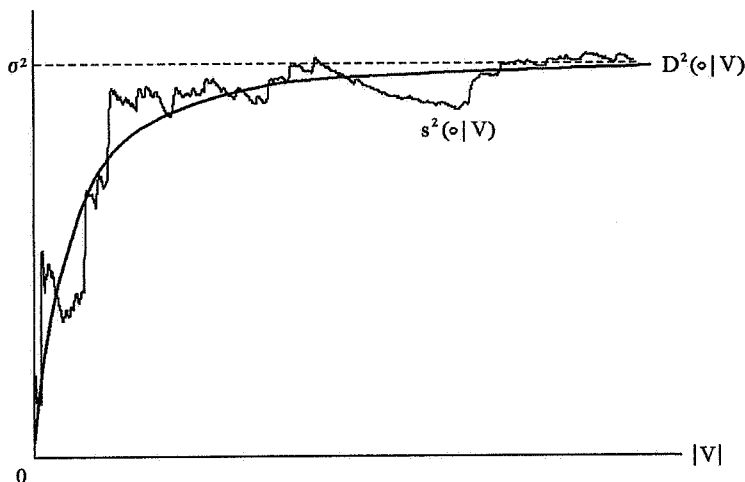


figure 4.3 : dispersion statistique, variance de dispersion et variance *a priori* en fonction de la taille du domaine dans lequel se distribuent les valeurs ponctuelles

La différence entre la variance de dispersion d'un point dans V et la variance *a priori* peut être importante si le domaine V n'est pas très étendu, auquel cas la variance de $Z(V)$ n'est pas négligeable. Au contraire, si V est très grand vis-à-vis des dimensions caractéristiques de C (portée, ou portée intégrale), on aura :

$$D^2(\circ|V) \approx D^2(\circ|\infty) = C(\mathbf{0}) = \sigma^2.$$

De même, un fort effet de pépite ou une microstructure prépondérante dans la fonction de covariance C sera responsable d'une différence importante entre la variance ponctuelle et la variance de $Z(V)$, et donc d'un écart faible entre variance *a priori* et variance de dispersion ; cet effet est maximal dans le cas d'une covariance purement pépitique, pour laquelle on a :

$$\overline{C}(V, V) = 0 \text{ et } D^2(\circ|V) = C(\mathbf{0}).$$

Lorsque seule l'*hypothèse intrinsèque* est vérifiée, on obtient l'expression de la variance de dispersion en remplaçant formellement la covariance C par $-\gamma$:

$$D^2(\circ|V) = \bar{\gamma}(V, V) = \frac{1}{|V|^2} \int_V \int_V \gamma(\mathbf{x} - \mathbf{y}) d\mathbf{x} d\mathbf{y}.$$

Variance de dispersion d'un bloc v dans un domaine V

On cherche à présent à généraliser la notion de variance de dispersion au cas de variables de support v non ponctuel. Cette généralisation, qui peut sembler évidente, se heurte toutefois à une difficulté : lorsque V est quelconque et que l'on cherche à calculer une moyenne de l'écart quadratique $[Z(v) - Z(V)]^2$, il se produit des "débordements" aux frontières de V quand le centre de v parcourt V , de sorte que la moyenne de $Z(v)$ ne coïncide plus avec $Z(V)$. Cet effet est important lorsque v n'est pas négligeable par rapport à V . Une définition rigoureuse de la dispersion de v dans V nécessite une contrainte supplémentaire de *partition* de V par v .

On suppose qu'il existe un pavage de V par des sous-domaines v_i qui se déduisent tous par translation du domaine v .

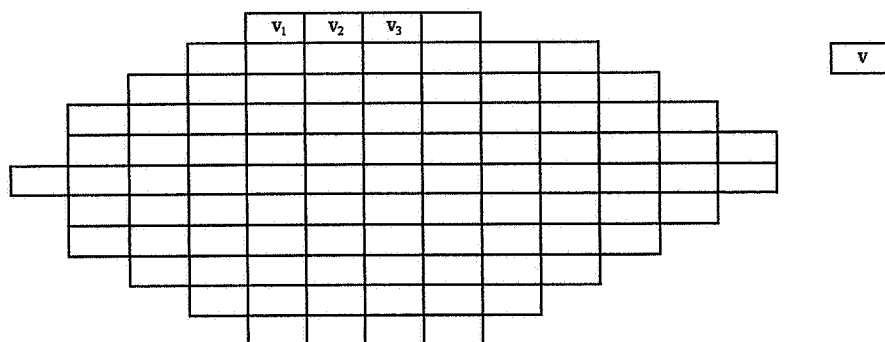


figure 4.4 : partition de V en domaines v_i

On désigne par p le nombre de sous-domaines v_i constituant V et par $z(v_i)$ la valeur moyenne de $z(\mathbf{x})$ sur v_i :

$$z(v_i) = \frac{1}{|v_i|} \int_{v_i} z(\mathbf{x}) d\mathbf{x}.$$

On appelle *dispersion statistique de v dans V* , et on note $s^2(v|V)$, la quantité

$$s^2(v|V) = \frac{1}{p} \sum_{i=1}^p [z(v_i) - z(V)]^2.$$

Cette définition généralise la notion de dispersion d'un point dans le domaine V . Si on fait tendre v vers le point (corrélativement, p tend vers l'infini), on retrouve bien la dispersion statistique de $z(\mathbf{x})$ dans V . On randomise à présent la dispersion $s^2(v|V)$, en remplaçant dans la formule précédente la variable régionalisée par la fonction aléatoire associée, d'où l'obtention d'une variable aléatoire $S^2(v|V)$:

$$S^2(v|V) = \frac{1}{p} \sum_{i=1}^p [Z(v_i) - Z(V)]^2.$$

On définit enfin la **variance de dispersion de v dans V** , notée $D^2(v|V)$, comme l'espérance mathématique de $S^2(v|V)$:

$$D^2(v|V) = E[S^2(v|V)].$$

De même que précédemment, cette formule peut être explicitée à l'aide des (co)variances des variables régularisées sur v et V .

En effet,

$$\begin{aligned} D^2(v|V) &= E\left\{\frac{1}{p} \sum_{i=1}^p [Z(v_i) - Z(V)]^2\right\} \\ &= \frac{1}{p} \sum_{i=1}^p \text{var}[Z(v_i) - Z(V)] \quad \text{car } Z(v_i) - Z(V) \text{ est d'espérance nulle} \\ &= \frac{1}{p} \sum_{i=1}^p \{\bar{C}(v_i, v_i) + \bar{C}(V, V) - 2\bar{C}(v_i, V)\} \end{aligned}$$

$$\text{Or } \forall i, \bar{C}(v_i, v_i) = \bar{C}(v, v) \text{ et } \sum_{i=1}^p \bar{C}(v_i, V) = \bar{C}\left(\sum_{i=1}^p v_i, V\right) = \bar{C}(V, V)$$

D'où

$$D^2(v|V) = \bar{C}(v, v) + \bar{C}(V, V) - 2\bar{C}(V, V) = \bar{C}(v, v) - \bar{C}(V, V).$$

En résumé,

$$\underline{D^2(v|V) = \bar{C}(v, v) - \bar{C}(V, V)}.$$

La variance de dispersion apparaît ainsi comme la *chute de variance* lorsque l'on passe de la distribution des valeurs sur les blocs v à celle sur les blocs V .

I.3.2. Définition générale de la variance de dispersion

L'expression ci-dessus peut servir de définition à la variance de dispersion de v dans V , à savoir, dans le **cadre stationnaire d'ordre deux** :

$$\underline{D^2(v|V) = \bar{C}(v, v) - \bar{C}(V, V)}.$$

Lorsque v est réduit à un point, cette formule se simplifie en : $D^2(\circ | V) = C(\mathbf{0}) - \bar{C}(V, V)$, où $C(\mathbf{0}) = \sigma^2$ est la variance *a priori* de $Z(\mathbf{x})$.

Cette définition "abstraite" de la variance de dispersion n'exige plus la contrainte de pavage de V par v , et permet également d'introduire la notion de covariance de dispersion de deux domaines v et v' dans V :

$$D^2(v, v'|V) = \bar{C}(v, v') - \bar{C}(V, V).$$

En contrepartie, on perd la signification physique de la variance de dispersion. Elle n'est plus qu'un outil de calcul et peut même prendre des valeurs négatives. De manière générale, lorsque v est inclus dans V , la variance de dispersion de v dans V est positive, ce qui traduit que les valeurs sont moins dispersées sur un gros bloc V que sur un petit bloc v (conséquence de l'effet de support), mais il existe des exceptions à cette règle pour des covariances présentant un effet de trou⁴. En fait, on n'est assuré de la positivité (et de la signification physique) de $D^2(v|V)$ que lorsqu'il existe un pavage de V par v .

Il est possible de généraliser ces définitions au **cadre intrinsèque** (il suffit de remplacer C par $-\gamma$). Il vient :

$$D^2(v|V) = \bar{\gamma}(V, V) - \bar{\gamma}(v, v)$$

$$D^2(\circ|V) = \bar{\gamma}(V, V)$$

$$D^2(v, v'|V) = \bar{\gamma}(V, V) - \bar{\gamma}(v, v')$$

Dans le cadre stationnaire, la variance de dispersion d'un point dans le champ tend vers la variance *a priori* lorsque la taille du champ devient infinie : $D^2(\circ|\infty) = C(0) = \sigma^2$. Mais si le variogramme n'est pas borné (fonction aléatoire intrinsèque stricte), la variance de dispersion croît indéfiniment à mesure que la taille du champ augmente ; la régionalisation présente alors une capacité de dispersion infinie.

1.3.3. Relations entre variances d'extension et de dispersion

La construction de la variance de dispersion (revenir à la première définition) fournit une première relation avec la variance d'extension :

$$D^2(\circ|V) = \frac{1}{|V|} \int_V E[Z(\mathbf{x}) - Z(V)]^2 d\mathbf{x} = \frac{1}{|V|} \int_V \sigma_E^2(\mathbf{x}, V) d\mathbf{x}.$$

La variance de dispersion d'un point dans un domaine V est la moyenne des variances d'extension des points de V dans V . Pour des supports v non ponctuels, tels que V peut être pavé par p blocs de la taille de v , on a de même :

$$D^2(v|V) = \frac{1}{p} \sum_{i=1}^p \sigma_E^2(v_i, V).$$

⁴ Par exemple, dans l'espace à une dimension, la covariance périodique $C(r) = \cos(r)$ donne, sur tout segment de longueur ℓ :

$$\bar{C}(\ell, \ell) = \frac{2}{\ell^2} [1 - \cos(\ell)]$$

qui n'est pas une fonction décroissante de ℓ . Ainsi, on pourra avoir $D^2(\ell|L) = \bar{C}(\ell, \ell) - \bar{C}(L, L) < 0$ bien que $\ell < L$. $D^2(\ell|L)$ ne sera positif à coup sûr que si L est un multiple de ℓ .

On peut également relier la variance d'extension de v dans v' à l'aide des (co)variances de dispersion dans un domaine V arbitraire :

$$\sigma_E^2(v, v') = D^2(v|V) + D^2(v'|V) - 2D^2(v, v'|V).$$

Toutes ces relations sont valables en dehors du cadre stationnaire d'ordre deux (hypothèse intrinsèque stricte).

1.3.4. Relation d'additivité (formule de Krige)

La formule de Krige ou relation d'additivité est une conséquence de la définition générale de $D^2(v|V)$. Pour trois domaines quelconques v , V et W , elle s'écrit :

$$D^2(v|W) = D^2(v|V) + D^2(V|W).$$

Conformément à la première définition de la variance de dispersion, cette formule n'a de sens physique que si v et V constituent des partitions emboîtées de W . En particulier, on a :

$$D^2(v|V) = D^2(\circ|V) - D^2(\circ|v).$$

Cette relation permet de déterminer la variance de blocs de taille v dans un champ V , à partir seulement des variances ponctuelles dans v et dans V ; si l'on évalue les variances de dispersion par les variances empiriques correspondantes, on évite même de recourir à un modèle variographique. On peut ainsi prévoir, en partie, la distribution des blocs à partir de celle des points.

II. Application à l'estimation globale

En géostatistique intrinsèque, l'estimation globale consiste à estimer la **valeur moyenne** de la variable régionalisée sur l'ensemble du champ $\mathcal{D}$. Cette moyenne constitue une *grandeur régionale*, calculable si la réalité était connue exhaustivement.

II.1. Remarques préliminaires

II.1.1. Contrainte d'homogénéité

En cas d'hétérogénéités dans la régionalisation, qui rendent les hypothèses stationnaire et intrinsèque inadéquates, il convient de diviser le champ en plusieurs sous-zones homogènes, de procéder à l'estimation globale de chaque sous-zone et d'additionner, en les pondérant, les différentes estimations et variances d'estimation, supposées indépendantes. Une alternative consiste à utiliser les techniques transitives (cf. chapitre 1), qui ne requièrent pas d'hypothèse de stationnarité ; dans ce cas, la quantité estimée n'est plus la moyenne mais l'abondance.

II.1.2. Choix de l'estimateur

Nous n'étudierons ici que l'estimation de la moyenne globale de la variable régionalisée par la moyenne arithmétique des valeurs échantillonnées. On pourrait éventuellement prendre comme estimateur une moyenne pondérée, donnant des poids différents aux observations, ce qui présente un intérêt lorsque l'échantillonnage est fortement irrégulier ou est préférentiel (par exemple, lorsque les zones densément échantillonnées sont celles où la régionalisation prend des valeurs fortes), mais ceci complique le calcul de la variance d'estimation. Une autre possibilité, plus simple, consiste à rechercher des zones reconnues de manière homogène, à estimer séparément chacune de ces zones, puis à supposer que les estimations obtenues sont indépendantes les unes des autres.

Les données présentent parfois des valeurs "aberrantes", soit très fortes, soit au contraire très faibles, qui peuvent avoir un impact important sur la moyenne expérimentale. Dans ce contexte, on sait que la médiane de l'échantillon⁵ est un estimateur *robuste*, i.e. peu sensible aux valeurs extrêmes. Cependant, la médiane est un estimateur biaisé de la moyenne globale (sauf dans certains cas très particuliers, comme celui où la distribution des valeurs est symétrique) et ne constitue donc pas un estimateur acceptable.

⁵ c'est-à-dire la valeur qui divise l'échantillon en deux parties d'effectifs égaux, les valeurs mesurées ayant été au préalable classées par ordre croissant ou décroissant. La médiane utilise donc peu l'information disponible, puisqu'elle n'est pas une mesure mais résulte d'un *classement*. Contrairement à la moyenne arithmétique, elle n'est pas affectée par les valeurs anormalement faibles ou élevées des mesures, d'où son intérêt potentiel.

II.1.3. Hypothèses et objectif

Dans toute la suite, nous supposons, pour simplifier la présentation, que la fonction aléatoire $Z(\mathbf{x})$, dont la variable régionalisée est une réalisation, est stationnaire d'ordre deux, mais tous les résultats restent valables sous l'hypothèse intrinsèque stricte. Comme dans le cas de la statistique classique, il est facile de voir que l'estimateur par la moyenne arithmétique des données est **sans biais**, c'est-à-dire que son espérance est égale à la grandeur à estimer. Si l'absence de biais de l'estimateur peut paraître rassurante, il faut bien voir cependant qu'elle n'est pas un critère infaillible de qualité et qu'elle ne garantit pas que l'erreur commise soit faible ; il est important de connaître la dispersion de l'estimateur autour de la valeur vraie. Cette dispersion est mesurée par la **variance** de l'erreur d'estimation, appelée *variance d'estimation* globale, et quantifie la **précision** de l'estimation.

L'apport de la géostatistique par rapport aux techniques statistiques classiques est la prise en compte de la position des observations et de la structure spatiale de la variable étudiée dans le calcul d'une variance d'estimation globale "réaliste". Il est en effet intuitif que la précision d'une estimation dépend de ces deux paramètres ; par exemple, une variable erratique donnera lieu à une plus grande imprécision qu'une régionalisation régulière. Dans la suite de ce paragraphe, nous allons examiner différentes stratégies d'échantillonnage, qui conduisent à des expressions simples de la variance d'estimation.

II.1.4. Quel sens objectif donner au biais et à la variance d'estimation globale ?

Au niveau mathématique, le biais et la variance d'estimation globale mesurent la précision de l'estimation et permettent de comparer la qualité de différents estimateurs et de différents modes d'échantillonnage. Cependant, ces notions peuvent sembler purement théoriques, puisqu'elles correspondent à l'espérance et la variance des erreurs d'estimation globale obtenues sur un *grand nombre de réalisations* de la fonction aléatoire. Or en pratique, l'erreur d'estimation globale est **unique**, et ne peut être randomisée car on ne dispose que d'une seule réalisation de la fonction aléatoire. Dans ce contexte, il paraît difficile de donner un sens objectif aux notions de biais et de variance d'estimation globale.

Pour sortir de cette impasse, il faut probabiliser le réseau d'échantillonnage, c'est-à-dire considérer qu'il a été implanté au hasard uniformément sur le domaine étudié. Ainsi, *si l'on pouvait déplacer la configuration des données sur le champ*, on pourrait former de multiples estimations, dont l'erreur moyenne serait nulle et la variance coïnciderait avec la variance d'estimation globale. Cette interprétation se réfère à l'approche transitive exposée au chapitre 1. Elle permet de donner un sens objectif à l'absence de biais et à la variance d'estimation globale, mais elle pose aussi un problème nouveau : la géométrie du champ n'est plus connue puisque ce dernier est modifié dès lors que l'on déplace le réseau des données (ce problème est atténué lorsque le champ est très étendu).

II.2. Estimation globale en support ponctuel

Nous étudions dans ce paragraphe le cas où le support des échantillons est assimilable à un point, ce qui implique en particulier que le nombre d'échantillons qu'il est possible de former est infini.

II.2.1. Echantillonnage aléatoire uniforme

Les emplacements des échantillons sont tirés au hasard pur, c'est-à-dire indépendamment les uns des autres, et uniformément dans le champ $\mathcal{D}$. On estime la moyenne exhaustive sur le champ $\mathcal{D}$

$$Z(\mathcal{D}) = \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} Z(\mathbf{x}) d\mathbf{x}$$

par la moyenne des observations :

$$Z^* = \frac{1}{n} \sum_{\alpha=1}^n Z(\mathbf{x}_{\alpha})$$

où les $\mathbf{x}_{\alpha}$ sont des réalisations de variables aléatoires $\mathbf{X}_{\alpha}$ indépendantes et uniformes dans $\mathcal{D}$.

Détaillons le calcul de la variance d'estimation globale. Elle s'exprime de la manière suivante :

$$\text{var}[Z^* - Z(\mathcal{D})] = \text{var}\left[\frac{1}{n} \sum_{\alpha=1}^n Z(\mathbf{X}_{\alpha}) - Z(\mathcal{D})\right]$$

où à la fois la fonction Z et les sites $\mathbf{X}_{\alpha}$ sont aléatoires. Les erreurs partielles $Z(\mathbf{X}_{\alpha}) - Z(\mathcal{D})$ étant d'espérance nulle (par stationnarité), on peut écrire :

$$\text{var}[Z^* - Z(\mathcal{D})] = E\left\{\frac{1}{n^2} \sum_{\alpha=1}^n [Z(\mathbf{X}_{\alpha}) - Z(\mathcal{D})]^2\right\}.$$

$$\text{D'après la formule de l'espérance totale : } \text{var}[Z^* - Z(\mathcal{D})] = E\left\{E\left[\frac{1}{n^2} \sum_{\alpha=1}^n [Z(\mathbf{X}_{\alpha}) - Z(\mathcal{D})]^2 \mid Z\right]\right\}.$$

Fixons provisoirement la réalisation z de la fonction aléatoire Z , et calculons la variance d'estimation conditionnelle à $Z = z$:

$$\left\{E\left[\frac{1}{n^2} \sum_{\alpha=1}^n [Z(\mathbf{X}_{\alpha}) - Z(\mathcal{D})]^2 \mid Z = z\right]\right\} = \frac{1}{n^2} E\left\{\sum_{\alpha=1}^n [z(\mathbf{X}_{\alpha}) - z(\mathcal{D})]^2\right\} = \frac{1}{n^2} \text{var}\left\{\sum_{\alpha=1}^n [z(\mathbf{X}_{\alpha}) - z(\mathcal{D})]\right\}.$$

Dans cette dernière expression, seuls les sites $\mathbf{X}_\alpha$ demeurent aléatoires. Or, par hypothèse, ceux-ci sont indépendants, donc les erreurs partielles $z(\mathbf{X}_\alpha) - z(\mathcal{D})$ sont indépendantes (leurs covariances sont nulles) et :

$$\frac{1}{n^2} \text{var} \left\{ \sum_{\alpha=1}^n [z(\mathbf{X}_\alpha) - z(\mathcal{D})] \right\} = \frac{1}{n^2} \sum_{\alpha=1}^n \text{var} [z(\mathbf{X}_\alpha) - z(\mathcal{D})].$$

Les $\mathbf{X}_\alpha$ étant uniformément distribués dans $\mathcal{D}$, il vient :

$$\text{var} [z(\mathbf{X}_\alpha) - z(\mathcal{D})] = \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} [z(\mathbf{x}) - z(\mathcal{D})]^2 d\mathbf{x} = s^2(\circ|\mathcal{D})$$

où, par définition, $s^2(\circ|\mathcal{D})$ est la dispersion statistique d'un point dans $\mathcal{D}$.

Par suite,
$$\frac{1}{n^2} \text{var} \left\{ \sum_{\alpha=1}^n [z(\mathbf{X}_\alpha) - z(\mathcal{D})] \right\} = \frac{s^2(\circ|\mathcal{D})}{n}$$

puis, en déconditionnant par rapport à Z et en prenant l'espérance, on obtient l'expression finale :

$$\text{var}[Z^* - Z(\mathcal{D})] = \frac{E[S^2(\circ|\mathcal{D})]}{n} = \frac{D^2(\circ|\mathcal{D})}{n}$$

En résumé :

$$\text{var}[Z^* - Z(\mathcal{D})] = \frac{D^2(\circ|\mathcal{D})}{n}$$

où $D^2(\circ|\mathcal{D})$ désigne la variance de dispersion d'un point dans le champ $\mathcal{D}$.

On peut comparer cette formule à celle donnée par la statistique classique, à savoir⁶ :

$$\text{var}[Z^* - m] = \frac{\sigma^2}{n}$$

où $m = E[Z(\mathbf{x})]$ pour tout $\mathbf{x}$ et σ^2 est la variance *a priori* de $Z(\mathbf{x})$.

Comme la statistique classique, la géostatistique donne une variance d'estimation globale qui décroît en $1/n$. En particulier, l'écart-type de l'estimation globale varie en $1/\sqrt{n}$; il faudra par exemple quatre fois plus d'échantillons pour diminuer de moitié cet écart-type, c'est-à-dire pour doubler la précision de l'estimation.

⁶ La statistique classique suppose que toutes les valeurs $z(\mathbf{x})$ de la variable régionalisée sont des réalisations indépendantes de variables aléatoires $Z(\mathbf{x})$ de même loi de probabilité (en particulier d'espérance m et de variance σ^2). En admettant les propriétés de stationnarité et d'ergodicité (cf. annexe B), m s'identifie à la moyenne spatiale à estimer

$$Z(\mathcal{D}) = \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} Z(\mathbf{x}) d\mathbf{x}.$$

En revanche, en géostatistique, le numérateur de l'expression n'est pas la variance *a priori* de $Z(\mathbf{x})$, mais la variance de dispersion d'un point dans le champ ; cette quantité dépend du domaine à estimer et ne coïncide avec la variance *a priori* que dans le cas particulier d'un domaine infiniment grand ou d'un effet de pépite pur. Par ailleurs, la variance de dispersion peut être définie dans le cadre intrinsèque strict, contrairement à la variance *a priori*. Ainsi, la géostatistique permet d'utiliser une gamme de modèles plus vaste que la statistique classique, limitée au cadre stationnaire du second ordre.

II.2.2. Echantillonnage aléatoire stratifié

Le champ $\mathcal{D}$ est partitionné en cellules $\{V_\alpha, \alpha = 1 \dots n\}$ toutes identiques à une même cellule de référence V . A l'intérieur de chaque cellule, on tire au hasard un échantillon, indépendamment des autres échantillons. Pour le calcul de la variance d'estimation globale, la démarche est la même que dans le cas d'un échantillonnage aléatoire uniforme : on commence par calculer la variance d'estimation conditionnelle à $Z = z$ (i.e. en fixant la réalisation de Z), puis on élimine le conditionnement relatif à la réalisation, c'est-à-dire que l'on restitue à Z son caractère aléatoire. La variance d'estimation globale que l'on obtient est :

$$\text{var}[Z^* - Z(\mathcal{D})] = \frac{D^2(\circ|V)}{n}$$

où $D^2(\circ|V)$ est la variance de dispersion d'un point dans la cellule de référence V .

En effet, l'erreur totale est la moyenne des erreurs partielles sur chaque cellule :

$$Z^* - Z(\mathcal{D}) = \frac{1}{n} \sum_{\alpha=1}^n [Z(\mathbf{X}_\alpha) - Z(\mathbf{V}_\alpha)].$$

Conditionnellement à $Z = z$, seules les positions $\mathbf{X}_\alpha$ sont aléatoires ; comme elles sont indépendantes, on en déduit que les erreurs partielles le sont également (pour z fixée), donc la variance d'estimation globale conditionnelle à z vaut :

$$\text{var}[Z^* - Z(\mathcal{D})|Z = z] = \frac{1}{n^2} \sum_{\alpha=1}^n \text{var}[z(\mathbf{X}_\alpha) - z(\mathbf{V}_\alpha)] = \frac{1}{n^2} \{n \times s^2(\circ|V)\} = \frac{s^2(\circ|V)}{n}$$

en reprenant la formule de la variance dans le cas aléatoire uniforme, avec cette fois chaque échantillon $\mathbf{X}_\alpha$ aléatoire dans un bloc égal à V . On obtient l'expression annoncée en déconditionnant par rapport à Z .

On établit par ailleurs que la variance d'estimation de l'échantillonnage aléatoire stratifié est toujours inférieure à celle de l'échantillonnage aléatoire uniforme, puisque, d'après la relation d'additivité ou *formule de Krige* :

$$D^2(\circ|V) - D^2(\circ|\mathcal{D}) = -D^2(V|\mathcal{D}) \leq 0.$$

On peut généraliser la formule précédente au cas où l'on tire p points au hasard (de façon indépendante) dans chaque cellule, au lieu d'un seul point. Si l'on note n le nombre total de points échantillonnés, alors l'expression de la variance d'estimation globale est inchangée :

$$\text{var}[Z^* - Z(\mathcal{D})] = \frac{D^2(\circ | V)}{n}$$

mais il faut bien voir qu'à effort d'échantillonnage fixé, c'est-à-dire à nombre d'échantillons n fixé, il est plus avantageux de diviser le champ en n cellules et tirer un point par cellule, car les cellules sont alors plus nombreuses, donc plus petites ; par suite, la variance de dispersion $D^2(\circ | V)$ est plus faible, et l'estimation plus précise.

II.2.3. Echantillonnage régulier

Paradoxalement, l'échantillonnage régulier est le cas le plus difficile à traiter, pour lequel des approximations sont nécessaires. On découpe le domaine $\mathcal{D}$ en cellules $\{V_\alpha, \alpha = 1 \dots n\}$ toutes identiques à une même cellule de référence V , et l'on échantillonne le point central $\mathbf{x}_\alpha$ de chacune de ces cellules.

L'expression rigoureuse de la variance d'estimation s'obtient à l'aide de la formule générale

$$\text{var}[Z^* - Z(\mathcal{D})] = \frac{1}{n^2} \sum_{\alpha=1}^n \sum_{\beta=1}^n C(\mathbf{x}_\alpha - \mathbf{x}_\beta) - 2 \sum_{\alpha=1}^n \frac{1}{n} \int_{\mathcal{D}} C(\mathbf{x}_\alpha - \mathbf{x}) d\mathbf{x} + \frac{1}{|\mathcal{D}|^2} \int_{\mathcal{D}} \int_{\mathcal{D}} C(\mathbf{x} - \mathbf{y}) d\mathbf{x} d\mathbf{y}$$

ou son équivalent en termes de variogramme.

Cette expression est souvent difficile à calculer lorsque le nombre d'échantillons est élevé ou lorsque la géométrie du champ est complexe. Pour la simplifier, l'idée est de décomposer l'erreur globale en une somme d'erreurs élémentaires, dont il est facile de calculer la variance, et de supposer ces erreurs élémentaires non corrélées. Plus précisément, on utilise l'un ou l'autre des deux principes d'approximation suivants :

- *principe de composition des termes de ligne, de section et de tranche*

Le réseau régulier de points de données présente des droites et des plans privilégiés. On estime alors chaque ligne (représentant par exemple un sondage) par la moyenne des échantillons qu'elle contient, puis chaque plan par une moyenne pondérée des lignes (supposées connues) qu'il contient, et enfin le champ par une moyenne pondérée des plans qu'il contient. On commet ainsi trois erreurs consécutives que l'on suppose indépendantes, ou du moins de covariances négligeables vis-à-vis de leurs variances. La variance d'estimation globale s'obtient alors en additionnant les variances correspondant à ces trois erreurs (extension des points aux lignes, des lignes aux plans et des plans au champ), qui sont respectivement appelées *terme de ligne*, *terme de section* et *terme de tranche*.

L'intérêt de ce principe de composition est qu'il permet de simplifier le calcul de la variance d'estimation à une succession de calculs de variances d'extension sur des volumes géométriquement simples ; ces derniers peuvent être résolus à l'aide de fonctions auxiliaires et d'abaques. Son inconvénient est qu'il peut être mis en défaut (on montre notamment qu'il est inadapté dans le cas des *grandes mailles*, c'est-à-dire dans le cas où les distances entre les données sont supérieures à la portée de la covariance).

- **principe de composition des variances d'extension élémentaires**

Comme $Z(\mathcal{D}) = \frac{1}{n} \sum_{\alpha=1}^n Z(V_{\alpha})$, on peut décomposer l'erreur d'estimation globale

$$Z^* - Z(\mathcal{D}) = \frac{1}{n} \sum_{\alpha=1}^n Z(\mathbf{x}_{\alpha}) - Z(\mathcal{D})$$

en la moyenne des erreurs "partielles" sur chaque cellule :

$$Z^* - Z(\mathcal{D}) = \frac{1}{n} \sum_{\alpha=1}^n [Z(\mathbf{x}_{\alpha}) - Z(V_{\alpha})].$$

On suppose ces erreurs partielles non corrélées⁷, de sorte que la variance d'estimation globale apparaît comme la somme des variances des erreurs partielles. Plus précisément :

$$\text{var}[Z^* - Z(\mathcal{D})] \approx \frac{1}{n^2} \sum_{\alpha=1}^n \text{var}[Z(\mathbf{x}_{\alpha}) - Z(V_{\alpha})] = \frac{\sigma_E^2(o, V)}{n}$$

où $\sigma_E^2(o, V)$ est la *variance d'extension* d'un échantillon $\mathbf{x}_{\alpha}$ dans sa cellule V_{α} :

$$\sigma_E^2(o, V) = \text{var}[Z(\mathbf{x}_{\alpha}) - Z(V_{\alpha})] \text{ indépendant de } \alpha = 1 \dots n$$

soit, de manière explicite, en se plaçant dans le cas plus général où il n'existe qu'un variogramme, et en notant $\mathbf{x}_0$ le point central de V :

$$\sigma_E^2(\mathbf{x}_0, V) = \frac{2}{|V|} \int_V \gamma(\mathbf{x} - \mathbf{x}_0) d\mathbf{x} - \frac{1}{|V|^2} \int_V \int_V \gamma(\mathbf{x} - \mathbf{y}) d\mathbf{x} d\mathbf{y}.$$

Cette formule ne fait intervenir que les valeurs du variogramme pour des distances inférieures au diamètre de V , qui ne sont pas directement accessibles à partir des données disponibles (le premier point du variogramme expérimental correspond à une distance égale à la maille d'échantillonnage). Par suite, l'ajustement du modèle variographique près de l'origine sera fondamental dans l'appréciation de la précision de l'estimation.

⁷ Plus exactement, il suffit que la somme des covariances d'extension soit négligeable vis-à-vis de la somme des variances d'extension.

II.2.4. Comparaison des différentes stratégies d'échantillonnage

On a déjà vu que la **variance d'estimation de l'échantillonnage aléatoire stratifié est toujours plus faible que celle de l'échantillonnage aléatoire uniforme**. L'objet de ce paragraphe est de comparer les performances (en termes de variance d'estimation) des échantillonnages régulier et aléatoire stratifié, lorsque les cellules – et par conséquent les nombres d'échantillons – sont identiques. Pour cela, il faut comparer la variance de dispersion d'un point dans la cellule V

$$D^2(\circ|V) = \bar{\gamma}(V, V)$$

avec la variance d'extension du point central de V dans V :

$$\sigma_E^2(\circ, V) = 2\bar{\gamma}(\circ, V) - \underbrace{\gamma(\mathbf{0})}_0 - \bar{\gamma}(V, V) = 2\bar{\gamma}(\circ, V) - \bar{\gamma}(V, V).$$

Ces formules de définition des variances d'extension et de dispersion montrent qu'il est difficile de comparer directement $D^2(\circ|V)$ et $\sigma_E^2(\circ, V)$. Toutefois, on sait que la variance d'extension et la variance de dispersion sont liées par la relation

$$D^2(\circ|V) = \frac{1}{|V|} \int_V E[Z(\mathbf{x}) - Z(V)]^2 d\mathbf{x} = \frac{1}{|V|} \int_V \sigma_E^2(\mathbf{x}, V) d\mathbf{x}$$

c'est-à-dire que la variance de dispersion d'un point dans V est la moyenne des variances d'extension des points de V dans V . Or, en règle générale, le meilleur site pour estimer un bloc est son point central ; en d'autres termes, la variance d'extension d'un point dans V est minimale pour le point central de V .

Par conséquent, on a en général (mais pas toujours) :

$$\sigma_E^2(\circ, V) \leq D^2(\circ|V).$$

Cette inégalité traduit la **supériorité de l'échantillonnage régulier** du point de vue de la variance d'estimation globale.

Cette supériorité s'estompe lorsque la maille d'échantillonnage devient grande par rapport à la portée de la covariance. De plus, dans l'inférence de la covariance, l'échantillonnage aléatoire stratifié peut permettre d'identifier le comportement près de l'origine, car les sites d'observation ne sont pas régulièrement répartis et peuvent donc être séparés par des distances petites, alors que l'échantillonnage régulier ne permet pas de travailler à une échelle inférieure à la maille d'échantillonnage. Ainsi, dans certains cas, l'échantillonnage aléatoire stratifié facilite la modélisation de la covariance au voisinage de l'origine et est préféré à un échantillonnage régulier.

II.3. Support non ponctuel

On suppose à présent que les unités échantillonnées ne sont plus ponctuelles, mais sont des surfaces ou des volumes identiques à un même support de référence v . On fait l'hypothèse que le champ $\mathcal{D}$ peut être partitionné en cellules qui se déduisent toutes de v par translation, et que l'échantillonnage s'opère parmi ces cellules. En particulier, le nombre de localisations possibles à l'intérieur du champ est fini, contrairement au cas ponctuel, ce qui entraîne des modifications dans les définitions précédentes et des corrections dans certaines formules. En particulier, les moyennes sur le champ $\mathcal{D}$ ainsi que les variances de dispersion et d'extension sont définies par des sommes discrètes et non par des intégrales.

Par ailleurs, il faut introduire un facteur correctif dans la formule donnant la variance d'estimation dans le cas d'un échantillonnage aléatoire uniforme. En effet, comme le nombre d'unités dans le champ est fini, les échantillons ne peuvent plus être considérés comme indépendants : on les sélectionne par un tirage *sans remise*. Si on désigne par v le support de l'unité échantillonnée, n le nombre d'échantillons et N le nombre total d'unités dans $\mathcal{D}$, alors la variance d'estimation s'exprime par⁸ :

$$\text{var}[Z^* - Z(\mathcal{D})] = \frac{N - n}{N - 1} \frac{D^2(v|\mathcal{D})}{n}.$$

Le facteur correctif tend vers 1 lorsque N tend vers l'infini (support v petit par rapport à $\mathcal{D}$), auquel cas on retrouve la formule du cas ponctuel. Il s'annule si $n = N$ (échantillonnage exhaustif), ce qui signifie que l'estimation est parfaite.

Dans le cas des échantillonnages régulier et stratifié aléatoire avec un échantillon par cellule, les relations trouvées dans le cadre ponctuel restent valables, à condition de remplacer dans les formules des variances d'extension et de dispersion le point ($\circ$) par le support v des échantillons.

Dans le cas d'un échantillonnage aléatoire stratifié où l'on tire p unités par cellules et n unités au total, les positions des échantillons de chaque cellule ne sont plus indépendantes, et la variance d'estimation devient :

$$\text{var}[Z^* - Z(\mathcal{D})] = \frac{p(N - n)}{Np - n} \frac{D^2(v|V)}{n}.$$

Si $N \rightarrow \infty$, on retrouve asymptotiquement la formule donnée dans le cas d'échantillons ponctuels.

⁸ Le facteur correctif est le même que celui introduit en statistique classique, où la variance d'estimation est :

$$\text{var}[Z^* - Z(\mathcal{D})] = \frac{N - n}{N - 1} \frac{\sigma^2}{n}.$$

III. Conclusions sur l'estimation globale

III.1. Facteurs conditionnant la précision de l'estimation

La variance d'estimation globale dépend de plusieurs facteurs. Elle est d'autant plus faible que

- la régionalisation est régulière dans l'espace (covariance ou variogramme régulier à l'origine) ;
- les sites de données sont proches et répartis de manière uniforme ;
- les données sont nombreuses ; ce facteur devient prédominant lorsque les positions des échantillons sont tirées au hasard ou lorsque la structuration de la régionalisation est faible.

En revanche, la variance d'estimation est indépendante des valeurs prises par les données. Ce résultat se retrouve en permanence en géostatistique linéaire, aussi bien en estimation globale que locale.

III.2. Comparaison avec la statistique descriptive

Les formules de variances d'estimation obtenues en géostatistique présentent un lien étroit avec celles de la **statistique descriptive** ou "théorie des sondages" (tableau 4.1). Cette dernière n'interprète pas les valeurs observées comme des réalisations de variables aléatoires, mais utilise les moyennes et variances *statistiques* calculées sur les échantillons, ce qui permet de prendre en compte la non-indépendance des observations. Elle s'affranchit même de l'hypothèse de stationnarité, puisqu'il n'y a plus d'interprétation probabiliste des valeurs. Exactement comme la géostatistique transitive, la statistique descriptive travaille directement sur la variable régionalisée ; le calcul des variances d'estimation repose uniquement sur la randomisation du réseau d'échantillonnage.

statistique descriptive	support ponctuel (°)	support non ponctuel (v)
échantillonnage aléatoire uniforme	$\frac{s^2(° \mathcal{D})}{n}$	$\frac{N-n}{N-1} \frac{s^2(v \mathcal{D})}{n}$
échantillonnage aléatoire stratifié avec p unités par cellule (p > 1) et n/p cellules	$\frac{p}{n^2} \sum_{i=1}^{n/p} s^2(° V_i)$	$\frac{p(N-n)}{Np-n} \frac{p}{n^2} \sum_{i=1}^{n/p} s^2(v V_i)$
échantillonnage régulier	incalculable	

tableau 4.1 : variances d'estimation globale en statistique descriptive

Si l'on remplace les dispersions statistiques $s^2(\cdot)$ par leur version "théorique", à savoir les variances de dispersion $D^2(\cdot)$ correspondantes – supposées constantes sur l'ensemble du champ (par stationnarité) –, on retrouve exactement les mêmes formules que celles obtenues en géostatistique.

L'avantage de l'approche descriptive est sa généralité, puisque qu'aucune hypothèse de stationnarité n'est nécessaire pour mener à bien les calculs. Ainsi, en échantillonnage aléatoire stratifié, les variances de dispersion $s^2(\cdot | V_i)$ ou $s^2(v | V_i)$ peuvent différer d'une cellule à l'autre. En contrepartie, on est obligé de tirer plusieurs échantillons par cellule pour pouvoir évaluer la dispersion statistique, contrainte qui n'existe pas en géostatistique où l'évaluation des variances de dispersion se fait grâce à un *modèle* de covariance. D'autre part, dans le cas de l'échantillonnage régulier, il n'est pas possible de fournir une variance d'estimation par les méthodes descriptives, car on ne peut calculer de manière empirique une variance d'extension à partir d'un point seulement.

III.3. Les problèmes de délimitation du champ

Jusqu'à présent, on a supposé le champ $\mathcal{D}$ de la variable régionalisée connu. Dans le cas contraire, l'estimation globale est menée sur un champ *estimé*, et l'incertitude sur la géométrie du champ doit intervenir dans le calcul de la variance d'estimation globale. Si l'on suppose que cette incertitude est indépendante de l'erreur d'estimation sur le champ estimé, alors il suffit d'ajouter un terme correctif à la variance d'estimation :

$$\sigma_{\text{est}(\mathcal{D})}^2 = \sigma_{\text{est}(\mathcal{D}^*)}^2 + \sigma_{\mathcal{D}}^2 D^2(\cdot | \mathcal{D}) / |\mathcal{D}|^2$$

avec $\sigma_{\text{est}(\mathcal{D}^*)}^2$: variance d'estimation globale sur le champ estimé $\mathcal{D}^*$

$\sigma_{\mathcal{D}}^2$: variance d'estimation globale de la superficie (2D) ou du volume (3D) de $\mathcal{D}$, calculée par les méthodes transitives, à l'aide du covariogramme géométrique (cf. chapitre 1).

Même sous l'hypothèse stationnaire, où la variable régionalisée est "indépendante" du champ (donc sa moyenne est invariante lorsque l'on augmente ou rétrécit arbitrairement $\mathcal{D}$), l'erreur commise sur l'estimation de la surface ou du volume de $\mathcal{D}$ est responsable d'une incertitude supplémentaire sur l'estimation de la moyenne globale. Par ailleurs, l'hypothèse d'indépendance entre l'erreur d'estimation globale et l'erreur d'origine géométrique est mise en défaut si les limites du champ sont définies par référence aux valeurs de la variable régionalisée, par exemple si le champ est l'ensemble des points pour lesquels la variable régionalisée est non nulle. Dans ce cas, le calcul rigoureux de la variance d'estimation se complique notablement ; le problème géométrique interfère avec celui de l'estimation de la valeur moyenne, sans compter que des problèmes de stationnarité risquent de se poser (variable présentant des valeurs plus faibles en bordure du champ).

Pour s'affranchir de ces contraintes de géométrie, il est souvent nécessaire de recourir au formalisme de la **géostatistique transitive** exposé dans le premier chapitre de ce document. Rappelons que le formalisme transitif n'a besoin d'aucune hypothèse (de type stationnarité) sur la variable régionalisée et esquive les problèmes de délimitation du champ, par le recours à l'abondance, c'est-à-dire une *somme*, au lieu de la moyenne.

III.4. Récapitulatif sur l'estimation globale par les méthodes statistiques et géostatistiques

III.4.1. Rappel des notations

$\mathcal{D}$: champ de la variable régionalisée

n : nombre total de sites échantillonnés

V : cellule de référence pour les échantillonnages stratifiés (stratifié aléatoire et régulier)

◦ ou v : support de l'unité échantillonnée (point ou surface / volume)

p : nombre d'unités échantillonnées dans V (sauf mention contraire, $p = 1$)

C ou γ : covariance ou variogramme de la fonction aléatoire $Z(\mathbf{x})$

σ^2 : variance *a priori* du support des échantillons ($\sigma^2 = C(0)$ pour des supports ponctuels)

N : nombre d'unités dans le champ $\mathcal{D}$ (infini pour un échantillonnage ponctuel).

On estime la moyenne exhaustive sur $\mathcal{D}$ par la moyenne arithmétique des échantillons. L'estimateur est sans biais (son espérance est égale à la valeur vraie inconnue), mais l'expression de la variance d'estimation, qui mesure la précision de l'estimation, dépend du modèle envisagé et de la stratégie d'échantillonnage. Les tableaux ci-dessous donnent les variances d'estimation selon le mode d'échantillonnage et la méthode d'estimation choisie.

III.4.2. Support ponctuel

support ponctuel		statistique classique (stationnarité d'ordre 2)	statistique descriptive (pas d'hypothèse de stationnarité)	géostatistique intrinsèque (stationnarité d'ordre 2 ou hypothèse intrinsèque)
échantillonnage aléatoire uniforme		$\frac{\sigma^2}{n}$	$\frac{s^2(\circ \mathcal{D})}{n}$	$\frac{D^2(\circ \mathcal{D})}{n}$
échantillonnage stratifié aléatoire	1 point par cellule		incalculable	$\frac{D^2(\circ V)}{n}$
	p points par cellule et n/p cellules		$\frac{p}{n^2} \sum_{i=1}^{n/p} s^2(\circ V_i)$	
échantillonnage régulier			incalculable	$\approx \frac{\sigma_E^2(\circ, V)}{n}$

tableau 4.2 : variances d'estimation globale – support ponctuel

où $D^2(\circ|\mathcal{D})$ est la variance de dispersion d'un point dans le champ $\mathcal{D}$;

$D^2(\circ|V)$ est la variance de dispersion d'un point dans la cellule de référence V ;

$s^2(\circ|\mathcal{D})$ et $s^2(\circ|V)$ sont les versions statistiques des variances précédentes ;

$\sigma_E^2(\circ, V)$ est la variance d'extension du point central de V dans V .

On a toujours : $\sigma^2 \geq D^2(\circ|\mathcal{D}) \geq D^2(\circ|V)$, et très souvent $D^2(\circ|V) \geq \sigma_E^2(\circ, V)$.

III.4.3. Support non ponctuel

L'unité échantillonnée est une surface ou un volume de support v .

support v non ponctuel		statistique classique (stationnarité d'ordre 2)	statistique descriptive (pas d'hypothèse de stationnarité)	géostatistique intrinsèque (stationnarité d'ordre 2 ou hypothèse intrinsèque)
échantillonnage aléatoire uniforme		$\frac{N - n}{N - 1} \frac{\sigma^2}{n}$	$\frac{N - n}{N - 1} \frac{s^2(v \mathcal{D})}{n}$	$\frac{N - n}{N - 1} \frac{D^2(v \mathcal{D})}{n}$
échantillonnage stratifié aléatoire	1 unité par cellule		incalculable	$\frac{D^2(v V)}{n}$
	p unités par cellule et n/p cellules		$\frac{p(N - n)}{Np - n} \frac{p}{n^2} \sum_{i=1}^{n/p} s^2(v V_i)$	$\frac{p(N - n)}{Np - n} \frac{D^2(v V)}{n}$
échantillonnage régulier			incalculable	$\approx \frac{\sigma_E^2(v, V)}{n}$

tableau 4.3 : variances d'estimation globale – support non ponctuel

où σ^2 représente la variance *a priori* de $Z(v)$;

$D^2(v|\mathcal{D})$ est la variance de dispersion du support v dans le champ $\mathcal{D}$;

$D^2(v|V)$ est la variance de dispersion du support v dans la cellule de référence V ;

$s^2(v|\mathcal{D})$ et $s^2(v|V)$ sont les versions statistiques des variances précédentes ;

$\sigma_E^2(v, V)$ est la variance d'extension du support v central de V dans V .

Par rapport au cas ponctuel, les facteurs correctifs qui apparaissent dans les formules sont d'autant plus importants que N est petit et le rapport n/N (appelé *taux de sondage*) élevé.

Le calcul des variances d'estimation globale en géostatistique ne requiert qu'un modèle variographique, qui est établi à l'issue de l'analyse variographique des données. Quant aux méthodes classique et descriptive, pour lesquelles l'analyse variographique est absente, on estime la variance *a priori* σ^2 et les dispersions statistiques $s^2(\cdot)$ par les variances empiriques associées, correctement normalisées. Plus précisément, des estimateurs sans biais sont fournis par :

$$\begin{aligned} \text{support ponctuel : } \hat{\sigma}^2 &= \hat{s}^2(o|\mathcal{D}) = \frac{1}{n-1} \sum_{\alpha=1}^n [z(\mathbf{x}_\alpha) - z^*]^2 \\ \hat{s}^2(o|V_i) &= \frac{1}{p-1} \sum_{\mathbf{x}_\alpha \in V_i} [z(\mathbf{x}_\alpha) - z^*(V_i)]^2 \quad \text{où } z^*(V_i) = \frac{1}{p} \sum_{\mathbf{x}_\alpha \in V_i} z(\mathbf{x}_\alpha) \\ \text{support non ponctuel : } \hat{\sigma}^2 &= \hat{s}^2(v|\mathcal{D}) = (1 - \frac{1}{N}) \frac{1}{n-1} \sum_{\alpha=1}^n [z(v_\alpha) - z^*]^2 \\ \hat{s}^2(v|V_i) &= \frac{Np-n}{Np} \frac{1}{p-1} \sum_{v_\alpha \in V_i} [z(v_\alpha) - z^*(V_i)]^2 \quad \text{où } z^*(V_i) = \frac{1}{p} \sum_{v_\alpha \in V_i} z(v_\alpha). \end{aligned}$$

Dans l'approche classique, la stratégie d'échantillonnage est indifférente. Cela est dû au fait que la position spatiale des échantillons est absente du formalisme, où toutes les observations sont considérées *indépendantes* les unes des autres. **Le modèle classique est peu réaliste et doit être abandonné dans l'étude des données régionalisées.**

IV. Application aux données agronomiques

Nous allons revenir sur l'exemple introduit aux chapitres 1 et 3, concernant la répartition de surfaces de canne à sucre. La zone d'étude a été conventionnellement circonscrite et couvre 16050 ha (cf. chapitre 3) ; la structure de la régionalisation sur ce champ est décrite par un modèle isotrope de variogramme. On suppose que, la réalité n'étant pas connue, on souhaite estimer l'*abondance*, c'est-à-dire la superficie totale de canne à sucre dans la zone étudiée, à partir d'un échantillon. Le taux d'échantillonnage est fixé à 1%. Nous allons examiner les répercussions du choix du mode d'échantillonnage et de la méthode d'estimation sur la précision de l'estimation.

On forme d'abord l'estimateur de $Z(\mathcal{D})$, superficie moyenne de canne à sucre par cellule de 1 ha, par la moyenne des échantillons :

$$Z^* = \frac{1}{n} \sum_{\alpha=1}^n Z(x_{\alpha}).$$

L'estimateur de la superficie totale s'en déduit en multipliant par la surface du champ, à savoir $|\mathcal{D}| = 16\,050$ ha. Quant à la variance d'estimation de l'abondance, elle est égale à celle de la moyenne multipliée par $|\mathcal{D}|^2$.

Les variances d'estimation de $Z(\mathcal{D})$ que l'on obtient suivant la stratégie d'échantillonnage sont résumées dans le tableau 4.4. Le taux de sondage étant fixé à 1%, on a pris $n = 160$ échantillons. Pour les échantillonnages régulier et stratifié aléatoire, on a supposé en première approximation que le champ était divisible en cellules de 100 ha. Pour faciliter les comparaisons entre les stratégies d'échantillonnage, les variances ont été calculées à l'aide du même modèle de variogramme, à savoir le modèle de variogramme régional, qui est le modèle "idéal" établi à partir de l'ensemble complet des données. En toute rigueur, il faudrait refaire l'analyse variographique au cas par cas, à partir des échantillons particuliers tirés, mais il serait alors plus difficile de comparer objectivement les résultats obtenus.

	statistique classique	géostatistique intrinsèque
aléatoire uniforme	6.25×10^{-4}	6.24×10^{-4}
aléatoire stratifié		5.69×10^{-4}
régulier		4.50×10^{-4}

tableau 4.4 : variances d'estimation globale de la moyenne (valeur réelle = 0.2125)

On voit combien il est important de prendre en compte les corrélations spatiales dans le calcul de la variance d'estimation. Ainsi, dans le cas de l'échantillonnage régulier, l'approche classique, qui néglige ces corrélations, surestime de 39% la variance d'estimation globale.

Les écarts-types d'estimation varient, selon le cas, entre 0.021 et 0.025, c'est-à-dire que l'erreur que l'on est susceptible de commettre est de l'ordre de grandeur du dixième de la valeur à estimer.

Le tableau 4.5 donne les variances correspondantes relatives à l'estimation de l'abondance ou superficie totale de canne à sucre. On y a ajouté les variances obtenues en géostatistique transitive (en se plaçant là aussi dans le cas idéal, i.e. en utilisant le vrai covariogramme transitif) :

	statistique classique	géostatistique intrinsèque	géostatistique transitive
aléatoire uniforme	161 000 ha ²	160 000 ha ²	168 000 ha ²
aléatoire stratifié		146 600 ha ²	128 300 ha ²
régulier		115 900 ha ²	80 119 ha ²

tableau 4.5 : variances d'estimation globale de l'abondance (valeur réelle = 3416 ha)

Ces variances correspondent aux écarts-types suivants :

	statistique classique	géostatistique intrinsèque	géostatistique transitive
aléatoire uniforme	401 ha	400 ha	410 ha
aléatoire stratifié		383 ha	358 ha
régulier		340 ha	283 ha

tableau 4.6 : écarts-types d'estimation globale de l'abondance

Les approches géostatistiques mettent en évidence la hiérarchie entre les stratégies d'échantillonnage, l'échantillonnage régulier étant le plus précis, suivi de l'aléatoire stratifié et finalement l'aléatoire uniforme. Dans ce dernier cas, les variances d'estimation globale sont comparables à celle fournie par la statistique classique. Cela signifie que le champ est suffisamment large pour pouvoir assimiler la variance *a priori* et la variance de dispersion d'un point dans $\mathcal{D}$.

En revanche, la géostatistique transitive donne des variances d'estimation notablement plus faibles dans les cas régulier et aléatoire stratifié. L'explication vient du fait que les variances transitives ne dépendent que du covariogramme et de la cellule de référence, et sont indifférentes au choix du champ $\mathcal{D}$, alors que ce choix se répercute dans les expressions classiques et intrinsèques. En effet, si l'on change le domaine d'étude $\mathcal{D}$, on modifie également le modèle de variogramme et le nombre d'échantillons. Dans notre exemple, le champ choisi n'est pas "optimal" car il comporte plus de la moitié des cellules non cultivées. On conçoit alors que les estimations classiques et intrinsèques auraient été plus précises si le champ avait été réduit de moitié (la superficie du champ intervient au carré dans les variances d'estimation, alors que le nombre d'échantillons n'intervient que par son inverse).

V. Complément : géostatistique transitive ou géostatistique intrinsèque ?

La géostatistique transitive ne fait aucune hypothèse sur la variable régionalisée étudiée. Au contraire, la géostatistique intrinsèque s'appuie sur des hypothèses de *stationnarité* et d'*ergodicité* de la fonction aléatoire associée, ou de ses accroissements ; moyennant ces hypothèses, il est possible de remplacer l'espérance mathématique en un point donné (inaccessible en pratique, puisque l'on dispose au mieux d'une seule réalisation en ce point) par une moyenne spatiale sur l'ensemble du champ.

L'approche intrinsèque n'est pertinente que si le phénomène étudié est suffisamment homogène dans le champ pour permettre un modèle de fonction aléatoire stationnaire ou à accroissements stationnaires. Le comportement de la variable régionalisée doit être "indépendant" de la plus ou moins grande proximité de la frontière du champ $\mathcal{D}$. Il faut notamment éviter la présence d'*effets de bord* (c'est-à-dire la présence de valeurs fortes situées préférentiellement au centre du champ, et de valeurs faibles en périphérie), faute de quoi une hypothèse de stationnarité ne serait pas acceptable. Cette condition n'est pas requise en géostatistique transitive ; l'outil structural utilisé (le covariogramme transitif) résume les traits essentiels à la fois de la variable régionalisée et de son champ (cf. chapitre 1), qui sont en quelque sorte étudiés simultanément.

Une analyse de la régionalisation est donc nécessaire pour choisir l'approche la plus adaptée : transitive ou intrinsèque. Il faut chercher à mettre en évidence une éventuelle interférence entre les propriétés de la variable régionalisée elle-même (partie "intrinsèque") et celles du champ (partie "géométrique"). Cette analyse repose sur l'étude conjointe du covariogramme transitif et du covariogramme géométrique, qui permet de tester expérimentalement l'hypothèse de l'indépendance de la variable et de son champ.

Supposons que la variable régionalisée $z(\mathbf{x})$ soit une réalisation d'une fonction aléatoire *stationnaire* et *ergodique* $Z(\mathbf{x})$ définie dans un champ $\mathcal{D}$ fixé. Alors la covariance non centrée

$$C_{nc}(\mathbf{h}) = E[Z(\mathbf{x} + \mathbf{h})Z(\mathbf{x})] \quad (\text{indépendante de } \mathbf{x} \text{ par hypothèse de stationnarité})$$

est estimée par la moyenne spatiale

$$\hat{c}_{nc}(\mathbf{h}) = \frac{1}{|\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}|} \int_{\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}} z(\mathbf{x}) z(\mathbf{x} + \mathbf{h}) d\mathbf{x}$$

où $\mathcal{D}_{-\mathbf{h}}$ représente le translaté du champ $\mathcal{D}$ par le vecteur $-\mathbf{h}$

$|\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}|$ est la mesure (longueur, aire, volume...) de $\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}$.

$\hat{c}_{nc}(\mathbf{h})$ est appelée **covariance régionale** ; elle correspond à l'estimateur classique de la covariance calculé à partir de la totalité des valeurs prises par $z(\mathbf{x})$. Comme le variogramme régional, il s'agit bien d'une *grandeur régionale*, définie de manière univoque par la connaissance complète de la variable régionalisée. Si, comme on le fait en géostatistique transitive, on prolonge $z(\mathbf{x})$ en dehors du champ $\mathcal{D}$ par zéro, la covariance régionale devient :

$$\hat{c}_{nc}(\mathbf{h}) = \frac{\int_{\text{espace}} z(\mathbf{x}) z(\mathbf{x} + \mathbf{h}) d\mathbf{x}}{K(\mathbf{h})} = \frac{g(\mathbf{h})}{K(\mathbf{h})}$$

où $g(\mathbf{h})$ représente le covariogramme transitif de $z(\mathbf{x})$ (cf. chapitre 1)

$K(\mathbf{h})$ est son covariogramme géométrique : $K(\mathbf{h}) = |\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}|$.

Cette estimation $\hat{c}_{nc}(\mathbf{h})$, calculée à partir d'une réalisation unique de Z , constitue une réalisation d'une variable aléatoire $\hat{C}_{nc}(\mathbf{h})$:

$$\hat{C}_{nc}(\mathbf{h}) = \frac{\int_{\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}} Z(\mathbf{x}) Z(\mathbf{x} + \mathbf{h}) d\mathbf{x}}{K(\mathbf{h})}$$

dont l'espérance est égale à la covariance centrée théorique et dont la variance tend vers 0 lorsque $\mathcal{D}$ est suffisamment grand (par ergodicité) :

$$\begin{aligned} E[\hat{C}_{nc}(\mathbf{h})] &= C_{nc}(\mathbf{h}) \\ \text{var}[\hat{C}_{nc}(\mathbf{h}) - C_{nc}(\mathbf{h})] &\xrightarrow{|\mathcal{D}| \rightarrow \infty} 0 \end{aligned}$$

Autrement dit, on peut identifier la covariance régionale à la fonction de covariance non centrée $C_{nc}(\mathbf{h})$. Or, pour un processus stationnaire ergodique, $C_{nc}(\mathbf{h})$ admet un palier égal au carré de l'espérance de Z , lorsque $|\mathbf{h}|$ tend vers l'infini :

$$E[\hat{C}_{nc}(\mathbf{h})] = C_{nc}(\mathbf{h}) \xrightarrow{|\mathbf{h}| \rightarrow +\infty} m^2.$$

Ainsi donc, sous réserve de *stationnarité* et d'*ergodicité* de Z , la covariance régionale, quotient du covariogramme transitif par le covariogramme géométrique, se stabilise autour d'un palier. En pratique, le champ $\mathcal{D}$ étant borné, on ne peut faire tendre $|\mathbf{h}|$ vers l'infini, et l'on observe toujours des fluctuations autour de ce palier. Pour qu'il y ait convergence (à un degré d'approximation satisfaisant), il est nécessaire que les dimensions du champ soient largement supérieures à la portée pratique des observations, qui traduit la dimension caractéristique de leur zone d'influence.

On peut interpréter ce résultat de la manière suivante : l'existence d'un palier pour $\hat{c}_{nc}(\mathbf{h})$ indique que le covariogramme transitif est proportionnel, aux grandes distances, au covariogramme géométrique. Or, ce dernier ne dépend pas des valeurs prises par la variable régionalisée. Par suite, le comportement aux grandes distances du covariogramme transitif ne dépend que de la géométrie du champ, ce qui implique en particulier l'absence d'effets de bord (dans le cas contraire, ce comportement dépendrait aussi des valeurs de la régionalisation, car il fait intervenir les sites du champ localisés en périphérie et diamétralement opposés). L'hypothèse d'indépendance entre la variable et son champ, i.e. la stationnarité de la fonction aléatoire $Z(\mathbf{x})$, est alors possible.

Ainsi, l'examen des covariogrammes peut permettre de décider si une hypothèse de stationnarité est acceptable. Si la covariance non centrée expérimentale $\hat{c}_{nc}(\mathbf{h}) = g(\mathbf{h}) / K(\mathbf{h})$ ne se stabilise pas autour d'un palier, l'approche intrinsèque risque de ne pas être adaptée au problème. Cela peut être dû à l'existence d'un effet de bord, de sorte qu'il n'y a pas indépendance entre la variable étudiée et le champ. Il convient toutefois de formuler quelques réserves : le test proposé n'est pas une méthode infaillible qui permettrait de décider à coup sûr de la validité ou non d'une hypothèse de stationnarité. L'étude exploratoire des données et le sens physique de l'utilisateur doivent toujours avoir leur mot à dire : en particulier, l'échelle de travail peut jouer un rôle décisif dans le choix d'une hypothèse stationnaire.

Un autre facteur à prendre en considération est le mode d'échantillonnage. La géostatistique transitive ne peut être mise en œuvre dans le cas d'un échantillonnage irrégulier quelconque, et est peu commode dans le cas d'un échantillonnage aléatoire, uniforme, préférentiel ou stratifié, car il n'est pas aisé d'évaluer le covariogramme transitif à partir d'un réseau irrégulier de données.

Application aux données de canne à sucre

La figure 4.5 donne les résultats du test précédent dans le cas des données de canne à sucre. La courbe $g(h) / K(h)$, calculée sur l'ensemble des données, est visualisée le long des directions nord-sud et est-ouest.

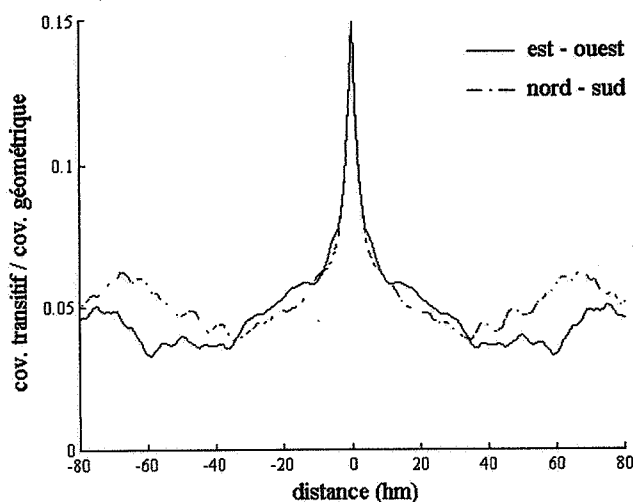


figure 4.5 : covariance régionale le long des directions nord / sud et est / ouest

La visualisation a volontairement été limitée aux distances inférieures à 80 hm, car l'inférence statistique n'est pas possible dans des conditions correctes pour des distances supérieures à la moitié du diamètre du champ, même lorsque l'on dispose de la réalité entière (cf. la discussion sur le variogramme régional au chapitre 3). A l'échelle d'observation, qui est relativement grande par rapport à celle de la structuration des données, on peut raisonnablement admettre que la covariance régionale admet un palier autour de 0.05, qui correspond d'ailleurs approximativement au carré de la valeur moyenne de la régionalisation (0.2125). Ce résultat indique qu'une l'hypothèse de stationnarité, ou tout au moins de stationnarité locale à l'échelle de 80 hm, est *a priori* envisageable.

Cette conclusion doit être nuancée. Chaque méthode, transitive ou probabiliste, présente ses avantages et ses limitations. Ainsi, en géostatistique transitive, la contrepartie de la simplicité et la généralité de la méthode est "l'interférence" qui se produit entre les propriétés de la variable et celles de son champ, en particulier l'anisotropie qui, dans cet exemple, est d'origine purement géométrique. Au contraire, l'approche probabiliste permet de séparer l'anisotropie des traits structuraux propres à la régionalisation, mais elle n'est pas pertinente aux frontières du champ, où l'on observe un effet de bord (valeurs préférentiellement plus faibles).

Résumé

Le formalisme du changement de support, ainsi que les concepts de variances d'extension et de dispersion, permettent de déduire les propriétés des variables définies sur des blocs, à partir des propriétés de la variable ponctuelle. Ils constituent aussi des outils efficaces pour résoudre les problèmes d'estimation globale.

L'estimation globale consiste à évaluer la valeur moyenne de la variable régionalisée sur l'ensemble du champ, et à calculer la variance d'estimation associée, qui sert à mesurer la précision de l'estimation. Souvent, il n'est pas utile de proposer un estimateur sophistiqué (dans ce chapitre, on se contente de prendre la moyenne arithmétique des données). L'intérêt de la géostatistique par rapport aux méthodes classiques, est que le calcul de la variance d'estimation global prend en compte de la position des observations et de la structure spatiale de la variable régionalisée.

Plusieurs stratégies d'échantillonnage conduisent à des expressions simples de la variance d'estimation et peuvent être comparées : ainsi, à taux de sondage égal, l'échantillonnage aléatoire uniforme donne une variance d'estimation supérieure à celle de l'échantillonnage aléatoire stratifié. Dans la plupart des cas, l'échantillonnage régulier est encore meilleur. Enfin, dans les cas où se posent des problèmes de délimitation du champ ou des problèmes de non-stationnarité, il est recommandé d'avoir recours aux méthodes transitives (*cf.* chapitre 1).

Chapitre 5 :

l'estimation locale : le krigeage

I. Introduction

Contrairement à l'estimation globale, l'estimation locale ne concerne qu'un secteur du champ étudié, et utilise en général un nombre limité des données disponibles. Dans ce chapitre, nous allons étudier le *krigeage* (baptisé ainsi en l'honneur d'un des précurseurs de la géostatistique, Daniel Krige), méthode qui permet d'estimer la valeur d'un site ou d'un bloc¹ à partir des valeurs observées sur les sites avoisinants, en prenant en compte leur configuration géométrique et la structure spatiale de la régionalisation.

Le krigeage s'appuie sur l'interprétation de la variable régionalisée comme la réalisation d'une fonction aléatoire, dont on a modélisé la fonction de covariance ou le variogramme. Il s'agit de rechercher, parmi les estimateurs formés par les combinaisons linéaires pondérées des données, celui qui présente les "meilleures" propriétés (en l'occurrence absence de biais et variance d'erreur minimale, qui constitue le critère de précision choisi).

Le krigeage présente plusieurs avantages sur les techniques d'interpolation déterministes (méthode de l'inverse des distances, splines, interpolation par plus proche voisin, régression polynomiale...). D'une part, il fournit une estimation précise et sans biais de la grandeur recherchée, qui prend en compte non seulement les informations de nature géométrique (nombre et configuration des sites de données), mais également les informations structurales contenues dans le modèle variographique. D'autre part, il permet d'apprécier quantitativement la précision de l'estimation, à l'aide d'une *variance d'estimation*, ce qui n'est pas possible sans le recours à un modèle stochastique. Signalons dès à présent qu'il ne faut pas confondre cette variance avec un intervalle de confiance sur l'estimation², auquel on a l'habitude de se référer.

¹ On appelle "bloc" une unité de taille supérieure à celle des échantillons, par opposition aux "points" qui sont de même taille que les échantillons.

² Un intervalle de confiance est défini par une borne inférieure, une borne supérieure et une marge d'erreur. C'est un outil bien plus significatif qu'une variance d'estimation, mais il nécessite d'identifier la loi de probabilité de l'erreur d'estimation (ce qui requiert des hypothèses supplémentaires).

II. Les quatre étapes du krigeage

La résolution d'un problème d'estimation locale par krigeage s'articule toujours autour des quatre mêmes étapes. Les différentes variantes que l'on rencontre tiennent seulement aux hypothèses faites sur la fonction aléatoire représentant le phénomène régionalisé (stationnarité d'ordre deux, avec moyenne connue ou non, ou hypothèse intrinsèque).

Le problème est le suivant : on souhaite estimer une valeur ξ_0 qui est une fonction linéaire de la variable régionalisée étudiée $z(\mathbf{x})$. Par exemple,

- 1) $\xi_0 = z(\mathbf{x}_0)$ valeur ponctuelle de z en $\mathbf{x} = \mathbf{x}_0$;
- 2) $\xi_0 = \frac{1}{|V_0|} \int_{V_0} z(\mathbf{x}) d\mathbf{x}$ valeur moyenne sur le bloc V_0 centré en $\mathbf{x} = \mathbf{x}_0$;
- 3) $\xi_0 = \int z(\mathbf{x}_0 + \mathbf{h}) p(\mathbf{h}) d\mathbf{h}$ moyenne mobile pondérée de z ("régularisée" de z par une fonction de pondération p de poids total unité)³ ;
- 4) $\xi_0 = \text{grad } z(\mathbf{x}) \big|_{\mathbf{x}=\mathbf{x}_0}$ composantes du gradient en $\mathbf{x} = \mathbf{x}_0$ (qui donnent l'inclinaison et la direction d'une strate géologique lorsque z représente la hauteur de cette strate).

Dans la suite de ce chapitre, nous n'allons détailler que les deux premiers cas, sans perte de généralité car nous verrons que toute estimation locale peut se ramener, par linéarité des opérations, à des estimations ponctuelles (c'est-à-dire au premier cas).

La quantité autour de laquelle repose le krigeage n'est pas l'estimation elle-même (notée ξ_0^*), mais l'**erreur d'estimation** $\xi_0^* - \xi_0$, différence entre l'estimation et la quantité à estimer. Cette erreur, qui est une variable aléatoire $\Xi_0^* - \Xi_0$ dans le modèle probabiliste, va devoir vérifier certaines contraintes.

- **contrainte de linéarité**

L'erreur de krigeage en un site donné doit être une combinaison linéaire pondérée de la fonction aléatoire Z étudiée. Cette contrainte est due à la décision de ne considérer que les deux premiers moments des lois de probabilité (espérance et variance), qui ne sont calculables que sur les combinaisons linéaires de Z . La construction d'estimateurs plus sophistiqués nécessiterait de spécifier la loi spatiale de Z au-delà de ses deux premiers moments ; les méthodes de géostatistique non linéaire permettent d'élaborer de tels estimateurs (krigeage disjonctif, krigeage lognormal, espérance conditionnelle).

³ La formulation générale d'une régularisée inclut les deux cas précédents (valeur ponctuelle et valeur moyenne sur un bloc).

La quantité à estimer étant déjà une fonction linéaire de $Z(\mathbf{x})$, la contrainte de linéarité est satisfaite en écrivant l'estimateur comme une combinaison linéaire pondérée des valeurs mesurées :

$$\Xi_0^* = \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha})$$

où les $\mathbf{x}_{\alpha}$ ($\alpha = 1 \dots n$) sont les sites échantillonnés utilisés pour l'estimation et où les poids λ_{α} sont les inconnues du problème de krigeage. Ces poids dépendront à la fois de la structure spatiale de la régionalisation et de la **configuration de krigeage**, c'est-à-dire de la configuration géométrique formée par les points de mesure et le site à estimer.

N. B. : l'estimateur ainsi défini est une variable aléatoire. Pour obtenir une estimation numérique, on applique les relations concernant la fonction aléatoire à la *réalisation* particulière que constituent les données expérimentales. Autrement dit, une fois les poids déterminés, l'estimation de ξ_0 sera

$$\xi_0^* = \sum_{\alpha=1}^n \lambda_{\alpha} z(\mathbf{x}_{\alpha}).$$

- **contrainte d'autorisation**

L'erreur d'estimation étant donnée (combinaison linéaire pondérée de la fonction aléatoire), il est nécessaire de *pouvoir* calculer son espérance et sa variance. Dans le cadre stationnaire d'ordre deux, il n'y a pas de contrainte effective d'autorisation car toutes les combinaisons linéaires admettent espérance et variance. En revanche, sous l'hypothèse intrinsèque, une combinaison linéaire pondérée est autorisée si et seulement si la somme de ses poids est nulle, faute de quoi les expressions que l'on écrira dans les étapes suivantes n'auront pas de sens.

- **contrainte de non-biais**

Cette étape consiste à exprimer que l'erreur d'estimation est d'espérance nulle, c'est-à-dire

$$E[\Xi_0^* - \Xi_0] = 0.$$

On ne peut écrire cette formule que si la contrainte précédente est satisfaite, qui assure l'existence de l'espérance de l'erreur. On peut interpréter cette contrainte de non biais en remplaçant l'espérance mathématique par une moyenne spatiale : *lorsqu'elle est calculée sur de nombreuses configurations de krigeage identiques, réparties sur tout le champ, la moyenne des erreurs d'estimation commises est proche de zéro*. L'absence de biais ne garantit pas que les erreurs soient faibles, mais seulement que leur moyenne globale est approximativement nulle.

- **contrainte d'optimalité**

A l'issue des étapes précédentes, l'estimateur est soumis à une ou plusieurs conditions mais n'est pas totalement spécifié. La dernière étape consiste à trouver les pondérateurs qui minimisent la variance de l'erreur d'estimation.

Intuitivement, la contrainte d'optimalité signifie que *lorsqu'elle est calculée sur de nombreuses configurations de krigeage identiques, réparties sur tout le champ, la variance statistique des erreurs d'estimation commises est aussi faible que possible*. Le critère de précision de l'estimation par krigeage correspond donc à la minimisation de l'erreur quadratique moyenne.

Conclusion

Le krigeage s'articule autour de quatre étapes hiérarchisées :

- la condition d'autorisation présuppose la condition de linéarité ;
- la condition de non biais ne peut être écrite que si l'erreur d'estimation vérifie la condition d'autorisation ;
- la condition d'optimisation concerne les combinaisons linéaires satisfaisant les contraintes d'autorisation et de non biais.

III. Choix du voisinage de krigeage

Contrairement aux problèmes globaux, les problèmes d'estimation locale ne mettent pas en jeu la totalité du champ et, en général, n'utilisent pas l'ensemble des données disponibles. On définit le **voisinage de krigeage** comme le domaine de l'espace qui contient le support de la grandeur à estimer et les données utilisées dans l'estimation. Plusieurs possibilités sont offertes à l'utilisateur.

III.1. Voisinage unique

On peut décider d'effectuer le krigeage en un point ou un bloc quelconque en conservant l'ensemble des points de données. On parle alors de krigeage en **voisinage unique**. Dans ce cas, les données très éloignées interviennent dans l'estimation. Cependant, sauf cas particulier, leur influence aura de fortes chances d'être faible ; (intuitivement, un site éloigné n'apporte pas beaucoup d'information sur le point ou le bloc à estimer, et sera par conséquent affecté d'un poids de krigeage faible).

Un krigeage en voisinage unique n'est pas adapté lorsque l'hypothèse stationnaire ou intrinsèque n'est vérifiée que *localement* (quasi-stationnarité). De même, lorsque les données sont très nombreuses, il est inutile de toutes les conserver pour une estimation locale, d'autant que cela risque d'augmenter notablement les temps de calcul. Il est alors nécessaire de réduire la taille du voisinage de krigeage.

III.2. Voisinage glissant

Le krigeage est effectué en **voisinage glissant** lorsqu'il n'utilise que les points de données "voisins" du site à estimer. En général, on ne se limite pas à une seule estimation locale, mais l'on cherche à estimer la variable régionalisée aux nœuds d'une grille régulière qui couvre la zone étudiée. Il reste à définir la taille et la forme du voisinage, que l'on centre sur le point ou le bloc à estimer et que l'on déplace à travers le champ à mesure que sont réalisées les estimations (d'où le qualificatif de *glissant*).

- **taille du voisinage**

L'hypothèse stationnaire ou intrinsèque doit être admissible à l'échelle du voisinage choisi. D'autre part, nous allons voir que, dans les équations du krigeage, la covariance ou le variogramme n'intervient que pour des distances inférieures au diamètre de la zone où l'on prend l'information qui sert à l'estimation. Ainsi, si le modèle variographique est peu fiable aux grandes distances, il est préférable de travailler avec un voisinage glissant de taille modérée.

En revanche, si l'on choisit un voisinage trop petit, contenant peu de données, l'estimation sera peu précise et sensible aux valeurs de ces données ; la carte de krigeage risque fort de présenter des artefacts (mauvais "raccordements" lorsque l'on passe d'un point à un autre et que les données utilisées ne sont pas les mêmes).

La taille du voisinage doit donc permettre un équilibre entre ces différents facteurs (précision et raccordement des estimations / temps de calcul, quasi-stationnarité et fiabilité du variogramme). Nous verrons ultérieurement un autre facteur — la minimisation du *biais conditionnel* — susceptible d'intervenir dans le choix de la taille du voisinage. Un critère objectif de décision est la **validation croisée** : on teste plusieurs tailles de voisinage et on retient celle qui donne les résultats les plus satisfaisants.

Remarque

Il n'y a pas de justification particulière pour limiter la taille du voisinage à la portée du modèle variographique, sous prétexte que les données situées au-delà de cette portée n'ont pas de corrélation avec le site à estimer. En fait, dans la plupart des cas, ces données interviennent indirectement dans l'estimation et en améliorent la précision, parfois de manière non négligeable. Le facteur à considérer dans le choix de la taille du voisinage est davantage la densité de l'échantillonnage que la portée du variogramme ou de la covariance.

- forme du voisinage

La forme du voisinage doit, dans la mesure du possible, tenir compte de l'anisotropie de la variable révélée par l'analyse variographique. Ainsi, dans le cas d'une anisotropie géométrique, on prendra un voisinage en forme d'ellipse (ou d'ellipsoïde) dont les caractéristiques – orientation et excentricité – sont identiques à celles de l'ellipse (ou de l'ellipsoïde) d'anisotropie. Souvent aussi, on divise cette ellipse en plusieurs secteurs, en général des quadrants ou des octants, dans chacun desquels on recherche un nombre fixé de données, afin de mieux répartir autour du point ou du bloc à estimer l'information que l'on va conserver.

La figure 5.1 présente un exemple, dans l'espace bidimensionnel, de voisinage glissant en forme d'ellipse centrée sur le site à estimer. Les données retenues, au maximum de trois par octant, sont indiquées.

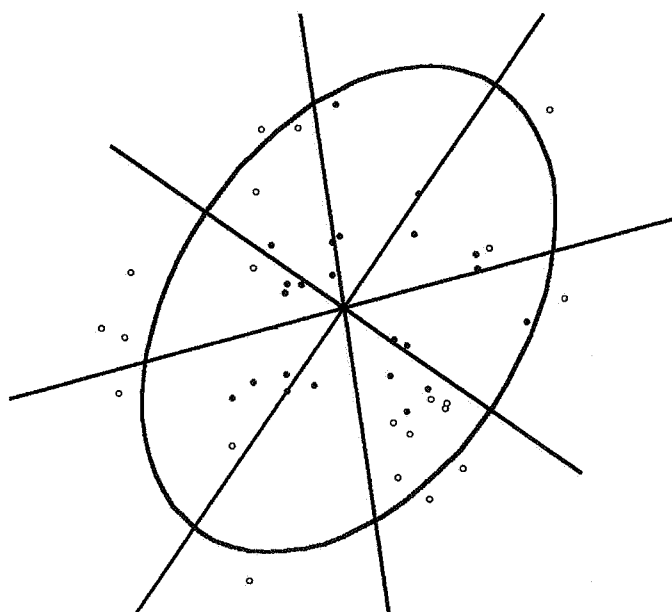


figure 5.1 : voisinage elliptique, divisé en octants
"=" = données retenues ; "o" = données écartées

En cas d'anisotropie plus complexe que l'anisotropie géométrique, on conserve souvent un voisinage en forme d'ellipse ou d'ellipsoïde, bien que l'on devrait idéalement choisir une forme plus sophistiquée (par exemple, un voisinage en forme de "bande" si l'on a affaire à une anisotropie zonale pure). Il faut donc chercher une ellipse qui se rapproche au mieux des courbes d'isovaleurs du modèle variographique, qui indiquent le niveau de corrélation, ou d'information, en fonction de la distance géographique.

IV. Le krigeage ponctuel

Nous allons d'abord nous intéresser au cas courant du krigeage ponctuel, c'est-à-dire que la quantité à estimer n'est autre que la valeur ponctuelle inconnue $z(\mathbf{x}_0)$ en un point non échantillonné $\mathbf{x}_0$ (voire en un point échantillonné). En règle générale, le krigeage ponctuel est effectué non pas en un seul point, mais aux nœuds d'une grille régulière couvrant la zone d'intérêt.

IV.1. Le krigeage à moyenne connue (krigeage simple)

On suppose ici que z est la réalisation d'une fonction aléatoire Z **stationnaire du second ordre** telle que

$$\begin{cases} \forall \mathbf{x} \in \mathcal{V}, E[Z(\mathbf{x})] = m \text{ connue} \\ \forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{V}, \text{cov}[Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})] = C(\mathbf{h}) \end{cases}$$

où $\mathcal{V}$ désigne le voisinage de krigeage (pour un krigeage à voisinage unique, $\mathcal{V}$ s'identifie au champ $\mathcal{D}$ de la variable régionalisée).

On suppose de plus dans un premier temps que m est nulle : $m = 0$.

Examinons une à une les quatre étapes du krigeage.

- **linéarité** : en prenant comme estimateur en $\mathbf{x}_0$

$$Z^*(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha})$$

on assure que l'erreur d'estimation satisfait la contrainte de linéarité. En toute rigueur, on devrait écrire $\lambda_{\alpha}(\mathbf{x}_0)$ au lieu de λ_{α} , car les poids optimaux dépendront bien sûr du point $\mathbf{x}_0$ à estimer.

- **autorisation** : sous l'hypothèse stationnaire du second ordre, toute combinaison linéaire est autorisée. La contrainte d'autorisation est automatiquement vérifiée.
- **non-biais** : elle est également vérifiée, car comme $m = 0$, il vient immédiatement :

$$E[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = \sum_{\alpha=1}^n \lambda_{\alpha} \underbrace{E[Z(\mathbf{x}_{\alpha})]}_{=m=0} - \underbrace{E[Z(\mathbf{x}_0)]}_{=m=0} = 0.$$

- **optimalité** : il reste à minimiser la variance de l'erreur d'estimation, qui se développe de la manière suivante à l'aide de la fonction de covariance C :

$$\text{var}[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = C(\mathbf{0}) + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{x}_{\alpha} - \mathbf{x}_0).$$

Comme il n'y a ni contrainte d'autorisation ni contrainte de non-biais, le minimum de cette expression est obtenu en annulant ses dérivées partielles par rapport aux poids inconnus $\{\lambda_{\alpha}, \alpha = 1 \dots n\}$. On obtient le système :

$$\sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) = C(\mathbf{x}_{\alpha} - \mathbf{x}_0) \quad \forall \alpha = 1 \dots n.$$

C'est un système linéaire où le nombre d'équations et le nombre d'inconnues sont égaux au nombre de points de données utilisés. En écriture matricielle, ce système s'écrit :

$$\begin{bmatrix} C(\mathbf{x}_1 - \mathbf{x}_1) & \dots & C(\mathbf{x}_1 - \mathbf{x}_n) \\ \vdots & & \vdots \\ C(\mathbf{x}_n - \mathbf{x}_1) & \dots & C(\mathbf{x}_n - \mathbf{x}_n) \end{bmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} C(\mathbf{x}_1 - \mathbf{x}_0) \\ \vdots \\ C(\mathbf{x}_n - \mathbf{x}_0) \end{pmatrix}$$

ce qui permet de déterminer les poids de krigage optimaux :

$$[\lambda_{\beta}] = [C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta})]^{-1} [C(\mathbf{x}_{\alpha} - \mathbf{x}_0)].$$

La valeur de la variance de l'erreur au point $\mathbf{x}_0$, appelée **variance de krigage**, se simplifie en reportant les équations du krigage dans l'expression précédente de la variance :

$$\sigma_{KS}^2(\mathbf{x}_0) = \sigma^2 - \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{x}_{\alpha} - \mathbf{x}_0)$$

où $\sigma^2 = C(\mathbf{0})$ est la variance *a priori* de la fonction aléatoire Z.

Dans le cas général où Z a une moyenne (espérance) **m non nulle**, mais connue, on se ramène au cas précédent en travaillant sur la variable $Z - m$, qui, elle, est de moyenne nulle. On écrira donc :

$$Z^*(\mathbf{x}_0) - m = \sum_{\alpha=1}^n \lambda_{\alpha} [Z(\mathbf{x}_{\alpha}) - m]$$

soit encore

$$Z^*(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha}) + \left(1 - \sum_{\alpha=1}^n \lambda_{\alpha}\right) m.$$

Le système de krigeage et la variance de krigeage sont donnés par les mêmes équations que précédemment. Ce krigeage stationnaire à moyenne connue a été baptisé **krigeage simple**. Cette méthode est relativement exigeante puisque la valeur de la moyenne intervient dans l'estimation, et son poids est d'autant plus élevé que le poids cumulé des données est faible. Ainsi, dans les zones peu échantillonnées, l'estimation va se calibrer sur la moyenne pour pallier le manque d'information.

Nous allons illustrer ces propos sur un exemple volontairement très simple ; il s'agit de huit données, réparties irrégulièrement le long d'une droite. On suppose que la moyenne est égale à 0.6 et que la covariance est formée d'un schéma sphérique de palier 1 et portée 1. La figure 5.2 présente les résultats du krigeage simple lorsque l'on estime chaque point de la droite à partir des huit données disponibles.

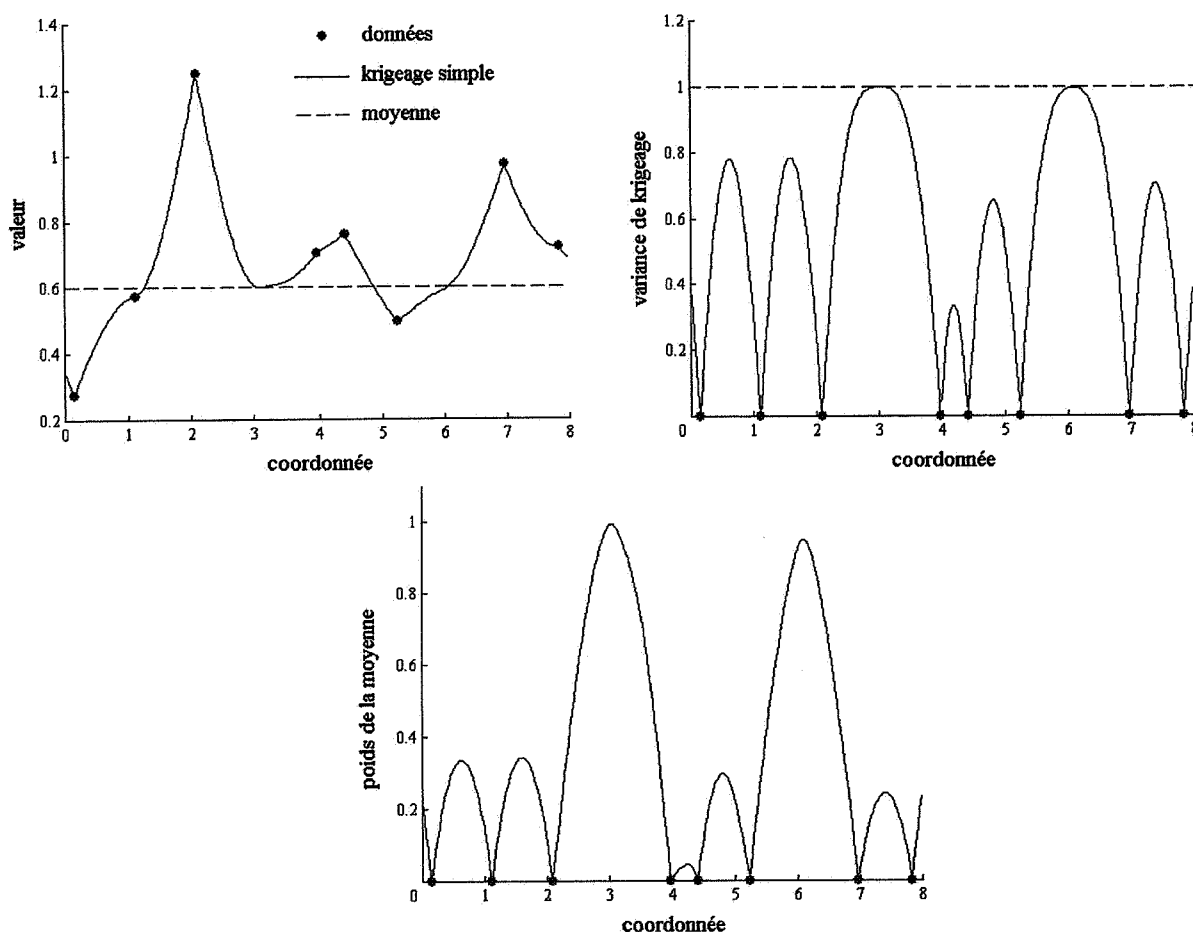


figure 5.2 : estimations par krigeage simple (en haut à gauche), variances d'estimation associées (en haut à droite) et poids attribué à la moyenne (en bas)

On observe que l'estimation d'un point de données redonne la valeur observée en ce point et que la variance d'estimation associée est nulle (propriété d'interpolation exacte du krigeage, qui est tout à fait générale).

A mesure que l'on s'éloigne des points de données, les valeurs krigées ont tendance à se rapprocher de la moyenne connue, dont le poids croît de manière significative, et la variance de krigeage augmente, indiquant une plus grande incertitude. A la limite, lorsque le site à estimer est séparé des observations d'une distance supérieure à la portée, tous les poids de krigeage sont nuls : l'estimation coïncide avec l'espérance et la variance d'estimation est égale à la variance *a priori* ($\sigma^2 = 1$) : les données n'ont plus d'influence, seule compte la connaissance de la moyenne.

Il convient également de signaler que ni la variance de krigeage ni le poids attribué à la moyenne ne dépendent des valeurs prises par les données : les variations que l'on apprécie sur les cartes correspondantes sont de nature purement géométrique.

Exemple important : cas d'un effet de pépité pur

Un exemple intéressant à examiner est celui où la fonction de covariance est purement pépitique : $C(\mathbf{0}) = \sigma^2$ et $C(\mathbf{h}) = 0$ si $|\mathbf{h}| > 0$. L'effet de pépité pur correspond à une absence totale de structuration des données. Dans ce cas, le système de krigeage simple en un point $\mathbf{x}_0$ non échantillonné se réduit à :

$$\begin{pmatrix} C(\mathbf{0}) & & 0 \\ & \ddots & \\ 0 & & C(\mathbf{0}) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

qui admet pour solution évidente $\lambda_1 = \dots = \lambda_n = 0$. Par suite, l'estimation est égale à la moyenne m et la variance d'estimation vaut σ^2 , variance *a priori* de Z . L'estimation et la variance d'estimation en un point non échantillonné sont indifférentes à la configuration de krigeage et aux valeurs mesurées aux sites de données. Tout se passe comme si les données n'apportaient aucune information sur le point à estimer⁴.

En outre, il se produit des discontinuités dans les valeurs de l'estimation et de la variance d'estimation au niveau des points de données. En effet, si le point à estimer $\mathbf{x}_0$ coïncide avec l'un des points de données $\mathbf{x}_\alpha$, le poids λ_α devient égal à 1 et les autres poids restent nuls, de sorte que l'estimation redonne la valeur observée en $\mathbf{x}_\alpha$ et la variance d'estimation est nulle (ce qui traduit que l'estimation est exacte en ce point).

En résumé, dans le cas d'un effet de pépité pur,

$$\text{si } \mathbf{x}_0 \notin \{\mathbf{x}_1, \dots, \mathbf{x}_n\}, Z^*(\mathbf{x}_0) = m \text{ et } \sigma_{KS}^2(\mathbf{x}_0) = \sigma^2 = C(\mathbf{0})$$

$$\text{si } \mathbf{x}_0 = \mathbf{x}_\alpha \in \{\mathbf{x}_1, \dots, \mathbf{x}_n\}, Z^*(\mathbf{x}_0) = Z(\mathbf{x}_\alpha) \text{ et } \sigma_{KS}^2(\mathbf{x}_0) = 0$$

⁴ On retrouve exactement le même système de krigeage lorsque la covariance, bien que non pépitique, admet une portée plus faible que les distances séparant les points de données et le point à estimer. L'estimation est alors égale à la moyenne et la variance d'estimation à la variance *a priori* du modèle.

IV.2. Le krigeage à moyenne inconnue (krigeage ordinaire)

Dans les applications, il est rare que l'on connaisse avec certitude la valeur de la moyenne. Il convient donc d'élargir le krigeage au cas où celle-ci est inconnue. Comme précédemment, on suppose que, à l'échelle du voisinage de krigeage, la régionalisation est une réalisation d'une fonction aléatoire Z **stationnaire d'ordre deux**, d'espérance m constante et de fonction de covariance C . Mais on suppose ici que m est **inconnue**. Les étapes du krigeage s'écrivent :

- **linéarité** : l'estimateur de Z en un point $\mathbf{x}_0$ est défini par :

$$Z^*(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha})$$

- **autorisation** : pas de contrainte effective dans le cadre stationnaire.
- **non-biais** : l'espérance de l'erreur d'estimation s'écrit :

$$\begin{aligned} E[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] &= \sum_{\alpha=1}^n \lambda_{\alpha} \underbrace{E[Z(\mathbf{x}_{\alpha})]}_m - \underbrace{E[Z(\mathbf{x}_0)]}_m \\ &= m \left(\sum_{\alpha=1}^n \lambda_{\alpha} - 1 \right) \end{aligned}$$

Or m étant inconnue, le seul moyen de garantir le non biais de l'erreur est d'imposer la condition (appelée parfois *condition d'universalité*) :

$$\sum_{\alpha=1}^n \lambda_{\alpha} = 1.$$

Cette égalité assure que, dans le cas où toutes les données sont égales à une même constante, la valeur estimée restituera cette constante.

- **optimalité** : la variance de l'erreur d'estimation se développe en :

$$\text{var}[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = C(0) + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{x}_{\alpha} - \mathbf{x}_0)$$

quantité qu'il faut minimiser sous contrainte de la condition de non-biais, à savoir

$$\sum_{\alpha=1}^n \lambda_{\alpha} = 1.$$

Pour résoudre ce problème de minimisation sous contrainte, on applique la technique des multiplicateurs de Lagrange. On introduit, dans l'expression de la variance, une inconnue auxiliaire μ , appelée *multiplicateur de Lagrange*, de la façon suivante :

$$\text{var}[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = C(0) + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{x}_{\alpha} - \mathbf{x}_0) + 2\mu \underbrace{\left(\sum_{\alpha=1}^n \lambda_{\alpha} - 1 \right)}_{=0}$$

et l'on minimise la fonction des $n+1$ variables $\lambda_1, \dots, \lambda_n, \mu$. En calculant les $n+1$ dérivées partielles de cette fonction puis en les annulant, on obtient le système :

$$\begin{cases} \frac{\partial}{\partial \lambda_{\alpha}} = 0 : \sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + \mu = C(\mathbf{x}_{\alpha} - \mathbf{x}_0) & \forall \alpha = 1 \dots n \\ \frac{\partial}{\partial \mu} = 0 : \sum_{\alpha=1}^n \lambda_{\alpha} = 1 & \text{(condition de non-biais)} \end{cases}$$

Ce système comporte une inconnue et une équation de plus que le système de krigeage simple. Ce krigeage est appelé **"krigeage ordinaire"**.

Le système de krigeage ordinaire écrit en notation matricielle est :

$$\begin{bmatrix} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) & \mathbf{1} \\ \mathbf{1} & 0 \end{bmatrix} \begin{bmatrix} \lambda_{\beta} \\ \mu \end{bmatrix} = \begin{bmatrix} C(\mathbf{x}_{\alpha} - \mathbf{x}_0) \\ 1 \end{bmatrix}$$

$$\begin{pmatrix} C(\mathbf{x}_1 - \mathbf{x}_1) & \dots & C(\mathbf{x}_1 - \mathbf{x}_n) & 1 \\ \vdots & & \vdots & \vdots \\ C(\mathbf{x}_n - \mathbf{x}_1) & \dots & C(\mathbf{x}_n - \mathbf{x}_n) & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} C(\mathbf{x}_1 - \mathbf{x}_0) \\ \vdots \\ C(\mathbf{x}_n - \mathbf{x}_0) \\ 1 \end{pmatrix}$$

Les poids de krigeage sont donnés par :

$$\begin{bmatrix} \lambda_{\beta} \\ \mu \end{bmatrix} = \begin{bmatrix} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) & \mathbf{1} \\ \mathbf{1} & 0 \end{bmatrix}^{-1} \begin{bmatrix} C(\mathbf{x}_{\alpha} - \mathbf{x}_0) \\ 1 \end{bmatrix}.$$

La variance de krigeage au point $\mathbf{x}_0$ s'écrit, en tenant compte des équations de krigeage :

$$\sigma_{\text{KO}}^2(\mathbf{x}_0) = \sigma^2 - \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{x}_{\alpha} - \mathbf{x}_0) - \mu.$$

Le variogramme étant équivalent à la covariance, d'après la relation $\gamma(\mathbf{h}) = C(0) - C(\mathbf{h})$, on peut choisir de l'utiliser au lieu de la fonction de covariance. Les équations de krigeage deviennent :

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) - \mu = \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_0) & \forall \alpha = 1 \dots n \\ \sum_{\alpha=1}^n \lambda_{\alpha} = 1 \end{cases}$$

soit :

$$\begin{bmatrix} \gamma(\mathbf{x}_\alpha - \mathbf{x}_\beta) & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_\beta \\ -\mu \end{bmatrix} = \begin{bmatrix} \gamma(\mathbf{x}_\alpha - \mathbf{x}_0) \\ 1 \end{bmatrix}$$

$$\begin{pmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_1) & \gamma(\mathbf{x}_1 - \mathbf{x}_n) & 1 \\ \gamma(\mathbf{x}_n - \mathbf{x}_1) & \gamma(\mathbf{x}_n - \mathbf{x}_n) & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_n \\ -\mu \end{pmatrix} = \begin{pmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_0) \\ \gamma(\mathbf{x}_n - \mathbf{x}_0) \\ 1 \end{pmatrix}$$

La variance de krigeage vaut :

$$\sigma_{KO}^2(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_\alpha \gamma(\mathbf{x}_\alpha - \mathbf{x}_0) - \mu.$$

Evaluation optimale de la moyenne (espérance mathématique)

Les quatre étapes du krigeage peuvent être utilisées pour évaluer la moyenne m inconnue (qui en toute rigueur est un *paramètre* du modèle, et non une grandeur régionale, c'est-à-dire une grandeur physiquement mesurable). On pose :

$$M^* = \sum_{\alpha=1}^n \lambda_\alpha Z(\mathbf{x}_\alpha)$$

comme estimateur de m .

Après écriture des étapes du krigeage, on aboutit au système :

$$\begin{cases} \sum_{\beta=1}^n \lambda_\beta C(\mathbf{x}_\alpha - \mathbf{x}_\beta) + \mu = 0 & \forall \alpha = 1 \dots n \\ \sum_{\alpha=1}^n \lambda_\alpha = 1 \end{cases}$$

qui ne diffère du système de krigeage ordinaire que par le second membre.

La variance d'estimation vaut $\sigma_m^2 = -\mu$, relation qui montre que le multiplicateur de Lagrange introduit ici est toujours négatif.

Illustration

Nous allons reprendre l'exemple des huit sites pour illustrer succinctement la technique du krigeage ordinaire (figure 5.3). Il est intéressant d'ajouter, sur le graphique des valeurs krigées, la valeur estimée de l'espérance mathématique.

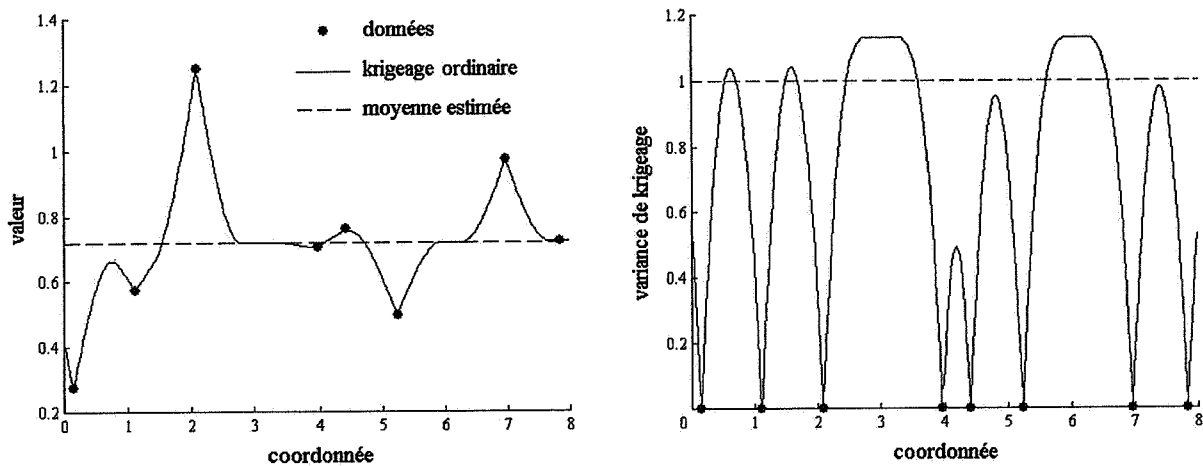


figure 5.3 : estimations par krigeage ordinaire (à gauche)
et variances d'estimation associées (à droite)

A l'instar du krigeage simple, on retrouve la propriété d'interpolation exacte des données, avec une variance de krigeage associée nulle, ce qui signifie que l'estimation est parfaite. Par contre, loin des données, l'estimation n'est plus "attirée" vers une moyenne fixée *a priori*, mais vers une valeur légèrement supérieure à 0.7. Cette valeur correspond en fait à la moyenne *estimée* de manière optimale à partir des données ; cette dernière est modifiée dès que l'on change de voisinage de krigeage, ce qui implique que l'estimation est calibrée autour d'une moyenne *locale*.

Par rapport au krigeage simple, on voit que la variance de krigeage peut être supérieure à la variance *a priori*. L'explication vient du fait que la méconnaissance de la moyenne introduit une incertitude supplémentaire dans l'estimation.

Cas particulier d'un effet de pépité pur

Lorsque la covariance est purement pépitique (c'est-à-dire nulle pour tout vecteur $\mathbf{h}$ non nul), il n'y a aucune structure spatiale entre les valeurs observées aux différents sites. On vérifie alors qu'en un point non échantillonné le système de krigeage attribue des poids égaux à tous les sites d'observation⁵. Autrement dit, en présence d'un effet de pépité pur, l'estimation ne privilégie aucun échantillon et coïncide avec la moyenne arithmétique des données :

$$\text{si } \mathbf{x}_0 \notin \{\mathbf{x}_1, \dots, \mathbf{x}_n\}, Z^*(\mathbf{x}_0) = \frac{1}{n} \sum_{\alpha=1}^n Z(\mathbf{x}_\alpha).$$

La variance de krigeage ordinaire en un site *non échantillonné* est légèrement supérieure à la variance *a priori* $\sigma^2 = C(\mathbf{0})$:

$$\sigma_{\text{KO}}^2(\mathbf{x}_0) = \underbrace{\sigma^2}_{\text{variance a priori de la fonction } Z} + \underbrace{\frac{\sigma^2}{n}}_{\text{variance d'estimation de la moyenne}}.$$

La forme de l'estimateur et de la variance d'estimation change en un point de données : le krigeage redonne la valeur observée : poids égal à 1 pour le point à estimer, poids nul pour les autres points, multiplicateur de Lagrange nul. En conséquence, la variance d'estimation est nulle. Il y a donc des discontinuités dans l'estimation et la variance d'estimation au niveau des points de données.

IV.3. Le krigeage intrinsèque ordinaire

Supposons maintenant que la fonction aléatoire Z soit **intrinsèque** (non nécessairement stationnaire). Les étapes de krigeage s'écrivent :

- **linéarité** : $Z^*(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_\alpha Z(\mathbf{x}_\alpha)$
- **autorisation** : l'erreur d'estimation $Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)$ doit être une combinaison linéaire de poids total nul, ce qui conduit à :

$$\sum_{\alpha=1}^n \lambda_\alpha = 1$$

⁵ Plus généralement, on obtient des pondérateurs de krigeage égaux lorsque les distances qui séparent les données et le point à estimer, ainsi que les distances entre les données elles-mêmes, sont supérieures à la portée de la covariance (problème des "grandes mailles"). Ce résultat est mis en défaut si l'on peut trouver des données séparées par une distance inférieure à la portée ; dans ce cas, l'estimation coïncide avec celle de la moyenne inconnue et attribue des poids différents aux données, de manière à corriger leurs redondances d'information.

- **non-biais** : sous l'hypothèse intrinsèque, toute combinaison linéaire autorisée est d'espérance nulle, donc la contrainte de non biais est automatiquement vérifiée ;
- **optimalité** : le calcul de la variance d'estimation se fait comme dans le cas stationnaire, mais en remplaçant formellement la covariance C par l'opposé du variogramme $-\gamma$:

$$\text{var}[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = -\underbrace{\gamma(\mathbf{0})}_0 - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + 2 \sum_{\alpha=1}^n \lambda_{\alpha} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_0).$$

On cherche à minimiser cette expression sous la contrainte d'autorisation. Le calcul est semblable à celui effectué dans le cas stationnaire, en introduisant un multiplicateur de Lagrange μ . On aboutit au système :

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) - \mu = \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_0) & \forall \alpha = 1 \dots n \\ \sum_{\alpha=1}^n \lambda_{\alpha} = 1 \end{cases}$$

soit, matriciellement,

$$\begin{bmatrix} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_{\beta} \\ -\mu \end{bmatrix} = \begin{bmatrix} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_0) \\ 1 \end{bmatrix}$$

$$\begin{pmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_1) & \dots & \gamma(\mathbf{x}_1 - \mathbf{x}_n) & 1 \\ \vdots & & \vdots & \vdots \\ \gamma(\mathbf{x}_n - \mathbf{x}_1) & \dots & \gamma(\mathbf{x}_n - \mathbf{x}_n) & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ -\mu \end{pmatrix} = \begin{pmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_0) \\ \vdots \\ \gamma(\mathbf{x}_n - \mathbf{x}_0) \\ 1 \end{pmatrix}$$

La variance de krigeage vaut : $\sigma_{\text{KO}}^2(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_0) - \mu$.

Ces résultats sont les mêmes que ceux obtenus dans le cas stationnaire d'ordre deux (en utilisant le variogramme au lieu de la fonction de covariance). Cette identité explique que les deux types de krigeage ont reçu le même nom de **krigeage ordinaire**. Cependant, la contrainte sur la somme des pondérateurs n'a pas la même signification : contrainte de non-biais dans le cas stationnaire, contrainte d'autorisation dans le cas intrinsèque.

IV.4. Remarques sur le système de krigeage

- Le système et la variance de krigeage prennent en compte :
 - 1) les distances entre le point à estimer et les sites d'observation, via les termes $C(\mathbf{x}_\alpha - \mathbf{x}_0)$ ou $\gamma(\mathbf{x}_\alpha - \mathbf{x}_0)$;
 - 2) la configuration géométrique des sites d'observation, et la possible redondance des informations qu'ils contiennent, par l'intermédiaire des termes $C(\mathbf{x}_\alpha - \mathbf{x}_\beta)$ ou $\gamma(\mathbf{x}_\alpha - \mathbf{x}_\beta)$;
 - 3) la structure spatiale de la régionalisation, décrite par la fonction de covariance C ou le variogramme γ .
- En général, le poids attribué à un site d'observation est d'autant plus élevé que ce site est proche du point à estimer. Mais des phénomènes plus complexes peuvent perturber ce constat "intuitif" :
 - **existence d'un fort effet de pépité** dans le modèle variographique, qui a tendance à donner le même poids à tous les sites d'observation (comme il y a absence de structure spatiale, un site proche apporte autant d'information qu'un site éloigné). On voit donc qu'à la proximité géographique s'ajoute un deuxième facteur : la régularité de la variable régionalisée.
 - **présence d'une anisotropie** : il existe dans ce cas une ou plusieurs directions dans l'espace où le phénomène est moins structuré et la corrélation moins forte. Ainsi, une donnée proche dans cette direction peut apporter "moins d'information" qu'une donnée plus éloignée dans une autre direction, et avoir de ce fait un poids de krigeage plus faible.
 - **redondances entre données** : lorsque plusieurs données sont "groupées" et proches les unes des autres, elles deviennent redondantes ; leur poids cumulé sera quasiment le même que celui que recevrait une seule donnée au lieu du groupe. En d'autres termes, le krigeage corrige les effets dûs aux irrégularités d'échantillonnage et ne surpondère pas les données groupées au détriment des données isolées.
 - **effet d'écran** : il peut arriver qu'un site d'observation fasse écran à un autre site par rapport au site à estimer. Le site "écranté", même s'il est proche du site à estimer, peut alors avoir un poids faible, voire nul (écran total⁶). L'estimation ignore quasiment la valeur observée sur le site écranté. Lorsque cet effet est indésirable, on peut l'atténuer en introduisant dans le modèle une constante pépitique. Une autre solution consiste à regrouper les deux sites et à leur affecter le même poids, au prix d'une perte de précision.

⁶ Dans l'espace à une dimension, on peut citer comme exemples d'effet d'écran total le cas d'une covariance exponentielle (en krigeage simple) et celui d'un variogramme linéaire (en krigeage intrinsèque ordinaire) ; pour des variogrammes linéaires à l'origine, comme le schéma sphérique, on a un effet d'écran presque total, en krigeage ordinaire, si toutes les distances mises en jeu restent dans la partie linéaire du variogramme. A deux dimensions, le krigeage simple produit un effet d'écran entre directions perpendiculaires lorsque la covariance peut se factoriser, comme par exemple la covariance gaussienne : si on pose $\mathbf{h} = (h_1, h_2)$, alors $C(\mathbf{h}) = C(h_1) \times C(h_2)$.

- **effet d'écran inverse** : en krigeage ordinaire, il arrive que des sites éloignés aient des poids plus forts que des sites proches. L'explication de ce phénomène fait appel au théorème d'additivité qui sera examiné ultérieurement ; disons simplement que les sites éloignés sont valorisés car ils participent à l'estimation de la *moyenne* de la variable, laquelle intervient indirectement dans l'estimation de la valeur inconnue.
- **effet de relais** : la présence d'un point de donnée proche du point à estimer permet à un point de donnée "lointain", situé à une distance supérieure à la portée pratique du modèle variographique, d'avoir un poids non négligeable dans l'estimation, alors qu'en l'absence du point proche son poids aurait été nul (krigeage simple) ou faible (krigeage ordinaire).
- **Les pondérateurs et la variance d'estimation ne dépendent pas des valeurs prises par les données**, mais seulement de la configuration de krigeage et du modèle variographique. On peut donc, connaissant ce modèle, prévoir la précision de l'estimation à partir d'une configuration fixée des sites de données. Cette propriété est néanmoins une limitation de la géostatistique linéaire, une contrepartie de la simplicité du modèle ; intuitivement, la précision d'une estimation est moindre dans les zones plus "erratiques" (souvent, elles correspondent aux zones de fortes valeurs) que dans celles de faible variabilité.
- La matrice du premier membre du système de krigeage ne dépend que de la position relative des sites de données. Par suite, lorsque plusieurs estimations utilisent des configurations de sites identiques, il suffit d'inverser cette matrice une seule fois. C'est le cas lorsque l'on travaille en voisinage unique, ou en voisinage glissant avec des données réparties sur une grille régulière.
- Multiplier la fonction de covariance ou le variogramme par une constante $\kappa > 0$ ne change pas l'estimation : les poids obtenus seront les mêmes. Par contre, la variance d'estimation est modifiée et multipliée par κ .
- La plupart des modèles variographiques⁷ permettent de trouver des pondérateurs négatifs ou supérieurs à 1, même lorsqu'il existe une condition de somme unité pour les poids de krigeage (krigeage ordinaire). L'avantage d'une telle procédure est qu'elle peut fournir des estimations sortant des limites données par les valeurs observées, c'est-à-dire supérieures à la plus grande valeur mesurée ou inférieures à la plus petite. Au contraire, les techniques d'interpolation par combinaisons linéaires pondérées qui imposent des poids compris entre 0 et 1 donnent des estimations toujours comprises entre le minimum et le maximum des valeurs observées (figure 5.4). Or, dans la plupart des cas, il n'y a pas de raison que les valeurs mesurées atteignent les valeurs extrêmes potentielles de la zone, d'où l'intérêt d'avoir des poids sortant de l'intervalle $[0,1]$. En contrepartie, dans les cas où la variable est, par exemple, toujours positive, il y a un risque de trouver des estimations négatives. Ce risque est d'autant plus important que la variable présente une distribution fort dissymétrique : un poids négatif, même faible, affecté à une valeur forte peut conduire à une estimation négative si les autres valeurs des points de données ne sont pas trop élevées.

⁷ Une exception notable est constituée par le modèle pépétique pur, qui fournit des pondérateurs nuls (krigeage simple) ou égaux à $1/n$ (krigeage ordinaire).

La figure 5.4 illustre cette propriété sur l'exemple unidimensionnel des huit sites, en comparant l'estimateur du krigeage ordinaire avec celui de l'inverse des distances. Le premier utilise un modèle de variogramme parabolique à l'origine (cubique), propice à l'obtention de pondérateurs négatifs. Quant au second, il attribue à chaque donnée un poids toujours positif, inversement proportionnel à sa distance au site à estimer, et ne permet pas d'obtenir une valeur estimée supérieure à la plus haute donnée.

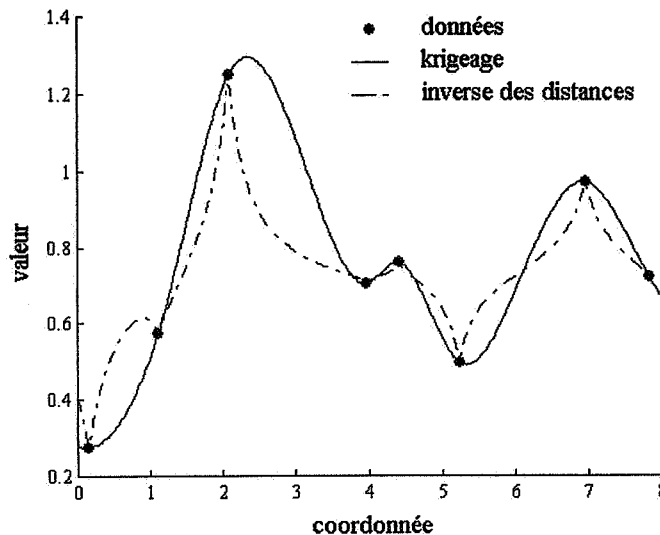


figure 5.4 : comparaison des estimations obtenues par krigeage ordinaire et par la méthode de l'inverse des distances

IV.5. Propriétés du krigeage

- **régularité du système**

On démontre que si les sites d'observation sont tous distincts, le système de krigeage est régulier, c'est-à-dire qu'il fournit une solution unique (sauf cas "pathologique"). Néanmoins, il peut arriver que des sites d'observation très proches entraînent une instabilité dans l'inversion de la matrice du premier membre du système, laquelle présente alors deux lignes quasiment égales. Ce problème se pose souvent pour des covariances ou des variogrammes très réguliers (schéma gaussien par exemple). Une solution acceptable consiste à ajouter au modèle un léger effet de pépité ; une autre solution consiste à regrouper, au moment de l'estimation, les sites très proches en un site "moyen".

- **interpolation exacte**

Le krigeage ponctuel est un interpolateur exact, c'est-à-dire que l'estimation en un site de mesure redonne la valeur mesurée et la variance de krigeage en ce point est nulle. Il suffit de vérifier que, si le point à estimer x_0 coïncide avec un point de donnée, l'estimateur défini par $Z^*(x_0) = Z(x_0)$ vérifie les contraintes du krigeage ; comme l'erreur d'estimation associée est nulle, elle est de toute évidence sans biais et de variance minimale (nulle).

- **propriété d'additivité**

Supposons que l'on désire estimer par krigeage la valeur moyenne d'un bloc à partir d'observations ponctuelles. Une possibilité est de kriger la valeur de chaque point du bloc, puis de faire la moyenne de toutes les estimations ponctuelles. Toutefois, pour que la moyenne de ces krigeages coïncide avec le krigeage de la moyenne du bloc (que l'on peut effectuer directement), il est indispensable de prendre les mêmes données pour tous les krigeages ponctuels. Cette propriété résulte de la linéarité des équations du krigeage et du fait qu'en utilisant les mêmes données, la matrice dans le premier membre du système est identique pour tous les krigeages ponctuels.

De manière générale, la propriété d'additivité du krigeage consiste à dire que, *sous réserve d'utiliser les mêmes données* pour toutes les estimations concernées, le krigeage de toute fonction *linéaire* $\mathcal{L}$ de la variable Z est égal à la fonction linéaire appliquée au krigeage de Z , soit :

$$\mathcal{L}^*(Z) = \mathcal{L}(Z^*).$$

Ainsi, tout krigeage d'une grandeur qui est une fonction linéaire de la variable régionalisée étudiée (moyenne sur un bloc, moyenne mobile, gradient...) peut se ramener à des estimations ponctuelles, d'où l'intérêt accordé au krigeage ponctuel. En revanche, il n'est pas possible de déduire la variance d'estimation de $\mathcal{L}(Z)$ à partir des variances d'estimations ponctuelles.

- **propriété d'orthogonalité**

En ce qui concerne le **krigeage simple**, les équations de krigeage s'écrivent encore :

$$0 = \sum_{\beta=1}^n \lambda_{\beta} \text{cov}[Z(\mathbf{x}_{\beta}), Z(\mathbf{x}_{\alpha})] - \text{cov}[Z(\mathbf{x}_0), Z(\mathbf{x}_{\alpha})] = \text{cov}\left[\sum_{\beta=1}^n \lambda_{\beta} Z(\mathbf{x}_{\beta}) - Z(\mathbf{x}_0), Z(\mathbf{x}_{\alpha})\right] \quad \forall \alpha = 1 \dots n$$

soit :

$$\text{cov}[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0), Z(\mathbf{x}_{\alpha})] = 0 \quad \forall \alpha = 1 \dots n.$$

L'erreur d'estimation n'est corrélée à aucune des données, ni, par conséquent, aux combinaisons linéaires des données. La non-corrélation est aussi appelée *orthogonalité*.

De même, le système de **krigeage ordinaire** peut s'écrire :

$$\text{cov}[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0), Z(\mathbf{x}_{\alpha})] = -\mu \quad \forall \alpha = 1 \dots n.$$

Considérons une combinaison linéaire des données $\sum_{\alpha=1}^n \omega_{\alpha} Z(\mathbf{x}_{\alpha})$ de poids total nul. On a :

$$\text{cov}\left[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0), \sum_{\alpha=1}^n \omega_{\alpha} Z(\mathbf{x}_{\alpha})\right] = -\mu \underbrace{\sum_{\alpha=1}^n \omega_{\alpha}}_{=0} = 0.$$

En résumé, l'erreur de krigeage ordinaire ponctuel est non corrélée aux combinaisons linéaires des données de poids total nul. L'erreur de krigeage simple est orthogonale à toute combinaison linéaire des données.

• **propriété de lissage**

La carte des valeurs estimées par krigeage est toujours plus lisse que celle des valeurs ponctuelles réelles, c'est-à-dire qu'elle présente moins de fluctuations. La recherche d'une estimation précise s'accompagne inévitablement de cet effet de lissage, car l'estimation ne peut pas "inventer" les détails qui n'apparaissent pas dans les observations. Si l'on veut reproduire la variabilité de la régionalisation, il sera nécessaire d'avoir recours aux techniques de simulation.

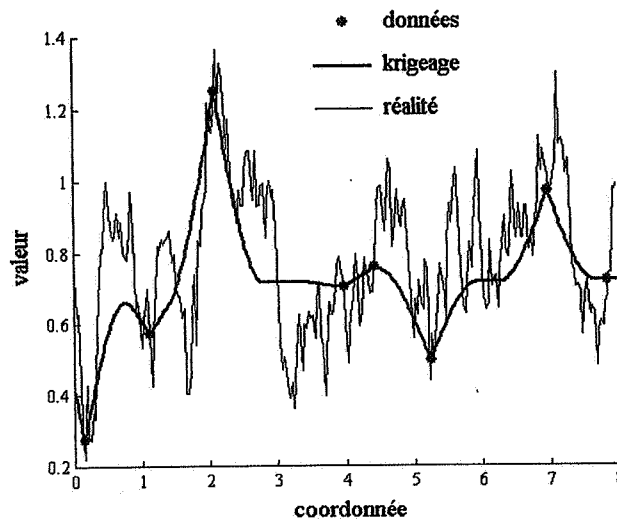


figure 5.5 : illustration de la propriété de lissage

On peut donner une traduction mathématique à la propriété de lissage dans le cas du krigeage simple (à moyenne connue). En effet, dans ce cas, l'erreur d'estimation est orthogonale à toute combinaison linéaire des données, donc en particulier à l'estimateur lui-même :

$$\text{cov}[Z^*(x_0) - Z(x_0), Z^*(x_0)] = 0$$

ou encore $\text{cov}[Z^*(x_0), Z(x_0)] = \text{var}[Z^*(x_0)]$.

$$\begin{aligned} \text{Par suite, } \sigma_{\text{KS}}^2(x_0) &= \text{var}[Z^*(x_0) - Z(x_0)] \\ &= \text{var}[Z^*(x_0)] + \text{var}[Z(x_0)] - 2\text{cov}[Z^*(x_0), Z(x_0)] \\ &= \text{var}[Z(x_0)] - \text{var}[Z^*(x_0)] \end{aligned}$$

soit encore $C(0) = \sigma^2 = \sigma_{\text{KS}}^2(x_0) + \text{var}[Z^*(x_0)]$.

En particulier, $\text{var}[Z^*(x_0)] \leq C(0)$: la variance de la valeur estimée est inférieure à la variance *a priori*. On notera au passage que l'estimateur de krigeage n'est pas stationnaire d'ordre deux, puisque sa variance dépend du point x_0 .

• **relation entre krigeage simple et krigeage ordinaire : théorème d'additivité**

Considérons une fonction aléatoire Z stationnaire d'ordre deux, d'espérance m , et Y son "résidu" : $Y(x) = Z(x) - m$ (par construction, Y est d'espérance nulle). On établit (*théorème d'additivité*) :

$$Z_{KO}^*(x_0) = Y_{KS}^*(x_0) + M_{KO}^* \left[1 - \sum_{\alpha=1}^n \lambda_{\alpha}^{KS}(x_0) \right]$$

où les indices KS et KO se réfèrent respectivement au *krigeage simple* et au *krigeage ordinaire*.

Cette relation montre que le krigeage ordinaire s'identifie au krigeage simple, à condition de remplacer la valeur inconnue de l'espérance m par son estimation optimale m^* obtenue par krigeage ordinaire. Qu'elle soit connue ou estimée, la moyenne est affectée du poids

$$\lambda_m(x_0) = 1 - \sum_{\alpha=1}^n \lambda_{\alpha}^{KS}(x_0).$$

En général, ce poids augmente lorsque l'information disponible dans le voisinage du point à estimer devient rare ; $\lambda_m(x_0)$ constitue donc un indicateur de la quantité d'information autour de x_0 . Par ailleurs, on voit qu'une condition nécessaire et suffisante pour qu'il y ait identité en un point donné entre krigeage simple et krigeage ordinaire est que la somme des poids de krigeage simple est égale à 1, c'est-à-dire que le poids de la moyenne est nul.

On a également un théorème d'additivité en ce qui concerne les variances d'estimation :

$$\sigma_{KO}^2(x_0) = \sigma_{KS}^2(x_0) + \sigma_m^2 \left[1 - \sum_{\alpha=1}^n \lambda_{\alpha}^{KS}(x_0) \right]^2 = \sigma_{KS}^2(x_0) + \sigma_m^2 \lambda_m^2(x_0)$$

qui montre que la variance de krigeage ordinaire est toujours supérieure à celle de krigeage simple. La différence mesure la perte de précision que l'on subit faute de connaître la vraie valeur de m .

Cette seconde partie du théorème d'additivité peut d'ailleurs aider à **choisir la taille du voisinage en krigeage ordinaire** (à moyenne inconnue). Le krigeage simple détermine les échantillons qui ne peuvent être omis dans l'estimation, à savoir ceux affectés d'un poids non négligeable ; ce sont en général ceux situés à une distance inférieure à la portée pratique du modèle de covariance. Quand la somme des poids de krigeage simple est proche de 1, i.e. $\lambda_m(x_0)$ proche de 0, le krigeage ordinaire est peu différent du krigeage simple, ce qui indique que l'on perd peu de précision en estimant la moyenne inconnue. Le voisinage de krigeage simple est alors suffisant pour réaliser un krigeage ordinaire.

Dans le cas contraire, il est nécessaire d'augmenter la taille du voisinage de krigeage pour mieux estimer la moyenne et accroître la précision de l'estimation. Ceci est également utile pour réduire ce que l'on appelle le "**biais conditionnel**" : même si le krigeage est *globalement* sans biais, l'espérance de la valeur vraie *conditionnellement à la valeur estimée* n'est pas nécessairement égale à la valeur estimée, soit en termes mathématiques :

$$E[Z(x_0) | Z^*(x_0)] \neq Z^*(x_0).$$

Ce phénomène est connu en géostatistique minière, car il dégrade les résultats d'exploitation : la teneur moyenne *au-dessus d'un certain seuil* n'est pas correctement évaluée si l'on se base sur les valeurs estimées (cet effet est encore plus marqué si l'on utilise au lieu du krigeage un autre estimateur linéaire, comme celui des polygones d'influence). Pour réduire au maximum le biais conditionnel, on doit choisir un voisinage tel que la *pente* de la régression entre valeurs vraies et estimées soit la plus proche possible de l'unité. Cette pente est calculable à l'aide du modèle variographique ; elle est égale à 1 en krigeage simple, tandis qu'en krigeage ordinaire elle est d'autant plus éloignée de 1 que le poids de la moyenne $\lambda_m(x_0)$ est élevé et que la moyenne est mal estimée (variance d'estimation de la moyenne σ_m^2 forte), d'où l'utilité d'augmenter la taille du voisinage pour diminuer σ_m^2 . En pratique, le **biais conditionnel** est faible lorsque le **voisinage de krigeage est suffisamment grand** (pour fixer les idées, lorsqu'il contient plus de 15 données).

IV.6. Application aux données agronomiques

On reprend l'exemple des données de canne à sucre présenté aux chapitres 1, 3 et 4. On s'intéresse à présent à des estimations locales de la superficie de canne à sucre, à partir d'un échantillonnage régulier au taux de 1% (soit 1 cellule sur 10 selon les directions nord-sud et est-ouest).

IV.6.1. Choix de la méthode d'estimation

La structure spatiale de la régionalisation a été modélisée par un variogramme isotrope admettant un palier⁸. Pour l'estimation locale, on a le choix entre un krigeage *simple* et un krigeage *ordinaire*. Toutefois, l'estimateur de krigeage simple s'appuie sur la connaissance de la moyenne de la variable, dont le poids dans l'estimation est d'autant plus important que celui des données est faible : loin des données, le krigeage simple calibre l'estimation autour de la moyenne. Or, on sait que l'hypothèse de stationnarité est contestable à l'échelle du champ, mais qu'une hypothèse de quasi-stationnarité est admissible. Par conséquent, définir une espérance constante sur le champ, sur laquelle s'appuierait l'estimation locale, n'a guère de signification. Ainsi, dans cet exemple, le krigeage simple n'est pas adapté. On préfère réaliser un krigeage ordinaire : le fait de considérer la moyenne inconnue autorise cette moyenne à varier d'un voisinage de krigeage à l'autre, tout en restant approximativement constante à l'échelle d'un voisinage. Le krigeage ordinaire permet d'affaiblir l'hypothèse de stationnarité par rapport au krigeage simple.

La variable étudiée étant isotrope, aucune direction ne doit être privilégiée dans la recherche des données à utiliser pour les estimations. On choisit donc un **voisinage de krigeage** de forme circulaire. On limite le rayon de ce voisinage, de façon à ce que chaque estimation utilise au plus 20 données.

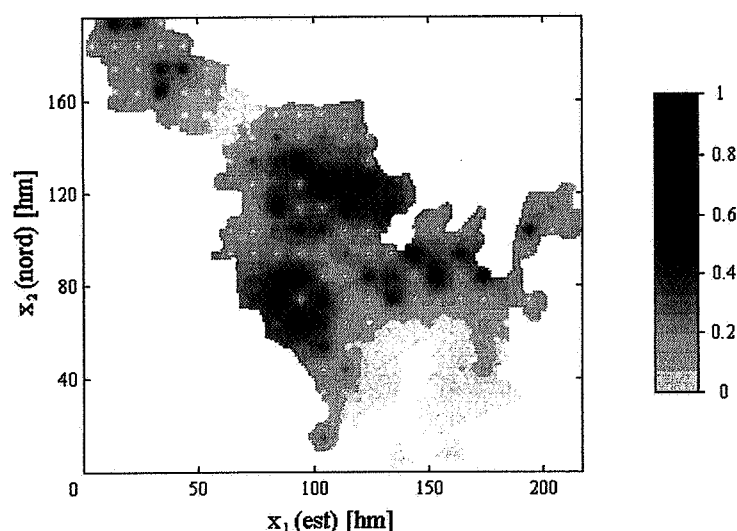


figure 5.6 : carte des estimations par krigeage ordinaire

⁸ Dans la suite, nous utiliserons le modèle établi à partir du variogramme *régional*, mais les résultats seraient similaires si l'on prenait l'un des modèles du variogramme *expérimental*, cf. chapitre 3.

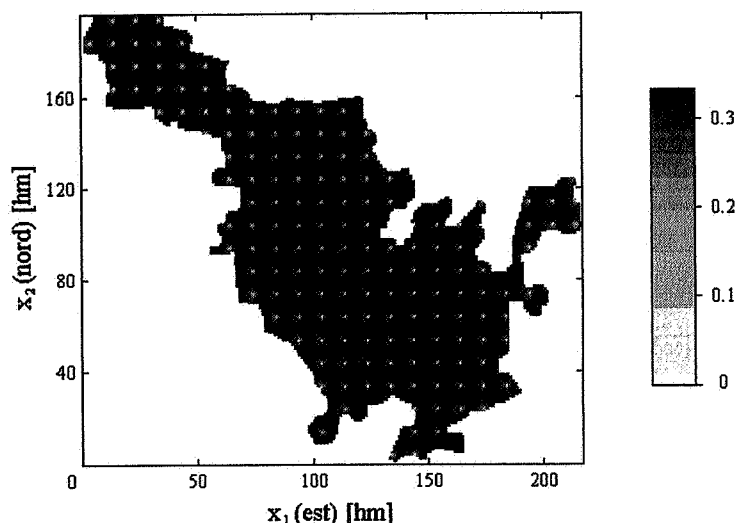


figure 5.7 : carte des écarts-types de krigeage
(racines carrées des variances de krigeage)

Les cellules échantillonnées apparaissent nettement sur les deux cartes précédentes. L'explication vient du fait que, dans le modèle de variogramme, 40% de la variabilité correspond à une structure de très courte portée (3 hm) à l'échelle considérée, qui se comporte pratiquement comme un effet de pépite, c'est-à-dire une absence de structuration spatiale. Cette microstructure produit des discontinuités (en réalité, des transitions très rapides) dans la carte de krigeage. Elle est également responsable d'une forte incertitude dans les estimations dès que l'on n'est plus en un site de données. Ainsi, les écarts-types d'estimation, nuls sur les cellules échantillonnées, valent 0.24 ha pour les cellules attenantes, ce qui est déjà considérable.

IV.6.2. Comparaison entre les valeurs vraies et estimées

Statistiques élémentaires sur les 16 050 cellules de la zone étudiée

	minimum	maximum	moyenne	médiane	variance	écart-type
valeurs vraies	0	1	0.2125	0	0.1044	0.3231
valeurs estimées	0	1	0.2127	0.1996	0.0203	0.1425

tableau 5.1 : statistiques monovariabiles des valeurs vraies et estimées par krigeage

Ici, les valeurs estimées se situent toutes dans l'intervalle $[0,1]$, mais cette propriété n'est pas garantie par le krigeage : même si toutes les données sont dans $[0,1]$, il peut arriver que des estimations sortent de cet intervalle.

Les valeurs moyennes sont remarquablement voisines, ce qui illustre la propriété de non-biais du krigeage. En revanche, il n'en est pas de même de la médiane, ni d'ailleurs des autres quantiles. La distribution des valeurs estimées est très différente de celle des valeurs vraies (figure 5.8).

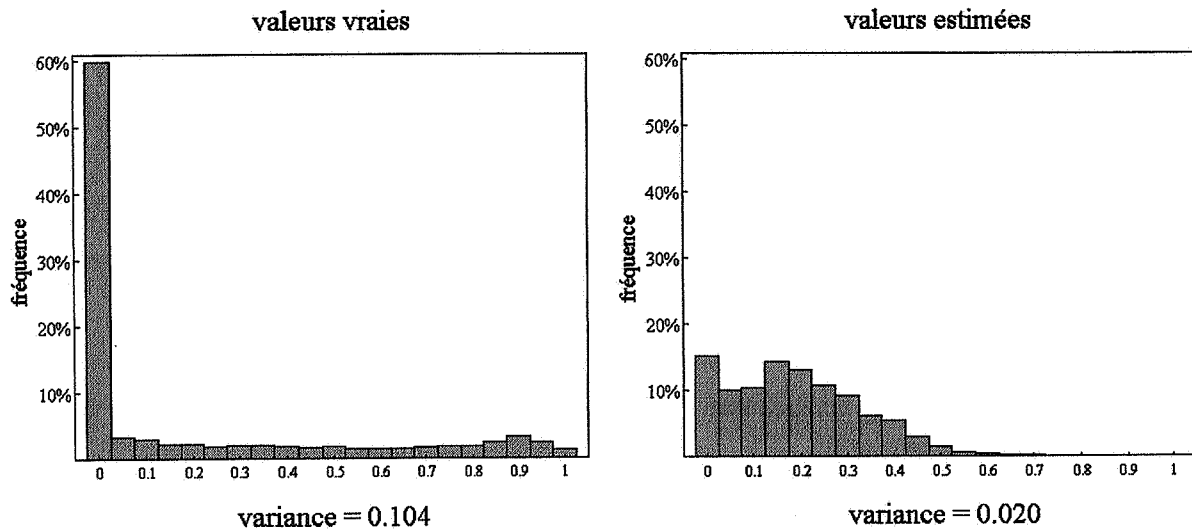


figure 5.8 : histogrammes des valeurs réelles (à gauche) et estimées (à droite)

L'histogramme des valeurs krigées est plus symétrique et plus resserré (variance cinq fois plus faible) que celui des valeurs vraies, ce qui est symptomatique de la **propriété de lissage** du krigeage : les estimations présentent moins de valeurs extrêmes, proches de 0 ou 1, et davantage de valeurs intermédiaires, proches de la moyenne.

Lissage et biais conditionnel

A cause de la propriété de lissage, le krigeage a tendance à surestimer les valeurs faibles et sous-estimer les valeurs fortes. Ceci se traduit par une régression

$$E[Z^*(\mathbf{x})|Z(\mathbf{x})]$$

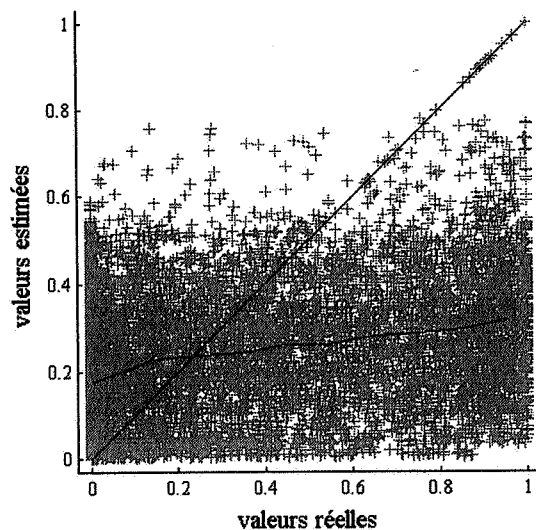
de pente inférieure à l'unité.

En ce qui concerne le biais conditionnel, c'est la régression inverse qui nous intéresse :

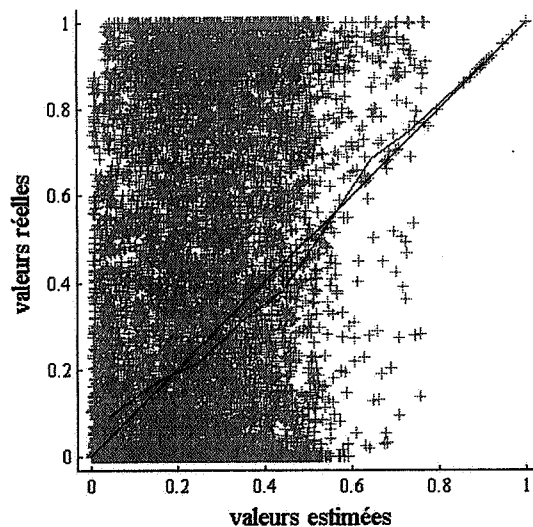
$$E[Z(\mathbf{x})|Z^*(\mathbf{x})].$$

On souhaite cette régression la plus proche possible de $Z^*(\mathbf{x})$, afin de minimiser le biais conditionnel.

Les figures 5.9 et 5.10 donnent les nuages de corrélation entre valeurs estimées et valeurs vraies, ainsi que les lignes de régression correspondantes. Elles montrent que si la propriété de lissage est très marquée, en revanche le biais conditionnel est faible.



*figure 5.9 : nuage de corrélation entre valeurs réelles et estimées
la courbe de régression (ligne brisée) a une pente plus faible
que la diagonale, ce qui illustre la propriété de lissage*



*figure 5.10 : nuage de corrélation entre valeurs estimées et réelles
la courbe de régression (ligne brisée) se confond presque avec
la diagonale, ce qui illustre la quasi-absence de biais conditionnel*

Comparaison avec un estimateur classique

Nous allons comparer sur cet exemple, les performances du krigeage avec celles de l'interpolation par plus proche voisin. Cette dernière attribue au site à estimer la valeur du point de données le plus proche. Comme le krigeage, il s'agit d'une combinaison linéaire, qui affecte un poids égal à 1 à la donnée la plus proche du site à estimer et un poids nul aux autres données. Contrairement au krigeage, l'interpolation par plus proche voisin ne conduit pas à un lissage. En effet, les points de données étant ici régulièrement répartis, l'histogramme des valeurs estimées est pratiquement identique à celui des échantillons, et aussi à celui des valeurs vraies.

	minimum	maximum	moyenne	médiane	variance	écart-type
valeurs vraies	0	1	0.2125	0	0.1044	0.3231
valeurs estimées	0	1	0.2111	0	0.0981	0.3132

tableau 5.2 : statistiques monovariabiles des valeurs vraies et estimées par plus proche voisin

Un premier reproche que l'on peut adresser à l'interpolation par plus proche voisin est qu'elle est purement géométrique, et que le résultat dépend seulement de l'implantation des échantillons. La carte des estimations présente en outre des discontinuités lorsque l'on passe de la zone d'influence d'un point de données à celle d'un autre point. Ces discontinuités sont totalement artificielles et n'ont aucun rapport avec le phénomène régionalisé étudié.

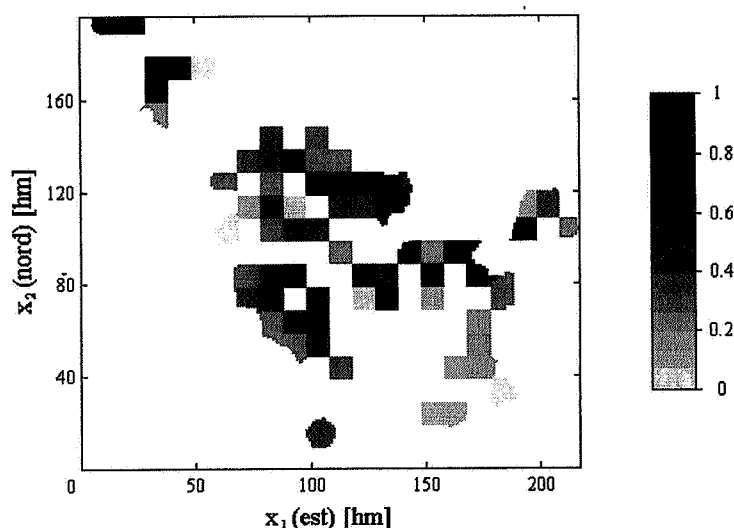


figure 5.11 : carte des valeurs estimées par plus proche voisin

Néanmoins, si l'on s'en tient à la recherche de la précision de l'estimation, l'interpolation par plus proche voisin donne des résultats certes moins bons que ceux du krigeage, mais du même ordre de grandeur. Ainsi, l'écart-type statistique des erreurs de krigeage sur l'ensemble du champ est égal à 0.306 ha, tandis que celui de l'interpolation par plus proche voisin est de 0.377 ha. Les coefficients de corrélation entre valeurs vraies et estimées sont respectivement de 0.339 (krigeage) et 0.299 (plus proche voisin). Ces "mauvais" résultats du krigeage sont dus à la faible structuration spatiale des valeurs.

Dans ce contexte, on pourrait penser ne pas s'embarrasser d'un estimateur sophistiqué comme le krigeage, et se contenter de l'interpolation par plus proche voisin, bien plus rapide à mettre en œuvre. Mais cette dernière possède un autre désavantage : elle conduit à un biais conditionnel bien plus important que le krigeage. Ainsi, elle a pour effet de sous-estimer les valeurs des sites présumés "pauvres" (ceux dont la valeur *estimée* est faible), et surestimer celles des sites présumés "riches" (ceux dont la valeur *estimée* est élevée), bien qu'en moyenne ces "biais" se compensent. En conclusion, on retiendra que la précision n'est pas le seul critère de qualité d'un estimateur, et que le krigeage évite d'avoir un grand biais conditionnel (pourvu que le voisinage d'estimation soit suffisamment grand), contrairement à d'autres estimateurs.

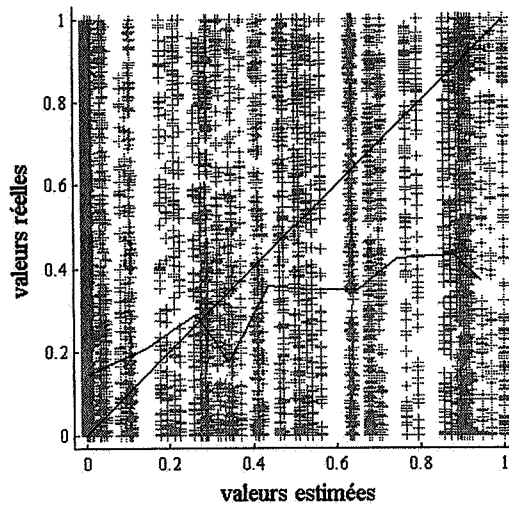


figure 5.12 : nuage de corrélation entre valeurs estimées par le plus proche voisin et valeurs réelles ; la courbe de régression (ligne brisée) diffère notablement de la diagonale, ce qui illustre la présence d'un fort biais conditionnel

Variances des erreurs réelles en fonction des variances théoriques

Stricto sensu, la variance de krigeage en un point quelconque correspond à la variance des erreurs d'estimation commises sur un grand nombre de réalisations de la fonction aléatoire. Une telle interprétation est purement théorique ; pour lui donner une signification *objective*, on remplace la répétition sur les réalisations, par une répétition dans l'espace : la variance d'estimation est celle des erreurs commises sur de nombreuses configurations de krigeage identiques réparties de manière homogène sur tout le champ.

Dans l'exemple qui nous occupe, il est possible de vérifier cette assertion. Le réseau d'échantillonnage étant régulier, il n'y a qu'un nombre limité de configurations de krigeage, et la carte des variances (ou celle des écarts-types) de krigeage dessine toujours le même motif, sauf en bordure de zone. A une même valeur de la variance d'estimation correspond donc un grand nombre de cellules, sur lesquelles il est possible de calculer les erreurs *réelles*, et leur variance. Le graphique 5.13 visualise les variances des erreurs réelles en fonction de celles des erreurs théoriques, et montre leur bonne adéquation. L'interprétation en termes objectifs de la variance d'estimation locale est donc bien fondée.

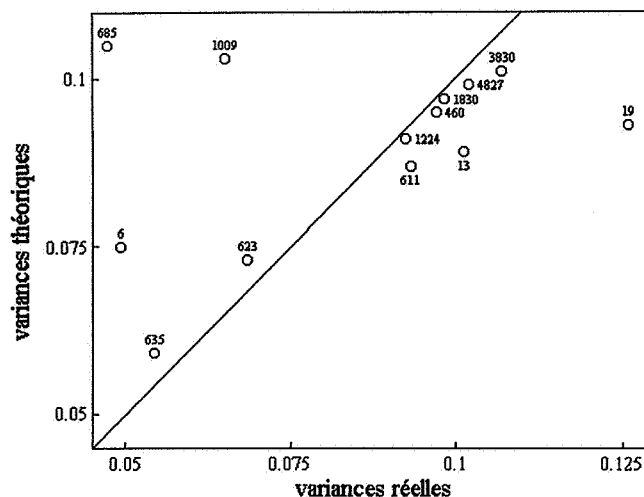
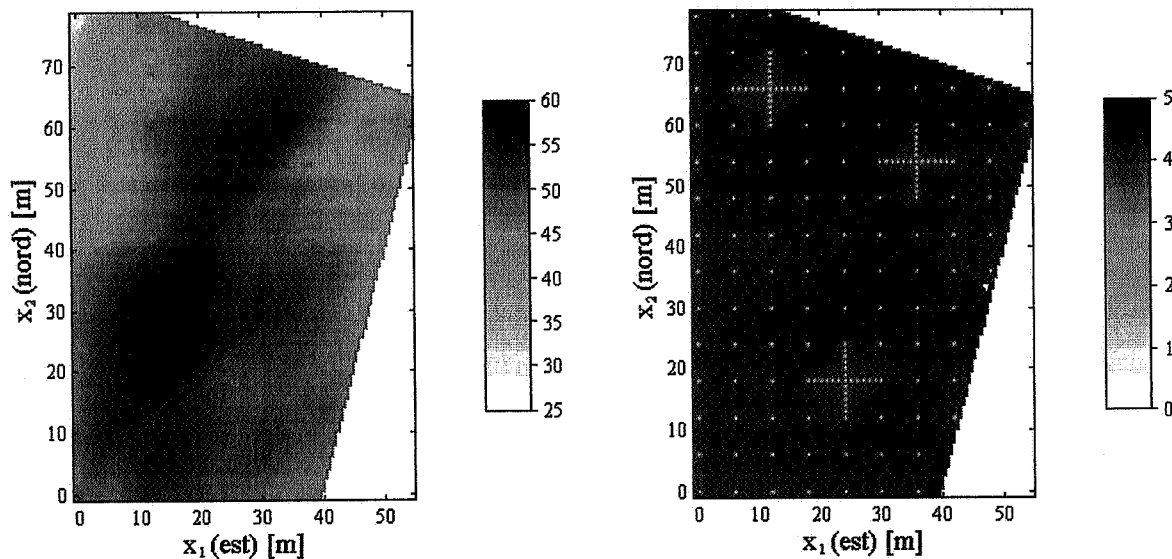


figure 5.13 : variances des erreurs réelles en fonction des variances théoriques la diagonale et les nombres de cellules concernées sont indiqués

IV.7. Application aux données de sol

Nous allons à présent reprendre l'exemple pédagogique des données de sol et utiliser le modèle de variogramme défini au chapitre 3 pour estimer, par krigeage intrinsèque ordinaire, le taux d'argile et limon aux nœuds d'une grille régulière de maille 0.5 m × 0.5 m. Le voisinage de krigeage choisi est un voisinage glissant, de forme elliptique, tel que les axes de l'ellipse coïncident avec les axes principaux de l'anisotropie introduite dans le modèle variographique. Chaque estimation utilise au plus 24 points de données (3 par octant d'ellipse). On visualise les cartes des valeurs estimées et des écarts-types d'estimation en niveaux de gris (figure 5.14). Le champ étudié est une parcelle d'allure trapézoïdale.



*figure 5.14 : cartes du krigeage (à gauche) et de l'écart-type de krigeage (à droite)
estimations ponctuelles aux nœuds d'une grille de maille 0.5 m × 0.5 m*

La carte des estimations met en évidence une dérive dans la direction ouest-nord-ouest / est-sud-est. L'effet de pépité introduit dans le modèle se manifeste par la présence de discontinuités au niveau des points de données, qui se détachent des autres nœuds de la grille d'estimation. Quant aux écarts-types d'estimation, ils sont nuls aux points de données, et d'autant plus élevés que l'on s'approche des bords du champ, où l'on travaille de plus en plus en extrapolation.

Il est intéressant de comparer la carte du krigeage avec celle que l'on obtiendrait par une méthode d'interpolation déterministe. Une méthode couramment employée est celle de l'**inverse des distances** ; elle consiste à pondérer chaque donnée par une puissance de l'inverse de la distance entre la donnée et le point à estimer :

$$\hat{z}(\mathbf{x}) = \frac{\sum_{\alpha=1}^n \frac{z(\mathbf{x}_{\alpha})}{|\mathbf{x} - \mathbf{x}_{\alpha}|^p}}{\sum_{\alpha=1}^n \frac{1}{|\mathbf{x} - \mathbf{x}_{\alpha}|^p}} \quad \text{avec } p > 0.$$

La figure 5.15 présente la carte des estimations obtenues par la méthode de l'inverse des distances, pour les puissances $p=1$, $p=2$ et $p=3$, en utilisant le même voisinage que dans le cas du krigeage.

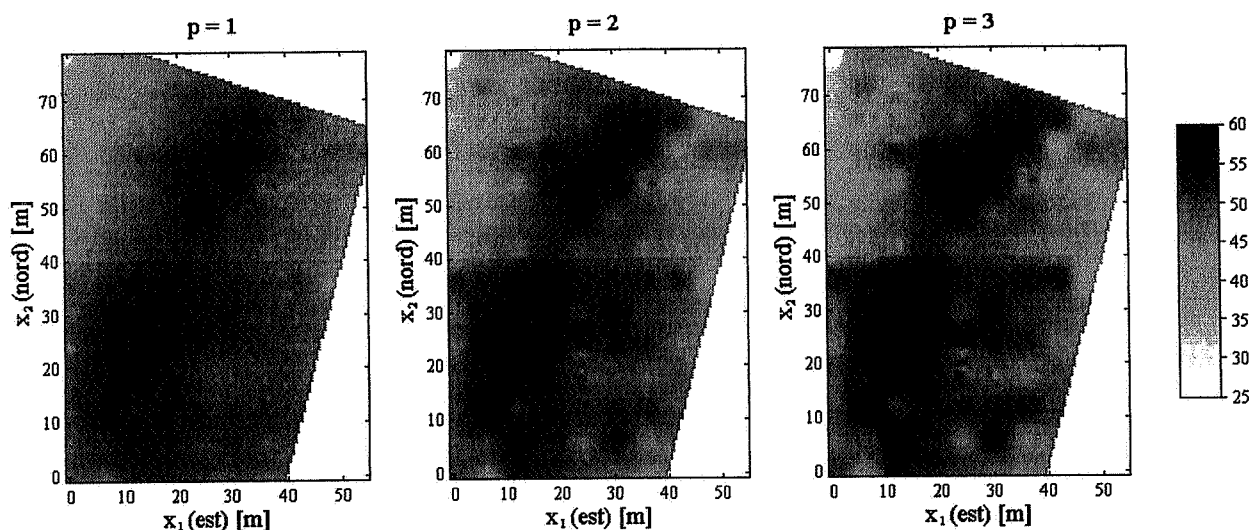


figure 5.15 : cartes des estimations par la méthode de l'inverse des distances
estimations ponctuelles aux nœuds d'une grille de maille $0.5 \text{ m} \times 0.5 \text{ m}$

Comme le krigeage, la méthode de l'inverse des distances interpole exactement les données. Mais elle conduit à des artefacts dans les cartes d'estimation. Notamment, le gradient de $\hat{z}$ est nul aux points de données, de sorte que les lignes d'isovaleurs sont en forme "d'œil de bœuf" autour de ces points ; cet effet est particulièrement visible sur les cartes des estimations correspondant à $p=2$ et $p=3$. Certains algorithmes diminuent ces artefacts à l'aide de coefficients de lissage, mais en contrepartie l'interpolation n'est plus exacte. Le choix de p peut se faire en utilisant la technique de validation croisée : on ôte tour à tour chaque point de données, et on estime sa valeur à l'aide des points restants ; on peut alors calculer en chaque point de données l'erreur commise. On itère la procédure pour plusieurs valeurs de p et on choisit celle qui fournit les meilleures estimations des données (variance des erreurs minimale). Le tableau suivant donne les résultats obtenus pour les valeurs $p=1$, $p=2$ et $p=3$, et pour le krigeage :

	moyenne des erreurs	variance des erreurs
$p=1$	0.039	19.30
$p=2$	0.063	19.35
$p=3$	0.063	20.13
krigeage	0.069	17.23

tableau 5.3 : résultats de validation croisée – inverse des distances et krigeage

Parmi les trois valeurs de p testées, la puissance $p=1$ donne la plus faible variance d'erreurs. Mais on note que le krigeage intrinsèque ordinaire donne des résultats encore meilleurs.

La principale critique que l'on peut adresser à la méthode de l'inverse des distances, ainsi qu'aux autres méthodes d'interpolation déterministes (splines, interpolation polynomiale...), est qu'elle ne tient pas compte de la structure spatiale de la variable régionalisée étudiée : effet de pépite, anisotropie, portée des observations, etc. Il s'agit d'une procédure automatique qui s'applique "aveuglément" à n'importe quel jeu de données et dont le but est de donner une carte plus esthétique que précise. Au contraire, le krigeage fait précéder l'interpolation par une analyse structurale, qui permet d'élaborer un modèle cohérent avec ce qui a été observé, de construire une estimation non biaisée de la quantité recherchée, et de mesurer la précision de cette estimation.

Il est également instructif de montrer sur cet exemple les différences entre le krigeage et la réalité. La figure 5.16 donne les histogrammes et les variogrammes expérimentaux, le long des directions principales d'anisotropie, des valeurs aux points de données puis des valeurs estimées par krigeage intrinsèque ordinaire.

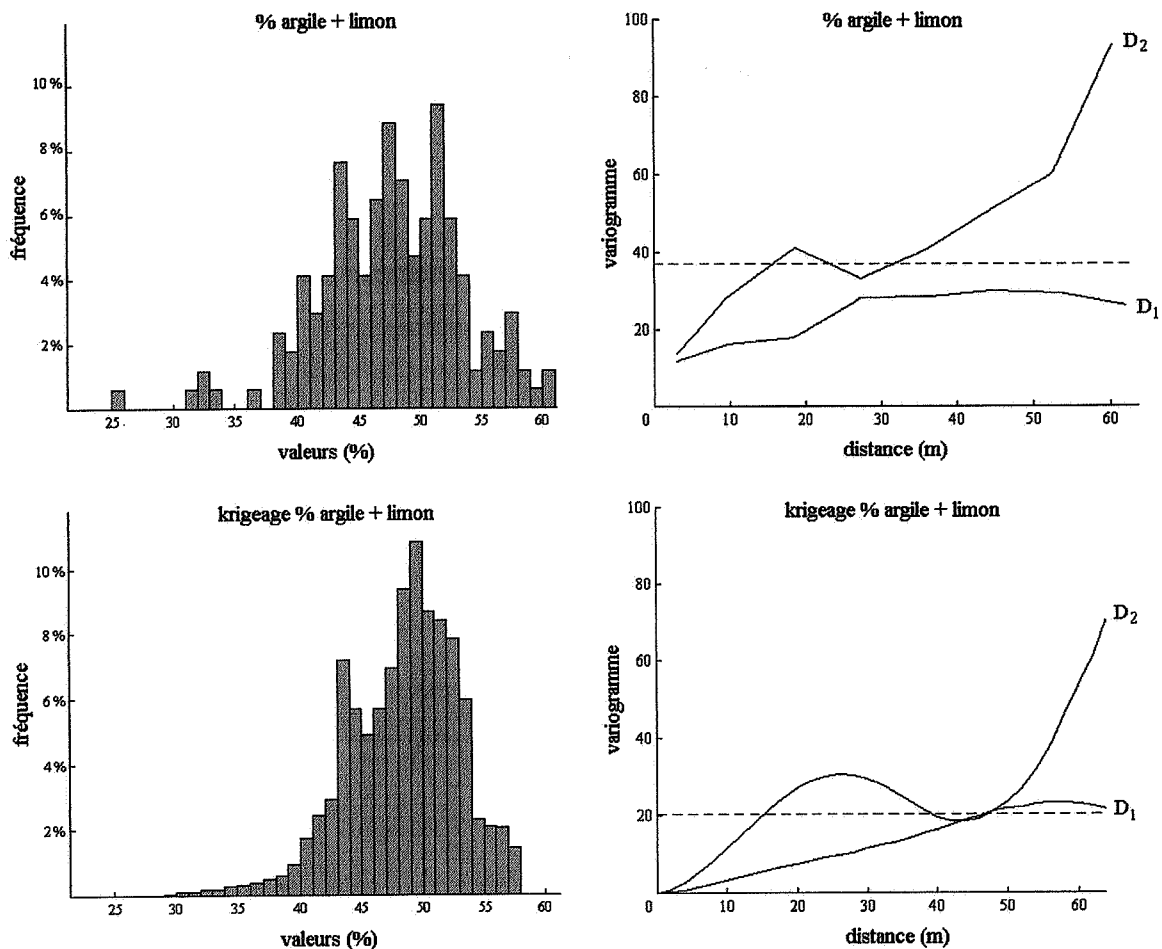


figure 5.16 : histogrammes et variogrammes des valeurs expérimentales (en haut) et des valeurs estimées par krigeage intrinsèque ordinaire (en bas)

L'histogramme des valeurs krigées est plus resserré que celui des données, avec moins de valeurs extrêmes et davantage de valeurs intermédiaires. La moyenne est à peu près la même (environ 48%), mais la variance est nettement moindre (20.2 pour les valeurs krigées contre 37.9 pour les valeurs expérimentales). En ce qui concerne les variogrammes expérimentaux, on retrouve dans les deux cas la non-stationnarité (variogrammes non bornés) et l'anisotropie, mais le variogramme des estimations n'a pas d'effet de pépité, présente un effet de trou le long de la direction D_2 , et ses valeurs sont plus faibles que celles du variogramme des données, ce qui traduit une variabilité moindre.

Ces constatations illustrent la propriété de lissage du krigeage et montrent en particulier que **le krigeage ne reproduit pas la structure de la variable initiale**. Le but du krigeage est la recherche de la précision dans l'estimation. Pour obtenir une carte qui respecte les traits structuraux de la variable régionalisée, il faut avoir recours aux techniques de simulations conditionnelles.

IV.8. Application aux données minières

Terminons ces études de cas par une application aux données minières introduites aux chapitres 2 et 3. D'après l'analyse variographique, la variable régionalisée étudiée (teneur en cuivre) présente une anisotropie zonale orientée le long de la direction verticale : la variabilité est moindre dans cette direction que dans le plan horizontal. En conséquence, pour mener l'estimation locale, on décide d'utiliser un voisinage de krigeage en forme d'ellipsoïde, dont le plus grand axe est vertical ; afin d'améliorer la recherche des données à utiliser, ce voisinage est divisé en octants, dans chacun desquels on recherche les trois données les plus proches.

Krigeage ordinaire ponctuel

La teneur en cuivre est d'abord estimée par krigeage ordinaire ponctuel, aux nœuds d'une grille de maille $10\text{m} \times 10\text{m} \times 12\text{m}$ couvrant la zone échantillonnée (figure 5.17) ; l'estimation de valeurs non ponctuelles est réservée à plus tard.

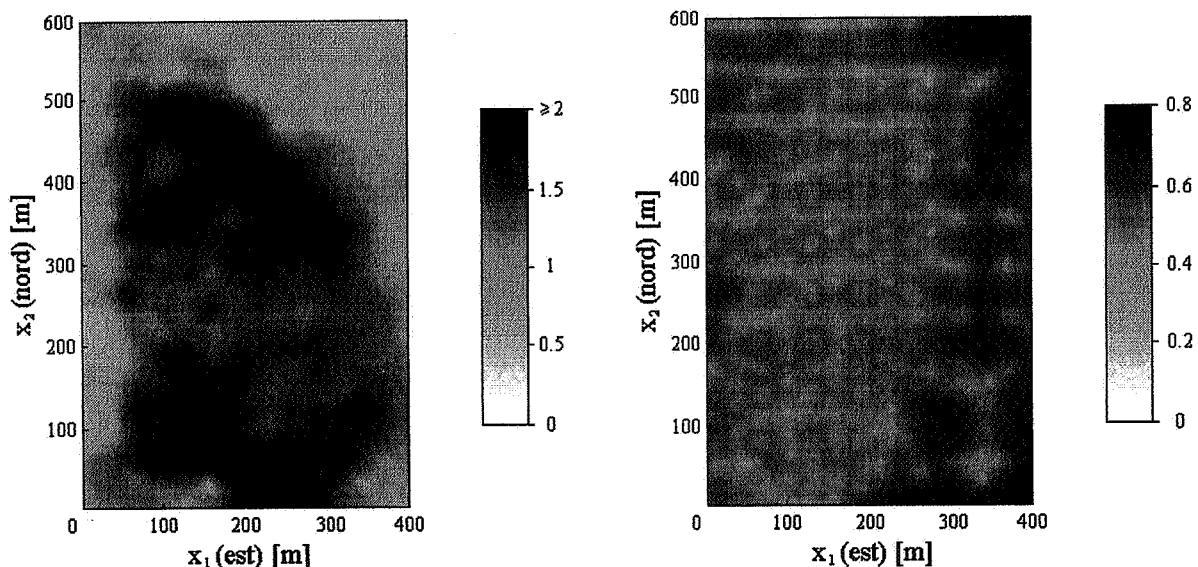


figure 5.17 : cartes des estimations par krigeage ordinaire ponctuel (à gauche) et des écarts-types d'estimation (à droite) ; représentation du banc de profondeur 65 m

On retrouve, sur la carte de krigeage, les zones de fortes valeurs détectées lors de l'étude exploratoire. La carte des écarts-types d'estimation est purement géométrique et ne dépend pas des valeurs prises par les données. Cette situation n'est pas réaliste, notamment si l'on se souvient de l'effet proportionnel mis en évidence sur les données expérimentales : les zones de fortes valeurs présentent une variabilité plus importante que celles de faibles valeurs, et sont donc sujettes à une incertitude plus élevée dans l'estimation.

Prise en compte de l'effet proportionnel

Une solution simple, pour prendre en compte l'effet proportionnel dans le calcul de la précision de l'estimation, consiste à utiliser, dans le système de krigeage, un variogramme quasi-stationnaire, en modulant son palier selon la valeur d'une moyenne locale :

$$\forall \mathbf{x} + \mathbf{h} \in V, \gamma(\mathbf{x} + \mathbf{h}, \mathbf{x}) = \phi[m^*(\mathbf{x})] \gamma_0(\mathbf{h})$$

où V est un bloc centré en $\mathbf{x}$

$m^*(\mathbf{x})$ est la moyenne expérimentale locale en $\mathbf{x}$ (locale = à l'échelle du bloc V)

γ_0 est le variogramme moyen (celui qui a été ajusté globalement).

Cette relation est en principe indépendante de la taille du support V choisi, pourvu que la moyenne locale puisse être considérée comme constante à l'échelle de V (hypothèse admissible dès lors que V est inclus dans le voisinage de krigeage, à l'échelle duquel l'hypothèse de quasi-stationnarité est censée être valide).

La fonction ϕ peut être modélisée à partir du nuage de corrélation entre moyennes et variances locales réalisé au chapitre 2. Il suffit d'écrire :

$$s^{*2}(\mathbf{x}) \approx \frac{1}{|V|^2} \int_V \int_V \gamma(\mathbf{u}, \mathbf{v}) d\mathbf{u} d\mathbf{v} = \phi[m^*(\mathbf{x})] \bar{\gamma}_0(V, V)$$

et d'identifier cette relation avec la régression établie au chapitre 2 : $s^{*2}(\mathbf{x}) \approx 0.152 [m^*(\mathbf{x})]^{2.29}$.

Cette modification du modèle variographique n'a pas d'impact sur la carte des valeurs estimées (à l'échelle du voisinage de krigeage, le variogramme n'est altéré que d'un facteur multiplicatif constant, donc l'estimation est inchangée). En revanche, la variance de krigeage en un point $\mathbf{x}$ sera multipliée par $\phi[m^*(\mathbf{x})]$ et reflètera donc plus fidèlement l'incertitude portant sur l'estimation. On obtient la carte suivante d'écarts-types de krigeage, qui combine à la fois l'information de nature géométrique et la connaissance des moyennes locales, donc les valeurs prises par les données.

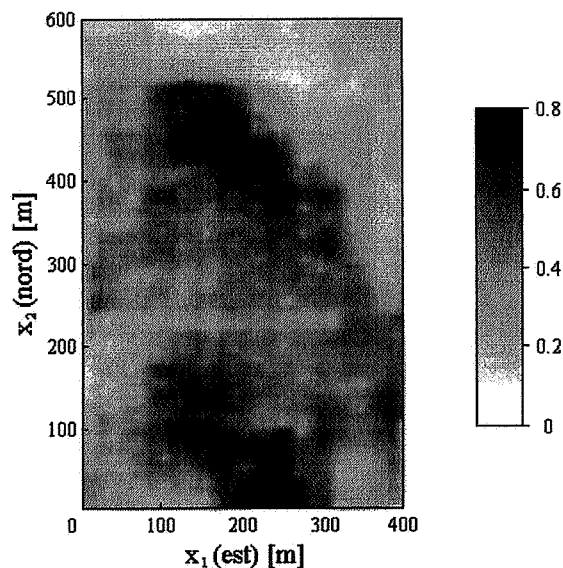


figure 5.18 : carte des écarts-types de krigeage, lorsque l'effet proportionnel est incorporé dans le modèle variographique

Comparaison entre krigeage ordinaire et krigeage simple

A titre comparatif, une seconde estimation a été réalisée par krigeage simple (à moyenne connue), en attribuant à la fonction aléatoire sous-jacente une moyenne égale à 0.9%, proche de la moyenne expérimentale pondérée des données. Les cartes d'estimation et d'écart-types d'estimation pour un banc particulier sont présentées sur la figure 5.19.

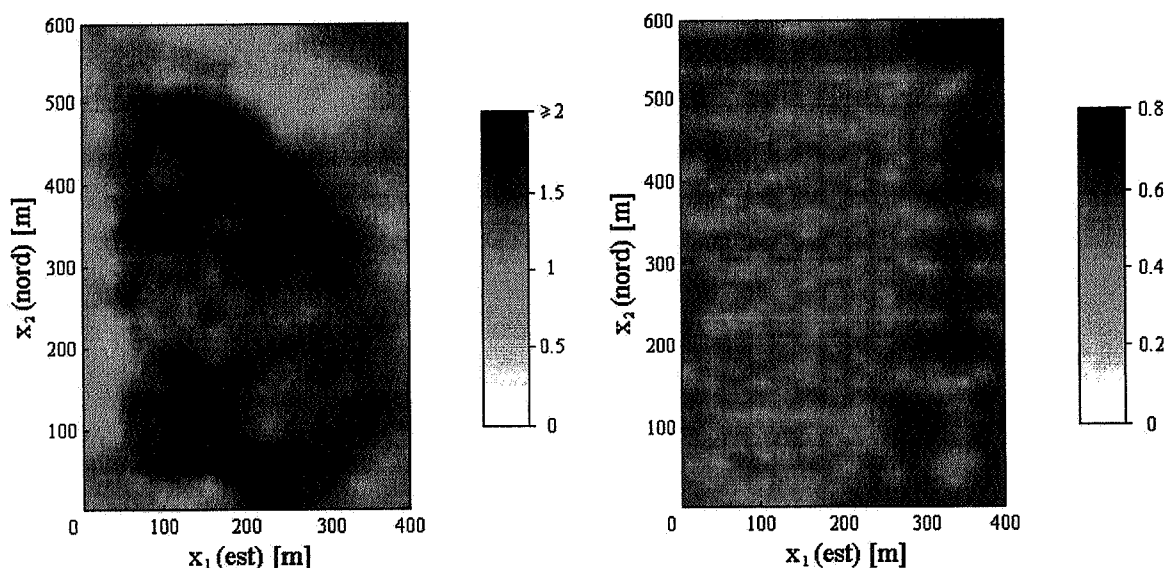


figure 5.19 : cartes des estimations par krigeage simple (à gauche) et des écarts-types d'estimation (à droite) ; représentation du banc de profondeur 65 m

La principale différence entre le krigeage simple et ordinaire se produit en périphérie de la zone estimée, qui possède peu d'échantillons (voir par exemple le coin nord-est des figures précédentes). Le krigeage simple calibre l'estimation sur la moyenne connue (0.9%), tandis que le krigeage ordinaire ne connaît pas cette moyenne et utilise seulement les données voisines, dont la valeur est nettement inférieure à 0.9%.

Il apparaît que l'ajustement de l'estimation sur la moyenne globale est artificielle et a tendance à surestimer la teneur des sites localisés en périphérie du gisement. En conclusion, l'hypothèse de stationnarité avec une moyenne invariante sur le domaine (cas du krigeage simple) n'est pas pertinente. Le krigeage ordinaire affaiblit cette hypothèse et permet que cette moyenne soit lentement variable.

En géostatistique minière, il est d'ailleurs fréquent que les zones de faibles valeurs soient moins densément échantillonnées que celles de fortes valeurs. Utiliser un krigeage simple, avec une moyenne calculée globalement, peut conduire à une surestimation des zones de faibles valeurs ; c'est pourquoi, on préfère en règle générale employer un krigeage ordinaire, qui permet davantage de flexibilité sans pour autant recourir à des modèles non stationnaires, nettement plus compliqués à mettre en œuvre.

V. Les effets des paramètres du modèle

On considère, dans la suite de ce paragraphe, un ensemble de 6 sites, notés A, B, C, D, E et F, répartis dans l'espace géographique selon la configuration de la figure 5.20. En chaque site, une variable a été mesurée. On désire estimer la valeur en un point non échantillonné, noté "?" sur la figure, en tenant compte des 6 valeurs mesurées.

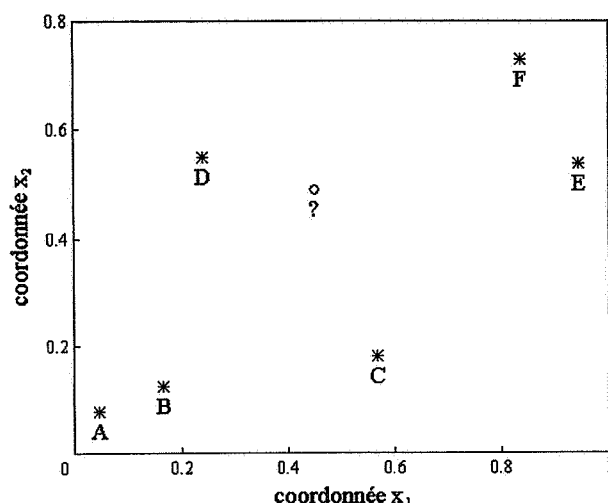


figure 5.20 : estimation ponctuelle au point "?"

On cherche à étudier les effets des paramètres du modèle (palier, portée, anisotropie, constante de pépité...) sur les poids affectés aux sites ainsi que sur l'écart-type de krigeage associé. Sauf avis contraire, les estimations sont effectuées par krigeage ordinaire, c'est-à-dire à moyenne inconnue.

V.1. Type de modèle

Nous allons examiner les répercussions du type de modèle sur l'estimation du point "?", en comparant les résultats obtenus avec un modèle sphérique isotrope et un modèle gaussien de même portée pratique 1 et de même palier (pris égal à 1).

Le modèle sphérique a pour équation

$$\gamma_1(\mathbf{h}) = \begin{cases} \frac{3}{2}|\mathbf{h}| - \frac{1}{2}|\mathbf{h}|^3 & \text{pour } 0 \leq |\mathbf{h}| \leq 1 \\ 1 & \text{pour } |\mathbf{h}| \geq 1 \end{cases}$$

et le modèle gaussien de même portée :

$$\gamma_2(\mathbf{h}) = 1 - \exp(-3|\mathbf{h}|^2).$$

La figure 5.21 montre que les poids de krigeage et l'écart-type d'estimation diffèrent nettement selon le modèle⁹. Dans le cas du modèle gaussien, les valeurs situées près du point à estimer (C et D notamment) ont des poids positifs bien plus forts que dans le cas du modèle sphérique, et l'écart-type d'estimation est réduit de moitié. L'explication tient au fait que le comportement parabolique à l'origine du modèle gaussien et son caractère indéfiniment dérivable traduisent un phénomène extrêmement régulier dans l'espace¹⁰, d'où une importance accrue des points proches dans l'estimation ; l'écart-type de krigeage est plus faible, car l'estimation est censée être plus précise que dans le cas du schéma sphérique qui modélise un phénomène plus erratique.

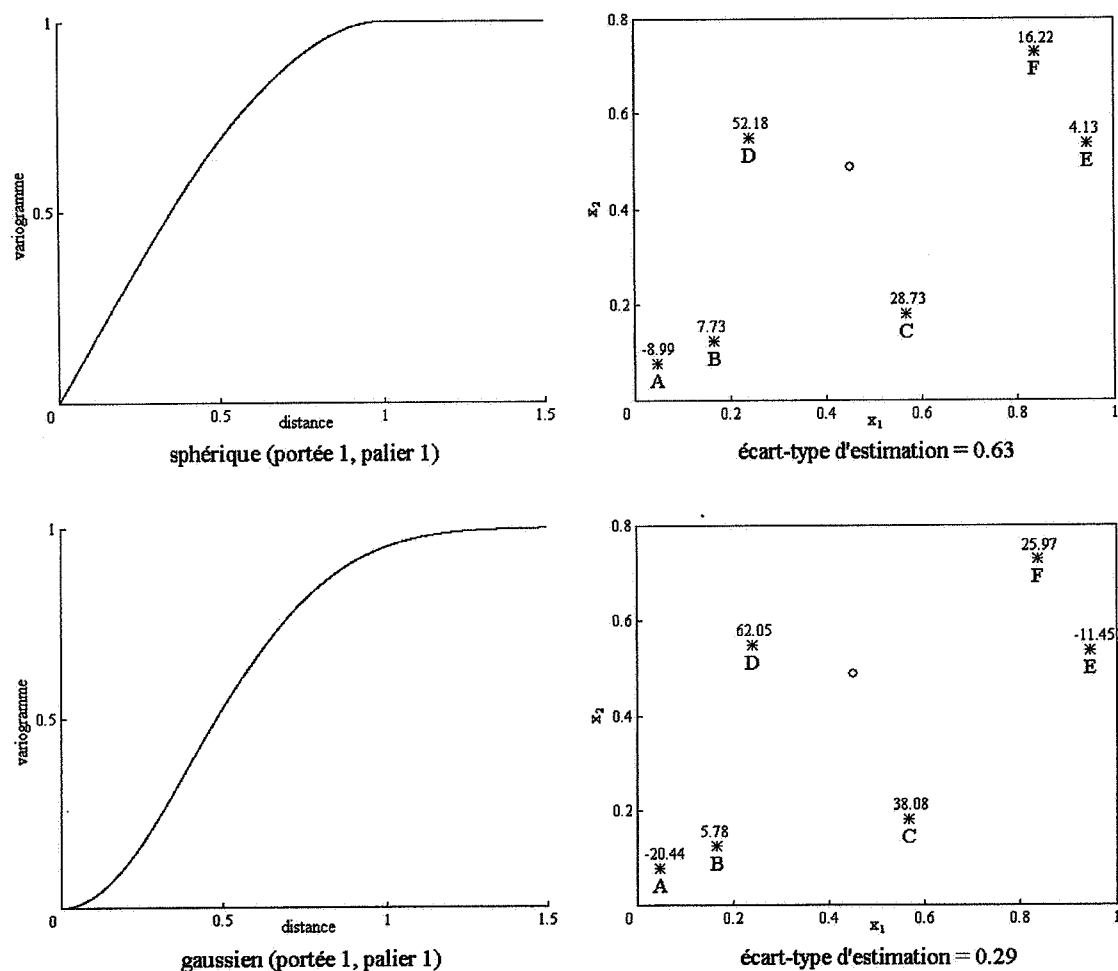


figure 5.21 : effet du type de modèle (sphérique / gaussien)

⁹ Pour des raisons d'affichage numérique, les poids de krigeage sont exprimés en pourcentage ; on vérifiera que leur somme est égale à 100% = 1.

¹⁰ L'expérience montre que le modèle gaussien est susceptible de conduire à des instabilités numériques lorsque des points de données sont proches : les valeurs mesurées risquent en effet d'être "incompatibles" avec l'extrême régularité du modèle (le krigeage avec un modèle gaussien est très proche d'un ajustement polynomial). L'introduction d'un effet de pépite, même léger, améliore significativement la robustesse des résultats : les poids de krigeage sont alors moins sensibles à de légères modifications de la configuration de krigeage.

Aussi bien dans le cas du modèle sphérique que dans celui du modèle gaussien, le site A, le plus extérieur, est affecté d'un poids négatif. On peut observer qu'avec un modèle gaussien cet effet est plus important que dans le cas sphérique. L'explication tient à la structure très continue (indéfiniment dérivable) du modèle gaussien. En effet, le site A est écranté par le site B par rapport au site à estimer. On peut imaginer que si la valeur mesurée sur le site A est forte et si celle du site B qui l'écrante est faible, une évolution régulière fournirait au site à estimer une valeur encore plus faible, d'où un poids négatif affecté au site extérieur A.

Une autre curiosité provient des sites localisés en haut à droite (E et F), pour lesquels on observe des différences notables entre les deux modèles. En particulier, le site E est affecté d'un poids faible ou négatif, bien qu'il soit situé relativement près du site à estimer. Cette constatation montre que le comportement des poids de krigeage est parfois peu intuitif et difficile à expliquer (en l'occurrence, les sites C et F interfèrent dans la détermination du poids assigné au site E).

V.2. Portée

Supposons que l'on ait trois variogrammes sphériques de même palier unité et de portées respectives 0.1, 1 et 2 :

$$\gamma_1(\mathbf{h}) = \begin{cases} \frac{3}{2} \frac{|\mathbf{h}|}{0.1} - \frac{1}{2} \left(\frac{|\mathbf{h}|}{0.1} \right)^3 & \text{pour } 0 \leq |\mathbf{h}| \leq 0.1 \\ 1 & \text{pour } |\mathbf{h}| \geq 0.1 \end{cases}$$

$$\gamma_2(\mathbf{h}) = \begin{cases} \frac{3}{2} |\mathbf{h}| - \frac{1}{2} |\mathbf{h}|^3 & \text{pour } 0 \leq |\mathbf{h}| \leq 1 \\ 1 & \text{pour } |\mathbf{h}| \geq 1 \end{cases}$$

$$\gamma_3(\mathbf{h}) = \begin{cases} \frac{3}{2} \frac{|\mathbf{h}|}{2} - \frac{1}{2} \left(\frac{|\mathbf{h}|}{2} \right)^3 & \text{pour } 0 \leq |\mathbf{h}| \leq 2 \\ 1 & \text{pour } |\mathbf{h}| \geq 2 \end{cases}$$

Les résultats du krigeage sont visualisés sur la figure 5.22. Dans le premier cas, la portée est très faible (0.1), et comme les distances entre tous les sites concernés sont supérieures à cette valeur, on obtient un résultat identique à un effet de pépite pur : chaque observation a le même poids (16.67%). Dans les deux autres cas (portées de 1 et 2), on trouve des valeurs de poids similaires, surtout pour les sites proches du site à estimer (C, D, E et F). L'explication tient au fait que, à l'échelle de travail, les deux modèles sphériques de portées 1 et 2 m apparaissent quasiment linéaires et ne diffèrent que d'un facteur multiplicatif, ce qui entraîne que les poids de krigeage ne changent pratiquement pas (cf. paragraphe suivant sur l'influence du palier). En revanche, l'écart-type de krigeage dépend fortement de la portée. Lorsque la portée augmente, la corrélation entre les sites de mesure et le site à estimer augmente : les valeurs observées apportent davantage d'information sur la valeur inconnue et par conséquent l'estimation est censée être plus précise.

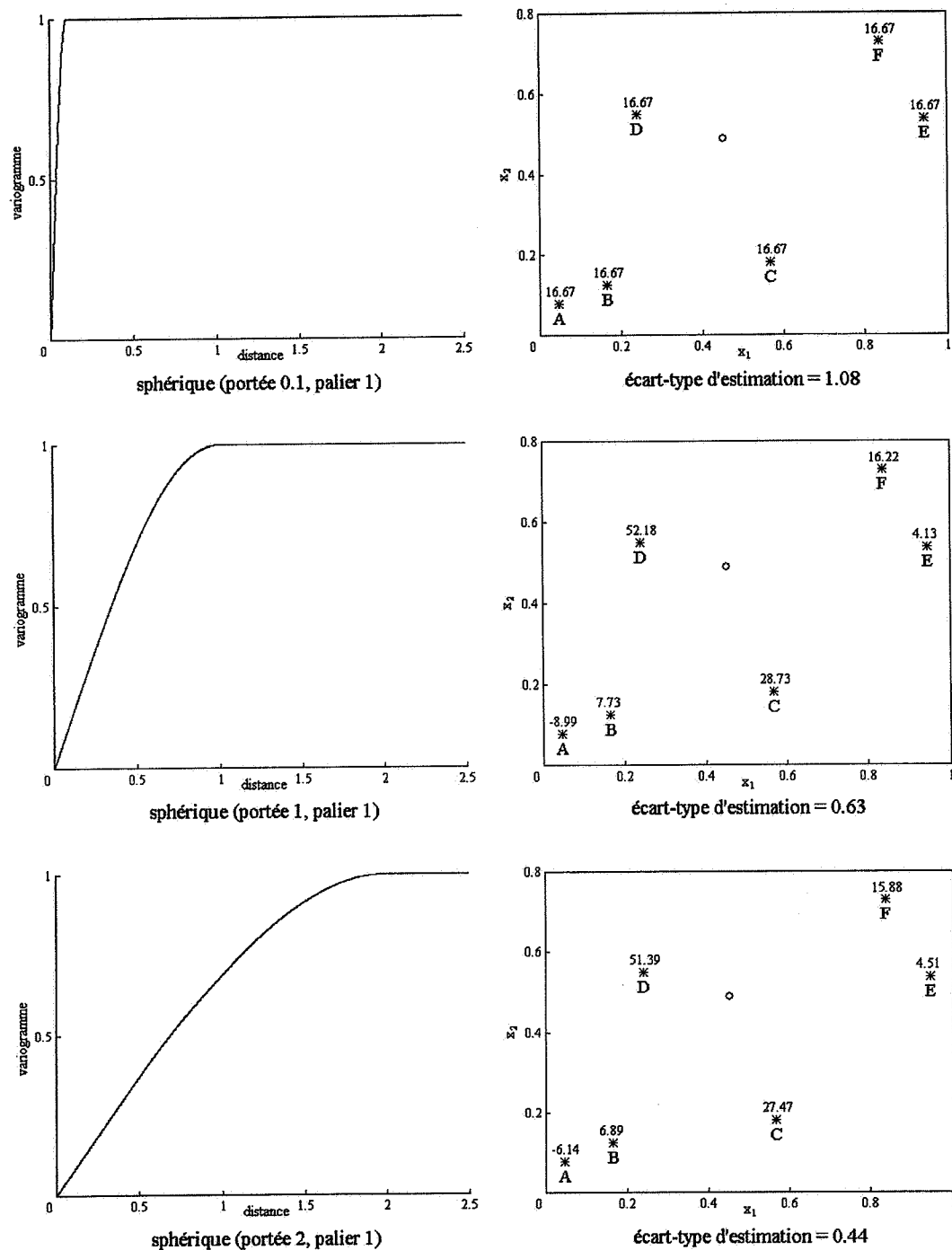


figure 5.22 : effet de la portée (0.1 / 1 / 2)

V.3. Palier

Voyons à présent comment se traduit sur l'estimation du point “?” l'influence du palier du modèle variographique. Les deux schémas sphériques isotropes ci-après ont la même portée unité et ne diffèrent que par la valeur du palier, égale à 1 ou à 2.

$$\gamma_1(\mathbf{h}) = \begin{cases} \frac{3}{2} |\mathbf{h}| - \frac{1}{2} |\mathbf{h}|^3 & \text{pour } 0 \leq |\mathbf{h}| \leq 1 \\ 1 & \text{pour } |\mathbf{h}| \geq 1 \end{cases}$$

et

$$\gamma_2(\mathbf{h}) = \begin{cases} 2 \left(\frac{3}{2} |\mathbf{h}| - \frac{1}{2} |\mathbf{h}|^3 \right) & \text{pour } 0 \leq |\mathbf{h}| \leq 1 \\ 2 & \text{pour } |\mathbf{h}| \geq 1 \end{cases}$$

La figure 5.23 montre que les poids de krigeage sont inchangés, donc la valeur estimée également : en d'autres termes, multiplier le variogramme par une constante positive n'altère pas l'estimation. En revanche, l'écart-type de krigeage a été modifié. On peut vérifier qu'il a été multiplié par $\sqrt{2}$ lorsque le palier a été doublé, c'est-à-dire que la variance d'estimation a été multipliée par le même facteur 2 que le palier du variogramme.

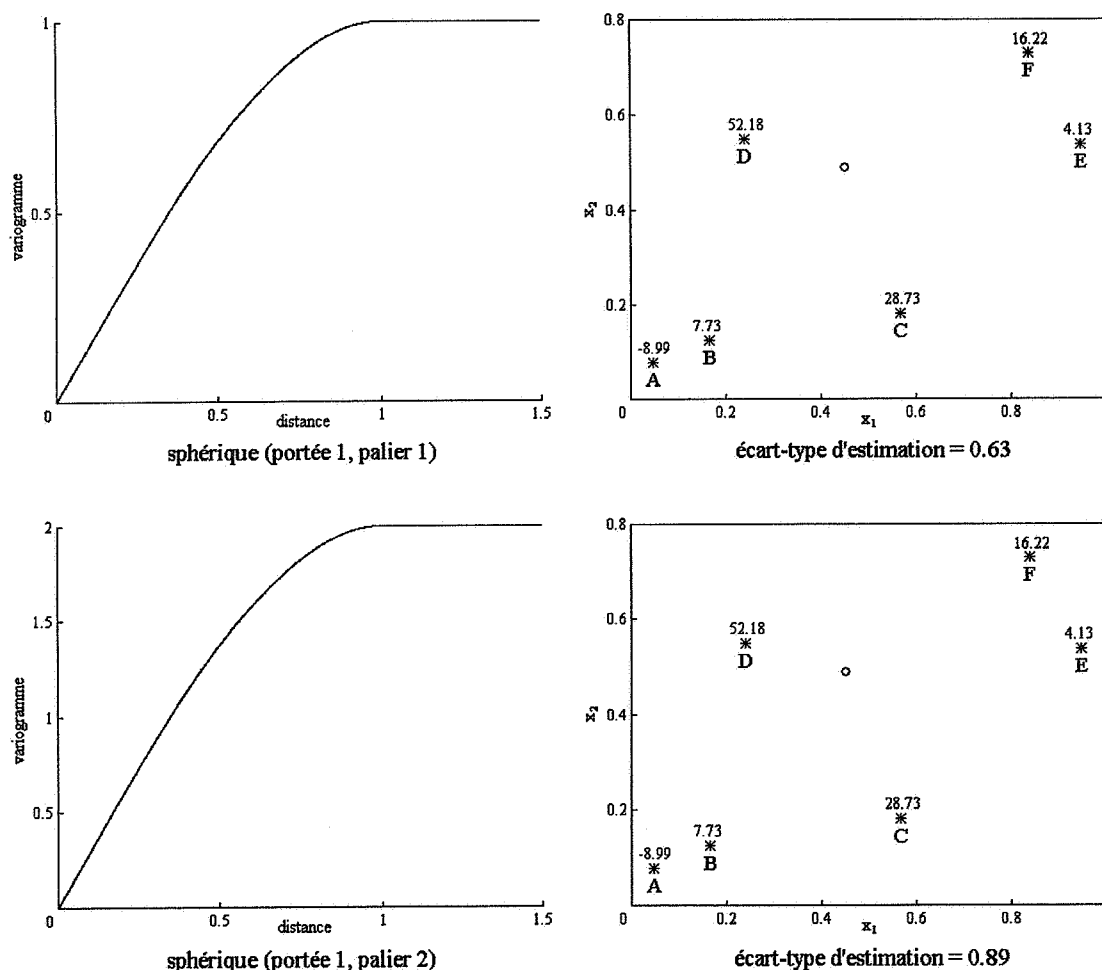


figure 5.23 : effet du palier (1 / 2)

V.4. Effet de pépite

Les deux schémas proposés ci-dessous ont même palier, égal à 1, mais l'un n'a pas de constante pépitique et l'autre a une constante pépitique égale à 0.5 :

$$\gamma_1(\mathbf{h}) = \begin{cases} \frac{3}{2}|\mathbf{h}| - \frac{1}{2}|\mathbf{h}|^3 & \text{pour } 0 \leq |\mathbf{h}| \leq 1 \\ 1 & \text{pour } |\mathbf{h}| \geq 1 \end{cases}$$

$$\gamma_2(\mathbf{h}) = \begin{cases} 0 & \text{pour } |\mathbf{h}| = 0 \\ \frac{1}{2} + \frac{1}{2} \left(\frac{3}{2}|\mathbf{h}| - \frac{1}{2}|\mathbf{h}|^3 \right) & \text{pour } 0 < |\mathbf{h}| \leq 1 \\ 1 & \text{pour } |\mathbf{h}| \geq 1 \end{cases}$$

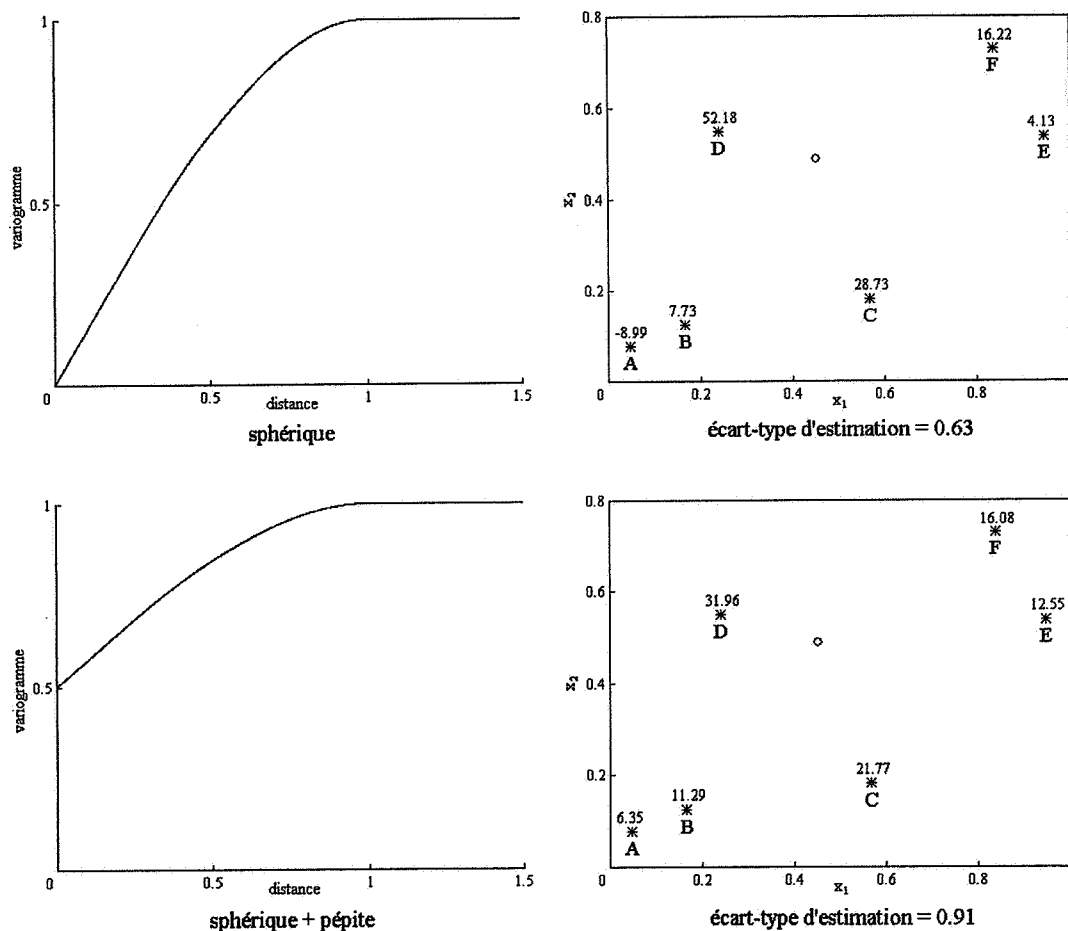


figure 5.24 : influence de l'effet de pépite (0 / 0.5)

L'effet de pépite modifie à la fois l'écart-type de krigeage, qui augmente lorsque la constante pépitique augmente, et l'estimation par l'intermédiaire des poids de krigeage. En présence d'un effet de pépite, l'étendue des poids (différence entre la plus forte et la plus faible valeur) se réduit par rapport à celle issue d'un variogramme sans effet de pépite. On remarquera aussi que les poids négatifs ont disparu en présence de l'effet de pépite.

Lorsque l'effet de pépite devient prépondérant, voire conduit à un variogramme purement pépitique :

$$\gamma(\mathbf{h}) = \begin{cases} 0 & \text{si } |\mathbf{h}| = 0 \\ 1 & \text{sinon} \end{cases}$$

il n'y a alors aucune structure spatiale entre les valeurs observées sur les différents sites. Les poids attribués sont tous égaux à $1/n$ (soit 16.67% ici), et l'estimation en tout point de l'espace différent d'un point de mesure est égale à la moyenne arithmétique des données (l'estimation d'un point de mesure redonne la valeur mesurée, car le krigeage est un interpolateur exact). La disposition des échantillons est indifférente, seul leur nombre compte.

V.5. Effet de trou

Nous allons examiner l'impact de l'introduction d'un effet de trou dans le variogramme. On choisit deux modèles pseudo-périodiques (sinus cardinal), de portées pratiques respectives 1 et 2 et de même palier unité. La valeur de la pseudo-période va être un paramètre important pour expliquer le signe et l'amplitude des poids de krigeage.

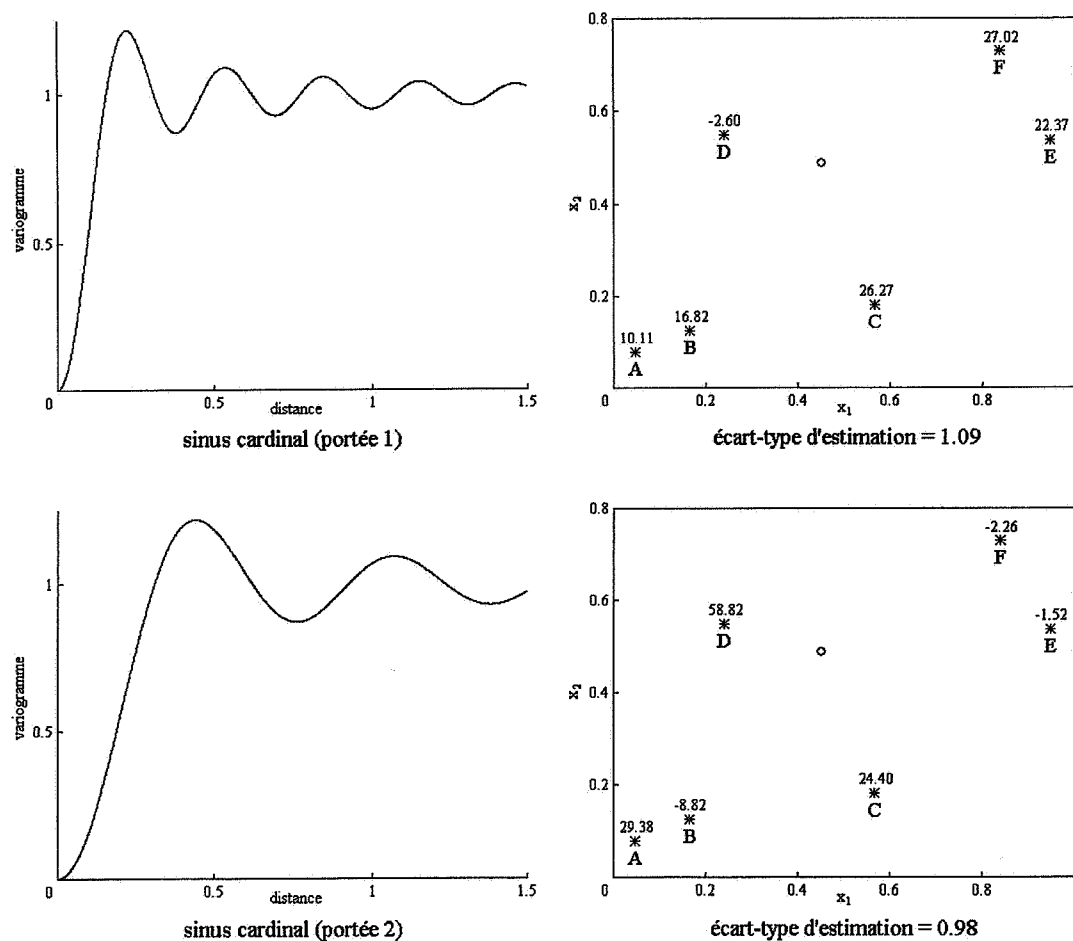


figure 5.25 : influence de l'effet de trou

Selon la valeur de la pseudo-période, l'effet de trou peut modifier radicalement les poids de krigeage et conduire à des situations curieuses où des sites proches ont des poids très faibles. Le premier exemple est particulièrement démonstratif : le site D, le plus proche du site à estimer, est affecté d'un poids négatif, car il est situé à une distance qui correspond au minimum de corrélation (environ 0.2) ; en quelque sorte, il est en opposition de phase par rapport au site à estimer. Dans le second cas, où la valeur de la pseudo-période est doublée, on remarque que le site A, le plus éloigné, est affecté d'un poids fort, alors que le site B qui l'écrante se voit attribuer un poids négatif. Dans les deux cas, l'écart-type d'estimation est plus important que celui obtenu avec un schéma sphérique de portée 1 et de palier unité. La raison est que le modèle sinus cardinal croît plus vite que le modèle sphérique jusqu'à dépasser le palier unité, donc la structuration des données se détériore plus rapidement.

V.6. Anisotropie

Jusqu'ici, nous avons supposé que les variations du phénomène observé étaient identiques dans toutes les directions. Nous allons maintenant comparer les résultats du krigeage dans le cas isotrope à ceux obtenus dans le cas d'une anisotropie géométrique de rapport 3 et d'angle 45° .

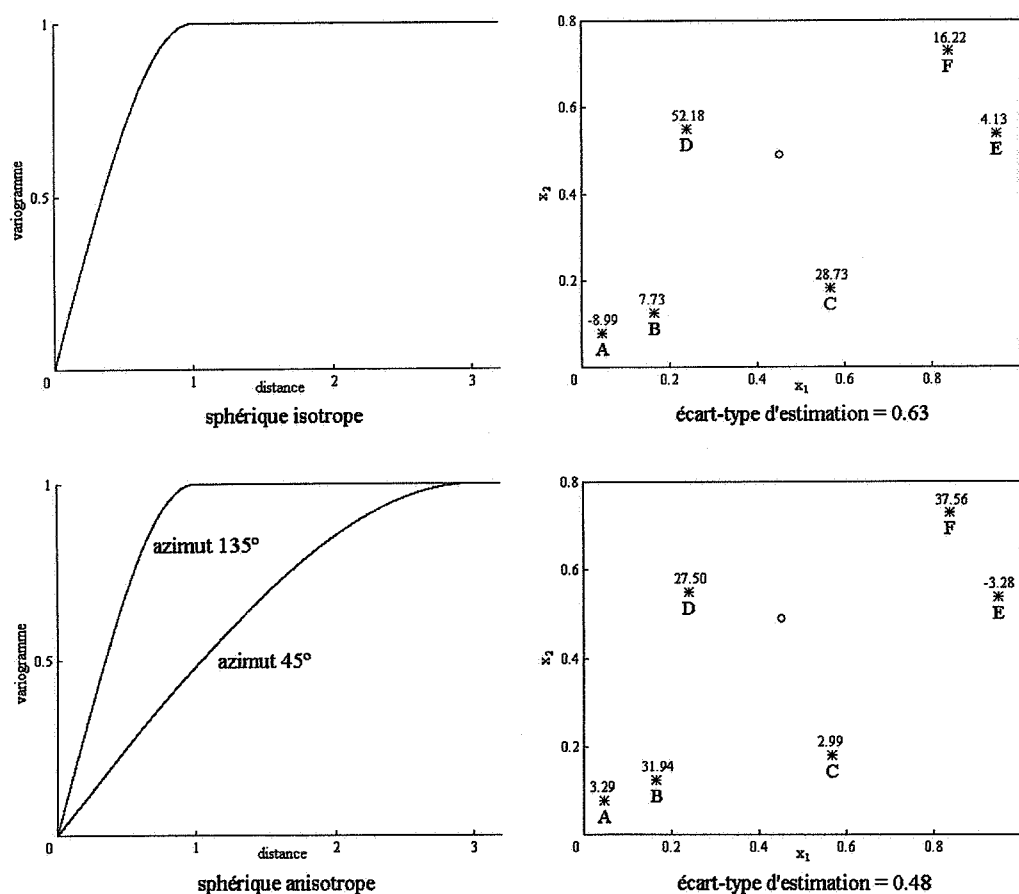


figure 5.26 : effet de l'anisotropie

Les poids de krigeage affectés aux sites situés dans la direction principale d'anisotropie ont augmenté (points A, B et F), ce qui traduit que la corrélation dans cette direction est plus forte que dans la direction orthogonale (points C et D).

V.7. Type de krigeage (simple / ordinaire)

Tous les essais effectués jusqu'à présent ont fait appel au krigeage ordinaire (c'est-à-dire à moyenne inconnue). On se propose ici d'examiner les différences sur les poids de krigeage qu'entraîne la connaissance de la moyenne. Deux modèles sphériques sont utilisés, de portées respectives 1 et 0.5 et de même palier unité (figure 5.27).

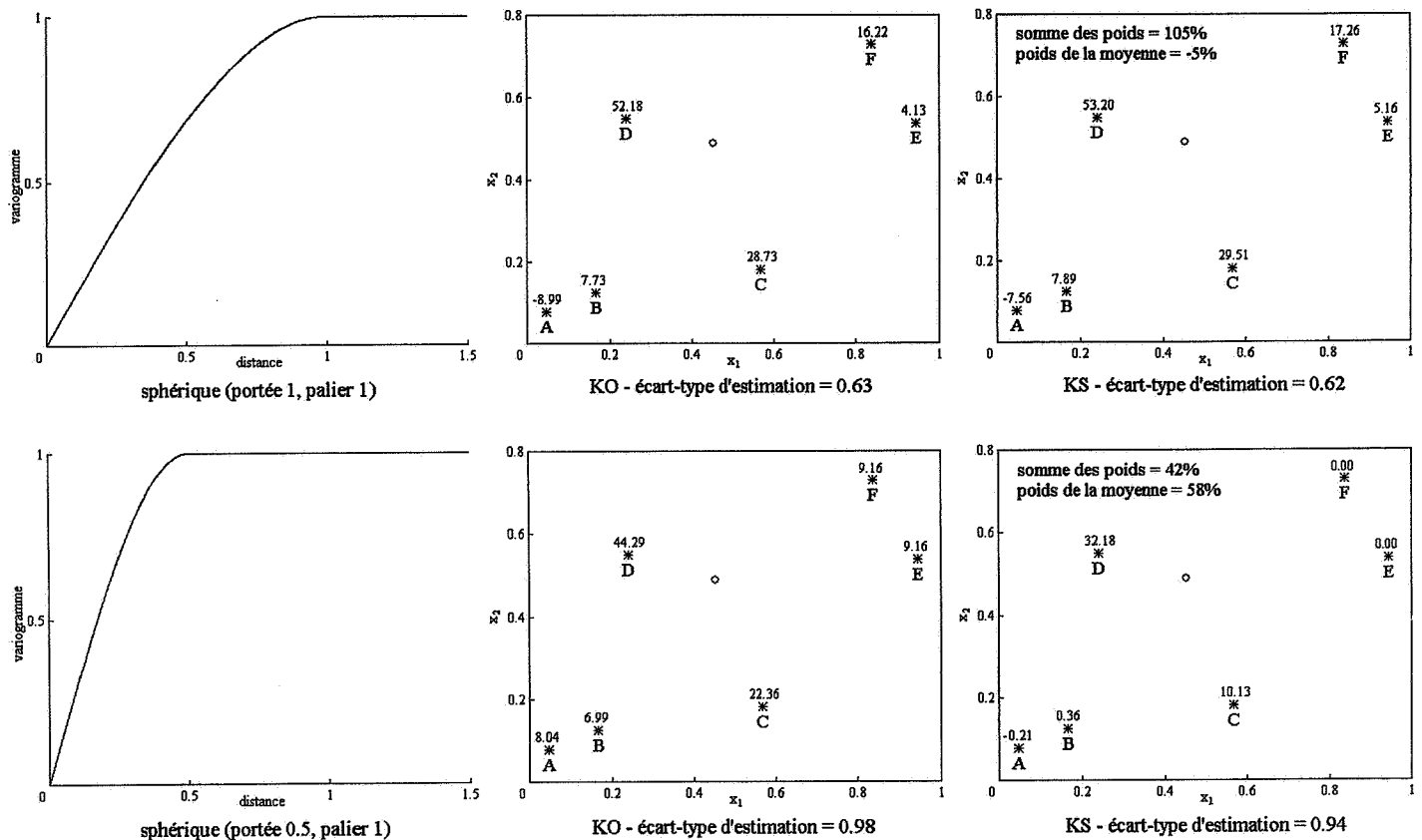


figure 5.27 : effet du type de krigeage, ordinaire (KO) ou simple (KS)
en haut : portée du modèle égale à 1, en bas : portée égale à 0.5

On sait (*théorème d'additivité*) que le krigeage ordinaire coïncide avec le krigeage simple à condition de remplacer la moyenne inconnue par son estimation optimale. Le fait de remplacer la moyenne par son estimation explique que le krigeage ordinaire donne toujours une variance d'estimation plus grande que le krigeage simple. Dans le premier cas (portée égale à 1), krigeages simple et ordinaire diffèrent peu. Cela signifie que l'information est suffisamment abondante pour que la moyenne, vraie ou estimée, soit affectée d'un poids faible. La perte de précision entre krigeages simple et ordinaire est alors faible. Dans le second cas où la portée des observations est deux fois plus petite, tout se passe comme si les informations disponibles étaient deux fois plus éloignées du site à estimer, d'où une plus grande imprécision dans l'estimation. Le krigeage simple compense la perte d'information en calant l'estimation à la moyenne (dont le poids passe de -5% à 58%), tandis que le krigeage ordinaire estime cette moyenne, dont on voit que le poids (58%) est "réparti" entre les données à raison d'environ 10% par donnée.

Il est intéressant d'examiner de manière plus détaillée les poids de krigeage attribués aux différents sites. Ainsi, dans le cas du krigeage simple avec une portée de 0.5, les données E et F, situées à une distance supérieure à la portée par rapport au site à estimer, ont des poids nuls ; il n'en est pas de même du site extérieur A, qui, bien que plus éloigné du site à estimer, admet un poids faible mais non nul. C'est une illustration de l'*effet de relais*, dû à la présence des sites B et C.

En ce qui concerne le krigeage ordinaire, le site A est affecté d'un poids plus fort que le site B qui l'écrante, situé pourtant plus près du point à estimer (*effet d'écran inverse*) : on peut décomposer le poids affecté à chaque site en une contribution du poids de krigeage simple et une contribution du poids de la moyenne estimée. La première a tendance à donner un poids supérieur aux sites proches, tandis que la seconde privilégie au contraire les points extérieurs et / ou isolés. Les deux contributions ont des effets opposés, ce qui explique que, lorsque la deuxième est prépondérante, des points éloignés reçoivent un poids plus fort que des points proches.

VI. Le krigeage par bloc

VI.1. Principe

Jusqu'à présent, nous nous sommes intéressés aux problèmes d'estimation en des localisations de même support que celui des sites d'observation. Nous allons voir maintenant comment estimer de plus grands supports, appelées blocs. La quantité à estimer est cette fois :

$$z(V) = \frac{1}{|V|} \int_V z(\mathbf{x}) d\mathbf{x}$$

où V est un bloc quelconque et |V| sa surface ou son volume.

Pour que les calculs aient un sens physique, il est nécessaire que la variable étudiée soit **additive**. Par exemple, l'estimation sur un bloc d'une variable comme le pH ne sera pas possible, car le pH d'un bloc n'est pas égal à la moyenne arithmétique des pH ponctuels sur ce bloc.

En vertu de la propriété d'additivité du krigeage, on peut se ramener à des estimations ponctuelles : on peut diviser chaque bloc selon une grille très fine, puis réaliser un krigeage ponctuel en chaque nœud de la grille, et enfin calculer la moyenne des valeurs krigées dans le bloc. Cette procédure, bien qu'assez simple, présente deux limitations :

- elle implique des temps de calculs importants dus à la résolution d'un grand nombre de systèmes de krigeage (autant qu'il y a de points à kriger dans le bloc) ;
- elle ne permet pas d'apprécier la précision de l'estimation du bloc.

Le krigeage par bloc supprime ces problèmes, puisqu'il ne nécessite que la résolution d'un seul système de krigeage, qui fournira une estimation et une variance d'estimation.

Dans le système de krigage ponctuel, le site à estimer n'intervient que dans le second membre du système. C'est donc ce second membre qu'il faut modifier dans le cadre de l'estimation par bloc.

A ce stade, il est utile de rappeler la notion de valeur moyenne d'une fonction aléatoire $Z(\mathbf{x})$ sur un bloc V . Il s'agit de la moyenne de toutes les variables ponctuelles $Z(\mathbf{x})$ définies dans V :

$$Z(V) = \frac{1}{|V|} \int_V Z(\mathbf{x}) d\mathbf{x}.$$

$Z(V)$ est une variable aléatoire, en tant que combinaison linéaire de variables aléatoires. Dans le cas où $Z(\mathbf{x})$ est stationnaire d'ordre deux, on peut calculer la covariance entre la valeur d'un point d'observation $\mathbf{x}_\alpha$ et celle du bloc V :

$$\overline{C}(\mathbf{x}_\alpha, V) = \text{cov}[Z(V), Z(\mathbf{x}_\alpha)] = \frac{1}{|V|} \int_V \text{cov}[Z(\mathbf{x}), Z(\mathbf{x}_\alpha)] d\mathbf{x} = \frac{1}{|V|} \int_V C(\mathbf{x} - \mathbf{x}_\alpha) d\mathbf{x}.$$

Ainsi, la covariance entre la valeur ponctuelle en $\mathbf{x}_\alpha$ et celle du bloc V est égale à la moyenne des covariances entre $Z(\mathbf{x}_\alpha)$ et $Z(\mathbf{x})$, où $\mathbf{x}$ parcourt V .

Si l'on écrit les quatre étapes du krigage, on s'aperçoit que, pour obtenir le système de krigage par bloc, il suffit de remplacer dans le second membre du système de krigage ponctuel la covariance entre $\mathbf{x}_\alpha$ et $\mathbf{x}_0$ par la covariance entre $\mathbf{x}_\alpha$ et V . Par exemple, dans le cas d'un krigage ordinaire (à moyenne inconnue), on obtient :

$$\begin{bmatrix} C(\mathbf{x}_\alpha - \mathbf{x}_\beta) & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_\beta \\ \mu \end{bmatrix} = \begin{bmatrix} \overline{C}(\mathbf{x}_\alpha, V) \\ 1 \end{bmatrix}$$

$$\begin{pmatrix} C(\mathbf{x}_1 - \mathbf{x}_1) & \cdots & C(\mathbf{x}_1 - \mathbf{x}_n) & 1 \\ \vdots & & \vdots & \vdots \\ C(\mathbf{x}_n - \mathbf{x}_1) & \cdots & C(\mathbf{x}_n - \mathbf{x}_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} \overline{C}(\mathbf{x}_1, V) \\ \vdots \\ \overline{C}(\mathbf{x}_n, V) \\ 1 \end{pmatrix}$$

Ce système peut d'ailleurs se déduire directement de celui du krigage ponctuel, à cause de la linéarité des équations. Pour information, la variance de krigage associée vaut :

$$\sigma_{KO}^2(V) = \overline{C}(V, V) - \sum_{\alpha=1}^n \lambda_\alpha \overline{C}(\mathbf{x}_\alpha, V) - \mu, \text{ où } \overline{C}(V, V) = \text{var}[Z(V)] = \frac{1}{|V|^2} \int_V \int_V C(\mathbf{x} - \mathbf{y}) d\mathbf{x} d\mathbf{y}$$

Le krigage par bloc peut être étendu au cas intrinsèque, à condition de remplacer dans les équations de krigage la covariance C par l'opposé du variogramme $-\gamma$.

VI.2. Exemple sur les données minières

Les unités sélectives d'exploitation correspondent à des blocs de taille $10\text{m} \times 10\text{m} \times 12\text{m}$ et sont plus volumineuses que les échantillons (composites de 12m de longueur) ; ce sont ces unités que le praticien cherche à évaluer, en vue de définir la stratégie d'exploitation et le devenir de chaque unité. La figure 5.28 présente les cartes de krigeage par bloc et des écarts-types de krigeage associés (la visualisation est limitée au banc médian, de profondeur 65m). Les estimations ont été réalisées par krigeage ordinaire, avec le même modèle de variogramme et le même voisinage de krigeage que ceux définis dans le cas ponctuel.

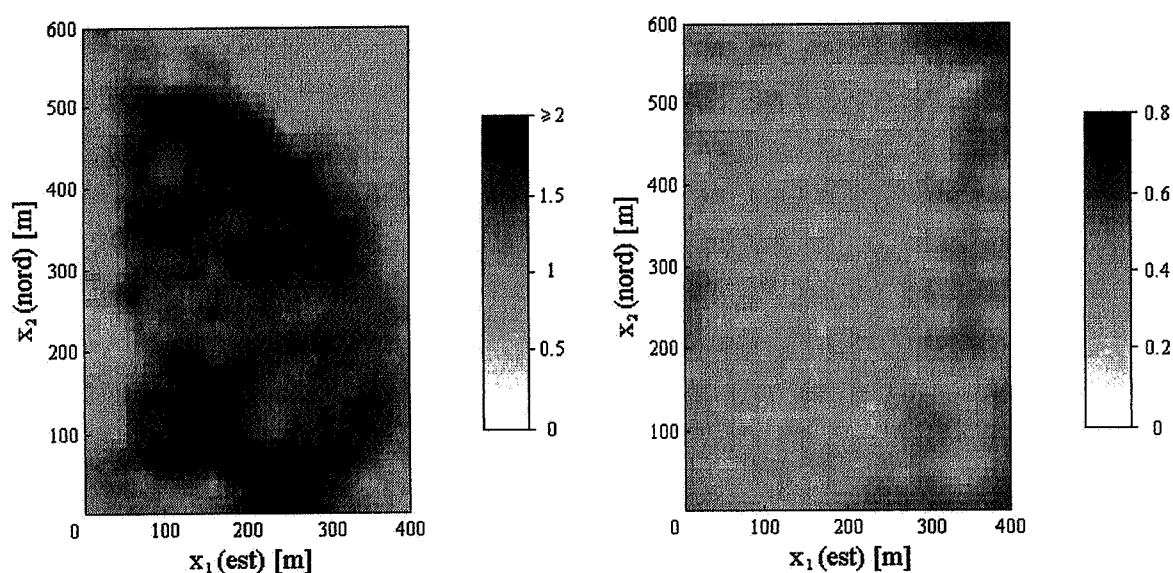


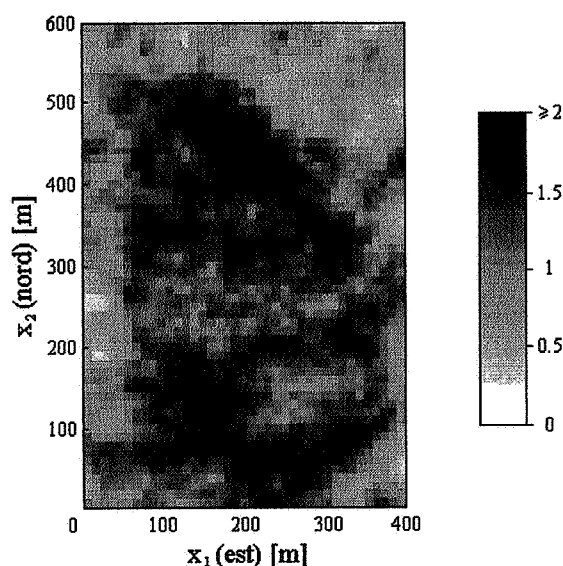
figure 5.28 : cartes du krigeage par bloc (à gauche) et de l'écart-type de krigeage par bloc (à droite) ; estimations sur des blocs de taille $10\text{m} \times 10\text{m} \times 12\text{m}$

La carte du krigeage par bloc ressemble fort à celle du krigeage ponctuel ; le coefficient de corrélation entre les estimations ponctuelles et celles des blocs est égal à 0.99. Autrement dit, l'estimation d'un bloc diffère peu de celle de son point central. Ceci s'explique par le fait que la taille des blocs est inférieure à la maille d'échantillonnage : l'estimation ponctuelle ne permet pas de beaucoup différencier les points d'un même bloc et est presque identique à l'estimation du bloc lui-même.

En revanche, la carte des écarts-types d'estimation présente des différences notables par rapport au cas ponctuel. En particulier, les écarts-types de krigeage par bloc ne sont jamais nuls (la valeur minimale est supérieure à 0.15), et leurs valeurs ont une bien moins grande amplitude que ceux du krigeage ponctuel. En fait, on vérifie que l'écart-type d'estimation d'un bloc est toujours inférieur à celui de son point central (sauf dans le cas exceptionnel où le point central coïncide avec un point de données). Cela est une conséquence de l'*effet de support*, selon lequel les valeurs des blocs sont moins dispersées et présentent une structure plus régulière que les valeurs ponctuelles ; comme on estime des valeurs moins variables, il est logique que les incertitudes d'estimation soient moindres dans le cas du krigeage par bloc.

Sélection sur les valeurs estimées

Il faut prendre garde que l'estimation, aussi précise soit-elle, n'est jamais identique à la réalité. La figure suivante propose une carte simulée, qui possède les mêmes caractéristiques que les valeurs réelles inconnues (même histogramme, même variogramme, etc.) et qui pourrait donc représenter les valeurs réelles des blocs. On note que cette carte présente davantage de variabilité que la carte des estimations : ceci signifie que la chute de variance entre les valeurs mesurées et les valeurs estimées — due à la propriété de lissage du krigeage — est plus importante que celle entre les valeurs mesurées et les valeurs réelles des blocs — due à l'effet de support.



*figure 5.29 : valeurs "réelles" des blocs
représentation du banc de profondeur 65m.*

Cette différence de variabilité entre les valeurs réelles et estimées peut être source de biais dans les prévisions si l'on effectue une sélection sur la carte des valeurs krigées. Ainsi, en géostatistique minière, tous les blocs ne sont pas intéressants, certains ayant une teneur trop faible pour que leur exploitation soit rentable ("stérile") ; seuls les blocs dont la teneur est supérieure à une *teneur de coupure* ("minéral") sont envoyés en laverie. Considérons à titre d'exemple une teneur de coupure égale à 0.5%. Les figures 5.30 et 5.31 présentent les résultats obtenus selon que la sélection s'effectue sur les valeurs réelles (cas idéal) ou les valeurs estimées (cas pratique).

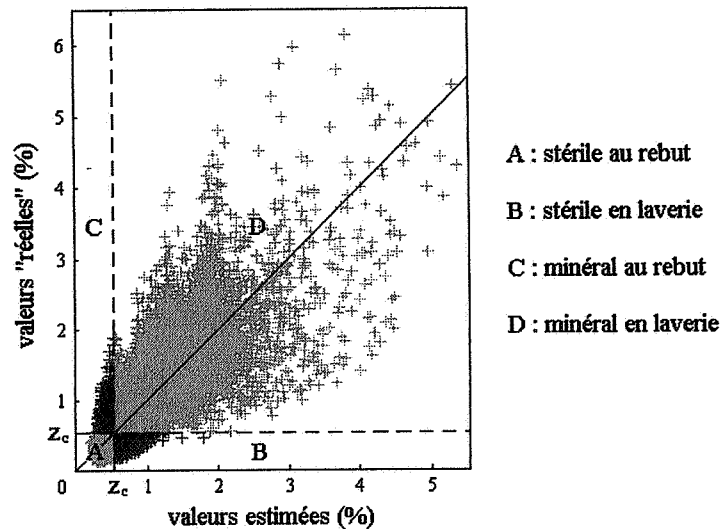


figure 5.30 : nuage de corrélation entre valeurs estimées et vraies, et classification des blocs en minéral / stérile et rebut / laverie

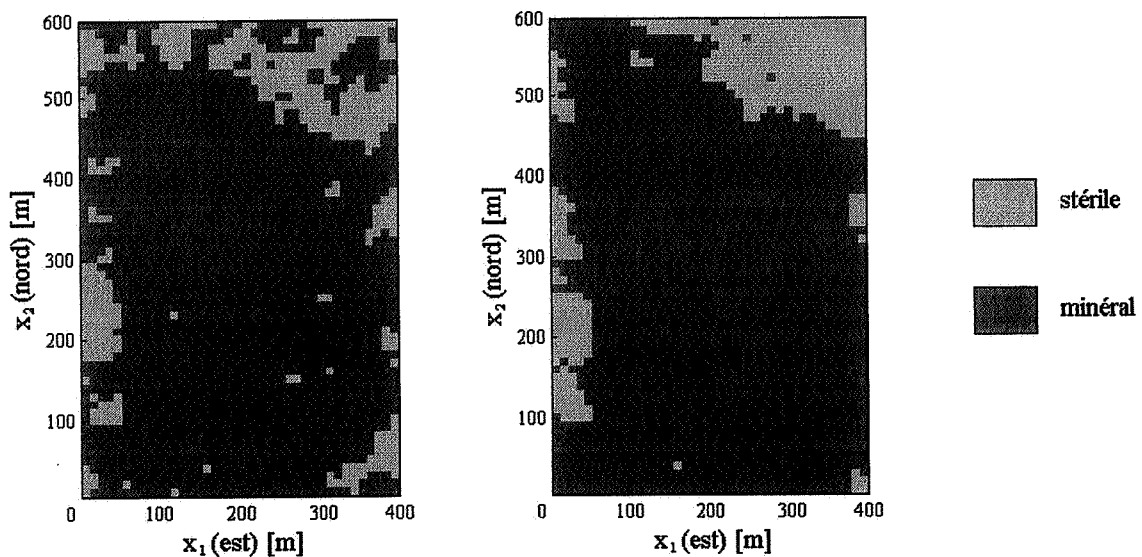


figure 5.31 : sélection sur les valeurs réelles (à gauche) et estimées (à droite) représentation du banc de profondeur 65m

Dans le premier cas (sélection sur les valeurs réelles), la proportion des blocs dépassant la teneur de coupure, appelée *tonnage*, est égale à 80.4%, tandis que dans le second cas, elle vaut 83.7%. Les valeurs moyennes au-dessus de la teneur de coupure sont respectivement égales à 1.09% et 1.05%. On visualise, sur la figure 5.32, l'évolution des tonnages et teneur moyenne en fonction de la teneur de coupure, aussi bien pour les valeurs krigées que réelles.

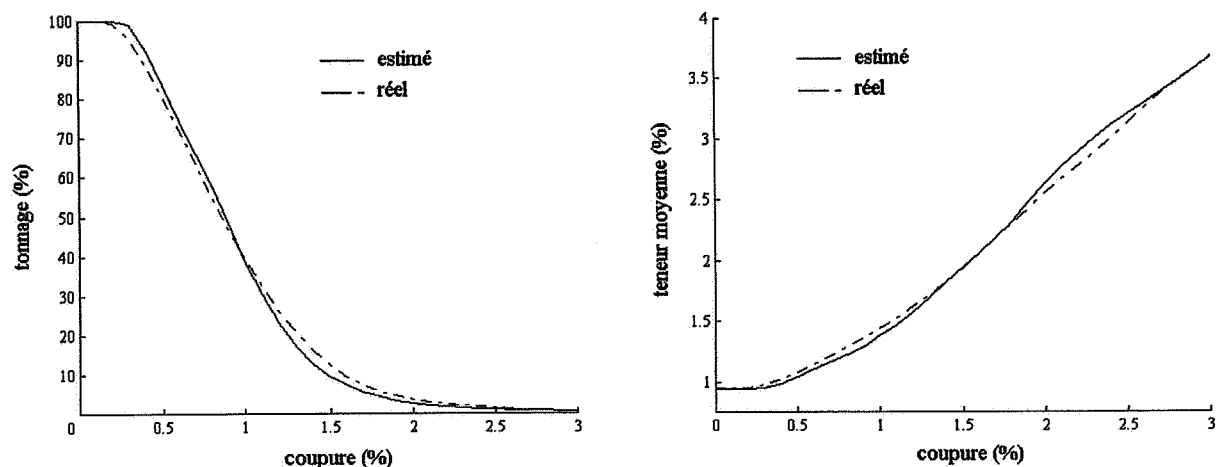


figure 5.32 : courbes tonnage – teneur de coupure et teneur moyenne – teneur de coupure pour les valeurs estimées (lignes continues) et réelles (pointillés)

Les courbes de tonnage sont équivalentes à la donnée de l'histogramme des valeurs. On voit ainsi que le krigeage surestime la proportion des blocs dépassant les teneurs de coupure faibles, et sous-estime la proportion dépassant les teneurs de coupure élevées, ce qui confirme que, dans cet exemple, les valeurs estimées sont moins variables que les valeurs vraies (le krigeage a “trop” lissé la réalité). L'écart maximal entre les tonnages réel et estimé vaut 4.13% et se produit pour une teneur de coupure égale à 0.4%.

Cet écart signifie en particulier que la sélection des blocs que l'on envoie en laverie n'est pas parfaite, et que l'on enverra des blocs trop pauvres et que l'on rejettera des blocs riches (“effet d'information”, voir figure 5.30).

Il est important à présent de savoir si, parmi les blocs que l'on envoie en laverie, la teneur moyenne estimée est égale à celle que l'on va effectivement récupérer.

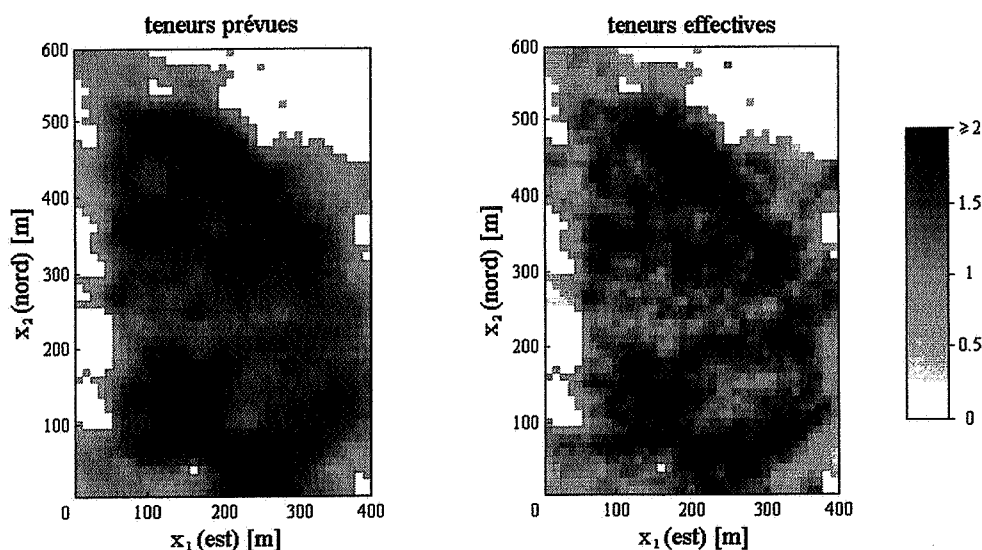


figure 5.33 : teneurs estimées et effectives des blocs qui seront envoyés en laverie

Si l'on remplace le concept de moyenne par celui d'espérance, ceci se traduit en termes mathématiques par :

$$E[Z^*(v) | Z^*(v) > z_c] = E[Z(v) | Z^*(v) > z_c]$$

où $Z(v)$, $Z^*(v)$ et z_c désignent respectivement la valeur réelle du bloc v , sa valeur estimée et la teneur de coupure.

Lorsque cette égalité est satisfaite, on dit qu'il n'y a pas de **biais conditionnel**. Un paramètre qui a un impact sur le biais conditionnel est le nombre d'échantillons utilisés dans l'estimation, et par suite la taille du voisinage de krigeage. Le tableau 5.4 donne les valeurs effectives et estimées des teneurs moyennes au-dessus de la teneur de coupure (égale à 0.5%) en fonction du nombre de données retenues dans l'estimation.

	teneur moyenne effective	teneur moyenne estimée
1 échantillon	1.08 %	1.14 %
4 échantillons	1.06 %	1.10 %
8 échantillons	1.06 %	1.08 %
16 échantillons	1.05 %	1.06 %
24 échantillons	1.05 %	1.05 %
32 échantillons	1.05 %	1.05 %
48 échantillons	1.04 %	1.04 %

tableau 5.4 : comparaison entre les teneurs prévues et effectives en fonction du nombre de données considérées dans le krigeage

Le biais conditionnel diminue lorsque la taille du voisinage croît, et devient négligeable à partir de 16 échantillons. Ainsi donc, bien que la sélection des blocs de minéral soit imparfaite à cause du recours à une estimation qui donne une vision déformée de la réalité (en l'occurrence trop lissée), la teneur moyenne effectivement récupérée est égale à celle prévue, dès lors que le biais conditionnel est négligeable, ce qui se produit lorsque le krigeage utilise un nombre suffisant d'échantillons.

Bénéfice conventionnel

Il est intéressant de compléter l'étude des courbes tonnage – teneur de coupure et teneur moyenne – teneur de coupure par celle du *bénéfice conventionnel* en fonction de la teneur de coupure. On définit, pour une teneur de coupure z , le bénéfice conventionnel par :

$$B(z) = [m(z) - z] \times T(z)$$

où $m(z)$ est la teneur moyenne au-dessus de z et $T(z)$ le tonnage associé.

Le bénéfice conventionnel est égal à la différence entre la quantité de métal que l'on récupère, mT , et la quantité minimale que l'on puisse récupérer, zT , qui, par définition de la teneur de coupure, compense les coûts de traitement du tonnage T . Il a donc la signification d'un gain économique.

Les courbes de bénéfice conventionnel présentent une hiérarchie intéressante, selon la grandeur à laquelle elles se rapportent.

Notons :

- $B(z)$: le bénéfice réel associé aux vraies valeurs ponctuelles ;
- $B_v(z)$: le bénéfice réel associé aux vraies valeurs des blocs de support v
- $B_v^*(z)$: le bénéfice estimé associé aux valeurs krigées des blocs de support v
- $B_v^{\text{eff}}(z)$: le bénéfice effectif associé aux vraies valeurs des blocs dont la valeur estimée est supérieure à z .

On a toujours :

$$B(z) \geq B_v(z) \geq B_v^{\text{eff}}(z).$$

La première inégalité est due à l'**effet de support** : certains blocs dont la teneur moyenne est supérieure à la teneur de coupure contiennent des parties "pauvres", qui sont rejetées dans le cas d'une sélection sur les valeurs ponctuelles ; inversement, certains blocs "pauvres" possèdent des morceaux "riches", lesquels sont sélectionnés si l'on considère les valeurs ponctuelles. Autrement dit, l'exploitation du gisement est d'autant moins *sélective* que la taille des unités de sélection est importante.

La seconde inégalité est une conséquence de l'**effet d'information** mentionné dans les paragraphes précédents. La sélection s'effectue en pratique à partir des valeurs estimées et non à partir des valeurs réelles (inconnues). Dès lors, il est inévitable que des blocs riches seront sous-estimés et envoyés au rebut, et des blocs pauvres surestimés et envoyés en laverie, ce qui se traduit par une baisse du bénéfice conventionnel par rapport au cas idéal où la sélection se réalise sur les valeurs réelles.

La figure 5.34 montre que, dans l'exemple des données de cuivre, les pertes de sélectivité due aux effets de support et d'information sont d'un ordre de grandeur comparable.

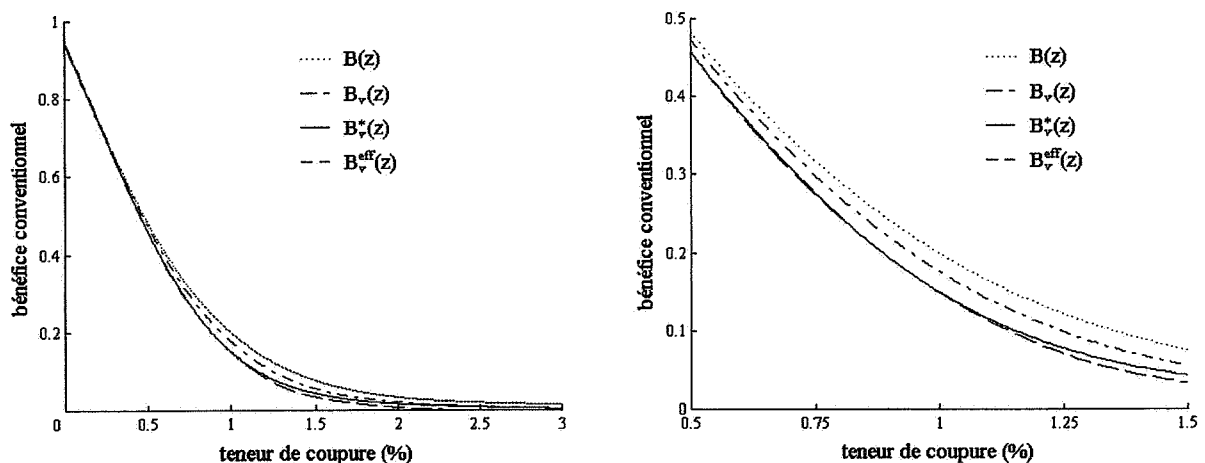


figure 5.34 : les différentes courbes de bénéfices conventionnels
à droite : restriction aux teneurs de coupure situées entre 0.5% et 1.5%

En l'absence de **biais conditionnel**, la courbe prédite par le krigage (B_v^*), calculée à partir des valeurs estimées des blocs, est très proche de la courbe effective (B_v^{eff}). Ainsi, bien que la sélection ne soit pas parfaite, l'absence de biais conditionnel permet une estimation correcte du bénéfice conventionnel, et de la teneur moyenne, des blocs qui seront sélectionnés.

VII. Complément : le krigeage aléatoire

Le krigeage aléatoire est utilisé lorsqu'il existe des **incertitudes dans la localisation des points de données**, problème fréquent en bathymétrie et en halieutique, ou lorsque l'échantillonnage est presque régulier (mais pas tout à fait). Dans ce second cas, les configurations de krigeage sont légèrement différentes d'un point à l'autre ; au lieu de résoudre un système de krigeage pour chacun des points à estimer, on peut préférer se référer à une "configuration moyenne" sur laquelle on calcule les pondérateurs (la contrepartie de cette approche est une perte de précision par rapport au krigeage classique).

Nous nous plaçons dans le cas où la fonction aléatoire étudiée est intrinsèque, de variogramme γ . On cherche à estimer la valeur inconnue $Z(\mathbf{x}_0)$ à l'aide d'une combinaison linéaire pondérée des données :

$$Z^*(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{X}_{\alpha})$$

où les $\mathbf{X}_{\alpha}$ désignent les positions des points de données, supposées aléatoires.

La condition d'autorisation est la même que celle du krigeage intrinsèque ordinaire :

$$\sum_{\alpha=1}^n \lambda_{\alpha} = 1.$$

La contrainte de non-biais est automatiquement satisfaite, car toute combinaison linéaire autorisée a une espérance nulle. La condition d'optimalité consiste donc à minimiser la variance d'estimation sous la contrainte que la somme des poids soit égale à 1.

La variance d'estimation s'écrit :

$$\sigma_{KA}^2(\mathbf{x}_0) = \text{var}[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = E \left\{ \left[\sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{X}_{\alpha}) - Z(\mathbf{x}_0) \right]^2 \right\}.$$

D'après la formule de l'espérance totale :

$$\sigma_{KA}^2(\mathbf{x}_0) = E \left\{ \sigma_{KA}^2(\mathbf{x}_0 | \mathbf{X}_1, \dots, \mathbf{X}_n) \right\}$$

$$\text{où } \sigma_{KA}^2(\mathbf{x}_0 | \mathbf{X}_1, \dots, \mathbf{X}_n) = E \left\{ \left[\sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{X}_{\alpha}) - Z(\mathbf{x}_0) \right]^2 | \mathbf{X}_1, \dots, \mathbf{X}_n \right\}.$$

$$\text{Or } \sigma_{KA}^2(\mathbf{x}_0 | \mathbf{X}_1 = \mathbf{x}_1, \dots, \mathbf{X}_n = \mathbf{x}_n) = 2 \sum_{\alpha=1}^n \lambda_{\alpha} \gamma(\mathbf{x}_0 - \mathbf{x}_{\alpha}) - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta})$$

d'où :

$$\sigma_{KA}^2(\mathbf{x}_0) = E \left\{ 2 \sum_{\alpha=1}^n \lambda_{\alpha} \gamma(\mathbf{x}_0 - \mathbf{X}_{\alpha}) - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \gamma(\mathbf{X}_{\alpha} - \mathbf{X}_{\beta}) \right\}.$$

Pour pouvoir calculer cette expression, il faut connaître la loi du vecteur aléatoire $\{\mathbf{X}_{\alpha}, \alpha = 1 \dots n\}$. Pour simplifier, nous supposons que chaque site $\mathbf{X}_{\alpha}$ est implanté uniformément dans un domaine v_{α} et indépendamment des autres sites. Cette hypothèse d'indépendance est souvent peu réaliste, car les erreurs de localisation sont en général corrélées.

Dans ce cas, on peut développer l'expression de la variance d'estimation :

$$\begin{aligned}
 \sigma_{KA}^2(\mathbf{x}_0) &= \int_{v_1} \dots \int_{v_n} \left[2 \sum_{\alpha=1}^n \lambda_{\alpha} \gamma(\mathbf{x}_0 - \mathbf{x}_{\alpha}) - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \right] \frac{d\mathbf{x}_1}{|V_1|} \dots \frac{d\mathbf{x}_n}{|V_n|} \\
 &= 2 \sum_{\alpha=1}^n \lambda_{\alpha} \underbrace{\frac{1}{|V_{\alpha}|} \int_{v_{\alpha}} \gamma(\mathbf{x}_0 - \mathbf{x}_{\alpha}) d\mathbf{x}_{\alpha}}_{\bar{\gamma}(\mathbf{x}_0, v_{\alpha})} - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \underbrace{\frac{1}{|V_{\alpha}|} \frac{1}{|V_{\beta}|} \int_{v_{\alpha}} \int_{v_{\beta}} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) d\mathbf{x}_{\alpha} d\mathbf{x}_{\beta}}_{\begin{matrix} 0 \text{ si } \alpha=\beta \\ \bar{\gamma}(v_{\alpha}, v_{\beta}) \text{ sinon} \end{matrix}} \\
 &= 2 \sum_{\alpha=1}^n \lambda_{\alpha} \bar{\gamma}(\mathbf{x}_0, v_{\alpha}) - \sum_{\alpha=1}^n \sum_{\beta=1, \beta \neq \alpha}^n \lambda_{\alpha} \lambda_{\beta} \bar{\gamma}(v_{\alpha}, v_{\beta})
 \end{aligned}$$

où $\bar{\gamma}(\mathbf{x}_0, v_{\alpha})$ représente la valeur moyenne de $\gamma(\mathbf{x}_0 - \mathbf{x})$ lorsque $\mathbf{x}$ parcourt v_{α} , et $\bar{\gamma}(v_{\alpha}, v_{\beta})$ la valeur moyenne de $\gamma(\mathbf{x} - \mathbf{y})$ lorsque $\mathbf{x}$ et $\mathbf{y}$ parcourent respectivement et indépendamment v_{α} et v_{β} .

Il reste à minimiser cette expression sous la contrainte d'autorisation. En introduisant un multiplicateur de Lagrange μ , on obtient le système :

$$\begin{cases} - \sum_{\beta=1, \beta \neq \alpha}^n \lambda_{\beta} \bar{\gamma}(v_{\alpha}, v_{\beta}) + \mu = -\bar{\gamma}(\mathbf{x}_0, v_{\alpha}) & \forall \alpha = 1 \dots n \\ \sum_{\beta=1}^n \lambda_{\beta} = 1 \end{cases}$$

Sous forme matricielle¹¹ :

$$\begin{pmatrix} 0 & \dots & -\bar{\gamma}(v_1, v_n) & 1 \\ \vdots & & \vdots & \vdots \\ -\bar{\gamma}(v_n, v_1) & \dots & 0 & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} -\bar{\gamma}(\mathbf{x}_0, v_1) \\ \vdots \\ -\bar{\gamma}(\mathbf{x}_0, v_n) \\ 1 \end{pmatrix}$$

La variance de krigeage, dans ce modèle particulier, est :

$$\sigma_{KA}^2(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} \bar{\gamma}(\mathbf{x}_0, v_{\alpha}) - \mu.$$

On retrouve les résultats du krigeage ponctuel classique lorsque v_{α} se réduit au point $\mathbf{x}_{\alpha}$.

Remarque : problèmes d'analyse variographique

La présence d'incertitudes dans la localisation des données altère le variogramme expérimental de $Z(\mathbf{x})$. En effet, la valeur associée au point $\mathbf{x}_{\alpha}$ a en réalité été prise en un autre point $\mathbf{X}_{\alpha} = \mathbf{x}_{\alpha} + \varepsilon_{\alpha}$, de sorte que la fonction aléatoire disponible n'est pas $Z(\mathbf{x})$ elle-même, mais

$$Z_{\varepsilon}(\mathbf{x}) = Z(\mathbf{x} + \varepsilon(\mathbf{x}))$$

où $\varepsilon(\mathbf{x})$ est une fonction aléatoire représentant l'erreur de localisation.

¹¹ Il ne faut pas confondre ce système de krigeage avec celui que l'on obtiendrait par krigeage de $Z(\mathbf{x}_0)$ à l'aide des valeurs moyennes $Z(v_{\alpha})$. La différence réside dans les termes diagonaux ($\alpha = \beta$) de la matrice du premier membre, qui sont nuls ici.

Nous supposons pour simplifier que ε est stationnaire ; en particulier la densité de probabilité d'un couple $(\varepsilon(\mathbf{x} + \mathbf{h}), \varepsilon(\mathbf{x}))$ ne dépendra que du vecteur $\mathbf{h}$.

Le variogramme de Z_ε peut s'exprimer en fonction de celui de Z :

$$\gamma_\varepsilon(\mathbf{h}) = \iint \frac{1}{2} E \{ [Z(\mathbf{x} + \mathbf{h} + \mathbf{u}) - Z(\mathbf{x} + \mathbf{v})]^2 \} f(\mathbf{u}, \mathbf{v}; \mathbf{h}) d\mathbf{u} d\mathbf{v}$$

où $f(\mathbf{u}, \mathbf{v}; \mathbf{h})$ est la densité bivariable du couple $(\varepsilon(\mathbf{x} + \mathbf{h}), \varepsilon(\mathbf{x}))$; elle ne dépend pas de $\mathbf{x}$ car ε est stationnaire.

Il vient :

$$\gamma_\varepsilon(\mathbf{h}) = \iint \gamma(\mathbf{h} + \mathbf{u} - \mathbf{v}) f(\mathbf{u}, \mathbf{v}; \mathbf{h}) d\mathbf{u} d\mathbf{v}.$$

Ainsi, même si γ est continu à l'origine, γ_ε présentera un effet de pépité d'amplitude

$$C_0 = \gamma_\varepsilon(\mathbf{0}^+) = \iint \gamma(\mathbf{u} - \mathbf{v}) f(\mathbf{u}, \mathbf{v}; \mathbf{0}^+) d\mathbf{u} d\mathbf{v}.$$

Cette constante pépitique n'est nulle que si les erreurs de localisation sont fortement corrélées aux petites distances (erreurs quasi égales en $\mathbf{x}$ et $\mathbf{x} + \mathbf{h}$ pour $|\mathbf{h}|$ infiniment petit) : $f(\mathbf{u}, \mathbf{v}; \mathbf{0}^+)$ est alors nulle, sauf si $\mathbf{u} = \mathbf{v}$ et, au moins aux petites distances, le variogramme n'est pas altéré. Réciproquement, γ_ε diffère d'autant plus de γ que les erreurs sont peu corrélées entre elles (*a fortiori* si elles sont indépendantes). La figure 5.35 illustre cette situation, pour des erreurs de localisation indépendantes et de loi uniforme sur un bloc de taille 2×2 ; le variogramme initial est composé d'un effet de pépité et d'un schéma sphérique de portée 10.

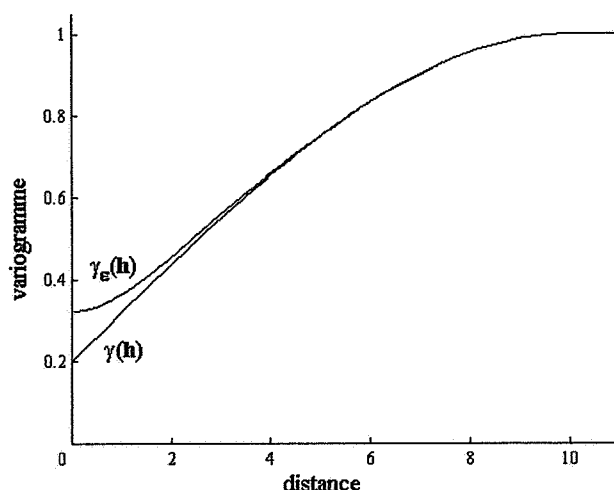


figure 5.35 : altération du variogramme due à des erreurs de localisation

Non seulement l'effet de pépité augmente (comme c'est le cas lorsqu'il y a des erreurs de mesure sur les valeurs), mais la forme du variogramme est modifiée : la partie continue passe d'un comportement linéaire à l'origine à un comportement parabolique. Pour retrouver le variogramme γ à partir de γ_ε , il sera nécessaire d'avoir identifié la loi de probabilité des incertitudes de localisation.

Résumé

Le krigeage permet d'estimer la valeur de la variable régionalisée en un point ou sur un bloc à l'aide d'une combinaison linéaire pondérée des données avoisinantes. Il permet aussi de mesurer la précision de la valeur estimée par l'intermédiaire d'une variance d'estimation. Sa mise en œuvre nécessite de spécifier un modèle de covariance ou de variogramme, ainsi qu'un *voisinage*, c'est-à-dire un domaine autour du point ou du bloc à estimer où sont prises les données qui servent à l'estimation.

S'il s'adapte à l'hypothèse de stationnarité choisie, le krigeage s'articule toujours autour des mêmes étapes. Plus précisément, l'erreur d'estimation, qui est aléatoire dans le modèle probabiliste, doit satisfaire quatre contraintes successives :

- **linéarité**
- **autorisation**
- **non-biais**
- **optimalité** (variance minimale).

Parmi les propriétés du krigeage, on retiendra l'interpolation exacte des données (en estimation ponctuelle) et la propriété de lissage, qui implique que les valeurs estimées sont moins dispersées que les valeurs vraies inconnues.

Chapitre 6 : les modèles non stationnaires

I. Introduction

La représentation d'une variable régionalisée par une fonction aléatoire stationnaire ou intrinsèque est une opération **subjective** qui incombe à l'utilisateur, au vu des données expérimentales, de l'information disponible sur le phénomène régionalisé, et de l'échelle de travail envisagée. La stationnarité indique une certaine homogénéité des caractéristiques de la variable dans l'espace ; elle est nécessaire pour mener à bien l'inférence statistique (calcul de la covariance ou du variogramme expérimental en vue de sa postérieure modélisation).

Les outils d'étude exploratoire, comme les nuages des valeurs prises par les données le long des directions de coordonnées, sont utiles pour décider de l'opportunité d'une hypothèse de stationnarité. L'analyse variographique est également un moyen de corroborer ou rejeter une telle hypothèse : *à l'échelle de travail*, selon que l'on considère l'hypothèse stationnaire ou intrinsèque, le variogramme expérimental doit se stabiliser autour d'un palier ou croître moins vite qu'une parabole. Parfois, l'hypothèse stationnaire est abandonnée pour des raisons physiques. Par exemple, en bathymétrie, on sait que la profondeur des fonds marins augmente lorsqu'on s'éloigne du rivage ; en géophysique, le gradient géothermique augmente quand on se rapproche du centre de la terre ; en environnement, la charge hydraulique diminue dans le sens de l'écoulement. De telles situations sont incompatibles avec la stationnarité, qui suppose que le comportement de la variable régionalisée est invariant dans l'espace.

Ce chapitre traite les problèmes d'**estimation locale** lorsque les modèles stationnaires sont insuffisants. Les problèmes d'estimation globale peuvent être résolus par les méthodes transitives exposées dans le premier chapitre de ce document, lesquelles ne requièrent aucune hypothèse sur la régionalisation étudiée.

Les modèles que nous allons développer ne seront valides qu'à l'échelle du voisinage glissant choisi pour l'estimation locale. Ces modèles travaillent sur des fonctions aléatoires plus générales ou plus complexes que les fonctions aléatoires stationnaires. Toutefois, leur principe consiste à extraire de ces fonctions aléatoires "quelque chose de stationnaire". Nous allons examiner deux approches permettant de traiter les processus non stationnaires. La première, connue sous le nom de **krigeage universel**, est basée sur une dichotomie de la régionalisation. Nous verrons que le problème posé est celui du bien-fondé de cette dichotomie. La seconde approche consiste à changer l'objet de l'étude et conduit au **krigeage intrinsèque d'ordre k** , qui généralise le passage de l'hypothèse stationnaire à l'hypothèse intrinsèque.

Il serait faux de croire que ces deux approches sont les seules méthodes permettant d'étudier des processus non stationnaires. Il existe d'autres techniques, comme par exemple le *krigeage transitif* (cf. chapitre 1), ou la méthode de la *dérive externe* qui sera présentée au chapitre 7 (géostatistique multivariable). Enfin, dans certains cas plus complexes, une modélisation *ad hoc* (ajustement d'une dérive calculée empiriquement) peut s'avérer une alternative plus performante que les méthodes standards de krigage universel ou intrinsèque d'ordre k .

II. Le krigage universel

II.1. La dichotomie du phénomène

II.1.1. Présentation du modèle

Si la variable régionalisée montre une tendance systématique, un modèle stationnaire ou intrinsèque est inadapté. Dans ce cas, l'approche la plus "naturelle" consiste à décomposer la fonction aléatoire $Z(x)$ en la somme de deux composantes :

- une **dérive** $m(x)$, supposée **déterministe**, qui décrit la "tendance" de la régionalisation
- un **résidu**¹ $Y(x)$ **aléatoire**

telles que

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

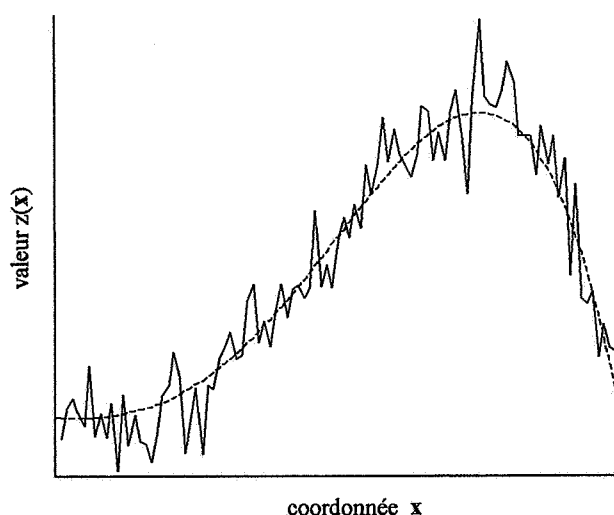


figure 6.1 : régionalisation raisonnablement non stationnaire
en pointillés est dessinée une dérive possible $m(x)$

¹ appelé aussi "résidu vrai". Tout ce qui concerne ce résidu est parfois qualifié de "sous-jacent".

II.1.2. Propriétés de la dérive

Représentation fonctionnelle de la dérive

La notion dérive n'est opératoire que si la fonction $m(\mathbf{x})$ varie régulièrement relativement à l'échelle de travail. Dans le cas contraire, on devrait la considérer comme la réalisation d'une nouvelle fonction aléatoire², et il serait impossible de séparer ce qui provient du résidu de ce qui provient de la dérive. On suppose la dérive suffisamment **régulière** pour pouvoir être approchée *localement* par une combinaison linéaire de fonctions des coordonnées f^ℓ (appelées dans la suite **fonctions de base**) :

$$m(\mathbf{x}) = \sum_{\ell} a_{\ell} f^{\ell}(\mathbf{x}).$$

La somme porte sur un nombre limité et identifiés d'indices. Les coefficients a_{ℓ} sont **déterministes**, mais inconnus, sinon l'hypothèse de dichotomie serait inutile puisque l'on pourrait alors travailler directement sur le résidu $Y(\mathbf{x}) = Z(\mathbf{x}) - m(\mathbf{x})$. Il arrive que le champ $\mathcal{D}$ soit trop étendu pour que la représentation de $m(\mathbf{x})$ sous la forme d'un développement en les fonctions f^{ℓ} soit admissible globalement. Il faut garder à l'esprit que cette décomposition n'est légitime que localement, i.e. à l'échelle d'un voisinage glissant, avec des coefficients a_{ℓ} qui peuvent différer d'un voisinage à l'autre.

Contraintes sur les fonctions de base

Des considérations mathématiques incitent à choisir les fonctions de base $f^{\ell}(\mathbf{x})$ parmi les exponentielles – polynômes³. Dans la plupart des cas, ces fonctions sont des puissances des coordonnées du point $\mathbf{x}$. Par exemple, dans $\mathbb{R}^2$, il s'agit des monômes $1, x_1, x_2, x_1^2, x_2^2, x_1 x_2 \dots$. On impose de plus que la première fonction de base, indexée par 0, soit constante et égale à 1 :

$$f^0(\mathbf{x}) \equiv 1 \quad \text{par convention.}$$

² Un modèle de dérive *aléatoire* est possible en théorie ; on pourrait même supposer cette dérive corrélée avec le résidu. Mais cette démarche est purement mathématique et difficile à mettre en pratique. L'hypothèse d'une dérive déterministe a pour but de ne pas compliquer inutilement le modèle (*principe d'économie*) et de rendre opératoire la dichotomie.

³ On souhaite que l'estimation *optimale* de la dérive soit indépendante du choix du repère des coordonnées, ce qui implique l'invariance par translation de l'espace vectoriel engendré par les fonctions de base. Seules les exponentielles (éventuellement complexes), les polynômes et les combinaisons linéaires de produits finis d'exponentielles et de polynômes, satisfont cette propriété (cf. annexe C). Dans certains cas toutefois, on transgresse cette règle et l'on modélise la dérive non pas par des exponentielles – polynômes, mais par une ou plusieurs variables auxiliaires connues partout, que l'on appelle *dérives externes* (cf. fin du chapitre 7).

II.1.3. Commentaires sur la dichotomie

Au niveau de la variable régionalisée

On s'attend que la dichotomie sépare la régionalisation en deux composantes différentes : la première, la "tendance", qui varie à grande échelle ; la seconde, dont les variations sont plus locales et plus erratiques, qui reflète les "anomalies" du phénomène régionalisé. Très souvent cependant, la décomposition $Z(\mathbf{x}) = Y(\mathbf{x}) + m(\mathbf{x})$ ne constitue qu'une commodité algébrique ; la dérive ne représente que la partie régulière de la variable régionalisée et n'a pas de signification physique particulière.

Au niveau de la fonction aléatoire

Pour pouvoir réaliser l'analyse variographique, des hypothèses de stationnarité seront nécessaires. Ces hypothèses vont naturellement porter sur le résidu $Y(\mathbf{x})$, la fonction $Z(\mathbf{x})$ étant non stationnaire. L'objectif de la dichotomie est d'ailleurs de faire apparaître "quelque chose" de stationnaire afin de permettre l'inférence statistique.

II.2. Krigeage universel ponctuel

II.2.1. Résidu stationnaire d'ordre deux

Dans ce modèle, on a la décomposition suivante :

$$Z(\mathbf{x}) = Y(\mathbf{x}) + m(\mathbf{x}) = Y(\mathbf{x}) + \sum_{\ell} a_{\ell} f^{\ell}(\mathbf{x})$$

où le résidu $Y(\mathbf{x})$ est stationnaire d'ordre deux. Plus précisément, on suppose :

- $E[Y(\mathbf{x})] = 0$ (l'éventuelle moyenne de Y peut être incluse dans le paramètre a_0 de la dérive, puisque $f^0(\mathbf{x}) \equiv 1$)
- $\text{cov}[Y(\mathbf{x} + \mathbf{h}), Y(\mathbf{x})] = \text{cov}[Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})] = C(\mathbf{h})$. Les fonctions aléatoires $Y(\mathbf{x})$ et $Z(\mathbf{x})$ admettent la même covariance car elles ne diffèrent que d'une quantité déterministe, invisible dans une covariance centrée. Nous reviendrons ultérieurement sur l'analyse variographique de $C(\mathbf{h})$.

La première hypothèse équivaut à : $E[Z(\mathbf{x})] = m(\mathbf{x})$. Cette écriture a le mérite de donner un sens précis à la notion de dérive : la dérive est l'espérance d'une fonction aléatoire non stationnaire, ce qui la distingue de la notion vague de *tendance*.

La stationnarité d'ordre deux du résidu implique que les moments d'ordre deux de $Z(\mathbf{x})$ sont invariants dans l'espace, mais pas le moment du premier ordre (l'espérance). Cela signifie que la variable régionalisée oscille avec une amplitude constante autour de la dérive $m(\mathbf{x})$, comme s'il existait une "force de rappel" régulatrice. Tous les phénomènes non stationnaires ne peuvent être représentés par un tel modèle, relativement exigeant.

Nous cherchons à estimer la valeur inconnue de Z en un point quelconque $\mathbf{x}_0$ à l'aide des données $\{ Z(\mathbf{x}_\alpha), \alpha = 1 \dots n \}$. Ecrivons le système de krigeage (baptisé **krigeage universel**) en $\mathbf{x}_0$:

- **contrainte de linéarité** : on pose $Z^*(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_\alpha Z(\mathbf{x}_\alpha)$ comme estimateur de $Z(\mathbf{x}_0)$;
- **contrainte d'autorisation** : cette contrainte a pour objectif de garantir l'existence des deux premiers moments (espérance et variance) de l'erreur d'estimation. Ici, elle est automatiquement satisfaite puisque pour tout point $\mathbf{x}$, $Z(\mathbf{x})$ admet une espérance et une variance finies (l'existence des moments de Z ne dépend pas de la présence d'une dérive *déterministe*) ; il en est de même de l'erreur d'estimation, combinaison linéaire sur Z .
- **contrainte de non-biais** : l'espérance de l'erreur d'estimation s'écrit :

$$\begin{aligned} E[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] &= \sum_{\alpha=1}^n \lambda_\alpha \sum_{\ell} a_{\ell} f^{\ell}(\mathbf{x}_{\alpha}) - \sum_{\ell} a_{\ell} f^{\ell}(\mathbf{x}_0) \\ &= \sum_{\ell} a_{\ell} \left[\sum_{\alpha=1}^n \lambda_{\alpha} f^{\ell}(\mathbf{x}_{\alpha}) - f^{\ell}(\mathbf{x}_0) \right] \end{aligned}$$

Les coefficients a_{ℓ} étant inconnus, on doit poser

$$\sum_{\alpha=1}^n \lambda_{\alpha} f^{\ell}(\mathbf{x}_{\alpha}) = f^{\ell}(\mathbf{x}_0) \quad \forall \ell$$

pour s'assurer que l'erreur d'estimation est d'espérance nulle. Ceci fournit un nombre de contraintes de non biais égal au nombre de fonctions de base. En particulier, pour $\ell = 0$, comme $f^0(\mathbf{x}) \equiv 1$, on retrouve la condition "classique"

$$\sum_{\alpha=1}^n \lambda_{\alpha} = 1.$$

- **contrainte d'optimalité** : la variance de l'erreur d'estimation se développe en :

$$\text{var}[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = C(\mathbf{0}) + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{x}_{\alpha} - \mathbf{x}_0).$$

Minimiser cette expression sous les contraintes précédentes requiert d'introduire des multiplicateurs de Lagrange μ_{ℓ} en nombre égal au nombre de contraintes de non biais, et d'exprimer la nullité des dérivées partielles par rapport aux λ_{α} et aux μ_{ℓ} de la quantité :

$$C(\mathbf{0}) + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{x}_{\alpha} - \mathbf{x}_0) + 2 \sum_{\ell} \mu_{\ell} \left[\sum_{\alpha=1}^n \lambda_{\alpha} f^{\ell}(\mathbf{x}_{\alpha}) - f^{\ell}(\mathbf{x}_0) \right].$$

L'annulation des dérivées par rapport aux μ_{ℓ} redonne les conditions de non-biais. Les dérivées par rapport aux λ_{α} fournissent les conditions d'optimalité. On obtient le système linéaire final :

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + \sum_{\ell} \mu_{\ell} f^{\ell}(\mathbf{x}_{\alpha}) = C(\mathbf{x}_{\alpha} - \mathbf{x}_0) & \forall \alpha = 1 \dots n \\ \sum_{\beta=1}^n \lambda_{\beta} f^{\ell}(\mathbf{x}_{\beta}) = f^{\ell}(\mathbf{x}_0) & \forall \ell \end{cases}$$

soit matriciellement :

$$\begin{pmatrix} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) & f^{\ell}(\mathbf{x}_{\alpha}) \\ f^{\ell}(\mathbf{x}_{\beta}) & 0 \end{pmatrix} \begin{pmatrix} \lambda_{\beta} \\ \mu_{\ell} \end{pmatrix} = \begin{pmatrix} C(\mathbf{x}_0 - \mathbf{x}_{\alpha}) \\ f^{\ell}(\mathbf{x}_0) \end{pmatrix}.$$

On observe que le point à estimer $\mathbf{x}_0$ n'intervient que dans le second membre. Ce système matriciel apparaît comme une généralisation du système de krigeage ordinaire, lequel ne comporte que la première fonction de base ($f^0(\mathbf{x}) \equiv 1$) et un seul multiplicateur de Lagrange μ .

Tous calculs faits, la variance de krigeage universel vaut :

$$\sigma_{\text{KU}}^2(\mathbf{x}_0) = C(\mathbf{0}) - \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{x}_{\alpha} - \mathbf{x}_0) - \sum_{\ell} \mu_{\ell} f^{\ell}(\mathbf{x}_0).$$

II.2.2. Résidu intrinsèque

Par hypothèse, on a toujours la dichotomie

$$Z(\mathbf{x}) = Y(\mathbf{x}) + m(\mathbf{x}) = Y(\mathbf{x}) + \sum_{\ell} a_{\ell} f^{\ell}(\mathbf{x})$$

mais cette fois, on suppose que le résidu $Y(\mathbf{x})$ est une fonction aléatoire **intrinsèque** :

- $E[Y(\mathbf{x} + \mathbf{h}) - Y(\mathbf{x})] = 0$;
- $\text{var}[Y(\mathbf{x} + \mathbf{h}) - Y(\mathbf{x})] = \text{var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = 2\gamma(\mathbf{h})$.

Dans ce modèle, la variance de $Y(\mathbf{x})$ ou $Z(\mathbf{x})$ n'existe pas forcément (ce qui indique que la régionalisation peut s'écarter infiniment de la dérive ; il n'y a plus de force de rappel régulatrice). Au niveau mathématique, il est impératif de travailler sur des combinaisons linéaires autorisées, c'est-à-dire telles que la somme des poids est nulle, pour pouvoir écrire des espérances et des variances, aussi bien pour $Y(\mathbf{x})$ que pour $Z(\mathbf{x})$.

Examinons les quatre étapes du krigeage ponctuel en $\mathbf{x}_0$:

- **linéarité** : $Z^*(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha})$.
- **autorisation** : l'erreur d'estimation doit être de poids total nul, soit : $\sum_{\alpha=1}^n \lambda_{\alpha} - 1 = 0$.
- **non-biais** : cette condition s'écrit

$$\begin{aligned}
 0 &= E[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] \\
 &= E\left[\underbrace{\sum_{\alpha=1}^n \lambda_{\alpha} Y(\mathbf{x}_{\alpha}) - Y(\mathbf{x}_0)}_{=0, \text{ car toute combinaison linéaire autorisée de } Y \text{ est d'espérance nulle}}\right] + \sum_{\alpha=1}^n \lambda_{\alpha} \sum_{\ell} a_{\ell} f^{\ell}(\mathbf{x}_{\alpha}) - \sum_{\ell} a_{\ell} f^{\ell}(\mathbf{x}_0) \\
 &= \sum_{\ell} a_{\ell} \left[\sum_{\alpha=1}^n \lambda_{\alpha} f^{\ell}(\mathbf{x}_{\alpha}) - f^{\ell}(\mathbf{x}_0) \right] \\
 &= \sum_{\ell \geq 1} a_{\ell} \left[\sum_{\alpha=1}^n \lambda_{\alpha} f^{\ell}(\mathbf{x}_{\alpha}) - f^{\ell}(\mathbf{x}_0) \right] \\
 &\quad \text{car la condition d'autorisation s'écrit encore } \sum_{\alpha=1}^n \lambda_{\alpha} f^0(\mathbf{x}_{\alpha}) - f^0(\mathbf{x}_0) = 0.
 \end{aligned}$$

Les coefficients a_{ℓ} étant inconnus, la contrainte de non-biais est assurée lorsque

$$\sum_{\alpha=1}^n \lambda_{\alpha} f^{\ell}(\mathbf{x}_{\alpha}) = f^{\ell}(\mathbf{x}_0) \text{ pour } \ell \geq 1.$$

- **optimalité** : il reste à minimiser la variance de l'erreur d'estimation sous les contraintes précédentes, qui sont les mêmes que celles obtenues dans le cas d'un résidu stationnaire. Par suite, d'après la règle de calcul des variances sous l'hypothèse intrinsèque, le système de krigeage sera formellement le même que précédemment, à condition de remplacer la covariance C par l'opposé du variogramme $-\gamma$:

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) - \sum_{\ell} \mu_{\ell} f^{\ell}(\mathbf{x}_{\alpha}) = \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_0) & \forall \alpha = 1 \dots n \\ \sum_{\beta=1}^n \lambda_{\beta} = 1 \\ \sum_{\beta=1}^n \lambda_{\beta} f^{\ell}(\mathbf{x}_{\beta}) = f^{\ell}(\mathbf{x}_0) & \forall \ell \geq 1 \end{cases}$$

Nous avons distingué la condition

$$\sum_{\beta=1}^n \lambda_{\beta} = 1$$

des autres

$$\sum_{\beta=1}^n \lambda_{\beta} f^{\ell}(\mathbf{x}_{\beta}) = f^{\ell}(\mathbf{x}_0) \quad (\ell \geq 1)$$

car la première est une condition d'autorisation, tandis que les autres sont des conditions de non-biais.

En écriture matricielle, le système devient :

$$\begin{pmatrix} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) & f^{\ell}(\mathbf{x}_{\alpha}) \\ f^{\ell}(\mathbf{x}_{\beta}) & 0 \end{pmatrix} \begin{pmatrix} \lambda_{\beta} \\ -\mu_{\ell} \end{pmatrix} = \begin{pmatrix} \gamma(\mathbf{x}_0 - \mathbf{x}_{\alpha}) \\ f^{\ell}(\mathbf{x}_0) \end{pmatrix}.$$

La variance de krigeage s'écrit, après simplifications :

$$\sigma_{\text{KU}}^2(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_0) - \sum_{\ell} \mu_{\ell} f^{\ell}(\mathbf{x}_0).$$

II.2.3. Propriétés du krigeage universel ponctuel

On retrouve les propriétés générales du krigeage ponctuel, à savoir :

- **interpolation exacte** des données ;
- **additivité** ;
- **propriété de lissage** : les valeurs krigées sont moins dispersées que les valeurs vraies ;
- l'estimation est inchangée si on multiplie la covariance ou le variogramme par une constante $\kappa > 0$; par contre, la variance d'estimation est multipliée par κ ;
- **propriété d'orthogonalité**

Les conditions d'optimalité

$$\sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + \sum_{\ell} \mu_{\ell} f^{\ell}(\mathbf{x}_{\alpha}) = C(\mathbf{x}_{\alpha} - \mathbf{x}_0) \quad \forall \alpha = 1 \dots n$$

peuvent encore s'écrire

$$\text{cov} \left[\sum_{\beta=1}^n \lambda_{\beta} Z(\mathbf{x}_{\beta}) - Z(\mathbf{x}_0), Z(\mathbf{x}_{\alpha}) \right] = - \sum_{\ell} \mu_{\ell} f^{\ell}(\mathbf{x}_{\alpha}) \quad \forall \alpha = 1 \dots n.$$

Cette égalité implique que l'erreur de krigeage ponctuel est non corrélée aux combinaisons linéaires des données

$$\sum_{\alpha=1}^n \omega_{\alpha} Z(\mathbf{x}_{\alpha})$$

telles que $\sum_{\alpha=1}^n \omega_{\alpha} f'(\mathbf{x}_{\alpha}) = 0$ pour tout ℓ :

$$\text{cov}\left[\sum_{\beta=1}^n \lambda_{\beta} Z(\mathbf{x}_{\beta}) - Z(\mathbf{x}_0), \sum_{\alpha=1}^n \omega_{\alpha} Z(\mathbf{x}_{\alpha})\right] = -\sum_{\ell} \mu_{\ell} \left[\sum_{\alpha=1}^n \omega_{\alpha} f'(\mathbf{x}_{\alpha})\right] = 0.$$

Lorsque la seule fonction de base est la constante $f^0 \equiv 1$, le krigeage universel s'identifie au krigeage ordinaire, et on retrouve bien la propriété d'orthogonalité telle qu'elle a été énoncée pour le krigeage ordinaire.

• régularité du système de krigeage

Deux conditions sont nécessaires et suffisantes pour garantir l'existence et l'unicité de la solution :

- condition dite *d'indépendance linéaire des fonctions de base sur les données*, c'est-à-dire :

$$\sum_{\ell} c_{\ell} f'(\mathbf{x}_{\alpha}) = 0 \quad \forall \alpha \Rightarrow c_{\ell} = 0 \quad \forall \ell.$$

Ainsi, il existe des configurations dégénérées : par exemple, si tous les points de données sont alignés dans l'espace géographique à deux dimensions, ou coplanaires dans l'espace à trois dimensions, le système devient singulier. Par ailleurs, il est nécessaire que le nombre de données soit au moins égal au nombre de fonctions de base, pour satisfaire la condition d'indépendance linéaire de ces fonctions sur les données.

- la sous-matrice $\mathbf{C} = [C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta})]_{\alpha, \beta=1 \dots n}$ ou $-\Gamma = [-\gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta})]_{\alpha, \beta=1 \dots n}$, selon que le résidu est stationnaire ou intrinsèque, est positive *conditionnelle* stricte, i.e. :

pour toute combinaison linéaire $(\lambda_{\alpha})_{\alpha}$ qui annule les fonctions de base sur les données, c'est-à-dire qui vérifie $\sum_{\alpha} \lambda_{\alpha} f'(\mathbf{x}_{\alpha}) = 0 \quad \forall \ell$, on a l'inégalité

$$\sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \geq 0 \quad \text{ou} \quad -\sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \geq 0$$

avec la nullité si et seulement si $(\lambda_{\alpha})_{\alpha} \equiv 0$.

II.3. Krigage universel non ponctuel

Examinons le problème de l'estimation d'une fonction linéaire de la variable régionalisée (valeur moyenne sur un bloc, moyenne mobile pondérée, gradient...) à partir des données ponctuelles. On peut certes toujours se ramener à des estimations ponctuelles, en vertu de la propriété d'additivité du krigage universel. Mais comme dans le cas du krigage simple ou ordinaire, il est possible d'écrire un système unique qui évite la résolution de nombreux systèmes de krigage ponctuel y qui fournit une variance d'estimation.

Ainsi, si l'on souhaite estimer, à partir des valeurs expérimentales ponctuelles $Z(\mathbf{x}_\alpha)$, une moyenne mobile pondérée du type

$$\Xi_0 = \int Z(\mathbf{x}_0 + \mathbf{h}) p(\mathbf{h}) d\mathbf{h}$$

où p est une fonction de pondération de poids total unité ($\int p(\mathbf{h}) d\mathbf{h} = 1$), on posera :

$$\Xi_0^* = \sum_{\alpha=1}^n \lambda_\alpha Z(\mathbf{x}_\alpha).$$

Considérons le cas plus général où le résidu $Y(\mathbf{x})$ admet un variogramme γ . Le système de krigage universel devient :

$$\begin{cases} \sum_{\beta=1}^n \lambda_\beta \gamma(\mathbf{x}_\alpha - \mathbf{x}_\beta) + \sum_{\ell} \mu_\ell f'(\mathbf{x}_\alpha) = \int \gamma(\mathbf{x}_\alpha - \mathbf{x}_0 - \mathbf{h}) p(\mathbf{h}) d\mathbf{h} & \forall \alpha = 1 \dots n \\ \sum_{\alpha=1}^n \lambda_\alpha f'(\mathbf{x}_\alpha) = \int f'(\mathbf{x}_0 + \mathbf{h}) p(\mathbf{h}) d\mathbf{h} & \forall \ell \end{cases}$$

et ne diffère du système de krigage ponctuel que par le second membre.

Quant à la variance de krigage, elle vaut :

$$\sigma_{\text{KU}}^2(\mathbf{x}_0) = - \iint \gamma(\mathbf{h} - \mathbf{k}) p(\mathbf{h}) p(\mathbf{k}) d\mathbf{h} d\mathbf{k} + \sum_{\alpha=1}^n \lambda_\alpha \int \gamma(\mathbf{x}_\alpha - \mathbf{x}_0 - \mathbf{h}) p(\mathbf{h}) d\mathbf{h} - \sum_{\ell} \mu_\ell \int f'(\mathbf{x}_0 + \mathbf{h}) p(\mathbf{h}) d\mathbf{h}.$$

Par exemple, pour l'estimation de la valeur moyenne d'un bloc de support V , p est l'indicatrice de V divisée par la mesure de V . Les équations de krigage deviennent :

$$\begin{cases} \sum_{\beta=1}^n \lambda_\beta \gamma(\mathbf{x}_\alpha - \mathbf{x}_\beta) + \sum_{\ell} \mu_\ell f'(\mathbf{x}_\alpha) = \frac{1}{|V|} \int_V \gamma(\mathbf{x}_\alpha - \mathbf{x}) d\mathbf{x} = \bar{\gamma}(\mathbf{x}_\alpha, V) & \forall \alpha = 1 \dots n \\ \sum_{\alpha=1}^n \lambda_\alpha f'(\mathbf{x}_\alpha) = \frac{1}{|V|} \int_V f'(\mathbf{x}) d\mathbf{x} & \forall \ell \end{cases}$$

et la variance de krigage universel par bloc est :

$$\begin{aligned} \sigma_{\text{KU}}^2(V) &= - \frac{1}{|V|^2} \int_V \int_V \gamma(\mathbf{x} - \mathbf{y}) d\mathbf{x} d\mathbf{y} + \frac{1}{|V|} \sum_{\alpha=1}^n \lambda_\alpha \int_V \gamma(\mathbf{x}_\alpha - \mathbf{x}) d\mathbf{x} - \frac{1}{|V|} \sum_{\ell} \mu_\ell \int_V f'(\mathbf{x}) d\mathbf{x} \\ &= - \bar{\gamma}(V, V) + \sum_{\alpha=1}^n \lambda_\alpha \bar{\gamma}(\mathbf{x}_\alpha, V) - \frac{1}{|V|} \sum_{\ell} \mu_\ell \int_V f'(\mathbf{x}) d\mathbf{x} \end{aligned}$$

II.4. Evaluation de la dérive

Telle qu'elle a été définie, la dérive est un *paramètre* du modèle, déterministe, et non une grandeur régionale que l'on pourrait calculer si l'on connaissait exhaustivement la réalité. C'est pourquoi le mot "évaluation" est plus adéquat que "estimation". L'évaluation de la dérive permet de représenter la "tendance" de la variable, mais elle a principalement pour but d'estimer le résidu $Y(\mathbf{x})$ aux points de données, en vue de réaliser son analyse variographique. Il est possible en théorie de rechercher un estimateur *optimal* de $m(\mathbf{x})$, à l'instar de ce qui se fait en krigeage stationnaire ordinaire pour l'évaluation optimale de la moyenne inconnue⁴. Toutefois, cette évaluation *optimale* fait appel à la covariance ou au variogramme du résidu, dont la connaissance est précisément le but cherché. En pratique, on doit se contenter d'estimateurs non optimaux, obtenus par des méthodes de moindres carrés ou de lissage.

II.4.1. Méthode des moindres carrés

On suppose connues les fonctions de base f^ℓ qui composent la dérive. La méthode des moindres carrés consiste à rechercher un estimateur de $m(\mathbf{x})$ de la forme

$$\hat{m}(\mathbf{x}) = \sum_{\ell} \hat{a}_{\ell} f^{\ell}(\mathbf{x})$$

avec les coefficients $\hat{a}_{\ell}$ qui minimisent la somme des écarts quadratiques

$$Q(\alpha_{\ell}) = \sum_{\alpha} [z(\mathbf{x}_{\alpha}) - \sum_{\ell} \alpha_{\ell} f^{\ell}(\mathbf{x}_{\alpha})]^2$$

où les $\mathbf{x}_{\alpha}$ désignent les points échantillonnés situés dans un voisinage du point $\mathbf{x}$ considéré.

En annulant les dérivées partielles par rapport aux α_{ℓ} , on en déduit que les coefficients $\hat{a}_{\ell}$ vérifient le système d'équations :

$$\sum_{\ell} \hat{a}_{\ell} \left[\sum_{\alpha} f^{\ell}(\mathbf{x}_{\alpha}) f^m(\mathbf{x}_{\alpha}) \right] = \sum_{\alpha} f^m(\mathbf{x}_{\alpha}) z(\mathbf{x}_{\alpha}) \quad \forall m.$$

Il s'agit d'un système linéaire, que l'on suppose régulier. Les coefficients $\hat{a}_{\ell}$ s'expriment alors linéairement comme combinaisons linéaires des données :

$$\hat{a}_{\ell} = \sum_{\alpha} \lambda_{\alpha}^{\ell} z(\mathbf{x}_{\alpha}).$$

⁴ L'examen de l'évaluation optimale de la dérive et de ses implications théoriques est présenté en annexe C. Son principal intérêt est de faire apparaître certains problèmes théoriques liés à la dichotomie de la régionalisation.

Au niveau du modèle probabiliste, les coefficients deviennent des variables aléatoires

$$\hat{A}_\ell = \sum_{\alpha} \lambda_{\alpha}^{\ell} Z(\mathbf{x}_{\alpha}).$$

On établit que, dans le cas où $Z(\mathbf{x})$ admet une fonction de covariance⁵, les estimateurs $\hat{A}_\ell$ sont sans biais, c'est-à-dire

$$\forall \ell, E[\hat{A}_\ell] = a_\ell.$$

Ainsi donc, la méthode des moindres carrés fournit un estimateur **sans biais** de la dérive. En revanche, il n'est en général pas optimal, i.e. la variance de l'erreur n'est pas minimale. On démontre cependant qu'il est optimal dans le cas particulier où le résidu est purement pépétique ; en pratique, il suffit pour que l'estimateur des moindres carrés soit optimal que la covariance existe et soit de portée faible relativement aux distances séparant les points de données (ce qui n'est pas avantageux pour une étude géostatistique...).

L'intérêt de la méthode des moindres carrés est qu'elle ne nécessite pas la connaissance de la fonction de covariance ou du variogramme du résidu. Cette méthode s'apparente aux techniques dites de *trend surface analysis*, dont le but est d'interpoler des données spatialisées à l'aide de polynômes.

II.4.2. Evaluation par moyenne mobile

L'idée directrice de la méthode d'évaluation par moyenne mobile est la suivante : dès lors que l'on admet la dichotomie $Z(\mathbf{x}) = Y(\mathbf{x}) + m(\mathbf{x})$, on peut considérer la moyenne de Z sur un voisinage de $\mathbf{x}$ (à la fois petit en comparaison de l'échelle de travail et contenant suffisamment de données) comme une estimation de la dérive au point $\mathbf{x}$. La dérive représente en effet l'espérance de Z , dont il semble naturel qu'elle puisse être approchée par une moyenne locale.

En pratique, on estime la dérive en $\mathbf{x}$ par la moyenne arithmétique des observations situées dans un voisinage de $\mathbf{x}$, soit :

$$\hat{M}(\mathbf{x}) = \frac{1}{v} \sum_{\alpha=1}^v Z(\mathbf{x}_{\alpha}).$$

Un inconvénient de l'estimateur ainsi construit est qu'il donne un poids égal à toutes les données du voisinage de $\mathbf{x}$, et donc ne prend pas en compte les irrégularités d'échantillonnage.

⁵ Dans le cas où il n'existe qu'un variogramme (résidu intrinsèque strict), la variance du premier coefficient $\hat{A}_0$ est infinie et son espérance peut ne pas exister. Le terme constant de la dérive pose alors un problème d'indétermination (pour plus de détail sur cette indétermination, on pourra se reporter, en annexe C, au paragraphe sur l'estimation optimale de chacun des coefficients de la dérive).

D'autres difficultés se posent du point de vue théorique. Si le résidu Y est stationnaire, $\hat{M}(x)$ est un estimateur sans biais de $m(x)$, à condition toutefois que les fonctions de base composant $m(x)$ varient peu à l'échelle du voisinage utilisé pour le calcul de la moyenne mobile. En revanche, lorsque Y est strictement intrinsèque, l'erreur d'estimation $\hat{M}(x) - m(x)$ n'est pas une combinaison linéaire autorisée puisque son poids total est égal à 1. Dans ce cas, on ne peut parler ni de biais ni de variance d'estimation.

Plus grave : on introduit une échelle privilégiée dans l'estimation, en l'occurrence la taille du voisinage sur lequel on effectue la moyenne mobile. Cette échelle ne correspond à aucune caractéristique propre du phénomène. Ainsi, l'évaluation de la dérive inconnue par une moyenne mobile n'est pas une opération anodine, et peut poser problème lors de l'analyse variographique du résidu car elle risque d'introduire des artefacts dans les variogrammes expérimentaux.

II.4.3. Les splines de lissage

Le mot anglais "spline" fait référence à une règle flexible utilisée par les dessinateurs et les cartographes pour matérialiser des lignes à courbure variable, passant par des points fixés *a priori* ou à proximité de ceux-ci, de telle manière que ces lignes soient suffisamment lisses. Par analogie, en mathématiques, ce terme désigne une famille de fonctions présentant des propriétés de régularité. On en distingue deux types : les fonctions d'*interpolation* qui passent exactement par les points d'observation et les fonctions de *lissage* qui passent "au plus près" de ces points. Ces dernières étant très régulières, on peut penser les utiliser comme une estimation de la dérive.

Prenons l'exemple, dans l'espace à deux dimensions, des splines "laplaciennes" de lissage. En tout point $x = (x_1, x_2)$, la fonction spline laplacienne est définie par :

$$\hat{m}(x) = a_0 + a_1 x_1 + a_2 x_2 + \sum_{\alpha=1}^n b_{\alpha} K(x - x_{\alpha})$$

où x_{α} ($\alpha = 1 \dots n$) sont les points échantillonnés utilisés dans l'estimation
 $K(h) = |h|^2 \ln(|h|)$

Les coefficients $a_0, a_1, a_2, b_1, \dots, b_n$ sont solutions du système d'équations :

$$\begin{cases} \sum_{\beta} b_{\beta} K(x_{\alpha} - x_{\beta}) + \rho b_{\alpha} + a_0 + a_1 x_{\alpha 1} + a_2 x_{\alpha 2} = z(x_{\alpha}) & \forall \alpha \\ \sum_{\beta} b_{\beta} = \sum_{\beta} b_{\beta} x_{\beta 1} = \sum_{\beta} b_{\beta} x_{\beta 2} = 0 \end{cases}$$

où ρ est un coefficient strictement positif (si $\rho = 0$, il y a interpolation exacte, ce que l'on ne souhaite pas).

On démontre que $\hat{m}$ est deux fois dérivable et est solution du problème de minimisation

$$\min_{\varphi} \left[\sum_{\alpha=1}^n [z(\mathbf{x}_{\alpha}) - \varphi(\mathbf{x}_{\alpha})]^2 + \rho \iint \left\{ \left(\frac{\partial^2 \varphi}{\partial x_1^2} \right)^2 + 2 \left(\frac{\partial^2 \varphi}{\partial x_1 \partial x_2} \right)^2 + \left(\frac{\partial^2 \varphi}{\partial x_2^2} \right)^2 \right\} dx_1 dx_2 \right]$$

où φ décrit l'ensemble des fonctions deux fois dérivables. $\hat{m}$ réalise un compromis entre un bon ajustement des données (mesuré par le premier terme) et la régularité de la fonction (mesuré par le second). Le coefficient ρ mesure ce compromis ; le lissage est d'autant plus marqué que ρ est élevé.

II.5. L'analyse variographique

La difficulté du krigeage universel réside dans l'analyse variographique, c'est-à-dire l'estimation et la modélisation du variogramme de $Z(\mathbf{x})$ ou de $Y(\mathbf{x})$, à partir des données expérimentales.

II.5.1. Analyse directe sur les données brutes

L'inférence statistique n'est en général pas possible à partir des données relatives à la variable brute $Z(\mathbf{x})$, à cause de la présence de la dérive. En effet, si l'on applique l'estimateur classique du variogramme à ces données, on obtient :

$$E[\hat{\gamma}(\mathbf{h})] = \frac{1}{2} E \{ [Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]^2 \} = \gamma(\mathbf{h}) + \frac{1}{2} [m(\mathbf{x} + \mathbf{h}) - m(\mathbf{x})]^2 \geq \gamma(\mathbf{h}).$$

Ainsi, le variogramme expérimental calculé à partir des données brutes est **biaisé** par rapport au véritable variogramme ; son espérance apparaît comme la somme du variogramme qui nous intéresse et d'un terme positif lié à la dérive. En quelque sorte, la dérive masque le variogramme vrai. Le seul cas où l'inférence est possible sans biais est celui où l'on sait que la dérive est constante selon une direction de l'espace : pour tout vecteur $\mathbf{h}$ orienté selon cette direction, $m(\mathbf{x} + \mathbf{h}) \approx m(\mathbf{x})$, et l'inférence de $\gamma(\mathbf{h})$ est envisageable. C'est par exemple le cas de la profondeur des fonds marins, où la dérive peut être considérée comme invariante dans la direction parallèle à la côte. On peut alors identifier le variogramme de $Z(\mathbf{x})$ dans cette direction, puis entreprendre une modélisation en supposant par exemple ce variogramme isotrope.

On conçoit cependant que cette démarche n'est pas applicable à tous les jeux de données. La dérive est inconnue, et il est difficile de déterminer des directions pour lesquelles elle est constante (si de telles directions existent...). Il convient de rechercher d'autres méthodes pour estimer $\gamma(\mathbf{h})$.

II.5.2. Analyse variographique du résidu estimé

Pour réaliser l'inférence statistique, une autre possibilité est de considérer le résidu $Y(\mathbf{x})$, qui possède le même variogramme que $Z(\mathbf{x})$ et vérifie de surcroît l'hypothèse stationnaire ou intrinsèque. Or ce résidu n'est pas directement accessible puisque les données disponibles portent sur la variable brute $Z(\mathbf{x})$.

On est alors contraint de recourir à une *estimation* de $Y(\mathbf{x})$, laquelle servira de base à l'analyse variographique. En pratique, il est plus facile d'estimer la dérive $m(\mathbf{x})$, puis d'en déduire une estimation du résidu aux sites de données $\{ \mathbf{x}_\alpha, \alpha = 1 \dots n \}$:

$$\hat{Y}(\mathbf{x}_\alpha) = Z(\mathbf{x}_\alpha) - \hat{M}(\mathbf{x}_\alpha).$$

Malgré sa simplicité apparente, cette méthode soulève deux difficultés.

La première est que pour réaliser l'estimation *optimale* de la dérive (donc du résidu), on a besoin de la fonction structurale (covariance ou variogramme) de Z , dont la détermination est précisément le but recherché. Pour sortir de cette impasse, on doit se contenter d'estimateurs non optimaux de la dérive, comme les moindres carrés, les splines de lissage ou l'évaluation par moyenne mobile.

Le deuxième problème, plus fondamental, concerne l'analyse variographique du résidu estimé. Il a en effet été démontré que, *même si l'on dispose de l'estimateur optimal de la dérive*, le variogramme théorique du résidu estimé $\gamma_R(\mathbf{x}, \mathbf{y})$ présente un **biais** par rapport au variogramme du résidu vrai⁶. De plus, il n'est en général pas stationnaire (c'est-à-dire qu'il dépend de l'implantation des points $\mathbf{x}$ et $\mathbf{y}$, et pas seulement de leur séparation) et présente toujours une portée très courte qui fait croire à tort que la régionalisation se réduit à la superposition d'une dérive et d'un résidu purement pépétique. Le biais ne provient pas de la qualité de la dérive estimée, mais du fait que l'on a recours à une estimation.

La figure 6.2 illustre cette situation, dans le cas unidimensionnel, avec un variogramme $\gamma(\mathbf{h})$ linéaire avec effet de pépite. Dans cet exemple, le variogramme moyen du résidu estimé

$$\bar{\gamma}_R(\mathbf{h}) = \frac{1}{|\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}|} \int_{\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}} \gamma_R(\mathbf{x}, \mathbf{x} + \mathbf{h}) d\mathbf{x}$$

présente un palier et un léger effet de trou, alors qu'il n'en est rien du variogramme vrai. Seul subsiste le comportement à l'origine au niveau du variogramme du résidu estimé.

⁶ Plus précisément, le variogramme du résidu estimé est toujours inférieur au variogramme vrai (cf. annexe C) : le variogramme expérimental que l'on obtient a par conséquent tendance à sous-estimer la variabilité réelle du phénomène. Cette propriété subsiste lorsqu'on estime la dérive (donc le résidu) par un procédé quelconque, non optimal, comme par exemple une méthode de moindres carrés.

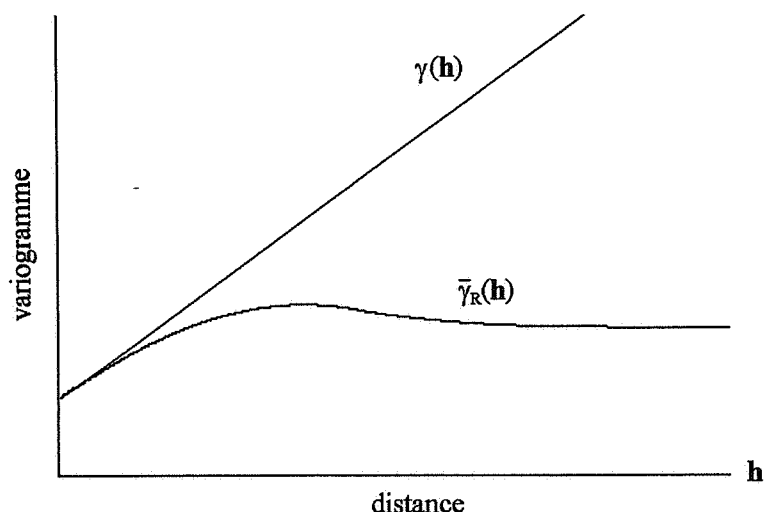


figure 6.2 : problèmes d'analyse variographique en krigeage universel, exemple dans le cas d'un variogramme linéaire + pépite

De manière générale, on montre que le biais n'est négligeable que pour les distances petites par rapport à la taille du voisinage qui a servi à estimer la dérive. Du variogramme vrai, on ne pourra donc espérer connaître que l'effet de pépite et le comportement au voisinage de l'origine.

Remarque

On peut exprimer le variogramme du résidu estimé à l'aide du variogramme vrai, mais cette relation n'est pas bijective. Le variogramme vrai reste donc théoriquement indéterminé. On peut chercher à réduire cette indétermination, en choisissant par exemple, parmi tous les variogrammes possibles $g(\mathbf{x} + \mathbf{h}, \mathbf{x})$, celui qui est "le plus stationnaire" :

$$\gamma(\mathbf{h}) = \bar{g}_0(\mathbf{h}) \text{ où } g_0 \text{ minimise } \iint [g(\mathbf{x} + \mathbf{h}, \mathbf{x}) - \bar{g}(\mathbf{h})]^2 d\mathbf{x} d\mathbf{h}.$$

Il apparaît par ailleurs que l'indétermination du variogramme n'a de répercussion que sur les variances des estimateurs optimaux (krigeage de Z ou de sa dérive), mais non sur les estimateurs eux-mêmes : tous les variogrammes théoriquement possibles conduisent aux mêmes estimateurs optimaux, pour la dérive comme pour la variable régionalisée.

Alors qu'elle semblait naturelle au départ, la séparation du phénomène en une dérive et un résidu soulève de sérieuses difficultés théoriques : le variogramme sous-jacent est irrémédiablement masqué et son inférence est biaisée. Une étude menée en krigeage universel demande une réelle subtilité pour sa mise en œuvre et doit donc se faire avec précautions. Il convient en tout cas d'éviter d'attribuer au résidu "vrai" (inaccessible expérimentalement) la covariance ou le variogramme du résidu estimé, qui sous-estime la variabilité réelle.

L'application du krigeage universel à un jeu de données réelles (à savoir, les données de sol) est reportée après l'examen du krigeage intrinsèque d'ordre k , qui va donner un nouvel éclairage aux problèmes de non-stationnarité.

III. Le krigeage intrinsèque d'ordre k

Le krigeage intrinsèque d'ordre k a été élaboré pour résoudre les problèmes d'analyse variographique rencontrés en krigeage universel. Il s'agit d'une généralisation du passage de l'hypothèse stationnaire à l'hypothèse intrinsèque. Dans le modèle intrinsèque, ce n'est pas la fonction aléatoire elle-même qui est manipulée, mais ses *accroissements*. Ce procédé annule les constantes (i.e. on ne travaille plus qu'à une constante additive près) et permet l'inférence du variogramme sans biais même lorsque l'espérance est inconnue, alors que l'estimateur de la covariance est biaisé (cf. chapitre 3). Cette approche va être élargie pour traiter les processus non stationnaires : l'idée est de travailler sur des accroissements d'ordre supérieur qui annulent non seulement les constantes, mais aussi les monômes de degré 1, 2..., bref tout ce qui pourrait composer la dérive.

III.1. Combinaisons linéaires autorisées d'ordre k

Etant donné une fonction aléatoire Z non stationnaire, l'idée directrice consiste à trouver des combinaisons linéaires de Z qui soient stationnaires d'ordre deux. On montre qu'une condition nécessaire⁷ pour pouvoir satisfaire cette propriété, est que les poids λ_i de ces combinaisons linéaires et les points $\mathbf{x}_i$ associés vérifient les conditions :

$$\sum_i \lambda_i f^\ell(\mathbf{x}_i) = 0 \quad \forall \ell$$

où les f^ℓ forment une famille *complète* (c'est-à-dire stable par translation) d'exponentielles – polynômes.

La plupart des applications pratiques se restreignent à la famille des monômes de degré inférieur ou égal à un entier k donné⁸. Les combinaisons linéaires vérifiant les égalités précédentes sont appelées **combinaisons linéaires autorisées d'ordre k** ; plus précisément, une combinaison linéaire autorisée d'ordre k est la donnée d'un système de poids et de points $\lambda = \{\lambda_i, \mathbf{x}_i\}_i$, tel que, pour tout monôme f^ℓ de degré inférieur ou égal à k , on ait

$$\sum_i \lambda_i f^\ell(\mathbf{x}_i) = 0.$$

On impose de plus que la première fonction, d'indice 0, soit constante : $f^0(\mathbf{x}) \equiv 1$.

⁷ Il faut en effet que la famille des combinaisons linéaires que l'on cherche à définir soit stable par translation, car la stationnarité d'ordre deux implique l'invariance par translation des deux premiers moments.

⁸ Dans certains cas plus rares (phénomènes temporels), on peut prendre une famille complète de fonctions trigonométriques.

On retrouve les contraintes introduites en krigeage universel, mais ici elles ont le sens de conditions d'autorisation. En ne manipulant que des combinaisons linéaires autorisées d'ordre k , on annule tout ce qui pourrait être une composante de la dérive. Par conséquent, on n'a plus besoin de définir une dichotomie dérive – résidu : la dérive devient "interne" au modèle et il suffit seulement de connaître son degré k . Le krigeage intrinsèque d'ordre k va permettre de regrouper les différents modèles rencontrés jusqu'ici (krigeage simple, ordinaire, universel) en un formalisme unique.

Exemples

Sur la figure 6.3, la combinaison linéaire en (a), correspondant à deux points auxquels sont affectés les poids -1 et 1 , est une combinaison linéaire autorisée d'ordre 0 (accroissement classique, qui annule les constantes). Le cas (b) définit une combinaison linéaire d'ordre 1 dans R , car elle annule non seulement les constantes mais aussi le monôme x_1 de degré 1. Des exemples de combinaisons linéaires d'ordre 1 et 2 dans l'espace bidimensionnel sont fournis en (c) et (d) respectivement : la combinaison linéaire en (c) annule les constantes et les monômes du premier degré x_1 et x_2 , celle en (d) annule de surcroît les monômes du second degré x_1^2 , $x_1 x_2$, x_2^2 . Enfin, le cas (e) donne un exemple de combinaison linéaire d'ordre 2 lorsque les points sont répartis de manière irrégulière dans l'espace R^2 .

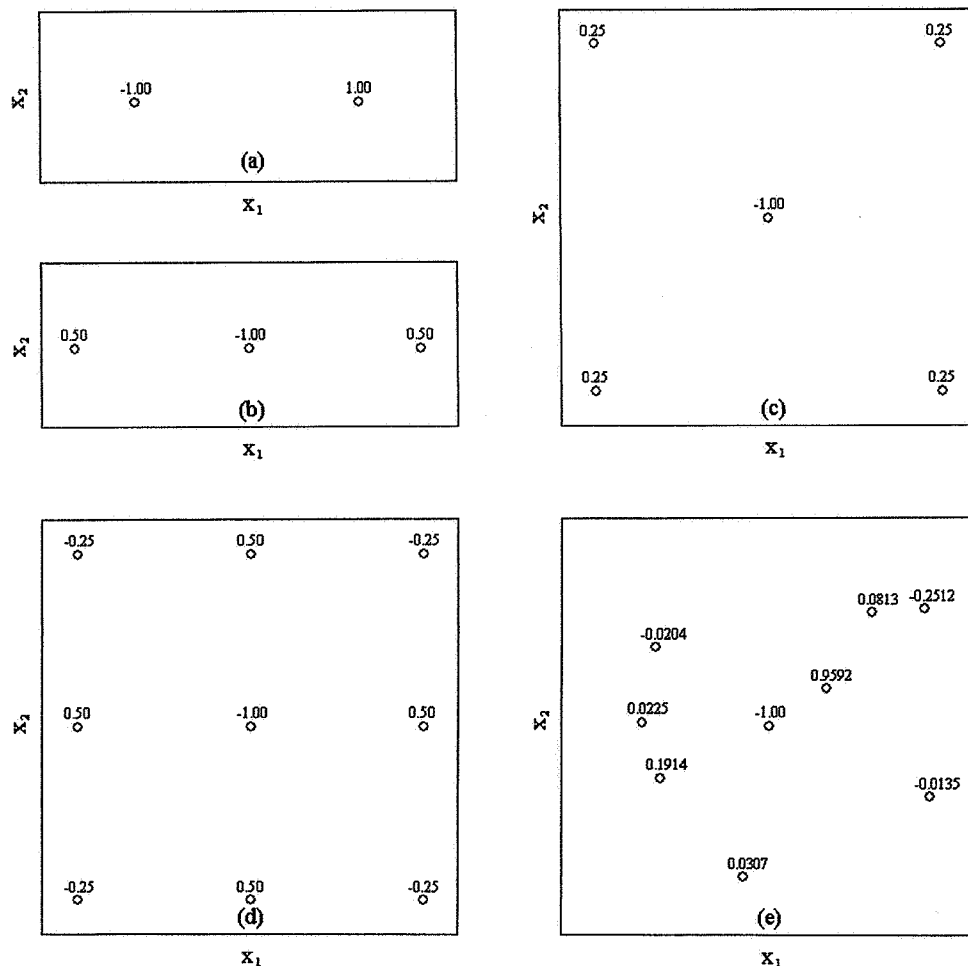


figure 6.3 : exemples de combinaisons linéaires autorisées d'ordre 0, 1 et 2

Remarques

- 1) Le nombre de points associés à une combinaison linéaire autorisée d'ordre k dans l'espace $\mathbb{R}^d$ doit être supérieur au nombre de contraintes caractérisant cette combinaison, à savoir

$$C_{d+k}^k = \frac{(d+k)!}{k! d!} \text{ (nombre de monômes de degré inférieur ou égal à } k \text{ dans } \mathbb{R}^d \text{).}$$

- 2) Pour construire une combinaison linéaire autorisée d'ordre k sur un nombre suffisant de sites, il suffit de réaliser le krigeage universel d'un de ces sites (noté $\mathbf{x}_0$) à partir des autres $\{\mathbf{x}_\alpha, \alpha = 1 \dots n\}$, à l'aide d'un modèle quelconque de variogramme et en prenant une dérive de degré k . L'erreur de krigeage, qui est une combinaison linéaire pondérée attribuant les poids de krigeage λ_α aux sites $\mathbf{x}_\alpha$ et le poids -1 au site $\mathbf{x}_0$, vérifie les contraintes d'autorisation et de non-biais, qui coïncident avec celles définissant une combinaison linéaire autorisée d'ordre k .
- 3) Les combinaisons linéaires d'ordre k sont une extension de la notion d'accroissement utilisé dans le formalisme intrinsèque classique ; pour cette raison, elles sont également appelées "**accroissements généralisés**" d'ordre k .

III.2. Fonctions aléatoires intrinsèques d'ordre k

Pour étendre l'hypothèse intrinsèque, nous allons définir une classe plus générale de fonctions aléatoires. Afin de ne pas alourdir la lecture avec des concepts trop abstraits, nous donnerons ci-dessous une définition "sommaire" et simplifiée de ces fonctions aléatoires. La formalisation théorique est présentée dans l'annexe D de ce document.

Une fonction aléatoire $Z(\mathbf{x})$ est intrinsèque à l'ordre k si ses "accroissements généralisés" d'ordre k possèdent une espérance nulle et une variance invariante par translation. Plus précisément : pour toute combinaison linéaire d'ordre k , $\lambda = \{\lambda_i, \mathbf{x}_i\}_i$, et tout vecteur $\mathbf{h}$ de $\mathbb{R}^d$, les moments de la combinaison linéaire translatée de $\mathbf{h}$ (qui est autorisée d'ordre k) existent et ne dépendent pas de $\mathbf{h}$, mais seulement des poids λ_i et de l'implantation *relative* des points $\mathbf{x}_i$:

$$\forall \lambda = \{\lambda_i, \mathbf{x}_i\}_i \text{ autorisée d'ordre } k, \forall \mathbf{h}, \begin{cases} E[\sum_i \lambda_i Z(\mathbf{x}_i + \mathbf{h})] = 0 \\ \text{var}[\sum_i \lambda_i Z(\mathbf{x}_i + \mathbf{h})] \text{ ne dépend pas de } \mathbf{h} \end{cases}$$

C'est une **généralisation de l'hypothèse intrinsèque**. On vérifie en effet qu'une fonction aléatoire intrinsèque correspond à une fonction aléatoire intrinsèque d'ordre 0, pour laquelle les combinaisons linéaires autorisées d'ordre 0 (celles de poids total nul) sont stationnaires d'ordre deux.

III.3. Fonction de covariance généralisée

Soit $Z(\mathbf{x})$ une fonction aléatoire intrinsèque d'ordre k , supposée continue en moyenne quadratique. On montre qu'il existe une fonction structurale appelée **covariance généralisée** et notée K , définie à un polynôme pair de degré $\leq 2k$ près, telle que :

- 1) $\forall \mathbf{h} \in \mathbb{R}^d, K(\mathbf{h}) = K(-\mathbf{h})$
- 2) pour toute combinaison linéaire autorisée d'ordre k , $\lambda = \{\lambda_i, \mathbf{x}_i\}_i$, on a :

$$\text{var} \left[\sum_i \lambda_i Z(\mathbf{x}_i) \right] = \sum_i \sum_j \lambda_i \lambda_j K(\mathbf{x}_i - \mathbf{x}_j).$$

Formellement, la covariance généralisée s'utilise comme une covariance dans le calcul des variances des combinaisons linéaires autorisées.

III.3.1. Propriétés théoriques

- comportement à l'infini : le rapport $K(\mathbf{h})/|\mathbf{h}|^{2k+2}$ tend vers 0 lorsque $|\mathbf{h}|$ tend vers l'infini ;
- le comportement à l'origine (dérivabilité, continuité, ou effet de pépite) est lié au degré de régularité spatiale de la régionalisation ; le lien est le même que celui vu dans le cas de la covariance usuelle et du variogramme ;
- K est de type positif conditionnel d'ordre k , c'est-à-dire que pour toute combinaison linéaire autorisée d'ordre k , $\lambda = \{\lambda_i, \mathbf{x}_i\}_i$, on a :

$$\sum_i \sum_j \lambda_i \lambda_j K(\mathbf{x}_i - \mathbf{x}_j) \geq 0.$$

Cette propriété découle de la définition de la covariance généralisée (l'expression ci-dessus ayant la signification d'une variance). C'est une condition nécessaire et suffisante : il y a identité entre les covariances généralisées d'ordre k et les fonctions de type positif conditionnel d'ordre k .

De plus, comme une combinaison linéaire autorisée d'ordre $(k+1)$ est aussi une combinaison linéaire autorisée d'ordre k , l'ensemble des fonctions de type positif conditionnel d'ordre $(k+1)$ contient l'ensemble des fonctions de type positif conditionnel d'ordre k . Ainsi, **lorsque l'ordre k augmente, l'ensemble des modèles structuraux s'enrichit**. On retrouve la généralisation du passage de la covariance stationnaire C au variogramme γ . Notons tout de même le caractère particulier du variogramme γ , puisque c'est $-\gamma$ (et non γ lui-même) qui fait partie des fonctions de type positif conditionnel d'ordre 0.

III.3.2. Exemples de covariances généralisées isotropes

Outre tous les modèles de variogramme, qui sont aussi (au signe près) des fonctions de covariance généralisée, on peut donner l'exemple du **modèle spline** :

$$K(\mathbf{h}) = |\mathbf{h}|^{2p} \ln(|\mathbf{h}|) \text{ pour } 1 \leq p \leq k.$$

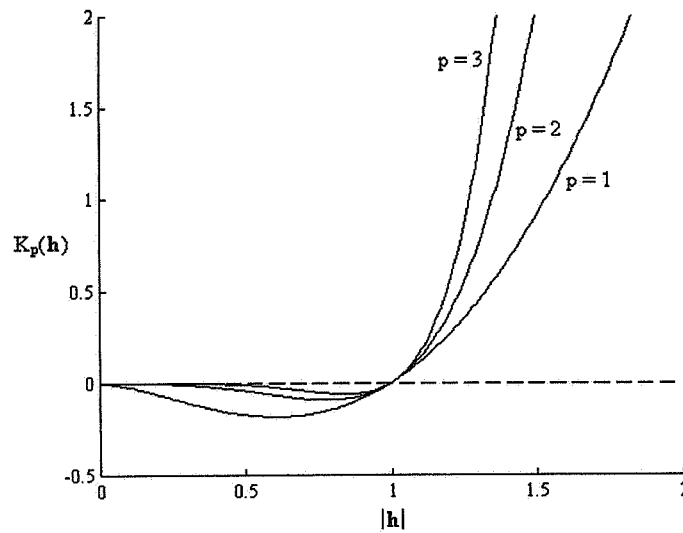


figure 6.4 : modèle de covariance généralisée spline

Citons également les modèles **puissances impaires**

$$K_p(\mathbf{h}) = (-1)^{p+1} |\mathbf{h}|^{2p+1} \text{ avec } 2p+1 < 2k+2, \text{ soit } p \leq k.$$

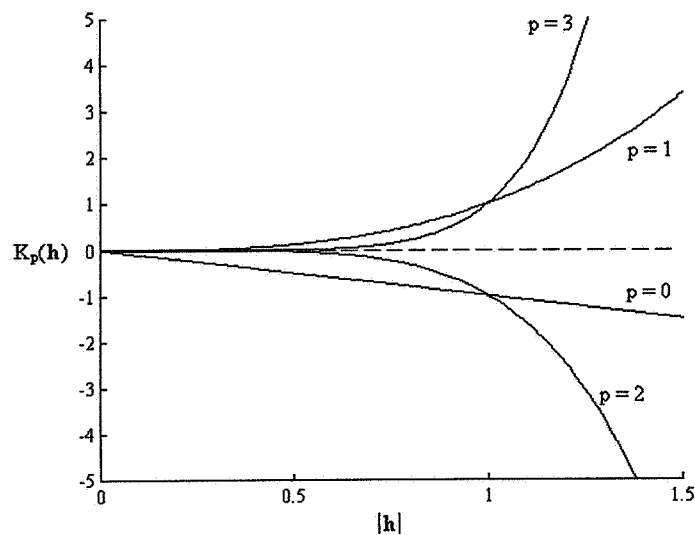


figure 6.5 : modèles de covariances généralisées puissances impaires

Une généralisation, peu usitée en pratique, du modèle puissance impaire est le **modèle "puissance non paire"** :

$$K_{\theta}(\mathbf{h}) = (-1)^{p+1} |\mathbf{h}|^{\theta} \quad \text{avec } p \in \mathbb{N} \text{ et } 2p < \theta < 2(p+1) \leq 2k+2$$

(le signe de $K_{\theta}(\mathbf{h})$ alterne avec la parité de la partie entière de $\theta/2$).

En pratique, les modèles les plus utilisés sont les **modèles polynomiaux** de la forme :

$$K(\mathbf{h}) = \sum_{p=0}^k b_p (-1)^{p+1} |\mathbf{h}|^{2p+1}.$$

Ce sont des polynômes par rapport au module $|\mathbf{h}|$ du vecteur $\mathbf{h}$; il serait superflu de rajouter des termes de degré pair, car une covariance généralisée est définie à un polynôme pair de degré $\leq 2k$ près.

On peut montrer que $K(\mathbf{h})$ est de type positif conditionnel d'ordre k moyennant certaines conditions sur les coefficients b_p : une condition suffisante est que les b_p soient positifs ou nuls, mais elle n'est pas nécessaire. Le tableau 6.1 donne les conditions nécessaires et suffisantes pour obtenir des modèles autorisés, dans les cas $k = 0, 1$ et 2 (on y a ajouté un effet de pépite $C_0 \delta(\mathbf{h})$ avec $C_0 \geq 0$ et $\delta(\mathbf{h}) = 1$ si $\mathbf{h} = \mathbf{0}$, 0 sinon).

k	covariances généralisées polynomiales
0	$K(\mathbf{h}) = C_0 \delta(\mathbf{h}) - b_0 \mathbf{h} \quad (b_1 = b_2 = 0)$
1	$K(\mathbf{h}) = C_0 \delta(\mathbf{h}) - b_0 \mathbf{h} + b_1 \mathbf{h} ^3 \quad (b_2 = 0)$
2	$K(\mathbf{h}) = C_0 \delta(\mathbf{h}) - b_0 \mathbf{h} + b_1 \mathbf{h} ^3 - b_2 \mathbf{h} ^5$
contraintes	dans $\mathbb{R}$: $C_0 \geq 0, b_0 \geq 0, b_2 \geq 0, b_1 \geq -2\sqrt{\frac{10b_0b_2}{3}}$ dans $\mathbb{R}^2$: $C_0 \geq 0, b_0 \geq 0, b_2 \geq 0, b_1 \geq -\frac{10}{3}\sqrt{b_0b_2}$ dans $\mathbb{R}^3$: $C_0 \geq 0, b_0 \geq 0, b_2 \geq 0, b_1 \geq -\sqrt{10b_0b_2}$ dans $\mathbb{R}^d$: $C_0 \geq 0, b_0 \geq 0, b_2 \geq 0, b_1 \geq -2\sqrt{\frac{5b_0b_2(d+3)}{3(d+1)}}$

tableau 6.1: contraintes sur les coefficients des covariances polynomiales

La covariance élémentaire en $|\mathbf{h}|^5$ correspond à un processus deux fois différentiable en moyenne quadratique, ce qui est une hypothèse très forte. On préfère souvent remplacer le terme en $|\mathbf{h}|^5$ par un terme en $|\mathbf{h}|^2 \ln(|\mathbf{h}|)$ (covariance spline) et utiliser une covariance de la forme :

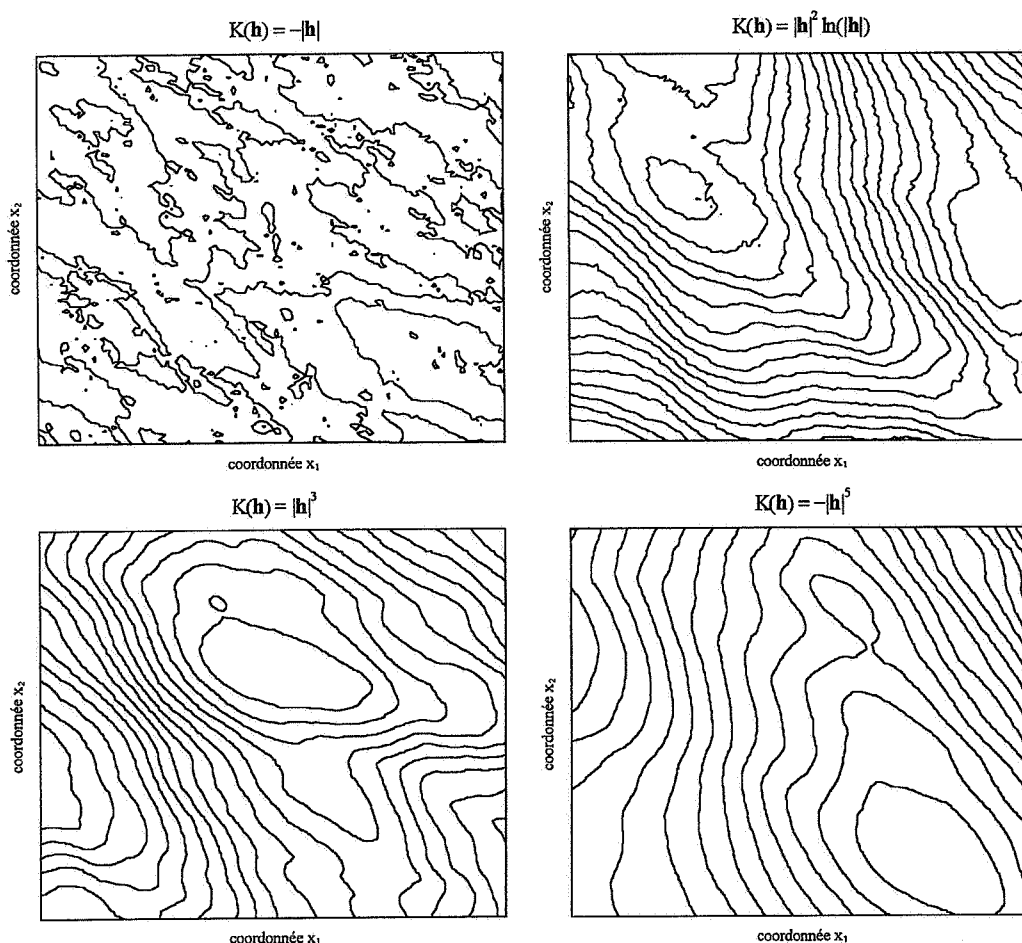
$$K(\mathbf{h}) = C_0 \delta(\mathbf{h}) - b_0 |\mathbf{h}| + b_s |\mathbf{h}|^2 \ln(|\mathbf{h}|) + b_1 |\mathbf{h}|^3 \quad \text{pour } k \geq 1.$$

Là encore, certaines contraintes sur les coefficients doivent être vérifiées pour que $K(\mathbf{h})$ soit bien une covariance généralisée admissible. Une condition suffisante est que tous les coefficients b_0, b_1 et b_s soient positifs, mais on peut réduire ces contraintes, comme le résume le tableau 6.2 (dans lequel Γ désigne la fonction gamma qui interpole la factorielle).

k	covariance généralisée
≥ 1	$K(\mathbf{h}) = C_0 \delta(\mathbf{h}) - b_0 \mathbf{h} + b_s \mathbf{h} ^2 \ln(\mathbf{h}) + b_1 \mathbf{h} ^3$
contraintes	dans $\mathbb{R}$: $C_0 \geq 0, b_0 \geq 0, b_1 \geq 0, b_s \geq -\frac{2}{\pi} \sqrt{6 b_0 b_1}$ dans $\mathbb{R}^2$: $C_0 \geq 0, b_0 \geq 0, b_1 \geq 0, b_s \geq -\frac{3}{2} \sqrt{b_0 b_1}$ dans $\mathbb{R}^3$: $C_0 \geq 0, b_0 \geq 0, b_1 \geq 0, b_s \geq -\frac{8}{\pi} \sqrt{\frac{b_0 b_1}{3}}$ dans $\mathbb{R}^d$: $C_0 \geq 0, b_0 \geq 0, b_1 \geq 0, b_s \geq -\frac{\Gamma(d)}{d 2^{d-2} \Gamma(d/2)^2} \sqrt{3 b_0 b_1 (d+1)}$

tableau 6.2: contraintes sur les coefficients des covariances polynomiales + spline

A titre d'illustration, la figure 6.6 présente, sous forme de lignes de niveau équiréparties, des cartes de réalisations de fonctions aléatoires intrinsèques d'ordre k , de covariances généralisées puissances impaires et spline. On observe que la covariance, associée à une fonction intrinsèque d'ordre 0, est assez erratique (les lignes de niveau sont très morcelées). Plus le degré de la covariance généralisée augmente, plus les fonctions aléatoires associées sont régulières et s'écartent de la stationnarité classique. On voit ainsi combien le formalisme des fonctions intrinsèques d'ordre k est plus général que celui des fonctions stationnaires.


 figure 6.6 : réalisations de fonctions aléatoires intrinsèques d'ordre k de covariances généralisées $K(\mathbf{h}) = -|\mathbf{h}|, |\mathbf{h}|^2 \ln(|\mathbf{h}|), |\mathbf{h}|^3$ et $-|\mathbf{h}|^5$

III.4. Le krigeage intrinsèque d'ordre k ponctuel

Nous supposons que la variable régionalisée étudiée est une réalisation d'une fonction aléatoire intrinsèque d'ordre k admettant une covariance généralisée K.

Pour estimer la valeur en un point $\mathbf{x}_0$ quelconque, on pose :

$$Z^*(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha})$$

où $\{ \mathbf{x}_{\alpha}, \alpha = 1 \dots n \}$ sont les sites d'observation voisins de $\mathbf{x}_0$.

On impose que l'erreur d'estimation soit une combinaison linéaire autorisée d'ordre k pour pouvoir calculer son espérance et sa variance, ce qui donne les conditions d'autorisation

$$\sum_{\alpha=1}^n \lambda_{\alpha} f^{\ell}(\mathbf{x}_{\alpha}) = f^{\ell}(\mathbf{x}_0) \quad \forall \ell.$$

La condition de non biais est automatiquement satisfaite, car toute combinaison linéaire autorisée est d'espérance nulle.

Il ne reste plus qu'à minimiser la variance de l'erreur d'estimation sous les conditions d'autorisation. En se souvenant que la variance d'une combinaison linéaire autorisée d'ordre k se calcule comme s'il existait une covariance stationnaire, mais en remplaçant formellement celle-ci par la covariance généralisée K, on aboutit au système :

$$\begin{pmatrix} K(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) & f^{\ell}(\mathbf{x}_{\alpha}) \\ f^{\ell}(\mathbf{x}_{\beta}) & 0 \end{pmatrix} \begin{pmatrix} \lambda_{\beta} \\ \mu_{\ell} \end{pmatrix} = \begin{pmatrix} K(\mathbf{x}_0 - \mathbf{x}_{\alpha}) \\ f^{\ell}(\mathbf{x}_0) \end{pmatrix}$$

où les μ_{ℓ} sont des multiplicateurs de Lagrange.

Il s'agit du même système que celui du krigeage universel, à ceci près que les conditions de non-biais sont devenues des conditions d'autorisation, et que les outils variographiques (les covariances généralisées) forment une famille plus vaste que celle des fonctions de covariance ou des variogrammes.

La variance de krigeage intrinsèque s'écrit :

$$\sigma_{KI-k}^2(\mathbf{x}_0) = K(0) - \sum_{\alpha=1}^n \lambda_{\alpha} K(\mathbf{x}_{\alpha} - \mathbf{x}_0) - \sum_{\ell} \mu_{\ell} f^{\ell}(\mathbf{x}_0).$$

On montre que les pondérateurs de krigeage et la variance de krigeage sont indépendants du choix arbitraire de la covariance généralisée K parmi la classe des covariances généralisées admissibles (K est définie à un polynôme pair de degré $\leq 2k$ près).

III.5. Propriétés du krigeage intrinsèque d'ordre k

On retrouve les propriétés générales du krigeage rencontrées dans le cadre stationnaire, en particulier :

- **interpolation exacte** : la valeur estimée en un point de mesure est égale à la valeur expérimentale ;
- **additivité** ;
- **propriété d'orthogonalité** : l'erreur d'estimation ponctuelle est non corrélée avec les combinaisons linéaires autorisées d'ordre k des données ;
- **propriété de lissage** : les valeurs estimées sont moins variables que les valeurs vraies.

Cette propriété ne peut pas se traduire par une inégalité entre les variances de Z et Z^* , car ces variances n'existent pas. On peut toutefois montrer que, pour toute combinaison linéaire autorisée d'ordre k, ou plus généralement pour toute *mesure linéaire* autorisée ϕ , i.e. telle que

$$\int f^\ell(\mathbf{x}) \phi(d\mathbf{x}) = 0 \quad \forall \ell$$

on a :

$$\text{var}[\phi(Z^*)] \leq \text{var}[\phi(Z)]$$

la différence étant égale à la variance de krigeage de $\phi(Z)$.

- **conditions de régularité** : le système de krigeage intrinsèque d'ordre k est régulier si et seulement si

- 1) les fonctions de base sont linéairement indépendantes sur les données :

$$\left(\sum_{\ell} c_{\ell} f^{\ell}(\mathbf{x}_{\alpha}) = 0 \quad \forall \alpha = 1 \dots n \right) \Rightarrow c_{\ell} = 0 \quad \forall \ell$$

- 2) la sous-matrice $\mathbf{K} = [K(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta})]_{\alpha, \beta=1 \dots n}$ est positive *conditionnelle* stricte :

pour toute combinaison linéaire autorisée d'ordre k $\{\lambda_{\alpha}, \mathbf{x}_{\alpha}\}_{\alpha}$ définie sur les points échantillonnés,

$$\sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} K(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \geq 0$$

avec la nullité si et seulement si $(\lambda_{\alpha})_{\alpha} \equiv 0$.

Signalons enfin que le système de krigeage ponctuel peut être adapté au **krigeage par bloc** ou plus généralement au krigeage d'une régularisée, en modifiant le second membre. On retrouve le même système que celui du krigeage universel non ponctuel, à condition de remplacer la covariance C ou l'opposé du variogramme $-\gamma$ par la covariance généralisée K .

III.6. L'analyse variographique

On se souvient que l'analyse variographique a été le point d'achoppement du krigeage universel. La question qui se pose naturellement est celle de la modélisation de la fonction de covariance généralisée K . La classe des covariances généralisées d'ordre k étant bien plus vaste que la classe des covariances ou variogrammes habituels et n'étant définie que sur une famille restreinte de combinaisons linéaires, l'inférence statistique est certes plus complexe, mais néanmoins possible.

III.6.1. Détermination de l'ordre k

L'ordre que l'on cherche à déterminer représente le degré de la "tendance" polynomiale locale, c'est-à-dire au niveau du voisinage de travail envisagé. Il faut prendre garde au fait qu'une tendance locale de degré k peut ne pas correspondre à une tendance de degré k à l'échelle de tout le champ $\mathcal{D}$.

En pratique, on utilise un critère de validation croisée. Le principe consiste à supprimer tour à tour chaque point de données x_α ($\alpha = 1 \dots n$), et à ajuster des polynômes de degré $k = 0, 1, 2 \dots$ à partir des données restantes, en utilisant un algorithme de moindres carrés. Pour chaque ordre k , on peut alors calculer au point supprimé "l'erreur", notée $\varepsilon_{\alpha,k}$, entre la valeur expérimentale et la valeur donnée par le polynôme de moindres carrés. L'ordre k_0 optimal est celui qui produit, en moyenne, la plus petite variance d'erreurs :

$$k_0 \text{ réalisant } \min_k [\text{var}(\varepsilon_{\alpha,k})].$$

L'inconvénient de cette méthode est son manque de robustesse face aux éventuelles données "aberrantes", qui vont à elles seules être responsables de grandes variances d'erreurs quel que soit le degré k choisi. Un critère plus efficace de sélection consiste à ranger, pour chaque point x_α , les erreurs en valeurs absolues $|\varepsilon_{\alpha,k}|$ par ordre croissant et à choisir le degré k_0 donnant le plus petit rang moyen sur l'ensemble des points x_α .

	x_1	...	x_n	rang moyen	} choix de k
$k = 0$	$r_0 = 3$		$r_0 = 1$	$\bar{r}_0$	
$k = 1$	$r_1 = 1$		$r_1 = 2$	$\bar{r}_1$	
$k = 2$	$r_2 = 2$		$r_2 = 3$	$\bar{r}_2$	

Habituellement, on ne dépasse pas l'ordre $k = 2$.

III.6.2. Inférence statistique de la covariance généralisée

Une fois l'ordre k déterminé, il faut ajuster un modèle de covariance généralisée.

En général, on choisit un modèle formé par une combinaison linéaire de modèles de base, que l'on sait admissibles. Plus précisément, on suppose que la covariance généralisée est de la forme :

$$K(\mathbf{h}) = \sum_p b_p K_p(\mathbf{h})$$

avec K_p modèles de covariance connus.
 b_p coefficients inconnus à déterminer "au mieux".

Par exemple, on peut prendre comme modèles de base la famille des puissances impaires, à savoir $K_p(\mathbf{h}) = (-1)^p |\mathbf{h}|^{2p+1}$ (avec $0 < 2p+1 < 2k+2$). La covariance gigue correspondante $K(\mathbf{h})$ contient des structures avec des comportements différents à l'origine : le terme linéaire est adapté pour des variables régionalisées continues, tandis que les termes d'ordre supérieur permettent de modéliser des variables régionalisées de plus en plus régulières. En rajoutant une covariance péritique, on décrit également des régionalisations plus erratiques (discontinues).

On doit avoir, pour toute combinaison linéaire autorisée d'ordre k sur les données $(\omega_\alpha, \mathbf{x}_\alpha)_\alpha$:

$$\text{var} \left[\sum_\alpha \omega_\alpha Z(\mathbf{x}_\alpha) \right] = \sum_p b_p \sum_\alpha \sum_\beta \omega_\alpha \omega_\beta K_p(\mathbf{x}_\alpha - \mathbf{x}_\beta).$$

La combinaison linéaire

$$\sum_\alpha \omega_\alpha Z(\mathbf{x}_\alpha)$$

étant d'espérance nulle, sa variance s'identifie au carré de son espérance. Au niveau des valeurs expérimentales, cela signifie que l'on calcule, pour chaque combinaison linéaire $(\omega_\alpha, \mathbf{x}_\alpha)_\alpha$:

- sa variance expérimentale : $\left[\sum_\alpha \omega_\alpha z(\mathbf{x}_\alpha) \right]^2$
- les variances théoriques associées aux covariances élémentaires $K_p(\mathbf{h})$

$$\sum_\alpha \sum_\beta \omega_\alpha \omega_\beta K_p(\mathbf{x}_\alpha - \mathbf{x}_\beta)$$

et que ces différentes variances sont liées par la relation approximative

$$\left[\sum_\alpha \omega_\alpha z(\mathbf{x}_\alpha) \right]^2 = \sum_p b_p \sum_\alpha \sum_\beta \omega_\alpha \omega_\beta K_p(\mathbf{x}_\alpha - \mathbf{x}_\beta).$$

On écrit cette relation pour de nombreuses combinaisons linéaires $(\omega_\alpha, x_\alpha)_\alpha$ autorisées d'ordre k . Un algorithme de régression permet alors d'ajuster "au mieux" les coefficients b_p : ils seront par exemple déterminés par une méthode des moindres carrés pondérés. Les pondérateurs ont pour but de ne pas attribuer trop de poids aux termes qui ont une forte variance.

III.6.3. Remarques sur la procédure d'ajustement

Si la forme du modèle gigogne est correcte, l'estimateur de la covariance généralisée est **sans biais**. Plusieurs difficultés subsistent, que nous nous contentons d'évoquer : combien de modèles de base introduire dans le schéma gigogne ? combien de combinaisons linéaires autorisées construire ? quels pondérateurs choisir dans l'algorithme de régression déterminant les b_p ? comment s'assurer que le modèle trouvé satisfait les contraintes sur les covariances généralisées ? comment évaluer la qualité de l'ajustement ? Certains choix relèvent de l'utilisateur, d'autres se font via des algorithmes automatiques.

En cas d'**hétérogénéité de la régionalisation**, le bon sens veut qu'on subdivise le champ en zones homogènes (sur chacune desquelles le modèle de fonction aléatoire intrinsèque d'ordre k est acceptable) et que l'on ajuste un modèle dans chaque zone. Il est toutefois illusoire de chercher à différencier des zones contenant peu de données ; dans ce cas, on devra se contenter d'un modèle global qui fournira une description moyenne de la structure du phénomène.

Les modèles de covariance généralisée couramment utilisés (covariances polynomiales et splines) sont **isotropes**. L'inférence statistique est bien plus délicate lorsque l'on cherche des modèles anisotropes, à cause du nombre plus important de paramètres à estimer. La restriction à des modèles isotropes n'est toutefois pas si contraignante qu'il n'y paraît, car on admet en général que l'éventuelle anisotropie disparaît en manipulant les combinaisons linéaires autorisées d'ordre k . Cependant, pour $k = 0$, cette limitation à des modèles isotropes est gênante, puisque les combinaisons d'ordre 0 n'annulent que les constantes, et ne peuvent par conséquent faire disparaître l'anisotropie ; il convient alors de se ramener au formalisme du variogramme puisque l'ordre $k = 0$ correspond à l'hypothèse intrinsèque habituelle.

L'avantage de la procédure présentée est sa généralité : contrairement à d'autres méthodes, elle ne requiert aucune hypothèse sur la loi de probabilité des combinaisons linéaires d'ordre k et elle permet l'inférence lorsque les échantillons sont disposés irrégulièrement. Son principal défaut réside dans son **manque de robustesse** en présence d'anomalies dans les données, ou de changements de structuration au sein de la régionalisation. Toutefois, on observe que l'estimation par krigeage intrinsèque d'ordre k est peu sensible à de légères modifications du modèle de covariance généralisée. Par contre, la variance de krigeage est davantage affectée ; il convient donc de l'interpréter avec prudence, à plus forte raison qu'on en déduit souvent, à tort, un intervalle de confiance sur l'estimation (± 2 écarts-types autour du krigeage par exemple), alors que la loi de l'erreur de krigeage est inconnue, et en particulier non gaussienne.

III.7. Exemple sur les données de sol

Reprenons l'exemple des données de sol introduit aux chapitres précédents. La structure de la variable % argile + limon a été modélisée par un variogramme gigogne, comportant un effet de pépite, un schéma sphérique et un schéma puissance en $|h|^{1.8}$. L'utilisation d'un tel modèle est limitée pour une fonction aléatoire intrinsèque classique, car le schéma puissance atteint pratiquement la borne maximale autorisée $|h|^2$. D'autre part, la présence d'une dérive dans la direction ouest-nord-ouest / est-sud-est est particulièrement marquée (voir la carte obtenue par krigeage ordinaire). Le recours à un modèle non stationnaire peut s'avérer une alternative plus performante qu'un modèle intrinsèque.

Envisageons d'abord l'approche du **krigeage universel**. Le variogramme des données brutes, presque parabolique, suggère la présence d'une dérive linéaire, dont on peut supposer d'ailleurs que la ligne de plus grande pente est dans la direction ouest-nord-ouest / est-sud-est (direction d'anisotropie présentant la plus forte variabilité). Nous allons travailler en voisinage glissant, de forme circulaire et de rayon 40 m. En chaque point d'observation, on estime la dérive par moindres carrés, à l'aide des données contenues dans le voisinage du point considéré. On en déduit alors une estimation du résidu aux points de données, dont on calcule le variogramme expérimental dans les quatre directions principales du plan : 0° , 45° , 90° et 135° par rapport à l'axe est-ouest.

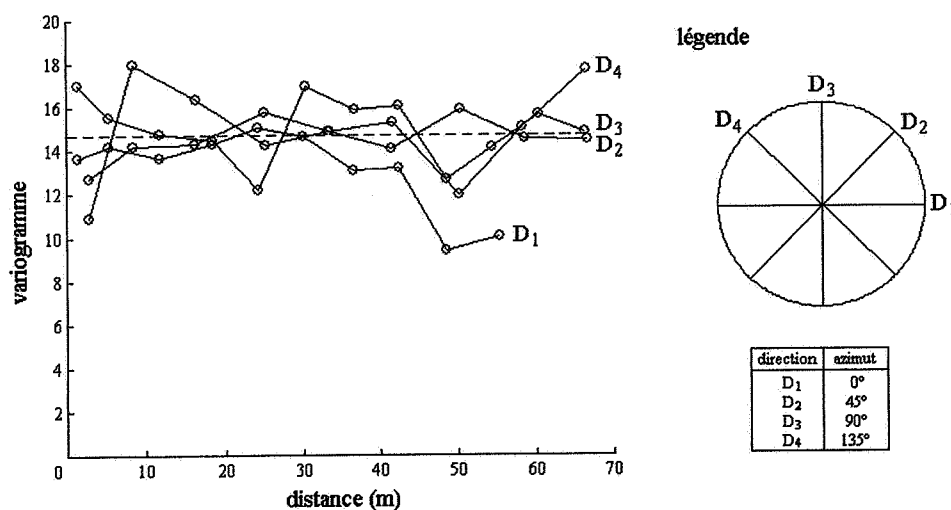


figure 6.7 : variogrammes expérimentaux directionnels du résidu estimé de % argile + limon (dérive linéaire)
la variance expérimentale est indiquée en pointillés

Le résidu *estimé* est purement pépitique. Il en est de même si l'on ajuste une dérive plus complexe (quadratique au lieu de linéaire), comme l'illustre la figure 6.8.

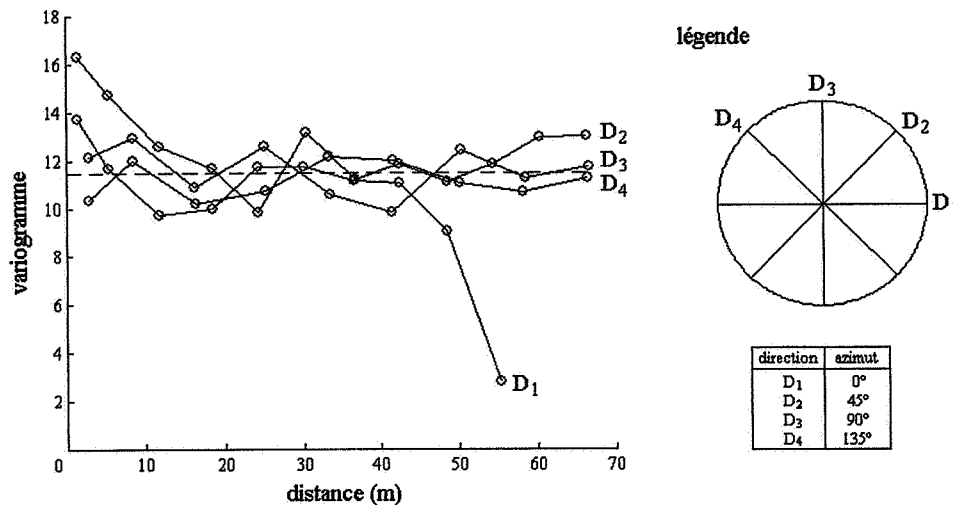


figure 6.8 : variogrammes expérimentaux directionnels
du résidu estimé de % argile + limon (dérive quadratique)
la variance expérimentale est indiquée en pointillés

En conclusion, les variogrammes expérimentaux du résidu *estimé* ne font apparaître aucune structure. Si l'on ne prend pas garde au fait que ces variogrammes sont biaisés, on a l'impression que le phénomène se réduit à la somme d'une dérive et d'un effet de pépité pur.

Pour résoudre ces problèmes d'analyse variographique, on décide de refaire l'étude en **krigeage intrinsèque d'ordre k**. On interprète la variable % argile + limon comme une réalisation d'une fonction aléatoire intrinsèque d'ordre k. Il s'agit de déterminer l'ordre k — qui représente en quelque sorte le degré de la "dérive" — puis de modéliser la covariance généralisée. En supposant que l'anisotropie, si elle existe, sera prise en compte par la dérive, on recherche un modèle de covariance isotrope, de forme polynomiale avec un éventuel effet de pépité. L'inférence de l'ordre k et de la covariance généralisée est effectuée en voisinage glissant de forme circulaire et de rayon 40 m. La procédure d'ajustement automatique diagnostique alors une dérive linéaire (i.e. $k = 1$), avec les fonctions de base 1 et x_1 (l'axe des x_1 est orienté dans la direction est-ouest), et une covariance généralisée pépité + puissance d'ordre 1 :

$$K(h) = 12.8 \delta(h) - 0.503|h|.$$

Muni de ce modèle, on réalise un krigage intrinsèque ponctuel de la variable % argile + limon, pour obtenir les cartes d'estimations et d'écarts-types d'estimation ci-après.

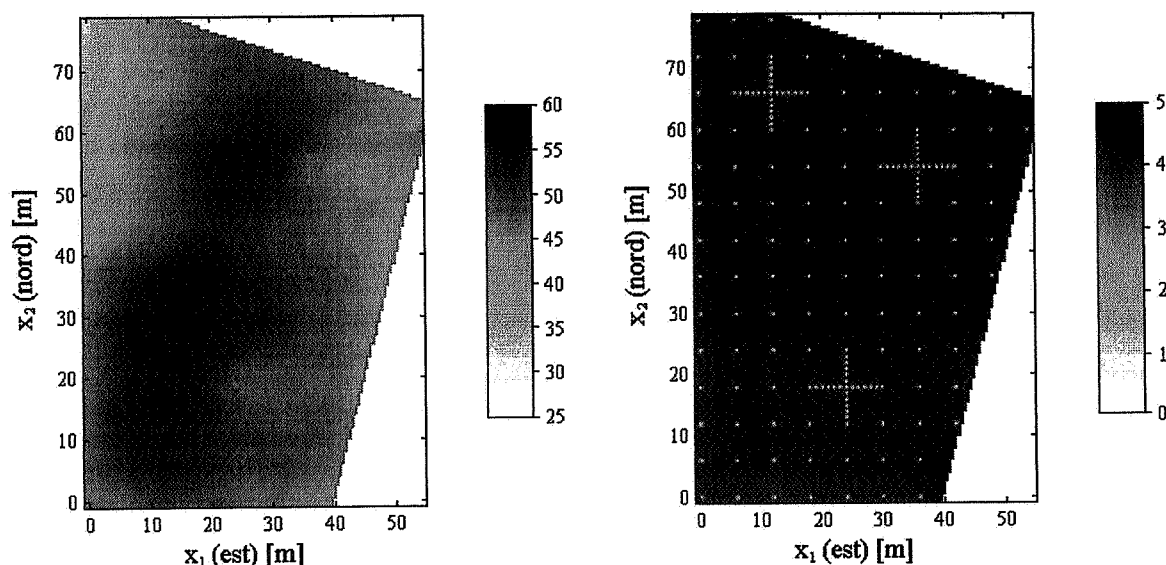


figure 6.9 : cartes des estimations (à gauche) et écarts-types d'estimation (à droite) par krigeage intrinsèque d'ordre 1 ponctuel, variable % argile + limon

Les cartes obtenues par krigeage intrinsèque d'ordre 1 sont sensiblement différentes de celles obtenues par krigeage intrinsèque ordinaire (cf. chapitre 5). Pour déterminer de manière objective laquelle des deux techniques fournit les meilleurs résultats, on utilise la méthode de validation croisée.

Les résultats sont résumés dans le tableau ci-dessous.

	moyenne		variance	
	krigeage intrinsèque ordinaire	krigeage intrinsèque d'ordre 1	krigeage intrinsèque ordinaire	krigeage intrinsèque d'ordre 1
erreurs	0.069	-0.209	17.23	16.50
erreurs standardisées	0.017	-0.050	1.14	1.01

tableau 6.3 : résultats de validation croisée

En outre, le nombre de données “robustes”⁹ est égal à 170 (sur un total de 170) pour le krigeage intrinsèque d'ordre 1, alors qu'il est de 166 pour le krigeage intrinsèque ordinaire. Au vu de la validation croisée, le krigeage intrinsèque d'ordre 1 donne de meilleurs résultats que le krigeage intrinsèque ordinaire. En particulier, la variance des erreurs est moindre (16.50 au lieu de 17.23, soit une réduction de 4.4%).

⁹ i.e. dont les erreurs standardisées associées sont situées (conventionnellement) dans l'intervalle $[-2.5 ; 2.5]$.

On peut reprendre l'analyse en **krigeage universel** à la lumière des résultats obtenus en krigeage intrinsèque d'ordre 1. Il apparaît en effet que les fonctions de base ne dépendent que de la première coordonnée x_1 . En revenant au formalisme du krigeage universel, c'est-à-dire en admettant une dichotomie dérive – résidu, cela implique que la dérive est constante le long de l'axe des x_2 (axe nord-sud), et on peut donc estimer *directement* le variogramme de Z le long de cet axe, à l'aide de l'estimateur classique du variogramme.

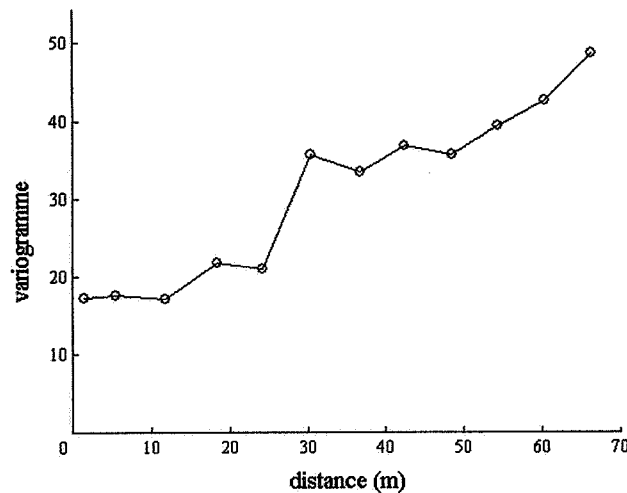


figure 6.10 : variogramme expérimental de % argile + limon dans la direction nord-sud

Moyennant une hypothèse d'isotropie (on espère en général que l'éventuelle anisotropie du phénomène sera prise en compte par la dérive), on peut modéliser le variogramme dans toutes les directions de l'espace, et donc résoudre les problèmes d'analyse variographique rencontrés précédemment.

Incidemment, la covariance généralisée estimée en krigeage intrinsèque d'ordre 1, et réécrite en termes de variogramme, s'ajuste très bien à ce variogramme expérimental (figure 6.11).

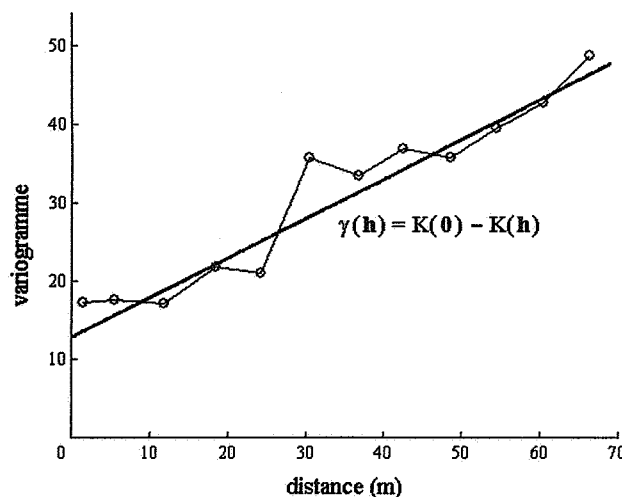


figure 6.11 : variogramme expérimental de Z le long de la direction nord-sud et modèle de covariance généralisée obtenu en krigeage intrinsèque d'ordre 1

Les deux approches (krigeage universel – krigeage intrinsèque d'ordre k) sont donc convergentes. Il est à noter cependant que si la dérive avait été linéaire dans sa forme la plus générale (fonctions de base 1, x_1 , x_2), on aurait été incapable de déterminer de manière précise dans quelle direction cette dérive aurait été constante. D'où l'intérêt du krigeage intrinsèque d'ordre k , qui ne fait pas appel au concept de dérive.

On pourrait objecter que l'analyse variographique en krigeage intrinsèque d'ordre k est peu flexible, car automatisée. La procédure est certes plus automatique que l'analyse structurale classique sur les covariances ou les variogrammes, mais pas entièrement. Les degrés de liberté de l'utilisateur sont le choix du degré de la dérive, des covariances de base introduites dans le modèle gigogne de covariance généralisée, et du voisinage de travail (taille, forme...); ce dernier élément est essentiel lorsque les données ne sont pas réparties de manière homogène dans le champ. Enfin, l'analyse variographique doit rester un travail interactif : toute information sur la physique du phénomène doit être incorporée au modèle, par exemple pour déterminer le comportement à courte échelle, ou encore suggérer un modèle de dérive.

Par ailleurs, il existe des possibilités de contrôle, aussi bien pour le choix de l'ordre k (critères statistiques de qualité : rang moyen et variance des erreurs) que pour la détermination de la covariance généralisée (les coefficients sont calculés par un algorithme de régression, et l'on peut hiérarchiser les différentes régressions possibles). Il est également toujours possible de tester les paramètres du modèle par **validation croisée** : la variance des erreurs et celle des erreurs standardisées fournissent un critère de décision pour le choix final de l'ordre k , de la covariance généralisée et du voisinage de krigeage.

Résumé

Certaines variables régionalisées, présentant une “tendance” systématique dans leur variation spatiale, sont incompatibles avec une hypothèse stationnaire ou même intrinsèque. La démarche naturelle consiste à décomposer la fonction aléatoire modélisant le phénomène en une *dérive* régulière, représentant la “tendance”, et un *résidu* sur lequel portent une hypothèse de stationnarité. Cette dichotomie du phénomène conduit au **krigeage universel**. La principale difficulté concerne l’analyse variographique, car on se heurte à des problèmes de biais lorsque l’on essaie d’évaluer le variogramme du résidu. La mise en œuvre du krigage universel doit donc se faire avec précaution.

La théorie des **fonctions aléatoires intrinsèques d’ordre k** permet de contourner les problèmes rencontrés en krigage universel. Il s’agit d’une approche plus synthétique que celle de la dichotomie, dans laquelle on renonce à séparer dérive et résidu. On s’astreint à ne travailler que sur des “accroissements généralisés” qui sont stationnaires. L’outil structural est la **covariance généralisée**, plus générale que la covariance ou le variogramme. Cette approche conduit au krigage intrinsèque d’ordre k . L’analyse variographique (inférence de la covariance généralisée) est plus commode qu’en krigage universel, quoique plus automatisée.

Chapitre 7 : géostatistique multivariable

Il est fréquent que plusieurs variables régionalisées se réfèrent à un phénomène unique et fassent l'objet d'une même étude géostatistique : citons à titre d'exemples l'évaluation des ressources minières dans un gisement polymétallique, l'analyse des concentrations de divers éléments polluants, ou encore celle de paramètres pédologiques mesurés sur des échantillons de sol. On aura tout intérêt à réaliser l'étude *conjointe* de toutes les variables régionalisées, de façon à tenir compte de leurs liens et de l'information apportée par les variables auxiliaires sur la variable d'intérêt. Dans le cadre de l'étude exploratoire des données, les méthodes d'analyse des données, telle l'analyse en composantes principales (*cf.* annexe F), permettent de compléter utilement l'approche géostatistique. Toutefois, ces techniques ne prennent pas en compte la position spatiale des observations, d'où l'intérêt des méthodes géostatistiques.

Du point de vue théorique, l'extension de la géostatistique linéaire au cas multivariable ne pose pas de problèmes insurmontables. En général, les difficultés ne tiennent ni aux concepts ni aux méthodes mathématiques, mais aux notations, plus complexes. Pour éviter une prolifération d'indices, on sera souvent conduit à adopter des écritures vectorielles ou matricielles.

Précisons le vocabulaire et les notations employés dans la suite de ce chapitre. On appelle **corégionalisation**¹ la donnée de l'ensemble des variables régionalisées $\{z_i, i = 1 \dots N\}$; celles-ci sont définies sur un même domaine borné de $\mathbb{R}^d$, appelé **champ** de la corégionalisation et noté $\mathcal{D}$. On parle d'**homotopie** ou d'**isotopie** lorsque toutes les variables sont mesurées en tous les points d'échantillonnage, d'**hétérotopie totale** lorsqu'elles sont mesurées sur des ensembles de points disjoints, d'**hétérotopie partielle** lorsqu'une partie seulement des points de mesure sont communs à toutes les variables (souvent aussi, l'ensemble des points de mesure d'une variable est inclus dans celui d'une autre variable). Enfin, conformément à l'usage, les fonctions aléatoires associées aux variables régionalisées seront notées en majuscules, à savoir : $\{Z_i, i = 1 \dots N\}$.

¹ Par abus de langage, ce terme désigne parfois l'ensemble des fonctions aléatoires associées aux différentes variables régionalisées.

I. Outils structuraux

Considérons une corégionalisation $\{z_i, i = 1 \dots N\}$, interprétée comme une réalisation d'un processus multivarié $\{Z_i, i = 1 \dots N\}$. Nous allons définir, comme pour le cas monovarié, des outils variographiques (fonctions de covariance et variogrammes), destinés à analyser la structure spatiale *conjointe* des variables. Ces outils ne verront cependant que les relations linéaires entre les variables.

Afin de permettre l'inférence statistique, des hypothèses de stationnarité et d'ergodicité seront nécessaires. Nous allons considérer deux types de stationnarité : la stationnarité des deux premiers moments des variables étudiées (stationnarité du second ordre conjointe) et la stationnarité de deux premiers moments des accroissements des variables (hypothèse intrinsèque conjointe). Quant à l'hypothèse d'ergodicité, elle sera toujours admise et ne sera pas étudiée ici.

I.1. Fonctions de covariance simples et croisées

I.1.1. Hypothèse de stationnarité conjointe

L'hypothèse de **stationnarité du second ordre conjointe** correspond à l'invariance par translation des deux premiers moments de la loi spatiale *conjointe* des fonctions aléatoires, et s'écrit de la manière suivante :

$$\begin{cases} E[Z_i(\mathbf{x})] = m_i & \forall i = 1 \dots N, \forall \mathbf{x} \in \mathcal{D} \\ \text{cov}\{Z_i(\mathbf{x} + \mathbf{h}), Z_j(\mathbf{x})\} = E\{[Z_i(\mathbf{x} + \mathbf{h}) - m_i][Z_j(\mathbf{x}) - m_j]\} = C_{ij}(\mathbf{h}) & \forall i, j = 1 \dots N, \forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{D} \end{cases}$$

La seconde relation définit les fonctions de covariances simples² (ou directes) $C_{ii}(\mathbf{h})$ et croisées $C_{ij}(\mathbf{h})$, $i \neq j$.

L'espérance de chaque variable est constante dans le champ et la covariance entre deux valeurs (d'une même variable ou non) dépend seulement du vecteur $\mathbf{h}$ séparant leurs points d'appui. On peut adopter une présentation matricielle, en définissant, pour tout vecteur $\mathbf{h}$, la matrice $\mathbf{C}(\mathbf{h})$ des fonctions de covariances directes et croisées :

$$\mathbf{C}(\mathbf{h}) = [C_{ij}(\mathbf{h})]_{i,j=1 \dots N}$$

² Les covariances simples ne sont autres que les fonctions de covariance usuelles de chacune des variables étudiées. L'hypothèse stationnaire conjointe est donc bien une généralisation de la stationnarité du second ordre introduite dans le cas monovarié.

I.1.2. Propriétés mathématiques

- **symétries** : la fonction de covariance croisée $C_{ij}(\mathbf{h})$ n'est symétrique ni en $\mathbf{h}$ ni en i, j :

$$\text{en général, } C_{ij}(\mathbf{h}) \neq C_{ji}(\mathbf{h}) \text{ et } C_{ij}(-\mathbf{h}) \neq C_{ij}(\mathbf{h}) \text{ pour } i \neq j$$

mais en revanche on a toujours $C_{ij}(\mathbf{h}) = C_{ji}(-\mathbf{h})$.

Soit, matriciellement³ : $\mathbf{C}(\mathbf{h}) \neq \mathbf{C}^t(\mathbf{h})$ et $\mathbf{C}(\mathbf{h}) \neq \mathbf{C}(-\mathbf{h})$, mais $\mathbf{C}(\mathbf{h}) = \mathbf{C}^t(-\mathbf{h})$.

- **signe et extrema**

On ne peut faire aucune remarque sur le signe d'une fonction de covariance croisée. Sa valeur à l'origine peut même être négative, ce qui traduit alors une corrélation négative ou un "antagonisme" entre les variables correspondantes. Ainsi, une inégalité du type $|C_{ij}(\mathbf{h})| \leq C_{ij}(\mathbf{0})$ est fausse pour une covariance croisée ($i \neq j$).

En particulier, le maximum d'une fonction de covariance croisée peut être décalé de l'origine $\mathbf{h} = \mathbf{0}$. Ce décalage est fréquent dans les phénomènes temporels où une variable peut avoir un effet non instantané sur une autre variable, ce qui produit un déplacement "spatial" de la corrélation entre les deux variables ; l'évolution des variables n'est pas synchrone. Ce phénomène est connu sous le nom d'*effet de retard* et est d'un grand intérêt dans l'analyse de la structure spatiale de la corégionalisation.

- **inégalité de Cauchy-Schwarz**

$$\forall i, j = 1 \dots N, \forall \mathbf{h}, C_{ii}(\mathbf{0})C_{jj}(\mathbf{0}) \geq |C_{ij}(\mathbf{h})|^2.$$

Cette inégalité résulte directement de la définition de la covariance croisée. Il faut prendre garde au fait que $\mathbf{h}$ n'intervient que dans le second membre ; une inégalité comme $C_{ii}(\mathbf{h})C_{jj}(\mathbf{h}) \geq |C_{ij}(\mathbf{h})|^2$ est fausse, car $C_{ii}(\mathbf{h})$ peut être négatif.

- **caractère de type positif**

Les covariances simples et croisées d'un ensemble de variables conjointement stationnaires vérifient une propriété de positivité, au sens où :

$$\text{var} \left[\sum_{i=1}^N \sum_{\alpha=1}^p \lambda_{\alpha}^i Z_i(\mathbf{x}_{\alpha}) \right] = \sum_{i=1}^N \sum_{j=1}^N \sum_{\alpha=1}^p \sum_{\beta=1}^p \lambda_{\alpha}^i C_{ij}(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \lambda_{\beta}^j \geq 0$$

pour tous systèmes de points $\mathbf{x}_{\alpha}$ ($\alpha = 1 \dots p$) et de poids λ_{α}^i ($i = 1 \dots N, \alpha = 1 \dots p$)

³ L'exposant ^t indique l'opérateur de transposition.

• **représentation spectrale**

Le **théorème de Bochner** se généralise au cas multivariable : une matrice de fonctions *continues* $C(\mathbf{h})$ est une matrice de fonctions de covariance si et seulement si elle admet une représentation spectrale de la forme :

$$C(\mathbf{h}) = \int_{\mathbb{R}^d} e^{-i \langle \mathbf{k} | \mathbf{h} \rangle} F(d\mathbf{k})$$

où i est le complexe dont le carré vaut -1 et F est une matrice hermitienne⁴ de type positif :

$$\forall \lambda_1 \dots \lambda_N \in \mathbb{C}, \forall \mathbf{k} \in \mathbb{R}^d, \sum_{i=1}^N \sum_{j=1}^N \lambda_i \bar{\lambda}_j F_{ij}(d\mathbf{k}) \geq 0$$

($\bar{\lambda}_j$ désigne le complexe conjugué de λ_j).

Pour les fonctions de covariances intégrables, on peut écrire, pour toute paire d'indices (i, j) :

$$\begin{cases} C_{ij}(\mathbf{h}) = \int_{\mathbb{R}^d} e^{-i \langle \mathbf{k} | \mathbf{h} \rangle} f_{ij}(\mathbf{k}) d\mathbf{k} & \forall \mathbf{h} \in \mathbb{R}^d \\ f_{ij}(\mathbf{k}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{i \langle \mathbf{k} | \mathbf{h} \rangle} C_{ij}(\mathbf{h}) d\mathbf{h} & \forall \mathbf{k} \in \mathbb{R}^d \end{cases}$$

La fonction f_{ij} est appelée *densité interspectrale* ou *densité spectrale croisée* ou encore *interspectre* entre les variables Z_i et Z_j . Cette fonction ne possède pas les propriétés remarquables de la densité spectrale directe f_{ii} . En particulier, elle n'est ni réelle ni *a fortiori* positive. En revanche, la matrice des densités spectrales d'un ensemble de fonctions de covariances continues doit être de type positif pour toute valeur de $\mathbf{k}$. Cela implique la relation de Cauchy-Schwarz

$$|f_{ij}(\mathbf{k})|^2 \leq f_{ii}(\mathbf{k}) f_{jj}(\mathbf{k}) \quad \forall \mathbf{k} \in \mathbb{R}^d.$$

La fonction de densité spectrale croisée s'écrit de manière générale :

$$f_{ij}(\mathbf{k}) = c_{ij}(\mathbf{k}) + i q_{ij}(\mathbf{k})$$

où c_{ij} , partie réelle de f_{ij} , est appelée *cospectre*
 q_{ij} , partie imaginaire de f_{ij} , est appelée *spectre de quadrature*.

Le cospectre est la transformée de Fourier inverse de la partie paire de la fonction de covariance, tandis que le spectre de quadrature est la transformée de Fourier inverse de la partie impaire.

Si le spectre de quadrature s'annule pour toutes les fréquences, la densité interspectrale est réelle et sa transformée de Fourier (i.e. la fonction de covariance croisée) est une fonction paire. Au contraire, les fréquences responsables de l'asymétrie de la covariance croisée et des effets de retard, sont celles pour lesquelles le spectre de quadrature est non nul. L'analyse spectrale permet ainsi d'interpréter les effets de retard en termes de fréquences, ce qui est utile lorsque l'on travaille dans l'espace à une dimension, le long de l'axe temporel par exemple.

⁴ Une matrice complexe F est hermitienne si $\bar{F}^t = F$: les éléments diagonaux sont réels et les éléments non diagonaux sont égaux à la conjuguée complexe de leur transposée.

I.2. Variogrammes simples et croisés

I.2.1. L'hypothèse intrinsèque conjointe

On dit que les fonctions aléatoires Z_i , $i = 1 \dots N$, sont conjointement intrinsèques si elles vérifient⁵ :

$$\begin{cases} E[Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{x})] = 0 & \forall i = 1 \dots N, \forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{D} \\ \text{cov}\{Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{x}), Z_j(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{x})\} = 2\gamma_{ij}(\mathbf{h}) & \forall i, j = 1 \dots N, \forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{D} \end{cases}$$

La seconde relation définit les variogrammes simples (γ_{ii}) et croisés (γ_{ij} , $i \neq j$) des variables. Compte tenu de la première relation, ces variogrammes s'écrivent aussi :

$$2\gamma_{ij}(\mathbf{h}) = E\{[Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{x})][Z_j(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{x})]\} \quad \forall i, j = 1 \dots N, \forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{D}.$$

Là encore, il s'agit d'une généralisation du cas monovarié, puisque le variogramme simple $\gamma_{ii}(\mathbf{h})$ n'est autre que le variogramme habituel de la variable Z_i .

I.2.2. Propriétés théoriques

Le variogramme simple ou croisé $\gamma_{ij}(\mathbf{h})$ est une fonction symétrique en $\mathbf{h}$, et en les indices i, j :

$$\forall i, j, \gamma_{ij}(\mathbf{h}) = \gamma_{ij}(-\mathbf{h}) \quad \text{et} \quad \gamma_{ij}(\mathbf{h}) = \gamma_{ji}(\mathbf{h})$$

soit, en introduisant la matrice $\Gamma(\mathbf{h}) = [\gamma_{ij}(\mathbf{h})]_{i,j=1 \dots N}$ des variogrammes simples et croisés :

$$\Gamma(\mathbf{h}) = \Gamma(-\mathbf{h}) = \Gamma^t(\mathbf{h}).$$

⁵ L'hypothèse intrinsèque complète correspond à la stationnarité conjointe des accroissements des fonctions aléatoires. Elle s'énonce comme suit :

1) $E[Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{x})] = m_i(\mathbf{h})$ indépendant de $\mathbf{x}$

2) $\text{cov}\{Z_i(\mathbf{x}_1 + \mathbf{h}_1) - Z_i(\mathbf{x}_1), Z_j(\mathbf{x}_2 + \mathbf{h}_2) - Z_j(\mathbf{x}_2)\} = C_{ij}(\mathbf{x}_1 - \mathbf{x}_2; \mathbf{h}_1, \mathbf{h}_2)$ ne dépend que de $\mathbf{x}_1 - \mathbf{x}_2$, $\mathbf{h}_1$ et $\mathbf{h}_2$.

Mais on montre que la seconde propriété est équivalente à celle (plus simple) de l'existence des variogrammes simples et croisés ne dépendant que du seul paramètre $\mathbf{h}$. Par ailleurs, comme dans le cas monovarié, on supposera toujours l'absence de dérive : $\forall i, \forall \mathbf{h}, m_i(\mathbf{h}) = 0$.

Les variogrammes simples et croisés passent par l'origine :

$$\forall i, j, \gamma_{ij}(\mathbf{0}) = 0$$

mais les variogrammes croisés ne sont pas forcément positifs.

Une propriété liant les variogrammes simples et croisés est que la matrice $\Gamma(\mathbf{h})$ est symétrique et de type positif pour tout vecteur $\mathbf{h}$, c'est-à-dire que :

$$\forall \mathbf{h}, \forall \lambda_1, \dots, \lambda_N \in \mathbb{R}, \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j \gamma_{ij}(\mathbf{h}) \geq 0.$$

En effet,

$$\begin{aligned} \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j \gamma_{ij}(\mathbf{h}) &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j \text{cov} \{ Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{x}), Z_j(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{x}) \} \\ &= \frac{1}{2} \text{var} \left\{ \sum_{i=1}^N \lambda_i [Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{x})] \right\} \geq 0 \end{aligned}$$

En particulier, on a l'inégalité de Cauchy-Schwarz : $\gamma_{ii}(\mathbf{h}) \gamma_{jj}(\mathbf{h}) \geq |\gamma_{ij}(\mathbf{h})|^2$.

I.2.3. Lien avec la fonction de covariance croisée

L'hypothèse intrinsèque conjointe est une généralisation de l'hypothèse stationnaire conjointe. Dans le cas où cette dernière est satisfaite, on a la formule

$$\gamma_{ij}(\mathbf{h}) = C_{ij}(\mathbf{0}) - \frac{1}{2} [C_{ij}(\mathbf{h}) + C_{ij}(-\mathbf{h})]$$

qui montre que le variogramme croisé prend la moyenne de la fonction de covariance croisée aux valeurs $+\mathbf{h}$ et $-\mathbf{h}$. Matriciellement, cette relation s'écrit :

$$\Gamma(\mathbf{h}) = \mathbf{C}(\mathbf{0}) - \frac{1}{2} [\mathbf{C}(\mathbf{h}) + \mathbf{C}(-\mathbf{h})]$$

et est analogue à celle obtenue dans le cas monovarié : $\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$.

En décomposant la fonction de covariance croisée en sa partie paire et sa partie impaire, on voit que le variogramme croisé ne contient que la partie paire de la fonction de covariance croisée et "perd" de l'information :

$$C_{ij}(\mathbf{h}) = \underbrace{\frac{1}{2} [C_{ij}(-\mathbf{h}) + C_{ij}(\mathbf{h})]}_{\text{terme pair}} + \underbrace{\frac{1}{2} [C_{ij}(\mathbf{h}) - C_{ij}(-\mathbf{h})]}_{\text{terme impair}}$$

Le variogramme croisé n'est pas adéquat pour modéliser des données lorsque le terme impair de la covariance croisée joue un rôle significatif, auquel cas ce terme est souvent responsable d'un effet de retard (décalage du maximum des fonctions de covariance croisée par rapport à l'origine). Cependant, la connaissance partielle du phénomène et le nombre restreint d'observations empêchent souvent de mesurer l'importance du décalage, surtout s'il a lieu dans plusieurs directions de l'espace. En outre, la modélisation d'une fonction de covariance non paire est délicate (cf. annexe H), si bien que l'on se contente le plus souvent de ne modéliser que la partie paire, ce qui revient *in fine* à travailler sur les variogrammes simples et croisés.

Ainsi, on préfère souvent utiliser les variogrammes simples et croisés, d'autant qu'ils peuvent être définis hors du cadre stationnaire du second ordre. L'hypothèse stationnaire conjointe implique en effet l'hypothèse intrinsèque, tandis que la réciproque n'est pas certaine.

I.2.4. Pseudo-variogramme croisé

Un autre outil structural a été introduit sous le nom de *pseudo-variogramme croisé*, noté $\pi_{ij}(\mathbf{h})$. Sous l'hypothèse que les accroissements simples et croisés des fonctions aléatoires sont stationnaires d'ordre deux, il est défini par :

$$\pi_{ij}(\mathbf{h}) = \frac{1}{2} \text{var}[Z_i(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{x})]$$

et ne dépend que de $\mathbf{h}$.

Il s'agit bien d'une généralisation du variogramme, puisque l'on retrouve $\pi_{ii}(\mathbf{h}) = \gamma_{ii}(\mathbf{h})$. Mais déjà, une complication tient à la nature différente des variables : que signifie la différence $Z_i(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{x})$ entre deux variables qui ne sont *a priori* pas exprimées dans des unités identiques ? Il faudrait pouvoir ramener les variables à une même échelle, adimensionnelle par exemple.

Dans le cas stationnaire d'ordre deux, on a la relation suivante avec la covariance croisée :

$$\pi_{ij}(\mathbf{h}) = \frac{1}{2} [C_{ii}(\mathbf{0}) + C_{jj}(\mathbf{0})] - C_{ij}(\mathbf{h}).$$

Ainsi, à l'instar de la covariance croisée, le pseudo-variogramme croisé peut servir à identifier et modéliser des effets de retard.

En revanche, dans le cadre intrinsèque, pour pouvoir définir le pseudo-variogramme croisé, on doit supposer que les accroissements croisés $Z_i(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{x})$ (et pas seulement les accroissements directs) sont intrinsèques, c'est-à-dire admettent une espérance et une variance finies. C'est une hypothèse assez forte et hasardeuse lorsque les variables sont de natures différentes. Le pseudo-variogramme croisé a toutefois l'avantage de pouvoir être calculé en cas d'hétérotopie totale, alors que le variogramme croisé classique nécessite de connaître les variables concernées aux mêmes points $\mathbf{x}$ et $\mathbf{x} + \mathbf{h}$, comme nous allons le voir.

II. L'analyse variographique multivariable

II.1. Calcul des covariances et variogrammes simples et croisés expérimentaux

Comme dans le cas monovarié, les covariances et variogrammes simples et croisés expérimentaux ne sont calculables que pour certains vecteurs $\mathbf{h}$, selon les positions des échantillons. Ce ne sont donc pas des fonctions, mais des suites discrètes de points, ce qui implique la nécessité d'une modélisation future.

Notons $\{\mathbf{x}_\alpha^i, \alpha = 1 \dots n_i\}$ les sites de mesure de la variable z_i ; leur nombre et leur position peuvent différer d'une variable à l'autre (cas d'hétérotopie, totale ou partielle).

La fonction de covariance (simple ou croisée) $C_{ij}(\mathbf{h})$ est habituellement estimée par la méthode des moments :

$$\hat{C}_{ij}(\mathbf{h}) = \frac{1}{|N_{ij}(\mathbf{h})|} \sum_{N_{ij}(\mathbf{h})} [z_i(\mathbf{x}_\alpha^i) - \bar{z}_i][z_j(\mathbf{x}_\beta^j) - \bar{z}_j]$$

où $\bar{z}_i = \frac{1}{n_i} \sum_{\alpha=1}^{n_i} z_i(\mathbf{x}_\alpha^i)$ et $\bar{z}_j = \frac{1}{n_j} \sum_{\beta=1}^{n_j} z_j(\mathbf{x}_\beta^j)$ sont des estimateurs des espérances de Z_i et Z_j

$$N_{ij}(\mathbf{h}) = \{ (\alpha, \beta) \text{ tel que } \mathbf{x}_\alpha^i - \mathbf{x}_\beta^j = \mathbf{h} \} ;$$

$|N_{ij}(\mathbf{h})|$ est le nombre de paires distinctes de l'ensemble $N_{ij}(\mathbf{h})$.

Quant au variogramme simple ou croisé $\gamma_{ij}(\mathbf{h})$, il est estimé classiquement par :

$$\hat{\gamma}_{ij}(\mathbf{h}) = \frac{1}{2|N_{ij}(\mathbf{h})|} \sum_{N_{ij}(\mathbf{h})} [z_i(\mathbf{x}_\alpha) - z_i(\mathbf{x}_\beta)][z_j(\mathbf{x}_\alpha) - z_j(\mathbf{x}_\beta)]$$

où cette fois $N_{ij}(\mathbf{h}) = \{ (\alpha, \beta) \text{ tel que } \mathbf{x}_\alpha - \mathbf{x}_\beta = \mathbf{h}, \text{ à la fois } z_i \text{ et } z_j \text{ étant mesurées en } \mathbf{x}_\alpha \text{ et } \mathbf{x}_\beta \}$. Le variogramme croisé nécessite d'avoir les données concernant les différentes variables aux mêmes points. Il n'est donc pas calculable en cas d'hétérotopie totale (auquel cas l'ensemble $N_{ij}(\mathbf{h})$ est vide pour tout vecteur $\mathbf{h}$), contrairement à la covariance croisée.

Remarquons que $\hat{\gamma}_{ij}(\mathbf{h}) = \hat{\gamma}_{ij}(-\mathbf{h})$, tandis que $\hat{C}_{ij}(\mathbf{h}) \neq \hat{C}_{ij}(-\mathbf{h})$ mais $\hat{C}_{ij}(\mathbf{h}) = \hat{C}_{ji}(-\mathbf{h})$.

Pour le calcul pratique, on utilise des critères analogues au cas monovarié (c'est-à-dire que l'on introduit des tolérances à la fois sur les distances et les angles) afin de robustifier les estimateurs.

II.2. Le modèle linéaire de corégionalisation

Les covariances ou variogrammes simples et croisés d'un ensemble de fonctions aléatoires $Z_1 \dots Z_N$ ne peuvent être modélisés indépendamment, car il existe entre eux des contraintes mathématiques. La matrice $\Gamma(\mathbf{h})$ des variogrammes simples et croisés doit notamment être symétrique de type positif pour tout vecteur $\mathbf{h}$. En pratique, pour satisfaire ces contraintes, on a très souvent recours au *modèle linéaire de corégionalisation*, qui permet de modéliser une grande classe de phénomènes⁶. Ce modèle généralise le modèle linéaire de régionalisation vu dans le cas monovarié. Nous reviendrons sur son interprétation lors de l'étude des techniques d'analyse krigéante.

II.2.1. Cas stationnaire du second ordre

Le modèle linéaire de corégionalisation de plusieurs fonctions aléatoires conjointement stationnaires d'ordre deux, fait l'hypothèse que toutes les covariances simples et croisées sont des combinaisons linéaires des mêmes structures de base :

$$\forall i, j \in [1, N], C_{ij}(\mathbf{h}) = \sum_{u=1}^S b_{ij}^u \rho_u(\mathbf{h})$$

en écriture matricielle,

$$\mathbf{C}(\mathbf{h}) = \sum_{u=1}^S \mathbf{B}_u \rho_u(\mathbf{h})$$

où $\mathbf{C}(\mathbf{h}) = [C_{ij}(\mathbf{h})]_{i,j}$ est la matrice des covariances simples et croisées

$\rho_u(\mathbf{h})$ est une covariance (fonction paire, de type positif) telle que $\rho_u(\mathbf{0}) = 1$

$\mathbf{B}_u = (b_{ij}^u)_{i,j}$ est appelée *matrice de corégionalisation*.

Ce modèle suppose en particulier que les fonctions de covariance croisées sont paires, en tant que combinaisons linéaires de fonctions ρ_u ayant cette propriété :

$$\mathbf{C}(\mathbf{h}) = \mathbf{C}(-\mathbf{h}) = \mathbf{C}^t(\mathbf{h}).$$

On ne pourra donc pas décrire des effets de retard à l'aide d'un tel modèle.

On sait de plus que pour tout système de points $\mathbf{x}_\alpha$ ($\alpha = 1 \dots p$) et tout système de poids λ_α^i ($i = 1 \dots N, \alpha = 1 \dots p$), on doit avoir :

$$\sum_{i=1}^N \sum_{j=1}^N \sum_{\alpha=1}^p \sum_{\beta=1}^p \lambda_\alpha^i C_{ij}(\mathbf{x}_\alpha - \mathbf{x}_\beta) \lambda_\beta^j \geq 0.$$

⁶ D'autres modèles multivariés, d'usage plus spécialisé, sont présentés en annexe H.

Cette inégalité est satisfaite si chacune des matrices de corégionalisation $\mathbf{B}_u$ est symétrique et de type positif.

En effet, si $\mathbf{B}_u$ est de type positif, il existe une matrice $\mathbf{A}_u$ telle que : $\mathbf{B}_u = \mathbf{A}_u^t \mathbf{A}_u$, ce qui revient à écrire le terme générique de $\mathbf{B}_u$ sous la forme :

$$b_{ij}^u = \sum_{k=1}^N a_{ki}^u a_{kj}^u.$$

On en déduit, compte tenu du modèle :

$$\begin{aligned} \sum_{i=1}^N \sum_{j=1}^N \sum_{\alpha=1}^p \sum_{\beta=1}^p \lambda_{\alpha}^i C_{ij}(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \lambda_{\beta}^j &= \sum_{i=1}^N \sum_{j=1}^N \sum_{\alpha=1}^p \sum_{\beta=1}^p \sum_{u=1}^S \lambda_{\alpha}^i b_{ij}^u \rho_u(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \lambda_{\beta}^j \\ &= \sum_{u=1}^S \sum_{k=1}^N \sum_{\alpha=1}^p \sum_{\beta=1}^p \underbrace{\left(\sum_{i=1}^N \lambda_{\alpha}^i a_{ki}^u \right)}_{\eta_{\alpha}^{u,k}} \underbrace{\left(\sum_{j=1}^N \lambda_{\beta}^j a_{kj}^u \right)}_{\eta_{\beta}^{u,k}} \rho_u(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \\ &= \sum_{u=1}^S \sum_{k=1}^N \underbrace{\left(\sum_{\alpha=1}^p \sum_{\beta=1}^p \eta_{\alpha}^{u,k} \eta_{\beta}^{u,k} \rho_u(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \right)}_{\geq 0} \geq 0 \end{aligned}$$

car ρ_u est une fonction de type positif

L'inégalité précédente est donc bien toujours vérifiée lorsque toutes les matrices $\mathbf{B}_u$ sont de type positif. Ceci est plus restrictif mais très facile à contrôler : toutes les valeurs propres des $\mathbf{B}_u$ doivent être positives. En particulier, on aura les relations (nécessaires mais non suffisantes) :

$$\forall i, j \in [1, N], \forall u \in [1, S], |b_{ij}^u| \leq \sqrt{b_{ii}^u b_{jj}^u}.$$

Ces inégalités impliquent qu'une structure ρ_u dont le palier est non nul sur une covariance croisée présente nécessairement un palier non nul sur les deux covariances simples correspondantes.

II.2.2. Cas intrinsèque strict

Soient à présent N fonctions aléatoires conjointement intrinsèques. Le modèle linéaire de corégionalisation suppose cette fois que les *variogrammes* simples et croisés sont des combinaisons linéaires des mêmes structures de base :

$$\forall i, j \in [1, N], \gamma_{ij}(\mathbf{h}) = \sum_{u=1}^S b_{ij}^u g_u(\mathbf{h})$$

matriciellement

$$\Gamma(\mathbf{h}) = \sum_{u=1}^S \mathbf{B}_u g_u(\mathbf{h})$$

où $\Gamma(\mathbf{h}) = [\gamma_{ij}(\mathbf{h})]_{i,j}$ est la matrice des variogrammes simples et croisés

$g_u(\mathbf{h})$ sont des variogrammes de base

$\mathbf{B}_u = (b_{ij}^u)_{i,j}$ sont les *matrices de corégionalisation*.

Comme précédemment, $\Gamma(\mathbf{h})$ devant être une matrice symétrique de type positif pour tout $\mathbf{h}$, on impose (condition suffisante) que *chacune* des matrices de corégionalisation $\mathbf{B}_u$ soit symétrique de type positif.

II.3. Ajustement pratique

Nous allons présenter brièvement les méthodes utilisées pour ajuster un modèle linéaire de corégionalisation aux variogrammes expérimentaux. Ces méthodes restent valables dans le cas des covariances simples et croisées expérimentales, à condition de travailler sur leur partie paire, puisque le modèle linéaire de corégionalisation impose que les covariances théoriques, simples et croisées, soient symétriques en $\mathbf{h}$.

II.3.1. Cas bivariable

Dans le cas de deux variables Z_1 et Z_2 , il est facile de définir une procédure d'ajustement du modèle aux variogrammes expérimentaux. On commence par choisir les structures de base qui interviendront dans le modèle bivariable, au vu des variogrammes expérimentaux et de la connaissance que l'on a du phénomène. On ajuste ensuite les variogrammes simples, et enfin le variogramme croisé en utilisant, pour chaque structure de base, un palier vérifiant l'inégalité de Cauchy-Schwarz :

$$|b_{12}^u| \leq \sqrt{b_{11}^u b_{22}^u}.$$

Les paliers peuvent être ajustés manuellement, ou automatiquement à l'aide de techniques d'optimisation sous contraintes.

L'extension de cette procédure à plus de deux variables ne garantit pas l'obtention d'un modèle autorisé, car l'inégalité de Cauchy-Schwarz ne suffit pas pour assurer que la matrice de corégionalisation soit de type positif.

II.3.2. Cas multivariable

On rappelle que le modèle de corégionalisation s'écrit : $\Gamma(\mathbf{h}) = [\gamma_{ij}(\mathbf{h})]_{i,j=1\dots N} = \sum_{u=1}^S \mathbf{B}_u g_u(\mathbf{h})$.

Les inconnues du problème sont les coefficients des matrices de corégionalisation $\mathbf{B}_u$, à déterminer de sorte que la matrice $\Gamma(\mathbf{h})$ soit un modèle autorisé le plus "conforme" possible aux matrices de variogrammes expérimentaux $\hat{\Gamma}(\mathbf{h}_k)$, $k = 1 \dots K$. Tous les autres paramètres du modèle (nombre, type, portée, anisotropie des structures de base g_u) doivent être spécifiés par l'utilisateur.

On présente ci-après un algorithme d'ajustement automatique basé sur un critère de moindres carrés pondérés. On note $\{\hat{\sigma}_i, i = 1 \dots N\}$ les écarts-types empiriques des variables $\{Z_i, i = 1 \dots N\}$. On cherche un modèle $\Gamma(\mathbf{h})$ qui minimise :

$$\sum_{k=1}^K \omega_k \|\hat{\Gamma}(\mathbf{h}_k) - \Gamma(\mathbf{h}_k)\|^2 = \sum_{k=1}^K \omega_k \frac{\sum_{i,j} [\hat{\gamma}_{ij}(\mathbf{h}_k) - \gamma_{ij}(\mathbf{h}_k)]^2}{\hat{\sigma}_i \hat{\sigma}_j}$$

sous la contrainte que les matrices de corégionalisation $\mathbf{B}_u$ définissant $\Gamma(\mathbf{h})$ soient de type positif. Le poids ω_k , $k = 1 \dots K$, peut être choisi proportionnel au numéro de paires de données qui interviennent dans le calcul de $\hat{\Gamma}(\mathbf{h}_k)$, et inversement proportionnel à la longueur de $\mathbf{h}_k$ (de cette façon, on privilégie l'ajustement des petites distances, qui sont les plus intéressantes). La division par les écarts-types empiriques permet de donner la même importance à chaque variable.

La procédure proposée est itérative. Etant donné $S-1$ matrices $\mathbf{B}_u$ ($u \in [1, S] - \{v\}$) de type positif, une étape élémentaire de cette procédure consiste à chercher une S -ième matrice $\mathbf{B}_v$ de type positif qui minimise (au sens du critère ci-dessus) l'écart avec les matrices expérimentales $\hat{\Gamma}(\mathbf{h}_k)$.

Pour cela, on définit une matrice $d\hat{\Gamma}_v$ par :

$$d\hat{\Gamma}_v = \sum_{k=1}^K \omega_k g_v(\mathbf{h}_k) [\hat{\Gamma}(\mathbf{h}_k) - \sum_{u=1, u \neq v}^S \mathbf{B}_u g_u(\mathbf{h}_k)].$$

On décompose $d\hat{\Gamma}_v$ en valeurs propres et vecteurs propres orthonormés par rapport à une métrique qui dépend des écarts-types empiriques :

$$d\hat{\Gamma}_v = \mathbf{Q}_v \Lambda_v \mathbf{Q}_v^t \quad \text{avec} \quad \mathbf{Q}_v^t \mathbf{M} \mathbf{Q}_v = \mathbf{I}, \quad \Lambda_v = \begin{pmatrix} \lambda_v^1 & & 0 \\ & \ddots & \\ 0 & & \lambda_v^N \end{pmatrix} \quad \text{et} \quad \mathbf{M} = \begin{pmatrix} \frac{1}{\hat{\sigma}_1^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\hat{\sigma}_N^2} \end{pmatrix}.$$

On obtient la matrice $\mathbf{B}_v$ cherchée en posant :

$$\mathbf{B}_v = \frac{\mathbf{Q}_v \Lambda_v^+ \mathbf{Q}_v^t}{\sum_{k=1}^K \omega_k [g_v(\mathbf{h}_k)]^2} \quad \text{avec} \quad \Lambda_v^+ = \begin{pmatrix} \max(0, \lambda_v^1) & & 0 \\ & \ddots & \\ 0 & & \max(0, \lambda_v^N) \end{pmatrix}.$$

Le fait de considérer Λ_v^+ au lieu de Λ_v assure que $\mathbf{B}_v$ soit de type positif.

On répète l'algorithme pour chaque indice $v \in [1, S]$, puis on l'itère jusqu'à obtenir stabilisation numérique des matrices de corégionalisation.

II.4. Illustration sur les données de sol

Nous allons étudier, sur les données de sol (cf. chapitres 2, 3, 5, 6), les variables

- taux d'argile et limon
- taux de terre fine (granulométrie < 2mm)
- pF4.2 (humidité au point de flétrissement permanent) mesurée sur la terre fine ; cette variable représente le taux d'humidité dans le sol que les plantes ne peuvent pas absorber.

Les tableaux ci-dessous résument les principales statistiques relatives à ces trois variables.

110 données	minimum	maximum	moyenne	médiane	variance	écart-type
% argile + limon	25.2	60.7	48.44	48.30	37.93	6.16
% terre fine	55.7	100	90.73	92.10	58.32	7.64
pF4.2	8.4	14.5	11.29	11.1	2.02	1.42

tableau 7.1 : statistiques monovariabes sur les 110 données issues de la grille régulière

170 données	minimum	maximum	moyenne	médiane	variance	écart-type
% argile + limon	25.2	60.7	47.62	47.70	33.76	5.81
% terre fine	55.7	100	90.25	91.00	56.40	7.51
pF4.2	8.4	14.5	11.19	11.1	2.00	1.41

tableau 7.2 : statistiques monovariabes sur les 170 données
(grille régulière + croix de sondage)

Les distributions des valeurs de % terre fine et pF4.2 peuvent être représentées par les histogrammes expérimentaux de ces variables. Ceux-ci ne mettent en évidence aucune valeur aberrante.

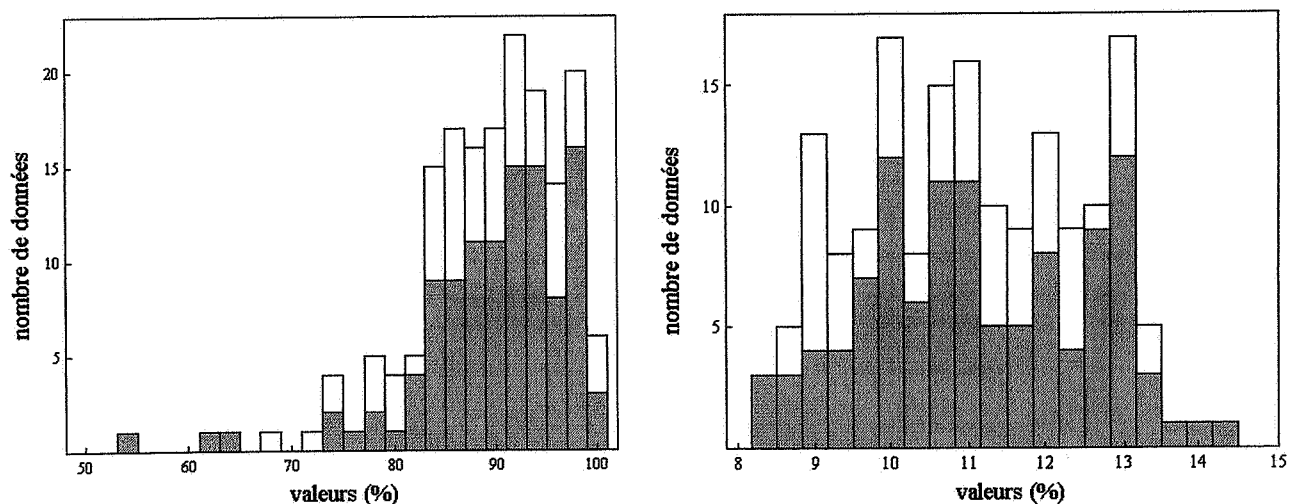


figure 7.1 : histogrammes des valeurs de % terre fine (à gauche) et pF4.2 (à droite)
gris = échantillons issus de la grille régulière, blanc = échantillons issus des croix de sondage

En ce qui concerne les statistiques multivariées (tableaux 7.3 et 7.4), on remarque que la variable % argile + limon est très corrélée avec la variable % terre fine (coefficient de corrélation linéaire égal à 0.82), et peu corrélée avec $pF4.2$ (coefficient de corrélation de 0.057). Cette analyse est confirmée par l'examen des nuages de corrélation des couples % argile + limon / % terre fine et % argile + limon / $pF4.2$ (figure 7.2) : le premier met en évidence une relation linéaire, tandis que le second est beaucoup plus diffus.

	% argile + limon	% terre fine	$pF4.2$
% argile + limon	38.3	38.7	0.50
% terre fine	38.7	58.9	-3.30
$pF4.2$	0.50	-3.30	2.04

tableau 7.3 : matrice de variance-covariance des trois variables calculée sur les 110 données de la grille régulière

	% argile + limon	% terre fine	$pF4.2$
% argile + limon	1	0.82	0.057
% terre fine	0.82	1	-0.30
$pF4.2$	0.057	-0.30	1

tableau 7.4 : matrice de corrélation des trois variables calculée sur les 110 données de la grille régulière

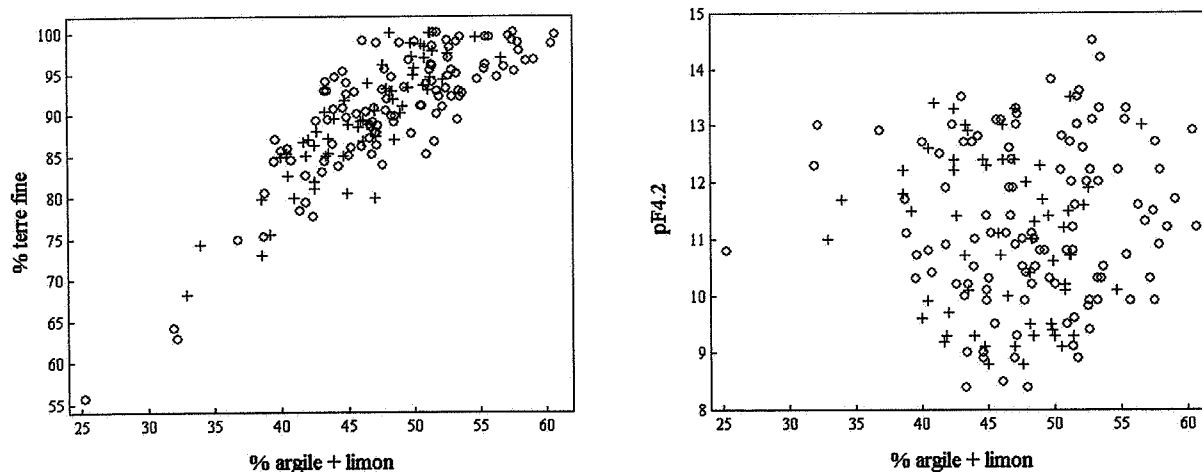


figure 7.2 : nuages de corrélation des couples

% argile + limon / % terre fine (à gauche) et % argile + limon / $pF4.2$ (à droite)

“o” = échantillons issus de la grille régulière, “+” = échantillons issus des croix de sondage

Les cartes variographiques des variables % terre fine et $pF4.2$ (figure 7.3) présentent une nette anisotropie, en première approximation de même angle que celle de la variable % argile + limon, soit environ 70° par rapport à la direction est-ouest.

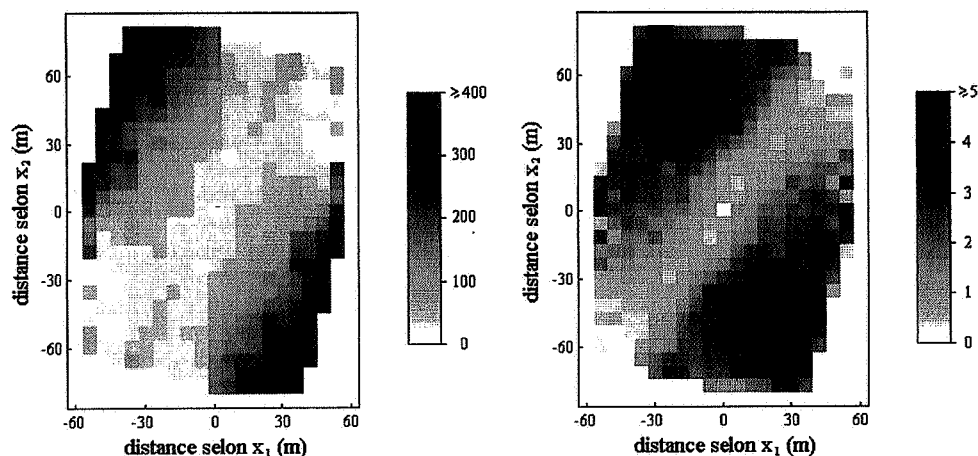


figure 7.3 : cartes variographiques de % terre fine (à gauche) et pF4.2 (à droite)

L'étude exploratoire de la stationnarité (chapitre 2) et l'analyse variographique (chapitre 3) ont montré que la corégionalisation n'est pas stationnaire (au moins en ce qui concerne la variable % argile + limon, mais la même conclusion prévaut pour les variables % terre fine et pF4.2). Le calcul des covariances simples et croisées n'a donc pas lieu d'être, la fonction de covariance relative au taux d'argile et limon n'existant pas. Pour analyser les liens spatiaux entre les trois variables, nous calculons donc les variogrammes expérimentaux simples et croisés, le long des directions principales d'anisotropie.

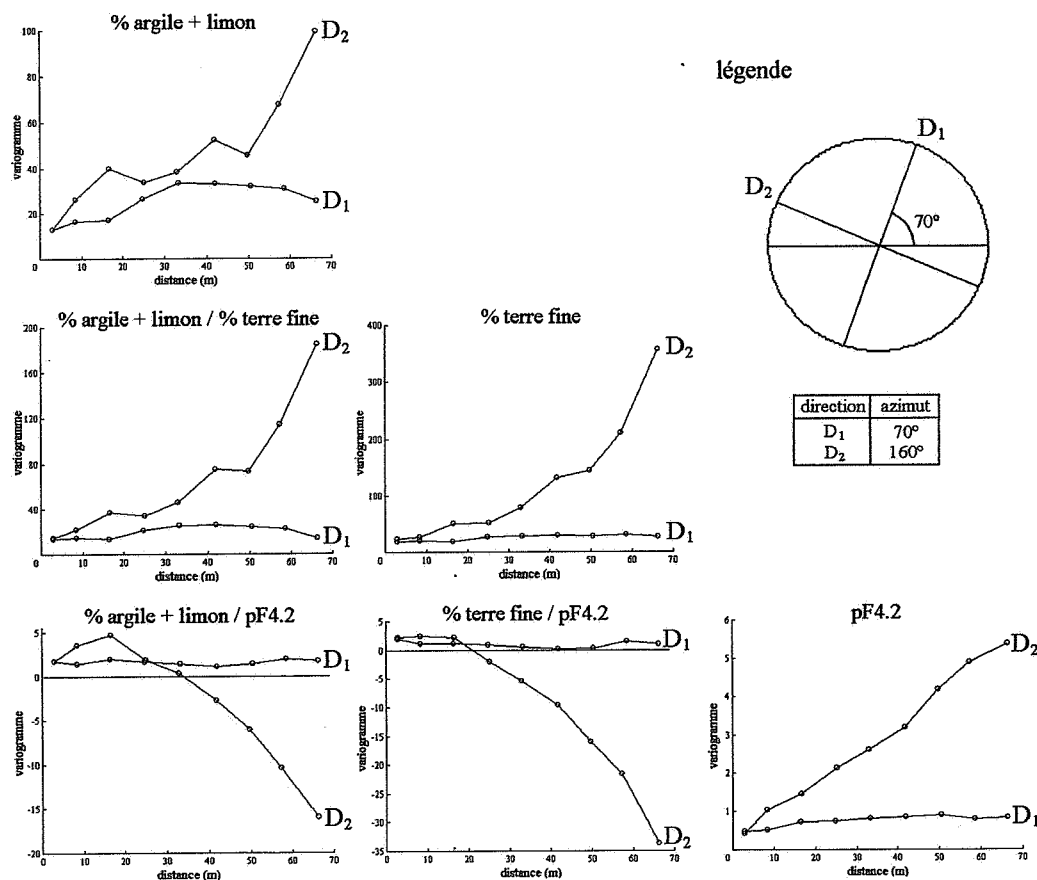


figure 7.4 : variogrammes simples et croisés anisotropes pour les données de sol

Ces variogrammes expérimentaux montrent que la structure des variables *% terre fine* et *pF4.2* est sensiblement la même que celle de la variable *% argile + limon* : les variogrammes simples sont anisotropes et, au moins le long de la direction D_2 (ouest-nord-ouest / est-sud-est), n'atteignent pas de palier, signe de non-stationnarité. D'autre part, les variogrammes croisés entre les trois variables sont porteurs d'information et quantifient les corrélations spatiales entre les accroissements de ces variables ; ils augmentent (en valeur absolue) avec la distance dans la direction D_2 , tandis qu'ils restent quasiment constants dans la direction orthogonale (D_1).

Les comportements des variogrammes simples et croisés aux courtes distances sont mis en évidence à l'aide des échantillons issus des croix de sondage. Ils présentent tous un effet de pépité à l'origine.

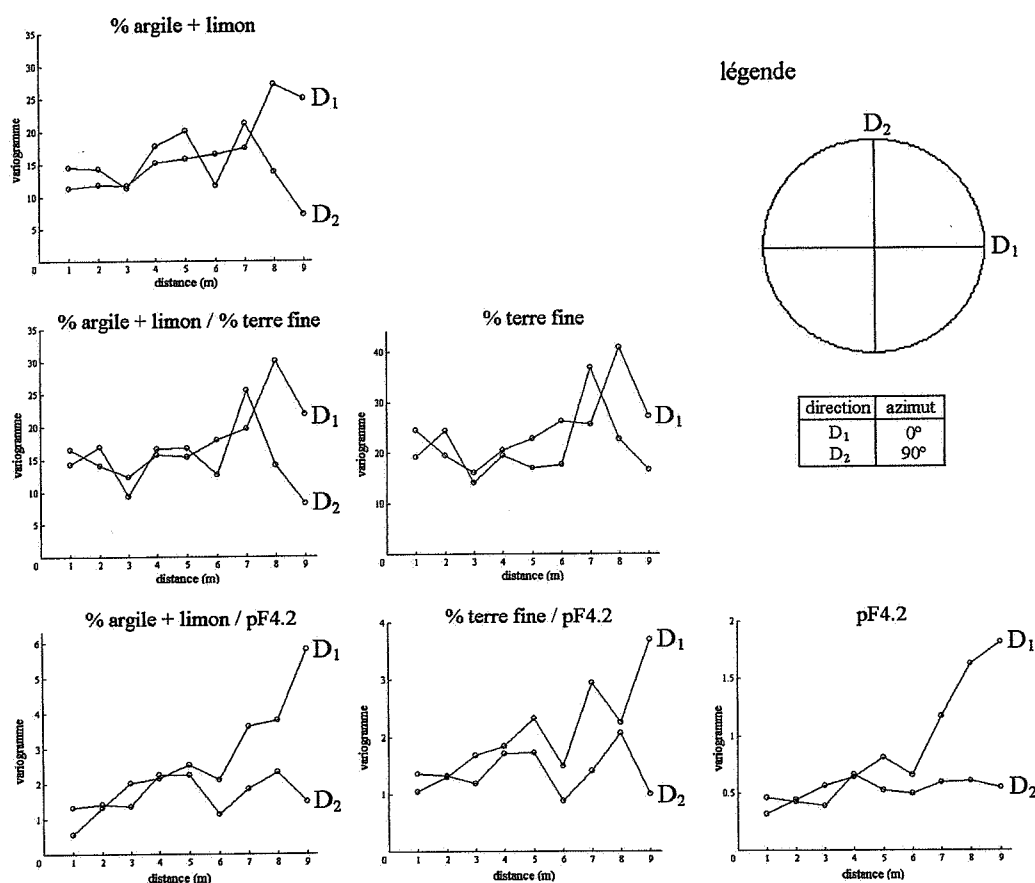


figure 7.5 : comportement aux courtes distances des variogrammes simples et croisés, calculés le long des croix de sondage

Les variogrammes simples et croisés ayant les mêmes caractères en ce qui concerne l'anisotropie, le comportement à l'origine (effet de pépité) et le comportement aux grandes distances (non-stationnarité), ils peuvent être modélisés par le même ensemble de structures de base (modèle linéaire de corégionalisation). Nous proposons la modélisation suivante, où interviennent les trois structures introduites dans le modèle monovarié relatif au taux d'argile et limon (cf. chapitre 3) : un effet de pépité, un schéma sphérique (avec anisotropie géométrique d'angle 70° par rapport à la direction est-ouest, portées 40m et 20m) et un modèle puissance $\gamma(\mathbf{h}) = |\mathbf{h}|^{1.8}$ (même angle d'anisotropie, coefficient d'anisotropie 1/5).

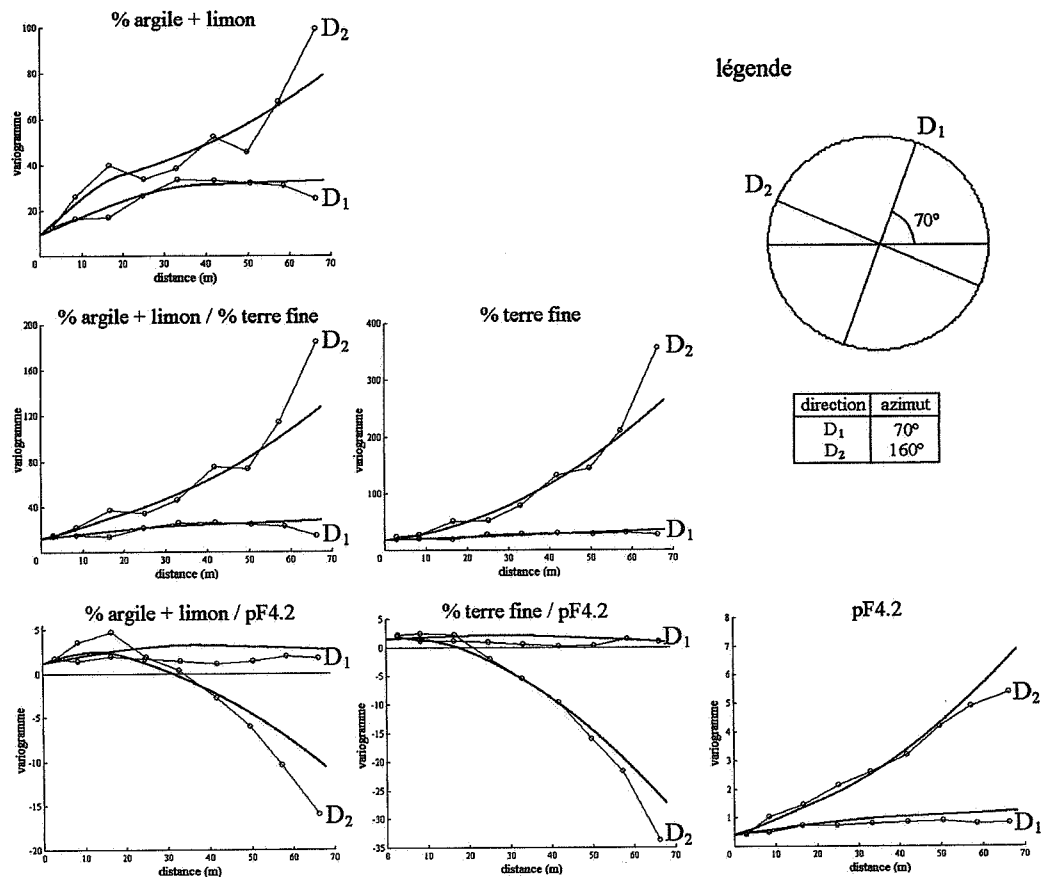


figure 7.6 : modèles simples et croisés anisotropes pour les données de sol

L'ajustement paraît satisfaisant. Comme dans le cas monovarié, on pourrait utiliser une technique de validation croisée pour tester la qualité du modèle. On aura tout intérêt à réaliser l'étude simultanée des trois variables plutôt que leur étude séparée, de façon à tenir compte des liens existant entre ces variables. A titre indicatif, nous donnons ci-dessous les paramètres du modèle.

Anisotropie géométrique d'angle 70° par rapport à l'axe est-ouest

Ordre des variables : % argile + limon - % terre fine - pF4.2

Structure 1 - Effet de pépité

Matrice de corégionalisation B_1

9.40	11.8	1.20
11.8	17.6	1.40
1.20	1.40	0.40

Structure 2 - Schéma sphérique de portées 40 m (D_1) et 20 m (D_2)

Coefficients d'anisotropie = (0.5 = 20/40 ; 1)

Matrice de corégionalisation B_2

20.9	9.80	2.30
9.80	4.70	1.15
2.30	1.15	0.50

Structure 3 - Modèle puissance - Exposant = 1.8

Coefficients d'anisotropie = (0.2 ; 1)

Matrice de corégionalisation B_3

0.0245	0.0540	-0.0071
0.0540	0.1220	-0.0150
-0.0071	-0.0150	0.0030

III. Le cokrigage

L'objectif du cokrigage est d'estimer la valeur d'une variable en un site ou un bloc à partir des mesures concernant non seulement cette variable, mais aussi une ou plusieurs variables additionnelles observées sur des sites voisins. L'estimation va tenir compte de la configuration géométrique des échantillons et du site à estimer, et de la structure spatiale *conjointe* de toutes les variables régionalisées, modélisée par les covariances ou variogrammes simples et croisés.

Il s'agit d'une extension de la technique du krigage au cas multivariable : l'estimation par cokrigage se construit comme une combinaison linéaire pondérée des données, sans biais et de variance d'erreur minimale. Bien qu'il nécessite des calculs plus volumineux que le krigage, le cokrigage est toujours préférable au krigage. En effet, l'information apportée par les variables secondaires améliore la précision de l'estimation de la variable cible ; cette propriété peut être décisive lorsqu'il est plus facile ou moins coûteux de mesurer des variables secondaires que la variable d'intérêt. D'autre part, l'estimation par cokrigage, en intégrant les liaisons spatiales entre variables, donne plus de cohérence aux résultats que ceux obtenus séparément par krigage.

III.1. Cokrigage ordinaire stationnaire

Nous ne présentons ici que le cokrigage ponctuel. Mais de même que dans le cas monovarié, on peut estimer toute fonction linéaire d'une variable donnée, les systèmes d'équations auxquels on aboutit étant linéaires ; il suffirait simplement de réécrire les seconds membres des systèmes de cokrigage.

III.1.1. Cas général

Nous cherchons à estimer une variable cible z_{i_0} (avec $i_0 \in [1, N]$) en un point $\mathbf{x}_0$ à l'aide des observations portant sur les variables z_i , $i = 1 \dots N$.

L'estimation est une combinaison linéaire des valeurs de toutes les variables, que l'on peut écrire sous la forme :

$$z_{i_0}^*(\mathbf{x}_0) = \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i z_i(\mathbf{x}_{\alpha}^i)$$

où $\{\mathbf{x}_{\alpha}^i, \alpha = 1 \dots n_i\}$ sont les points de mesure de la variable z_i , voisins de $\mathbf{x}_0$. Leur nombre dépend de l'indice de la variable, afin de prendre en compte le cas où les variables ne sont pas connues aux mêmes sites (*hétérotopie*).

La mise en œuvre du cokrigage est identique à celle du krigeage ordinaire. On suppose que les variables régionalisées z_i sont les réalisations de fonctions aléatoires Z_i ($i = 1 \dots N$) conjointement stationnaires d'ordre deux, de moyennes respectives m_i **inconnues** et de covariances simples et croisées $C_{ij}(\mathbf{h})$. Dans ce modèle, l'estimateur devient une variable aléatoire

$$Z_{i_0}^*(\mathbf{x}_0) = \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i Z_i(\mathbf{x}_{\alpha}^i).$$

La contrainte d'autorisation est toujours vérifiée, car on s'est placé dans le cadre stationnaire d'ordre deux (toute combinaison linéaire admet une espérance et une variance finies). Les poids de krigeage doivent être choisis de telle manière qu'ils satisfassent les deux conditions :

- l'estimation de l'erreur doit être sans biais
- la variance de l'erreur d'estimation doit être minimale (critère d'optimalité).

L'espérance de l'erreur d'estimation s'écrit :

$$\begin{aligned} E[Z_{i_0}^*(\mathbf{x}_0) - Z_{i_0}(\mathbf{x}_0)] &= \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i \underbrace{E[Z_i(\mathbf{x}_{\alpha}^i)]}_{m_i} - m_{i_0} \\ &= m_{i_0} \left(\sum_{\alpha=1}^{n_{i_0}} \lambda_{\alpha}^{i_0} - 1 \right) + \sum_{i \neq i_0} \left(m_i \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i \right) \end{aligned}$$

Les espérances étant inconnues, le seul moyen de garantir la condition de non biais est de poser

$$\sum_{\alpha=1}^{n_{i_0}} \lambda_{\alpha}^{i_0} = 1 \quad \text{et} \quad \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i = 0 \quad \text{pour } i \neq i_0.$$

Il reste à chercher les poids de cokrigage minimisant la variance d'estimation sous les contraintes précédentes. La variance de l'erreur d'estimation se développe en :

$$\text{var}[Z_{i_0}^*(\mathbf{x}_0) - Z_{i_0}(\mathbf{x}_0)] = \sum_{i=1}^N \sum_{j=1}^N \sum_{\alpha=1}^{n_i} \sum_{\beta=1}^{n_j} \lambda_{\alpha}^i \lambda_{\beta}^j C_{ij}(\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j) - 2 \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i C_{ii_0}(\mathbf{x}_{\alpha}^i - \mathbf{x}_0) + C_{i_0 i_0}(\mathbf{0})$$

La minimisation de cette quantité sous les contraintes de non-biais nécessite d'introduire des multiplicateurs de Lagrange μ_i , $i = 1 \dots N$.

On aboutit au système final

$$\begin{cases} \sum_{j=1}^N \sum_{\beta=1}^{n_j} \lambda_{\beta}^j C_{ij} (\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j) + \mu_i = C_{ii_0} (\mathbf{x}_{\alpha}^i - \mathbf{x}_0) & \forall i=1 \dots N, \forall \alpha=1 \dots n_i \\ \sum_{\beta=1}^{n_i} \lambda_{\beta}^i = \delta_{i_0}^i \quad (1 \text{ si } i=i_0, 0 \text{ sinon}) & \forall i=1 \dots N \end{cases}$$

On peut s'apercevoir de la relative complexité des notations. Il est plus commode d'écrire le système de cokrigage sous forme matricielle :

$$\begin{pmatrix} C_{11}(\mathbf{x}_{\alpha}^1 - \mathbf{x}_{\beta}^1) & 1 & \dots & C_{1i_0}(\mathbf{x}_{\alpha}^1 - \mathbf{x}_{\beta}^{i_0}) & 0 & \dots & C_{1N}(\mathbf{x}_{\alpha}^1 - \mathbf{x}_{\beta}^N) & 0 \\ 1 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{i_0 1}(\mathbf{x}_{\alpha}^{i_0} - \mathbf{x}_{\beta}^1) & 0 & \dots & C_{i_0 i_0}(\mathbf{x}_{\alpha}^{i_0} - \mathbf{x}_{\beta}^{i_0}) & 1 & \dots & C_{i_0 N}(\mathbf{x}_{\alpha}^{i_0} - \mathbf{x}_{\beta}^N) & 0 \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{N1}(\mathbf{x}_{\alpha}^N - \mathbf{x}_{\beta}^1) & 0 & \dots & C_{Ni_0}(\mathbf{x}_{\alpha}^N - \mathbf{x}_{\beta}^{i_0}) & 0 & \dots & C_{NN}(\mathbf{x}_{\alpha}^N - \mathbf{x}_{\beta}^N) & 1 \\ 0 & 0 & \dots & 0 & 0 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_{\beta}^1 \\ \mu_1 \\ \vdots \\ \lambda_{\beta}^{i_0} \\ \mu_{i_0} \\ \vdots \\ \lambda_{\beta}^N \\ \mu_N \end{pmatrix} = \begin{pmatrix} C_{i_0 1}(\mathbf{x}_{\alpha}^1 - \mathbf{x}_0) \\ 0 \\ \vdots \\ C_{i_0 i_0}(\mathbf{x}_{\alpha}^{i_0} - \mathbf{x}_0) \\ 1 \\ \vdots \\ C_{i_0 N}(\mathbf{x}_{\alpha}^N - \mathbf{x}_0) \\ 0 \end{pmatrix}$$

La matrice du membre gauche est construite avec des blocs rectangulaires $[C_{ij}(\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j)]$ de taille $n_i \times n_j$, complétés par des vecteurs de 0 ou de 1 ; de plus, elle est symétrique car $C_{ij}(\mathbf{h}) = C_{ji}(-\mathbf{h})$. Quant aux blocs $[\lambda_{\beta}^i]$ et $[C_{ii_0}(\mathbf{x}_{\alpha}^i - \mathbf{x}_0)]$, ils sont de taille $n_i \times 1$.

Le premier membre du système de cokrigage ne dépend pas de l'indice i_0 de la variable à estimer et sera donc le même pour chaque variable. Par conséquent, il suffit d'inverser une seule matrice pour obtenir l'estimation de toutes les variables au point $\mathbf{x}_0$. On remarque toutefois la taille importante de la matrice à inverser (par rapport au cas monovariable). Par exemple, cette taille est doublée lorsqu'on mesure deux variables aux mêmes sites. Ainsi, les temps de calcul deviennent vite importants lorsqu'on augmente le nombre de variables. C'est une des limites du cokrigage : pour fixer les idées, on ne pourra guère envisager de cokriger plus de cinq variables simultanément.

La variance de cokrigage est donnée par l'expression :

$$\sigma_{CKO}^2(\mathbf{x}_0, i_0) = C_{i_0 i_0}(\mathbf{0}) - \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i C_{ii_0}(\mathbf{x}_{\alpha}^i - \mathbf{x}_0) - \mu_{i_0}$$

où interviennent les covariances croisées entre la variable cible et les autres variables.

La variance de cokrigage est toujours inférieure à celle du krigeage monovarié, *quelle que soit la structure conjointe des Z_i* . En effet, le krigeage ordinaire, qui attribue des poids de somme unité à la variable cible et des poids nuls à toutes les autres variables, est un estimateur autorisé et sans biais. La variance de krigeage est donc supérieure à la variance de cokrigage, puisque, par construction, cette dernière minimise la variance d'estimation sur l'ensemble des estimateurs autorisés et sans biais. Intuitivement, les variables additionnelles apportent de l'information sur la variable cible et réduisent l'incertitude sur son estimation.

III.1.2. Cas homotopique

Les écritures matricielles se révèlent particulièrement simples dans le cas homotopique, que nous allons présenter ci-dessous.

Le vecteur des estimations est noté :

$$Z^*(x_0) = \sum_{\alpha=1}^n \Lambda^\alpha Z(x_\alpha)$$

soit

$$[Z_1^*(x_0), \dots, Z_N^*(x_0)]^t = \sum_{\alpha=1}^n \Lambda^\alpha [Z_1(x_\alpha), \dots, Z_N(x_\alpha)]^t$$

où $x_1 \dots x_n$ sont les sites de mesure (communs à toutes les variables), x_0 le site à estimer ;
 Λ^α est la matrice (N,N) des poids attribués au site x_α au cours des estimations.

Le système donnant l'ensemble des poids pour l'estimation de *toutes* les variables en x_0 peut s'écrire de manière très synthétique :

$$\begin{pmatrix} C(x_1 - x_1) & \dots & C(x_1 - x_n) & I \\ \vdots & \ddots & \vdots & \vdots \\ C(x_n - x_1) & \dots & C(x_n - x_n) & I \\ I & \dots & I & 0 \end{pmatrix} \begin{pmatrix} \Lambda^1 \\ \vdots \\ \Lambda^n \\ M \end{pmatrix} = \begin{pmatrix} C(x_1 - x_0) \\ \vdots \\ C(x_n - x_0) \\ I \end{pmatrix} \quad (\text{homotopie})$$

où C est la matrice des covariances simples et croisées

I est la matrice identité d'ordre N

M est la matrice (N,N) des multiplicateurs de Lagrange.

Ce système est analogue à celui du krigeage ordinaire (monovarié), à ceci près que les éléments du système sont ici des matrices au lieu de scalaires.

La matrice de variance - covariance des erreurs d'estimation s'écrit :

$$\Sigma^2(x_0) = C(0) - \sum_{\alpha=1}^n \Lambda_\alpha C(x_0 - x_\alpha) - M.$$

III.2. Cokrigage simple (moyennes connues)

Le cokrigage ordinaire ne peut être mis en œuvre quand, dans le voisinage de cokrigage, aucune donnée n'est disponible sur la variable d'intérêt : on ne peut dans ce cas assurer la condition de non-biais

$$\sum_{\alpha=1}^{n_{i_0}} \lambda_{\alpha}^{i_0} = 1.$$

Le cokrigage simple s'affranchit de cette contrainte et compense l'absence d'information sur la variable cible par la connaissance de sa moyenne.

En supposant les moyennes m_i **connues**, et en travaillant sur les variables centrées $Z_i - m_i$, l'estimateur de cokrigage simple s'écrit :

$$Z_{i_0}^*(\mathbf{x}_0) - m_{i_0} = \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i [Z_i(\mathbf{x}_{\alpha}^i) - m_i].$$

Les contraintes d'autorisation et de non-biais sont automatiquement satisfaites. Il ne reste qu'à écrire la contrainte d'optimalité (sans besoin d'introduire de multiplicateurs de Lagrange puisqu'il s'agit d'un problème de minimisation sans contrainte). Tous calculs faits, on obtient :

$$\sum_{j=1}^N \sum_{\beta=1}^{n_j} \lambda_{\beta}^j C_{ij}(\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j) = C_{ii_0}(\mathbf{x}_{\alpha}^i - \mathbf{x}_0) \quad \forall i = 1 \dots N, \forall \alpha = 1 \dots n_i$$

soit, sous forme matricielle :

$$\begin{pmatrix} C_{11}(\mathbf{x}_{\alpha}^1 - \mathbf{x}_{\beta}^1) & C_{12}(\mathbf{x}_{\alpha}^1 - \mathbf{x}_{\beta}^2) & \dots & C_{1N}(\mathbf{x}_{\alpha}^1 - \mathbf{x}_{\beta}^N) \\ C_{21}(\mathbf{x}_{\alpha}^2 - \mathbf{x}_{\beta}^1) & C_{22}(\mathbf{x}_{\alpha}^2 - \mathbf{x}_{\beta}^2) & & \\ \vdots & & \ddots & \\ C_{N1}(\mathbf{x}_{\alpha}^N - \mathbf{x}_{\beta}^1) & & & C_{NN}(\mathbf{x}_{\alpha}^N - \mathbf{x}_{\beta}^N) \end{pmatrix} \begin{pmatrix} \lambda_{\alpha}^1 \\ \lambda_{\alpha}^2 \\ \vdots \\ \lambda_{\alpha}^N \end{pmatrix} = \begin{pmatrix} C_{1i_0}(\mathbf{x}_{\alpha}^1 - \mathbf{x}_0) \\ C_{2i_0}(\mathbf{x}_{\alpha}^2 - \mathbf{x}_0) \\ \vdots \\ C_{Ni_0}(\mathbf{x}_{\alpha}^N - \mathbf{x}_0) \end{pmatrix}$$

Comme dans le cas du cokrigage ordinaire, le membre gauche ne dépend pas de l'indice i_0 de la variable cible. Il suffira donc d'inverser une seule matrice pour obtenir l'estimation de toutes les variables au point $\mathbf{x}_0$.

La variance de cokrigage simple s'écrit :

$$\sigma_{CKS}^2(\mathbf{x}_0, i_0) = C_{i_0 i_0}(\mathbf{0}) - \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i C_{i i_0}(\mathbf{x}_{\alpha}^i - \mathbf{x}_0).$$

III.3. Cokrigage intrinsèque ordinaire

Dans le cas où les fonctions aléatoires $Z_1 \dots Z_N$ sont conjointement intrinsèques, on peut réaliser un cokrigage à l'aide des variogrammes simples et croisés. Nous ne donnerons que le système final, déduit du cas stationnaire en remplaçant les covariances simples et croisées par l'opposé des variogrammes correspondants :

$$\begin{cases} \sum_{j=1}^N \sum_{\beta=1}^{n_j} \lambda_{\beta}^j \gamma_{ij}(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) - \mu_i = \gamma_{ii_0}(\mathbf{x}_{\alpha} - \mathbf{x}_0) & \text{pour } i=1 \dots N, \alpha=1 \dots n_i \\ \sum_{\beta=1}^{n_i} \lambda_{\beta}^i = \delta_{ii_0} = \begin{cases} 1 & \text{si } i=i_0 \\ 0 & \text{sinon} \end{cases} & \text{pour } i=1 \dots N \end{cases}$$

Pour obtenir ce système de cokrigage, on doit introduire l'hypothèse que les covariances croisées d'accroissements sont paires, pour pouvoir les remplacer par des variogrammes croisés dans le calcul de la variance d'estimation. Plus précisément, on doit supposer que :

$$\text{cov}\{Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{0}), Z_j(\mathbf{x}) - Z_j(\mathbf{0})\} = \text{cov}\{Z_i(\mathbf{x}) - Z_i(\mathbf{0}), Z_j(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{0})\}$$

où $\mathbf{0}$ représente une origine arbitraire. Pour pouvoir mettre en œuvre le cokrigage, le cas intrinsèque requiert donc une hypothèse supplémentaire de symétrie spatiale dans la structure de la corégionalisation.

La variance de cokrigage vaut : $\sigma_{CKO}^2(\mathbf{x}_0) = \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i \gamma_{ii_0}(\mathbf{x}_{\alpha} - \mathbf{x}_0) - \mu_{i_0}.$

III.4. Propriétés du cokrigage ponctuel

Les propriétés usuelles du krigage restent valables pour le cokrigage, en particulier :

- **interpolation exacte** : la valeur estimée redonne la valeur expérimentale en un point de mesure.
- **additivité** : l'estimation d'une fonction linéaire des variables coïncide avec la fonction linéaire des estimations de ces variables, à condition toutefois de prendre les mêmes données pour tous les cokrigages concernés.

En ce qui concerne la **régularité du système** de cokrigage, c'est-à-dire l'existence et l'unicité de la solution, il est nécessaire que les variables utilisées soient linéairement indépendantes aux points de données. Le système de cokrigage est singulier dès lors que plusieurs variables sont liées par une relation linéaire.

Par ailleurs, comme dans le cas du krigeage, les pondérateurs et la variance d'estimation ne dépendent que de la structure spatiale conjointe des variables et de la **configuration de cokrigeage** (c'est-à-dire la configuration géométrique formée par les points de données et le point à estimer). En particulier, la variance de cokrigeage n'est pas une variance conditionnelle, en ce sens qu'elle ne dépend pas des valeurs mesurées.

III.5. L'alternative krigeage – cokrigeage

Deux raisons justifient qu'il est toujours préférable de cokriger plusieurs variables plutôt que de les kriger séparément :

- **on profite**, pour estimer une variable, **de l'information apportée par les autres variables**. En particulier, le cokrigeage donne toujours une variance d'estimation plus faible que le krigeage. Son utilisation est surtout intéressante en cas d'**hétérotopie**, lorsque la variable d'intérêt est moins bien connue que des variables auxiliaires. Cependant, cet avantage s'estompe si la corrélation entre les variables est faible, auquel cas on risque d'augmenter les temps de calcul pour un gain de précision négligeable. De plus, en cas d'hétérotopie totale, l'inférence de la covariance croisée devient délicate, et celle du variogramme croisé est impossible.
- **on améliore la cohérence des résultats d'estimation**, en tenant compte des liens linéaires entre les variables.

Prenons l'exemple des données de sol, où l'on souhaite estimer les variables $Z_1(\mathbf{x})$: % terre fine (granulométrie < 2 mm), $Z_2(\mathbf{x})$: % graviers (2 mm ≤ granulométrie < 20 mm), et $Z_3(\mathbf{x})$: % cailloux (granulométrie ≥ 20 mm). Ces trois variables sont liées linéairement par la relation :

$$Z_1(\mathbf{x}) + Z_2(\mathbf{x}) + Z_3(\mathbf{x}) = 100\%.$$

Ainsi, connaître deux des trois variables $Z_1(\mathbf{x})$, $Z_2(\mathbf{x})$ et $Z_3(\mathbf{x})$ donne la troisième. Le krigeage fournit la meilleure estimation de chacune des variables prises séparément ; mais après krigeage, $100\% - Z_2^*(\mathbf{x}) - Z_1^*(\mathbf{x})$ ne donne en général pas le krigeage de $Z_3(\mathbf{x})$. Le cokrigeage, par contre, fournit la meilleure estimation conjointe des variables et de leurs combinaisons linéaires : en cokrigeant deux des trois variables⁷, on en déduira l'estimation de la troisième grâce à la relation :

$$Z_1^*(\mathbf{x}) + Z_2^*(\mathbf{x}) + Z_3^*(\mathbf{x}) = 100\%.$$

Toutefois, de même que le krigeage d'une variable positive ne garantit pas un résultat positif, le cokrigeage des variables ci-dessus n'assure pas que les valeurs estimées de Z_1 , Z_2 et Z_3 soient toujours positives, ni même inférieures à 100%. Il faudrait pour cela utiliser des méthodes plus sophistiquées, comme l'espérance conditionnelle.

⁷ On ne peut réaliser le cokrigeage des trois variables simultanément, car ces variables ne sont pas linéairement indépendantes. Le système de cokrigeage serait alors singulier. Il est nécessaire de ne retenir que deux variables parmi les trois et de déduire après cokrigeage l'estimation de la troisième variable ; ces estimations sont indépendantes du choix initial des deux variables.

III.6. Application aux données de sol

Les modèles de variogrammes simples et croisés anisotropes ajustés précédemment peuvent être utilisés pour réaliser l'estimation conjointe, par cokrigage intrinsèque ordinaire, des variables *% argile + limon*, *% terre fine* et *pF4.2*. Mais plutôt que d'étudier simultanément ces trois variables, nous allons examiner de manière précise l'apport de chacune des variables *% terre fine* et *pF4.2* sur l'estimation de la variable *% argile + limon* supposée être la variable d'intérêt. La question sous-jacente est la suivante : dans quelle mesure la connaissance d'une variable auxiliaire (*% terre fine* ou *pF4.2*) améliore-t-elle l'estimation de la variable cible (*% argile + limon*) ?

III.6.1. Un couple de variables corrélées : *% argile + limon* / *% terre fine*

Nous allons vérifier qu'il est avantageux de recourir au cokrigage lorsque la variable d'intérêt est fortement corrélée avec une variable auxiliaire *mieux reconnue sur le champ*. A ce titre, on prend l'exemple des variables *% argile + limon* et *% terre fine*, dont le coefficient de corrélation est égal à 0.82 (revoir au besoin la figure 7.2 donnant le nuage de corrélation entre ces variables).

Le modèle de variogrammes simples et croisés de ces deux variables est extrait du modèle linéaire de corégionalisation des trois variables *% argile + limon* – *% terre fine* – *pF4.2* introduit précédemment, soit graphiquement :

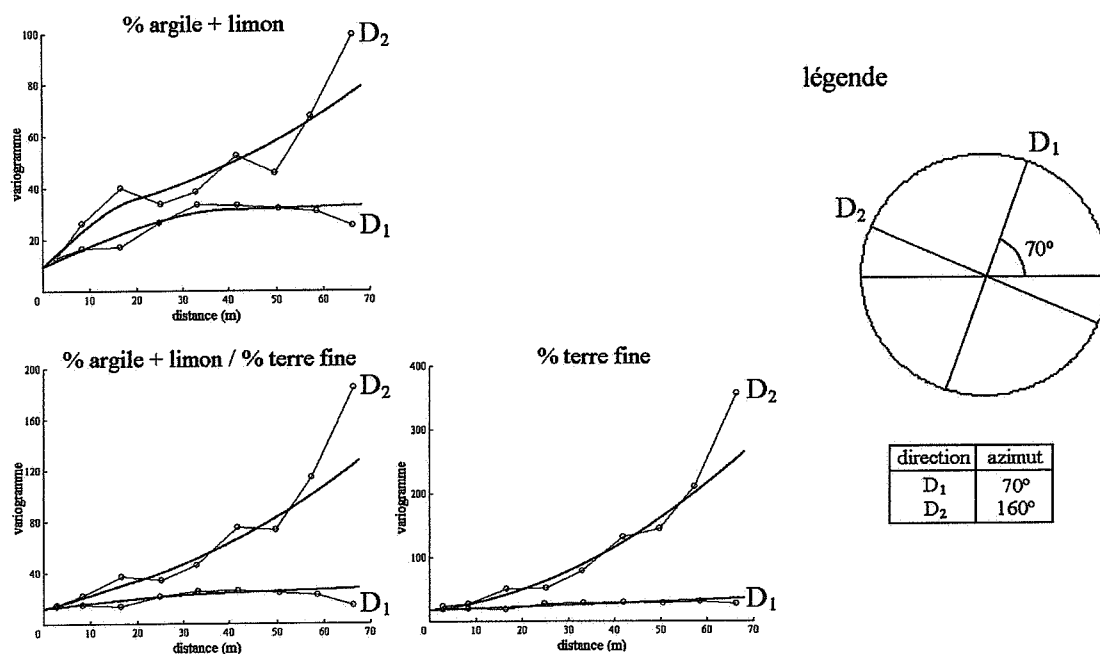


figure 7.7 : variogrammes simples et croisés expérimentaux et modélisés le long des directions principales d'anisotropie

Les figures 7.8 et 7.9 donnent les cartes de cokrigage et des écarts-types de cokrigage, calculés aux nœuds d'une grille régulière de maille $0.5\text{m} \times 0.5\text{m}$. Le voisinage de cokrigage utilisé est le même que celui du krigeage ordinaire (en forme d'ellipse dont le grand axe est orienté dans la direction D_1). Chaque estimation utilise au plus 24 points de données (trois par octant d'ellipse).

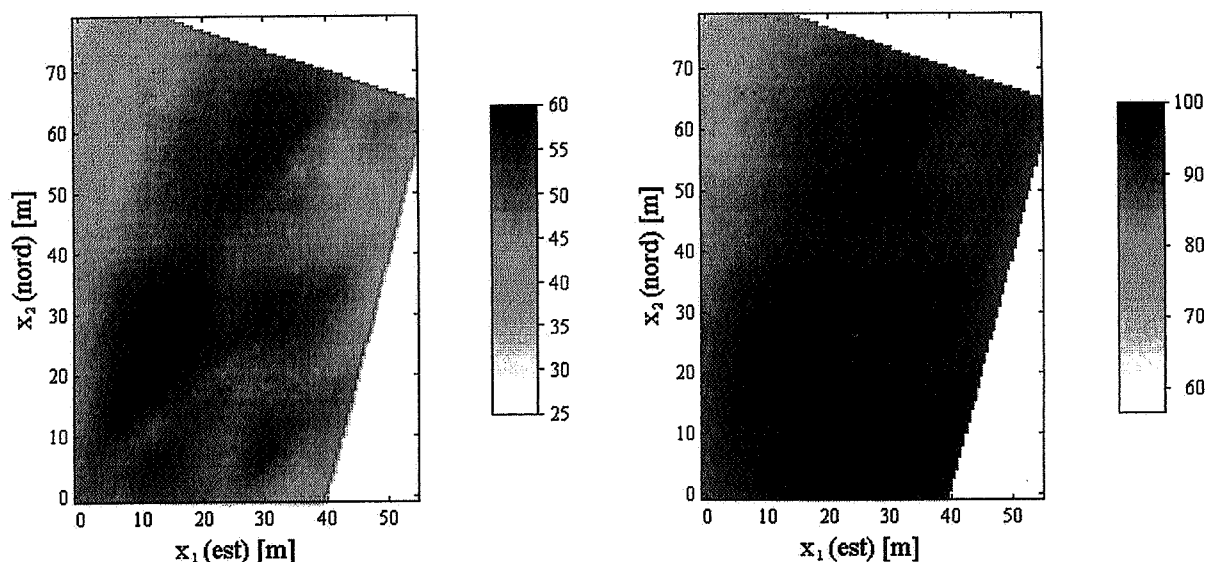


figure 7.8 : cartes du cokrigage de % argile + limon (à gauche) et % terre fine (à droite)

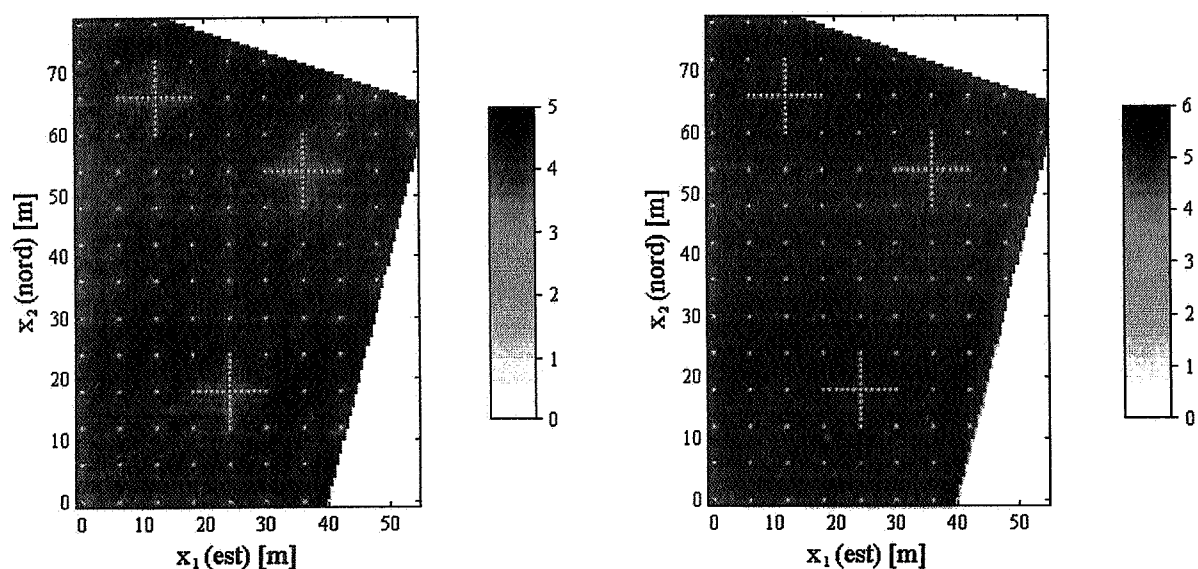


figure 7.9 : cartes des écarts-types de cokrigage de % argile + limon (à gauche) et % terre fine (à droite)

La carte de cokrigage est fort similaire à celle du krigeage obtenue au chapitre 5, bien que l'on puisse distinguer de légères différences entre les deux. Le coefficient de corrélation entre le krigeage et le cokrigage sur l'ensemble de la parcelle est d'ailleurs de 0.97, ce qui indique une très grande conformité entre les deux méthodes d'estimation. Quant aux écarts-types de cokrigage, ils deviennent très vite importants dès que l'on s'écarte quelque peu d'un point de données. Ceci s'explique par la présence d'un effet de pépité non négligeable dans les modèles simples et croisés.

Pour étudier plus précisément l'influence de la variable % *terre fine* sur l'estimation de la variable % *argile + limon*, on se propose d'effectuer le test suivant. On ne retient que les données situées sur la grille régulière (110 points), en oubliant les 60 points issus des croix de sondage. Ces dernières sont en effet très localisées, et des statistiques basées dessus risquent de ne pas être significatives de l'ensemble de la zone. Sur la grille régulière, on ne garde que la moitié des observations de la variable % *argile + limon*, réparties de manière aléatoire. Pour la variable % *terre fine*, on conserve au contraire les 110 observations ; on se place donc dans le cas d'un échantillonnage hétérotopique.

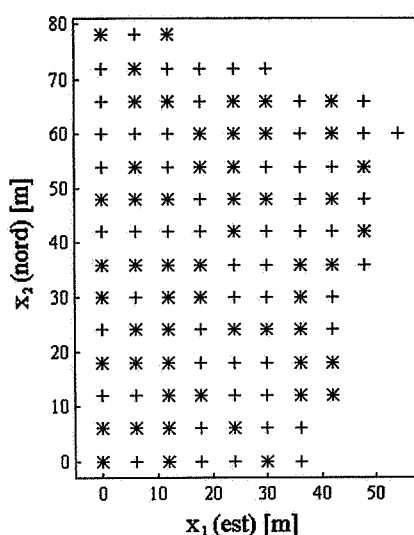


figure 7.10 : échantillonnage hétérotopique de la parcelle

“+” = points où l'on connaît % *terre fine*

“*” = points où l'on connaît % *argile + limon* et % *terre fine*

On estime alors la variable % *argile + limon* aux 55 nœuds de la grille régulière où les valeurs de cette variable ont été oubliées, selon trois méthodes :

- krigeage monovarié à l'aide des 55 données d'argile et limon;
- cokrigeage homotopique, à l'aide seulement des 55 points de données où les variables % *argile + limon* et % *terre fine* sont simultanément connues ;
- cokrigeage hétérotopique, à l'aide des 55 observations de % *argile + limon* et des 110 observations de % *terre fine*.

Enfin, on calcule les erreurs commises aux 55 nœuds où ont été menées les estimations. Les statistiques élémentaires de ces erreurs sont données dans le tableau suivant.

	minimum	maximum	moyenne	variance
krigeage monovarié	-9.04	8.61	1.14	18.66
cokrigeage homotopique	-9.02	8.94	0.82	19.26
cokrigeage hétérotopique	-6.53	5.30	0.01	7.19

tableau 7.5 : statistiques des erreurs commises sur 55 nœuds de la grille régulière

Le cokrigage homotopique et le krigeage donnent des résultats similaires, avec même un léger avantage pour le krigeage (variance d'erreurs moindre). En revanche, le cokrigage hétérotopique améliore de façon remarquable les estimations : la variance des erreurs d'estimation est pratiquement divisée par 3. Le calcul des coefficients de corrélation entre valeurs vraies et estimées sur les 55 points concernés corrobore cette conclusion.

	krigeage monovarié	cokrigage homotopique	cokrigage hétérotopique
coefficient de corrélation entre valeurs vraies et estimées	0.688	0.637	0.902

tableau 7.6 : coefficients de corrélation entre valeurs vraies et estimées sur 55 nœuds de la grille régulière

Les résultats précédents doivent cependant être nuancés : en effet, les points sur lesquels sont calculées les erreurs sont des points de mesure de la variable secondaire % *terre fine*, ce qui est sans doute à l'origine des excellents résultats du cokrigage hétérotopique. Il convient de regarder ce qui se passe en des points où *aucune* information n'est disponible. Pour cela, le moyen le plus efficace consiste à effectuer une **validation croisée** : on ôte tour à tour chacun des 110 points de la grille régulière, et on y réestime la variable % *argile + limon* à l'aide des autres points de données. Nous donnons ci-dessous les statistiques sur les erreurs commises par chacune des trois méthodes.

	minimum	maximum	moyenne	variance
krigeage monovarié	-9.03	20.81	0.71	22.19
cokrigage homotopique	-9.06	21.19	0.52	22.60
cokrigage hétérotopique	-10.52	9.59	-0.04	17.08

tableau 7.7 : statistiques sur les erreurs de validation croisée

Il apparaît à nouveau que le krigeage et le cokrigage homotopique donnent des résultats similaires, avec toujours une supériorité du krigeage. Ceci signifie que connaître le taux de terre fine en des points où l'on connaît déjà le taux d'argile et limon n'apporte guère d'information supplémentaire sur cette dernière variable, et ne permet pas d'en améliorer l'estimation : les deux variables sont "redondantes". En revanche, les résultats relatifs au cokrigage hétérotopique sont clairement meilleurs : **la connaissance d'une variable auxiliaire corrélée avec la variable d'intérêt et mieux reconnue, améliore de manière appréciable l'estimation de la variable d'intérêt.**

Remarque

Il est certain que les performances du cokrigage hétérotopique auraient été encore plus éloquentes si les modèles de variogrammes ne présentaient pas d'effet de pépité. L'effet de pépité correspond à une absence de structuration spatiale et empêche d'avoir une connaissance précise en dehors des points échantillonnés. Sans effet de pépité, les résultats du premier test, sur les 55 points de la grille régulière où seule la variable auxiliaire % *terre fine* est connue, resteraient représentatifs des sites localisés à proximité de ces points.

Pour confirmer la supériorité du cokrigage hétérotopique sur le krigeage, on peut également visualiser les cartes obtenues par les deux méthodes, et les comparer à la carte du krigeage calculée à l'aide de toutes les données d'argile et limon (*cf.* chapitre 5).

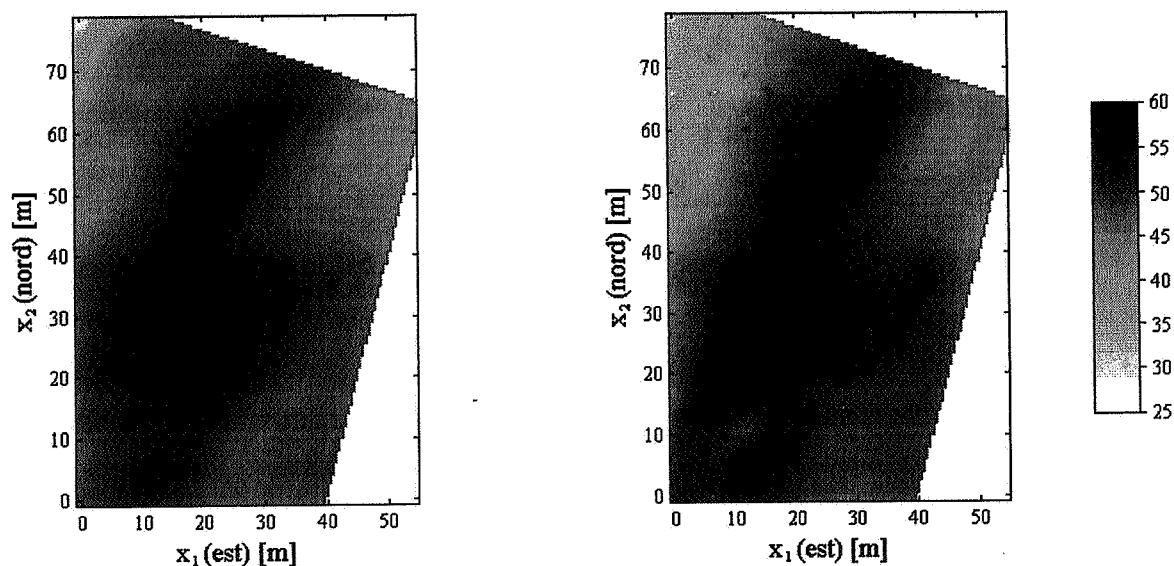


figure 7.11 : cartes des estimations de % argile + limon obtenues par krigeage à l'aide de 55 données (à gauche) et cokrigage hétérotopique (à droite)

La carte obtenue par krigeage, qui ne s'appuie que sur 55 points de données, apparaît plus sommaire que celle obtenue par cokrigage hétérotopique, laquelle incorpore les 55 données de la variable concernée et les 110 données de la variable auxiliaire. Les coefficients de corrélation entre ces estimations et celles obtenues par krigeage à l'aide des 170 données d'argile et limon (*cf.* chapitre 5) sont de 0.93 pour le krigeage à l'aide de 55 points, et 0.97 pour le cokrigage hétérotopique. Autrement dit, la connaissance de la variable % *terre fine* permet quasiment de retrouver l'estimation optimale de % *argile + limon* calculée à l'aide des 170 points de données, mais avec seulement 55 données d'argile et limon.

III.6.2. Un couple de variables peu corrélées : % argile + limon / *pF4.2*

On étudie à présent le couple de variables % *argile + limon* / *pF4.2*, dont le coefficient de corrélation est de 0.057 (c'est-à-dire très peu). Le nuage de corrélation (figure 7.2) ne révèle pas de lien entre ces deux variables. Par rapport au cas précédent, la situation est assurément plus défavorable pour le cokrigage, car la variable auxiliaire *pF4.2*, peu corrélée avec la variable d'intérêt, apporte moins d'information. On rappelle ci-après, sous forme graphique, le modèle de variogrammes simples et croisés entre les variables % *argile + limon* et *pF4.2*.

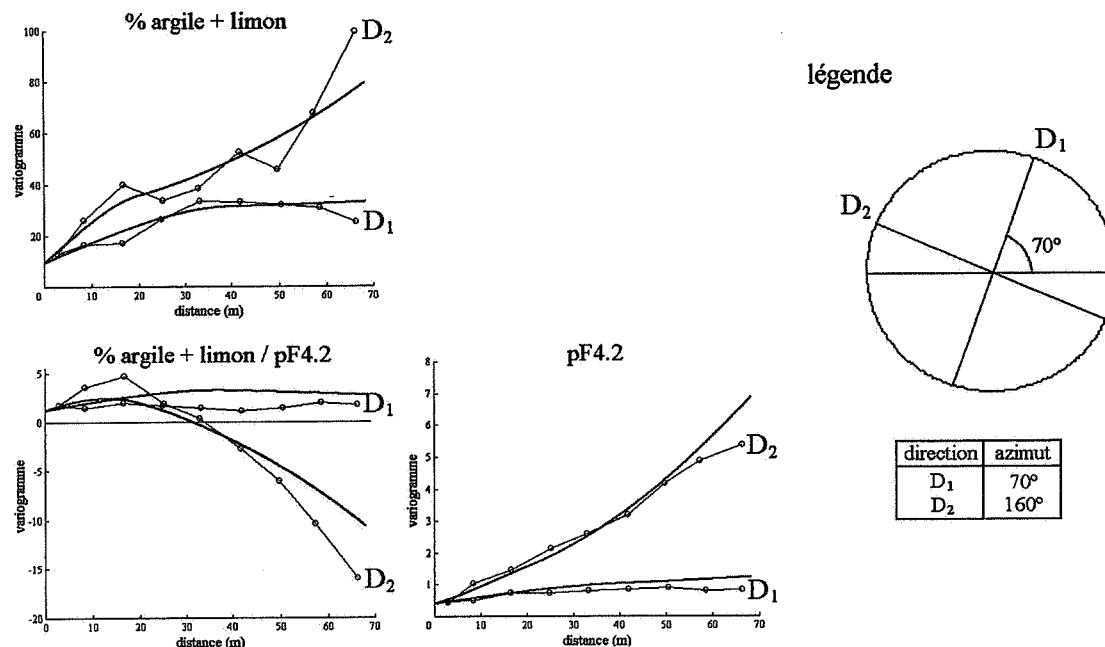


figure 7.12 : variogrammes simples et croisés expérimentaux et modélisés le long des directions principales d'anisotropie

On remarquera que la corrélation entre les accroissements des deux variables, mesurée par le variogramme croisé, est positive et pratiquement constante dans une direction (D₁, direction nord-nord-est / sud-sud-ouest) et varie fortement dans la direction orthogonale (D₂, direction ouest-nord-ouest / est-sud-est), où elle est positive pour les petites distances, et devient de plus en plus négative aux grandes distances. On voit ainsi que la structure conjointe des deux variables est plus complexe qu'il n'y paraît sur le nuage de corrélation, et peut difficilement être résumée par un seul coefficient de corrélation.

Comme précédemment, on effectue un cokrigage des deux variables aux nœuds d'une grille régulière de maille 0.5m × 0.5m, en utilisant un voisinage glissant de forme elliptique.

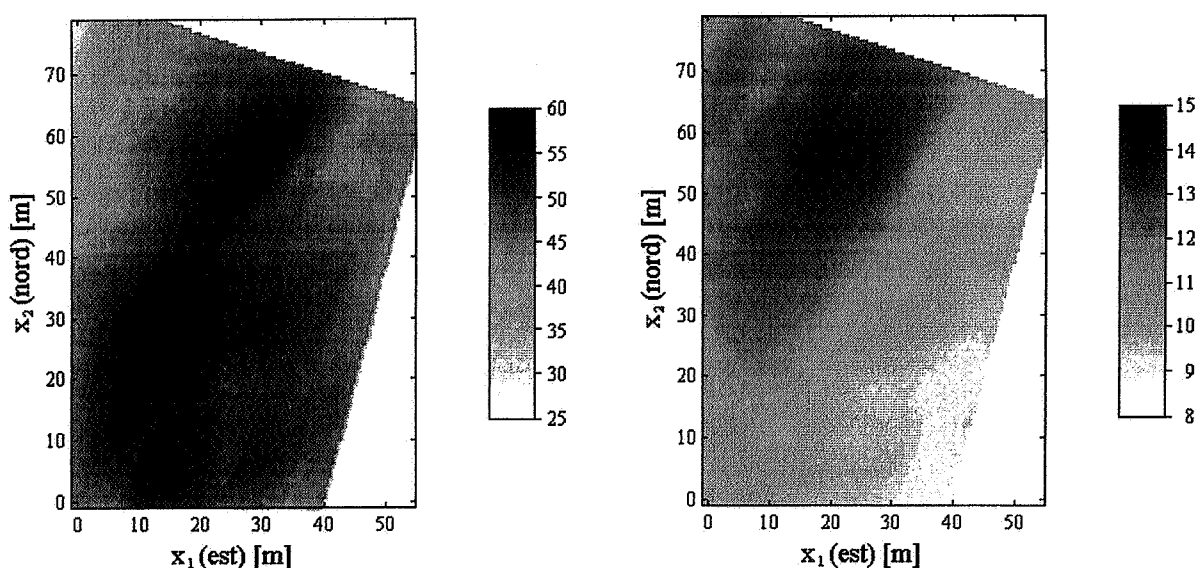


figure 7.13 : cartes du cokrigage de % argile + limon (à gauche) et pF4.2 (à droite)

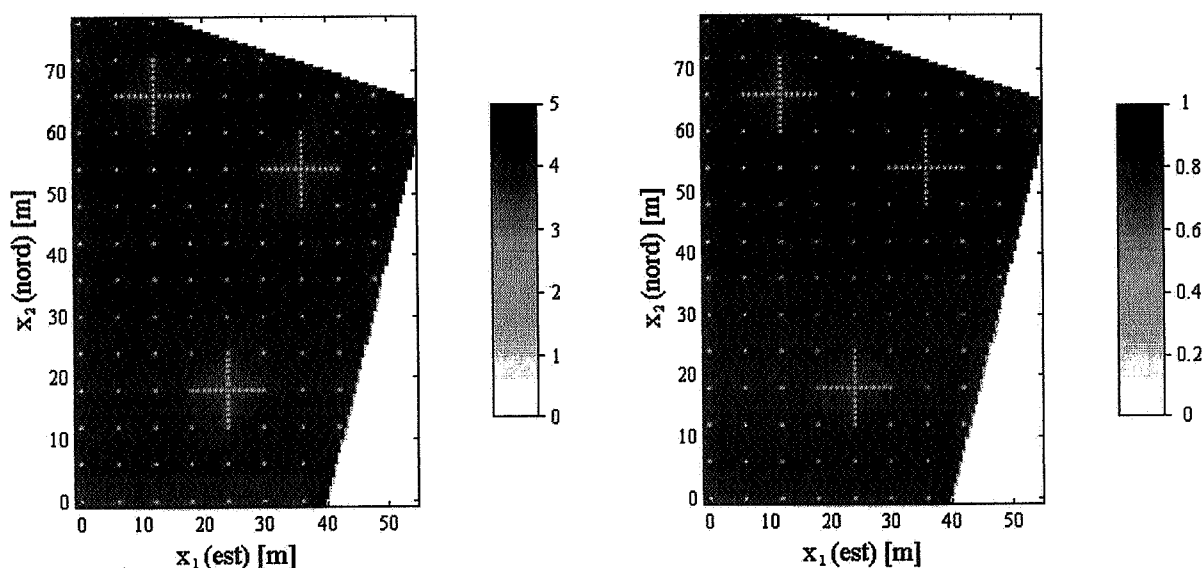


figure 7.14 : cartes des écarts-types de cokrigage de % argile + limon (à gauche) et pF4.2 (à droite)

Les cartes des estimations et des écarts-types d'estimation ne permettent pas de mettre en évidence des différences notables entre le cokrigage et le krigeage monovarié (cf. chapitre 5). Leur coefficient de corrélation est d'ailleurs de 0.99. Pour comparer les performances du krigeage et du cokrigage, on recommence les tests précédents. Les tableaux ci-dessous résument les résultats que l'on obtient.

- Erreurs commises sur les 55 nœuds de la grille régulière lorsque l'on oublie la valeur de % argile + limon en ces nœuds

	minimum	maximum	moyenne	variance
krigeage monovarié	-9.03	8.68	1.08	18.68
cokrigage homotopique	-8.98	8.66	1.12	18.80
cokrigage hétérotopique	-6.88	8.67	0.59	16.78

tableau 7.8 : statistiques des erreurs commises sur 55 nœuds de la grille régulière

	krigeage monovarié	cokrigage homotopique	cokrigage hétérotopique
coefficient de corrélation entre valeurs vraies et estimées	0.688	0.685	0.716

tableau 7.9 : coefficients de corrélation entre valeurs vraies et estimées sur 55 nœuds de la grille régulière

Le gain de précision aux points où la variable auxiliaire pF4.2 est connue est tangible, mais bien moindre que dans le cas précédent, ce qui laisse présager une détérioration des résultats en des points non échantillonnés (test suivant de validation croisée).

• Validation croisée sur les 110 points de la grille régulière

	minimum	maximum	moyenne	variance
krigeage monovariante	-9.03	20.81	0.71	22.19
cokrigeage homotopique	-8.98	21.01	0.75	22.15
cokrigeage hétérotopique	-8.49	20.63	0.44	22.05

tableau 7.10 : statistiques des erreurs de validation croisée

Dans ce cas précis, les résultats du cokrigeage sont décevants. La connaissance d'une variable peu corrélée avec la variable cible, même en davantage de points, n'améliore pas l'estimation de cette dernière. Avant d'entreprendre un cokrigeage, il vaut donc mieux s'assurer de sa qualité attendue, en visualisant le nuage de corrélation entre les variables, et en effectuant des tests de validation croisée. Il est inutile d'augmenter les temps de calculs pour un gain de précision négligeable.

III.7. Cas de simplification du système de cokrigeage : "autokrigeabilité"

Le nombre de variogrammes simples et croisés à ajuster augmente proportionnellement au carré du nombre de variables. La lourdeur du système de cokrigeage augmente également avec le nombre de variables. Il est donc intéressant de voir dans quelles conditions se simplifie le cokrigeage, c'est-à-dire dans quelles conditions il redonne l'estimation par krigeage.

Une variable est dite **autokrigeable** par rapport à un ensemble de variables si son krigeage propre coïncide avec son cokrigeage. Un cas trivial est celui où aucune variable auxiliaire n'est corrélée avec la variable d'intérêt : les covariances croisées ou variogrammes croisés correspondants sont identiquement nuls. Il est alors aisé de vérifier que tous les pondérateurs de cokrigeage sont nuls, mis à part ceux affectés aux échantillons de la variable d'intérêt. Intuitivement, les variables additionnelles n'apportent aucune information sur la variable cible.

Lorsqu'il existe une corrélation entre les variables de départ, on peut parfois se ramener à d'autres variables appelées *facteurs*, sans corrélation spatiale. La décomposition des Z_i à partir des facteurs est appelée factorisation des Z_i . Si les facteurs s'obtiennent par combinaisons linéaires des variables de départ, leur krigeage séparé (qui coïncide avec leur cokrigeage) permet de déduire le cokrigeage des Z_i . Cette approche est utilisée en géostatistique non linéaire pour la mise en œuvre de la technique du *krigeage disjonctif*, qui équivaut à un cokrigeage de variables indicatrices.

Dans le **cadre stationnaire du second ordre**, un cas plus général d'autokrigeabilité d'une variable intervient, *en cas d'homotopie*, lorsque les fonctions de covariances simples et croisées de cette variable sont proportionnelles à une même fonction $\rho(\mathbf{h})$:

$$C_{i_0 j}(\mathbf{h}) = \sigma_{i_0 j} \rho(\mathbf{h}) \quad \forall j = 1 \dots N.$$

Ceci est vérifié en particulier dans le modèle dit de **corrélation intrinsèque**, où *toutes* les fonctions de covariance simples et croisées sont proportionnelles entre elles :

$$C_{ij}(\mathbf{h}) = \sigma_{ij} \rho(\mathbf{h}) \quad \forall i, j = 1 \dots N.$$

Dans le **cadre intrinsèque**, on a les mêmes propriétés en remplaçant les covariances simples et croisées par les variogrammes simples et croisés. En particulier, le modèle de corrélation intrinsèque s'écrit :

$$\gamma_{ij}(\mathbf{h}) = p_{ij} g(\mathbf{h}) \quad \forall i, j = 1 \dots N$$

où $g(\mathbf{h})$ est un variogramme de base.

Nous reviendrons plus en détail sur le modèle de corrélation intrinsèque et son interprétation lors de l'étude des techniques d'analyse krigéante (cf. paragraphe suivant et annexe G).

En corrélation intrinsèque, en cas d'homotopie, toutes les variables sont autokrigeables : leur cokrigeage, simple ou ordinaire, coïncide avec leur krigéage séparé. Intuitivement, cela signifie que, comme toutes les variables présentent la même structure (directe et croisée), la connaissance de l'une d'elles équivaut, pour l'assignation des poids de cokrigeage, à celle de toutes ; tout se passe comme si l'information apportée par les variables additionnelles était totalement redondante. Il suffit même de résoudre un seul système de krigéage, car les pondérateurs affectés à un échantillon sont les mêmes pour toutes les variables. L'étude de la corégionalisation ($z_1 \dots z_N$) se réduit à celle d'une seule régionalisation, dont la structure est modélisée par la fonction de corrélation $\rho(\mathbf{h})$ ou le variogramme $g(\mathbf{h})$.

Il faut prendre garde au fait que cette propriété cesse d'être vraie en cas d'hétérotopie. La connaissance d'une variable auxiliaire, en corrélation intrinsèque avec la variable cible mais mieux échantillonnée, apporte une information précieuse, voire décisive, sur les valeurs non mesurées et permet d'améliorer significativement l'estimation de la variable d'intérêt.

IV. L'analyse krigéante

L'analyse krigéante, également appelée "krigeage factoriel multivariable" ou "analyse factorielle", permet d'étudier et d'interpréter des données multivariées situées dans l'espace géographique ; par rapport aux méthodes classiques comme l'analyse en composantes principales, elle prend en compte l'information spatiale et structurale contenue dans les données. Sa mise en œuvre s'articule autour de trois étapes :

- analyse variographique multivariable à l'aide du modèle linéaire de corégionalisation ;
- décomposition des fonctions aléatoires modélisant la corégionalisation en *facteurs* indépendants, synthétisant l'information multivariable ;
- estimation des valeurs de ces facteurs par cokrigeage.

La première étape a été largement étudiée dans la première partie de ce chapitre. Nous allons donner quelques éléments concernant les deux étapes suivantes ; le lecteur désireux d'approfondir sa connaissance du sujet trouvera une présentation détaillée de la théorie en annexe G. Pour alléger le texte, nous employerons, lorsque cela sera possible, des notations vectorielles ou matricielles.

IV.1. Décomposition des variables en facteurs

IV.1.1. Principe

A partir d'un ensemble de variables spatialement corrélées, on cherche à construire des variables synthétiques, ou *facteurs*, indépendantes entre elles, qui renferment chacune une partie de l'information multivariée. Plutôt qu'une indépendance des facteurs, il s'agit en fait d'une orthogonalité, c'est-à-dire d'une absence de corrélation spatiale, entre les facteurs eux-mêmes (cadre stationnaire d'ordre deux) ou leurs accroissements (cadre intrinsèque).

Bien que les facteurs définis soient des constructions purement mathématiques, ils peuvent trouver une interprétation d'ordre physique et mettre en évidence plusieurs processus dont le phénomène étudié est la superposition. On pourra aussi éliminer l'influence de facteurs jugés indésirables dans l'estimation des variables régionalisées, par exemple un effet de pépite ou une structure de courte portée qui correspond à du "bruit".

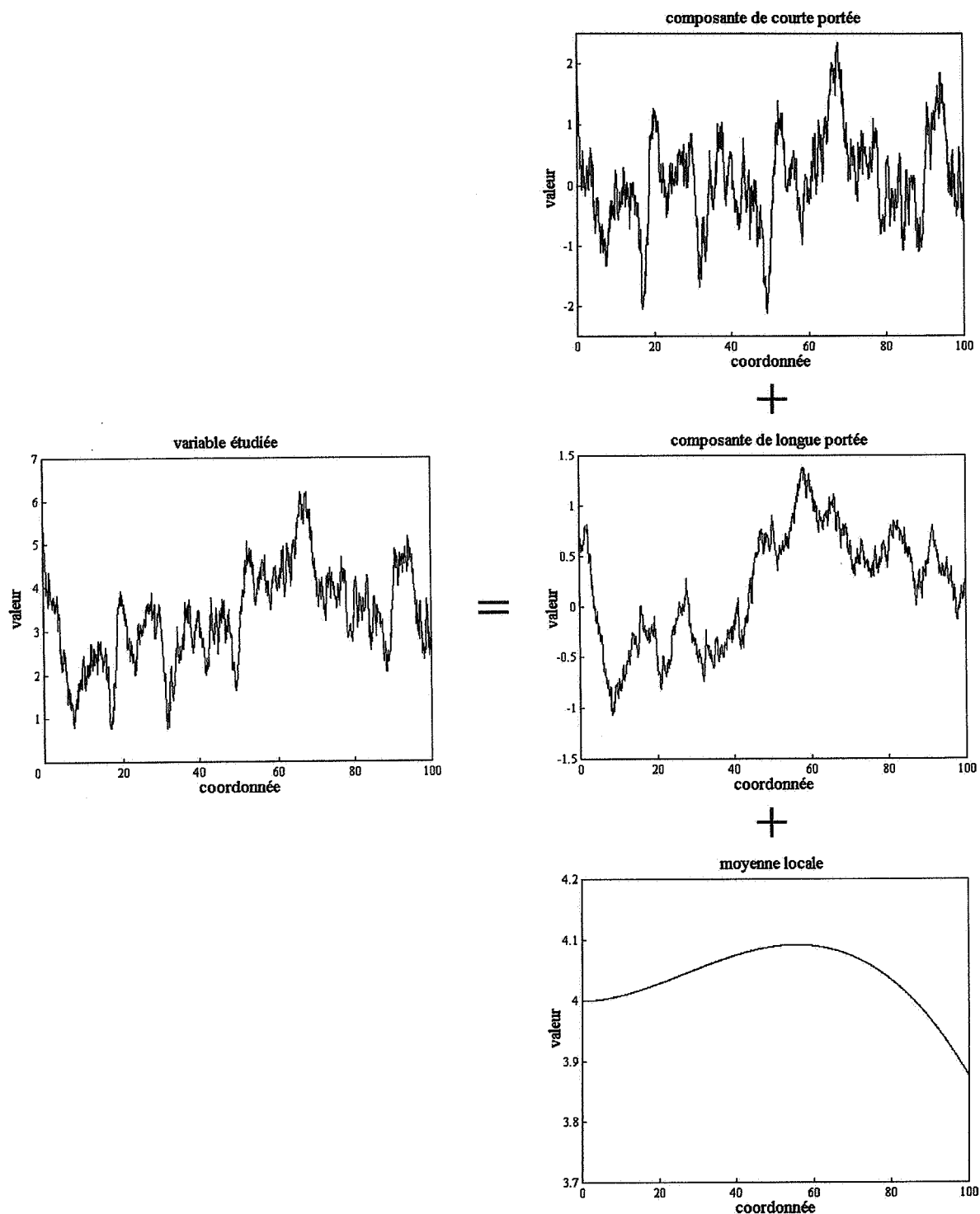


figure 7.15 : schéma de principe de l'analyse krigéante
(cas d'une seule variable dans un espace unidimensionnel)

Les domaines d'application de l'analyse krigéante sont variés : sciences du sol, géochimie, géomagnétisme, exploration pétrolière, analyse d'images... Une illustration de cette technique sur les données de sol est présentée ultérieurement.

IV.1.2. Définition des facteurs

Dans un premier temps, on s'intéresse à **une seule fonction aléatoire**, dont la covariance ou le variogramme est modélisé par un schéma gigogne (modèle linéaire de régionalisation). A chaque structure gigogne, on peut associer une composante élémentaire, de sorte que la variable de départ soit la somme de ces composantes, supposées indépendantes entre elles. Par exemple, une variable $Z(\mathbf{x})$ stationnaire, de moyenne locale $m(\mathbf{x})$ et de covariance gigogne

$$C(\mathbf{h}) = \sum_{u=1}^S b_u \rho_u(\mathbf{h})$$

peut être décomposée de la manière suivante :

$$Z(\mathbf{x}) = \sum_{u=1}^S X_u(\mathbf{x}) + m(\mathbf{x})$$

où les composantes X_u sont indépendantes, d'espérance nulle et de covariances respectives ρ_u . La variable étudiée apparaît ainsi comme la superposition de composantes de caractéristiques différentes (anisotropie, portée...) qui peuvent trouver une interprétation physique.

On s'intéresse ensuite au cas de **plusieurs variables** $\mathbf{Z}(\mathbf{x}) = [Z_1(\mathbf{x}), \dots, Z_N(\mathbf{x})]$ stationnaires, de moyennes locales $\mathbf{m}(\mathbf{x}) = [m_1(\mathbf{x}), \dots, m_N(\mathbf{x})]$ et de covariances simples et croisées proportionnelles entre elles (cas de **corrélation intrinsèque**) :

$$\mathbf{C}(\mathbf{h}) = \mathbf{V} \rho_0(\mathbf{h})$$

où $\rho_0(\mathbf{h})$ est une covariance telle que : $\rho_0(\mathbf{0}) = 1$

$\mathbf{V}$ est la matrice de variance-covariance de $\mathbf{Z}(\mathbf{x})$.

La diagonalisation de $\mathbf{V}$ fournit des facteurs $\mathbf{Y}(\mathbf{x}) = [Y_1(\mathbf{x}), \dots, Y_N(\mathbf{x})]$ d'espérance nulle et spatialement non corrélés, selon la relation :

$$\mathbf{Y}(\mathbf{x}) = [\mathbf{Z}(\mathbf{x}) - \mathbf{m}(\mathbf{x})] \times \mathbf{A}^{-1}$$

où $\mathbf{V} = \mathbf{A}^t \mathbf{A}$, avec $\mathbf{A} = \sqrt{\mathbf{D}} \mathbf{Q}^t$, $\mathbf{Q}^t \mathbf{Q} = \mathbf{I}$ et $\mathbf{D}$ matrice diagonale de taille $N \times N$.

Les facteurs spatiaux permettent de reconstituer les variables initiales :

$$\mathbf{Z}(\mathbf{x}) = \mathbf{Y}(\mathbf{x}) \times \mathbf{A} + \mathbf{m}(\mathbf{x}).$$

Chaque facteur est associé à une valeur propre (un élément de la matrice diagonale $\mathbf{D}$), qui mesure la quantité d'information qu'il contient. Il existe donc une hiérarchie entre facteurs, les premiers constituant un résumé de l'information totale.

Le **cas général** (plusieurs variables obéissant au modèle linéaire de corégionalisation) s'obtient en juxtaposant les résultats des deux étapes précédentes. On décompose d'abord les variables en plusieurs ensembles de composantes en corrélation intrinsèque dont les matrices de variance-covariance sont les matrices de corégionalisation, puis, dans chaque ensemble, on définit les facteurs à l'aide d'une diagonalisation de la matrice de corégionalisation associée. Par exemple, supposons que le modèle de corégionalisation soit de la forme :

$$C(\mathbf{h}) = \sum_{u=1}^S \mathbf{B}_u \rho_u(\mathbf{h}) .$$

La décomposition finale en facteurs orthogonaux s'écrit, sous forme matricielle :

$$\mathbf{Z}(\mathbf{x}) = \sum_{u=1}^S \mathbf{Y}_u(\mathbf{x}) \times \mathbf{A}_u + \mathbf{m}(\mathbf{x})$$

où $\forall u \in [1, S]$, $\mathbf{B}_u = \mathbf{A}_u^t \mathbf{A}_u$, avec $\mathbf{A}_u = \sqrt{\mathbf{D}_u} \mathbf{Q}_u^t$, $\mathbf{Q}_u^t \mathbf{Q}_u = \mathbf{I}$ et $\mathbf{D}_u$ matrice diagonale.

Chaque composante de $\mathbf{Y}_u$ est associée à un élément de $\mathbf{D}_u$, qui mesure son importance.

Tous ces résultats s'étendent au cadre intrinsèque, pour lequel les covariances n'existent pas forcément. On ne manipule que des *accroissements* des variables, ce qui explique que les constantes, en particulier les moyennes, disparaissent du formalisme. Le modèle multivariable (variogrammes simples et croisés) s'écrit comme une combinaison de variogrammes de base g_u , $u = 1 \dots S$:

$$\Gamma(\mathbf{h}) = \sum_{u=1}^S \mathbf{B}_u g_u(\mathbf{h}) .$$

Une décomposition intermédiaire consiste à poser :

$$\mathbf{Z}(\mathbf{x}) = \sum_{u=1}^S \mathbf{X}_u(\mathbf{x})$$

où, pour tout u , le groupe de composantes $\mathbf{X}_u(\mathbf{x})$ est associé à la matrice de corégionalisation $\mathbf{B}_u$ et au variogramme g_u .

On définit les facteurs spatiaux par :

$$\forall u \in [1, S], \mathbf{Y}_u(\mathbf{x}) = \mathbf{X}_u(\mathbf{x}) \times \mathbf{A}_u^{-1}$$

avec $\mathbf{B}_u = \mathbf{A}_u^t \mathbf{A}_u$, $\mathbf{A}_u = \sqrt{\mathbf{D}_u} \mathbf{Q}_u^t$, $\mathbf{Q}_u^t \mathbf{Q}_u = \mathbf{I}$ et $\mathbf{D}_u$ matrice diagonale.

La décomposition finale est : $\mathbf{Z}(\mathbf{x}) = \sum_{u=1}^S \mathbf{Y}_u(\mathbf{x}) \times \mathbf{A}_u$.

IV.2. Estimation des facteurs

Les facteurs spatiaux définis précédemment peuvent être estimés par cokrigage à partir des données multivariées initiales.

L'estimateur du facteur $Y_p^u(\mathbf{x})$, p-ième composante de $Y_u(\mathbf{x})$, sera de la forme :

$$Y_p^u(\mathbf{x}) = \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i Z_i(\mathbf{x}_{\alpha}^i).$$

Dans le cas où les variables sont conjointement stationnaires, de moyennes inconnues et de covariances

$$C(\mathbf{h}) = \sum_{u=1}^S \mathbf{B}_u \rho_u(\mathbf{h})$$

le système de cokrigage s'écrit :

$$\begin{cases} \sum_{j=1}^N \sum_{\beta=1}^{n_j} \lambda_{\beta}^j C_{ij}(\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j) + \mu_i = a_{pi}^u \rho_u(\mathbf{x}_{\alpha}^i - \mathbf{x}) & \text{pour } i=1 \dots N, \alpha=1 \dots n_i \\ \sum_{\beta=1}^{n_i} \lambda_{\beta}^i = 0 & \text{pour } i=1 \dots N \end{cases}$$

où a_{pi}^u désigne le terme générique de la matrice $\mathbf{A}_u$.

Dans le cadre du modèle gigogne suivant qui compte un schéma intrinsèque (sans palier) $g_S(\mathbf{h})$

$$\Gamma(\mathbf{h}) = \sum_{u=1}^{S-1} \mathbf{B}_u g_u(\mathbf{h}) + \mathbf{B}_S g_S(\mathbf{h})$$

seuls les facteurs associés aux variogrammes d'indices $u = 1 \dots S-1$ sont stationnaires d'ordre deux, les autres (ceux d'indice $u = S$) sont strictement intrinsèques. Le système de cokrigage est plus complexe :

$$\begin{cases} \sum_{j=1}^N \sum_{\beta=1}^{n_j} \lambda_{\beta}^j \gamma_{ij}(\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j) - \mu_i = a_{pi}^u g_u(\mathbf{x}_{\alpha}^i - \mathbf{x}) & \text{pour } i=1 \dots N, \alpha=1 \dots n_i \\ \sum_{i=1}^N a_{qi}^S \sum_{\beta=1}^{n_i} \lambda_{\beta}^i = \begin{cases} 0 & (\text{facteur stationnaire : } u \neq S) \\ \delta_p^q & (\text{facteur intrinsèque strict : } u = S) \end{cases} & \text{pour } q=1 \dots N \end{cases}$$

où δ_p^q est le symbole de Kronecker (égal à 1 si $p = q$, 0 sinon).

IV.3. Estimation des composantes spatiales intermédiaires

On peut aussi estimer $X_i^u(\mathbf{x})$, i -ème composante de $\mathbf{X}_u$, à partir des données initiales. Les systèmes de cokrigage ont la même structure que le cokrigage classique, à l'exception du second membre où les termes qui ne concernent pas la composante à estimer sont annulés. On trouvera une présentation de cette technique au cas monovarié ("krigeage avec filtrage de composantes") en annexe G.

Un moyen simple de retrouver les systèmes de cokrigage consiste à remarquer que :

$$X_i^u(\mathbf{x}) = \sum_{p=1}^N a_{pi}^u Y_p^u(\mathbf{x})$$

et à utiliser la linéarité des équations, ainsi que l'identité $\mathbf{B}_u = \mathbf{A}_u^t \mathbf{A}_u$.

Dans le cadre stationnaire d'ordre deux, le système de cokrigage de $X_{i_0}^u(\mathbf{x})$ se simplifie en :

$$\begin{cases} \sum_{j=1}^N \sum_{\beta=1}^{n_j} \lambda_{\beta}^j C_{ij}(\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j) + \mu_i = b_{ii_0}^u \rho_u(\mathbf{x}_{\alpha}^i - \mathbf{x}) & \text{pour } i=1 \dots N, \alpha=1 \dots n_i \\ \sum_{\beta=1}^{n_i} \lambda_{\beta}^i = 0 & \text{pour } i=1 \dots N \end{cases}$$

et dans le cadre du modèle intrinsèque précédent :

$$\begin{cases} \sum_{j=1}^N \sum_{\beta=1}^{n_j} \lambda_{\beta}^j \gamma_{ij}(\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j) - \mu_i = b_{ii_0}^u g_u(\mathbf{x}_{\alpha}^i - \mathbf{x}) & \text{pour } i=1 \dots N, \alpha=1 \dots n_i \\ \sum_{\beta=1}^{n_i} \lambda_{\beta}^i = \begin{cases} 0 & \text{(composante stationnaire : } u \neq S) \\ \delta_{i_0}^i & \text{(composante intrinsèque : } u = S) \end{cases} & \text{pour } i=1 \dots N \end{cases}$$

IV.4. Conclusion sur l'analyse krigéante

IV.4.1. Atouts et limitations

Le cokrigage des facteurs permet de réaliser des **cartes synthétiques** en utilisant les données multivariées situées dans un voisinage de chaque point de la carte. Les facteurs de l'analyse krigéante vérifient les propriétés suivantes :

- ils sont spatialement orthogonaux, i.e. les facteurs stationnaires et les *accroissements* des facteurs non stationnaires sont spatialement non corrélés ; en termes plus intuitifs, ils ne contiennent pas d'information redondante ;
- ils prennent en compte à la fois les valeurs des variables aux points d'observation et la structure spatiale du phénomène ;
- ils sont hiérarchisés ; les premiers facteurs relatifs à chaque modèle de base permettent de **“résumer” l'information multivariable** contenue dans les données et dans le modèle de corégionalisation.

L'analyse krigéante permet en outre une **étude à différents niveaux d'échelle** : on peut par exemple éliminer de l'analyse les facteurs correspondant à des structures pépétiques ou à très courtes portées, qui correspondent à du “bruit” (erreurs de mesure ou microstructures).

Ces affirmations appellent plusieurs remarques. On sait en effet qu'à cause de la propriété de lissage du krigage, les *estimateurs* des facteurs n'ont pas les mêmes propriétés statistiques que les “vrais” facteurs. En particulier, les facteurs estimés ne vérifient pas les propriétés d'orthogonalité, donc ils ne sont pas indépendants. D'autre part, la qualité de l'estimation des facteurs dépend de la taille du voisinage de krigage utilisé ; on observe en pratique qu'il est nécessaire d'utiliser des voisinages de taille fort supérieure aux portées des facteurs que l'on cherche à estimer.

Une autre objection, plus fondamentale, que l'on peut formuler à l'égard de l'analyse krigéante est que la détermination des facteurs est en partie arbitraire. Elle repose en effet sur l'identification du modèle corégionalisation, qui en aucun cas n'est défini de manière univoque par les données expérimentales. En fait, cet arbitraire est inhérent à toute méthode géostatistique ou d'analyse des données ; l'utilisateur est contraint d'injecter dans le modèle plus d'information que n'en contiennent les données.

Toutefois, il n'est pas possible ici de soumettre ces choix arbitraires à un contrôle expérimental. En effet, contrairement aux variables initiales, les facteurs définis ne sont pas des *grandeurs régionales*, c'est-à-dire que même une connaissance complète de la réalité ne suffirait pas à les déterminer. En d'autres termes, ces facteurs sont des entités théoriques, et se pose la question de leur signification physique. En fait, l'intérêt des méthodes d'analyse krigéante n'est pas l'estimation des facteurs en elle-même, dont l'objectivité est douteuse, mais l'interprétation que l'on peut faire de ces facteurs pour comprendre le phénomène étudié. En particulier, la décomposition en facteurs fait apparaître plusieurs niveaux de structuration, à des échelles différentes, qui peuvent avoir une explication physique (par exemple, anomalies locales / anomalies régionales).

IV.4.2. Exemple sur les données de sol

Nous allons illustrer les concepts de l'analyse krigéante sur les données de sol. Rappelons que la structure conjointe des trois variables étudiées (% argile + limon, % terre fine et $pF4.2$) a été ajustée selon le modèle linéaire de corégionalisation, à l'aide de trois structures de base : un effet de pépite, un schéma sphérique et un modèle puissance.

Indexons par i ($i = 1,2,3$) les variables, et par u ($u = 1,2,3$) les structures de base. L'analyse krigéante n'est pas menée directement, mais en trois étapes :

- estimation des composantes spatiales $X_i^u(\mathbf{x})$ associées à chaque variable ($i = 1,2,3$) et à chaque structure ($u = 1,2,3$), par cokrigéage avec filtrage de composantes.
- calcul des coefficients permettant de passer des composantes $X_i^u(\mathbf{x})$ aux facteurs spatiaux $Y_p^u(\mathbf{x})$.

Cette étape nécessite une attention particulière. Pour chacune des structures de base, il s'agit de diagonaliser la matrice de corégionalisation $\mathbf{B}_u$ associée, d'où l'obtention d'une matrice orthogonale de vecteurs propres $\mathbf{Q}_u$ et d'une matrice diagonale de valeurs propres $\mathbf{D}_u$. Cependant, bien que les variables étudiées soient sans unité (ce sont des pourcentages), elles ne représentent pas des grandeurs comparables ; il est alors avantageux de les réduire avant de calculer les facteurs spatiaux. Un critère de réduction souvent utilisé en analyse en composantes principales est celui des variances : on divise les variables étudiées par leur écart-type (racine carrée de la variance) pour les ramener à une échelle comparable. Dans le cas stationnaire, les variances *a priori* des variables sont données par les coefficients diagonaux de la matrice de variance-covariance, somme des matrices de corégionalisation. Ici, à cause de la structure puissance, non stationnaire, ces variances *a priori* sont infinies.

L'idée est de considérer les coefficients diagonaux de la matrice de corégionalisation $\mathbf{B}_3$ de la structure puissance, comme s'il s'agissait de variances. Ces coefficients indiquent en effet le "degré d'infinité" des variances *a priori* des variables. Il conviendrait donc de réduire les variables par les racines carrées de ces coefficients, avant de mener l'analyse krigéante proprement dite. En fait, cela revient à effectuer l'analyse krigéante des variables de départ, non réduites, mais en recherchant des facteurs spatiaux orthogonaux par rapport à la métrique

$$\mathbf{M} = \begin{pmatrix} b_{11}^3 & 0 \\ & b_{22}^3 \\ 0 & & b_{33}^3 \end{pmatrix} = \begin{pmatrix} 0.0245 & 0 \\ & 0.122 \\ 0 & & 0.003 \end{pmatrix}.$$

Les matrices $\mathbf{Q}_u$ et $\mathbf{D}_u$ cherchées ($u = 1, 2$ ou 3) sont alors solutions du problème de valeurs propres généralisé

$$\mathbf{Q}_u^t \mathbf{B}_u \mathbf{Q}_u = \mathbf{D}_u \quad \text{avec} \quad \mathbf{Q}_u^t \mathbf{M} \mathbf{Q}_u = \mathbf{I}.$$

Le tableau ci-après résume les résultats obtenus au cours de cette étape.

structure	effet de pépité	schéma sphérique	modèle puissance
matrice de corégionalisation	$\mathbf{B}_1 = \begin{pmatrix} 9.4 & 11.8 & 1.2 \\ 11.8 & 17.6 & 1.4 \\ 1.2 & 1.4 & 0.4 \end{pmatrix}$	$\mathbf{B}_2 = \begin{pmatrix} 20.9 & 9.8 & 2.3 \\ 9.8 & 4.7 & 1.15 \\ 2.3 & 1.15 & 0.5 \end{pmatrix}$	$\mathbf{B}_3 = \begin{pmatrix} 0.0245 & 0.054 & -0.007 \\ 0.054 & 0.122 & -0.015 \\ -0.007 & -0.015 & 0.003 \end{pmatrix}$
diagonalisation	$\mathbf{Q}_1 = \begin{pmatrix} 3.42 & -1.47 & 5.19 \\ -2.39 & -0.80 & 1.35 \\ -2.16 & 17.02 & 6.23 \end{pmatrix}$ $\mathbf{D}_1 = \begin{pmatrix} 16.2 & 0 & 0 \\ 0 & 77.0 & 0 \\ 0 & 0 & 568.0 \end{pmatrix}$	$\mathbf{Q}_2 = \begin{pmatrix} 1.23 & -2.05 & 5.92 \\ -2.81 & -0.06 & 0.56 \\ 0.81 & 17.28 & 5.82 \end{pmatrix}$ $\mathbf{D}_2 = \begin{pmatrix} 0.7 & 0 & 0 \\ 0 & 74.3 & 0 \\ 0 & 0 & 983.3 \end{pmatrix}$	$\mathbf{Q}_3 = \begin{pmatrix} 3.80 & -4.73 & -1.98 \\ 1.68 & 1.90 & -1.32 \\ -10.02 & -1.70 & -15.2 \end{pmatrix}$ $\mathbf{D}_3 = \begin{pmatrix} 2.736 & 0 & 0 \\ 0 & 0.009 & 0 \\ 0 & 0 & 0.254 \end{pmatrix}$

tableau 7.11 : diagonalisation des matrices de corégionalisation

(c) calcul des facteurs spatiaux estimés $Y_p^{u*}(\mathbf{x})$ ($p = 1,2,3$ et $u = 1,2,3$) à l'aide de la relation

$$[Y_1^{u*}(\mathbf{x}), Y_2^{u*}(\mathbf{x}), Y_3^{u*}(\mathbf{x})] = [X_1^{u*}(\mathbf{x}), X_2^{u*}(\mathbf{x}), X_3^{u*}(\mathbf{x})] \times \mathbf{A}_u^{-1}$$

avec $\mathbf{B}_u = \mathbf{A}_u^t \mathbf{A}_u$, $\mathbf{A}_u = \sqrt{\mathbf{D}_u} \mathbf{Q}_u^t \mathbf{M}$, $\mathbf{Q}_u^t \mathbf{M} \mathbf{Q}_u = \mathbf{I}$ et $\mathbf{D}_u$ matrice diagonale.

Pour chaque structure, on obtient trois facteurs spatiaux, hiérarchisés selon la quantité d'information qu'ils contiennent (mesurée par la valeur propre associée) ; nous n'allons retenir que le premier d'entre eux. Par ailleurs, les facteurs relatifs à l'effet de pépité ne sont guère intéressants, puisqu'ils sont nuls partout sauf aux points de données (ils peuvent représenter les erreurs de mesure). Nous ne donnons donc que les cartes des facteurs relatifs aux deux autres structures : modèle puissance et schéma sphérique.

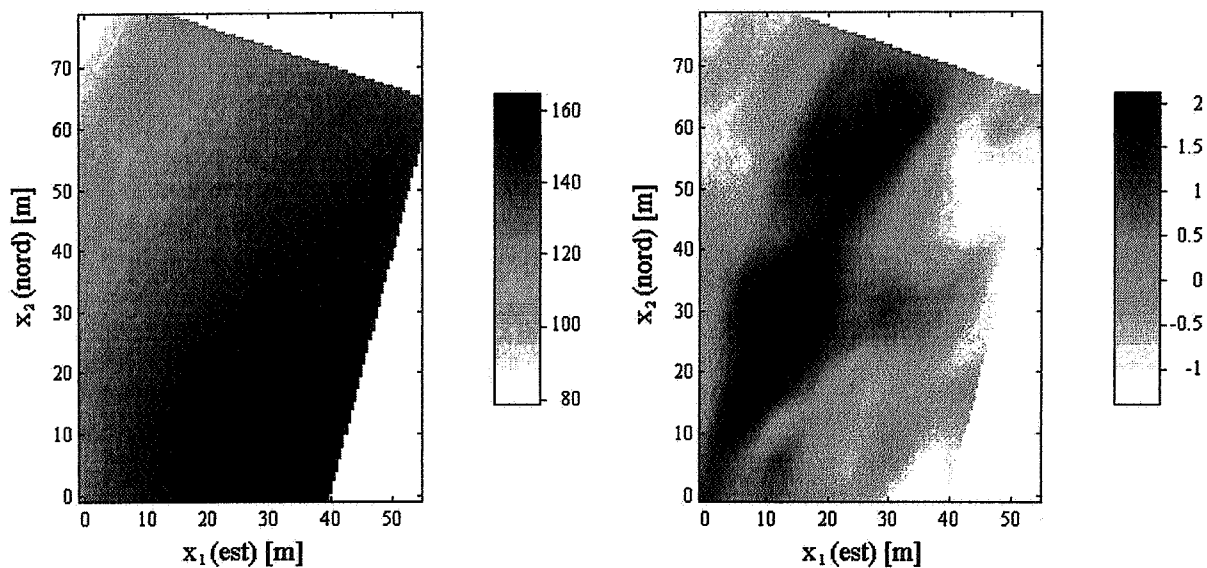


figure 7.16 : estimations des premiers facteurs relatifs au modèle puissance (à gauche) et au schéma sphérique (à droite)

Le premier facteur relatif au schéma puissance, qui correspond à une structure non stationnaire, affiche une nette dérive dans la direction nord-ouest / sud-est. En quelque sorte, il prend en charge la non-stationnarité du phénomène. Quant au premier facteur du schéma sphérique, il peut être vu comme un résumé de la structure sous-jacente, lorsque le phénomène est débarrassé de sa dérive. Le facteur estimé correspond en effet, par construction, à une structure stationnaire dont l'échelle de variation est de l'ordre de 20 à 40 mètres (portée du schéma sphérique dans les directions principales d'anisotropie).

De telles cartes peuvent servir à déterminer une partition du champ en zones homogènes, lesquelles pourront trouver une interprétation pédologique. Ainsi, la parcelle étudiée dans l'exemple a fait l'objet d'une étude au moment où les variables ont été mesurées. Une carte pédologique s'appuyant sur des observations qualitatives (comme la couleur, la structure et la texture des échantillons) a alors été établie et a permis d'identifier quatre unités pédologiques, à savoir :

- colluvions (C)
- colluvions sur tuf (C/T)
- colluvions sur limon (C/L)
- limon (L)

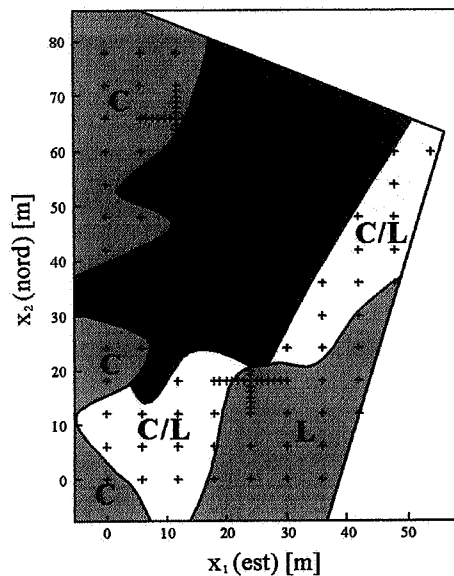


figure 7.17 : carte pédologique établie sur la base d'informations qualitatives

Cette carte n'est pas sans lien avec les cartes des facteurs précédemment présentées, notamment celle du premier facteur de la structure sphérique. En particulier, on peut distinguer sur cette dernière les unités "colluvions sur tuf" (C/T), "colluvions" (C) et "limon" (L). Les cartes des facteurs sont donc empreintes d'une certaine signification pédologique.

V. Modèles non stationnaires

V.1. Cokrigage universel

V.1.1. Hypothèses

Il est possible d'étendre le cokrigage à des corégionalisations non stationnaires, suivant la méthode du krigage universel. On suppose que chacune des fonctions aléatoires Z_i possède une dérive déterministe $m_i(\mathbf{x})$ représentable, à l'échelle de travail (voisinage glissant), par une combinaison linéaire de fonctions de base :

$$\forall i=1\dots N, Z_i(\mathbf{x}) = Y_i(\mathbf{x}) + m_i(\mathbf{x}) \quad \text{où} \quad m_i(\mathbf{x}) = \sum_{\ell} a_{\ell}^i f_{\ell}^i(\mathbf{x}).$$

Pour chaque variable Z_i , la somme ne porte que sur un nombre restreint et connu d'indices ℓ . On choisit les fonctions de base parmi les exponentielles – polynômes (en pratique, ce sont des monômes, parfois des fonctions trigonométriques) et on impose pour chaque variable que la première fonction, correspondant à $\ell = 0$, soit constante :

$$\forall i=1\dots N, f_i^0(\mathbf{x}) \equiv 1.$$

C'est naturellement sur les résidus $Y_i(\mathbf{x})$ que vont porter les hypothèses de stationnarité. Tout ce qui est relatif à ces résidus est qualifié de *sous-jacent*.

V.1.2. Cokrigage universel ponctuel

Nous nous limitons au cas où les résidus Y_i sont conjointement stationnaires du second ordre, c'est-à-dire où les fonctions aléatoires Y_i (et Z_i par la même occasion) admettent des fonctions de covariance simples et croisées, et nous ne présentons que le cokrigage ponctuel. On pourrait envisager le cokrigage d'une fonctionnelle linéaire des Z_i , mais les équations correspondantes se déduisent par linéarité du système de cokrigage ponctuel.

Pour estimer la variable Z_{i_0} au point $\mathbf{x}_0$, on forme la combinaison linéaire :

$$Z_{i_0}^*(\mathbf{x}_0) = \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i Z_i(\mathbf{x}_{\alpha}^i).$$

La condition d'autorisation n'est pas active (toutes les combinaisons linéaires sont autorisées lorsque les résidus Y_i sont stationnaires). Après écriture des conditions de non biais et d'optimalité, on trouve que les poids de cokrigage vérifient le système :

$$\begin{cases} \sum_{j=1}^N \sum_{\beta=1}^{n_j} \lambda_{\beta}^j C_{ij}(\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j) + \sum_{\ell} \mu_{\ell}^i f_{\ell}^{\ell}(\mathbf{x}_{\alpha}^i) = C_{ii_0}(\mathbf{x}_{\alpha}^i - \mathbf{x}_0) \quad \forall i=1 \dots N, \forall \alpha=1 \dots n_i \\ \sum_{\beta=1}^{n_i} \lambda_{\beta}^i f_{\ell}^{\ell}(\mathbf{x}_{\beta}^i) = \begin{cases} f_{i_0}^{\ell}(\mathbf{x}_0) & \text{si } i=i_0 \\ 0 & \text{sinon} \end{cases} \quad \forall \ell, \forall i=1 \dots N \end{cases}$$

et la variance de cokrigage vaut :

$$\sigma_{CKU}^2(\mathbf{x}_0, i_0) = C_{i_0 i_0}(\mathbf{0}) - \sum_{\ell} \mu_{\ell}^{i_0} f_{i_0}^{\ell}(\mathbf{x}_0) - \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i C_{ii_0}(\mathbf{x}_{\alpha}^i - \mathbf{x}_0).$$

On peut aussi proposer un estimateur optimal des dérivées (voire de chacun de leurs coefficients) des fonctions aléatoires Z_i . Mais on se doute que, comme dans le cas d'une seule variable, **l'analyse variographique concentrera l'essentiel des difficultés pour la mise en œuvre du cokrigage universel**. Les résultats concernant le krigeage universel restent valables pour le cokrigage : les covariances et variogrammes simples et croisés sous-jacents sont théoriquement indéterminés, car l'inférence à partir des données initiales ou des résidus *estimés* est biaisée (cf. chapitre 6).

V.1.3. Propriétés du cokrigage universel

Outre les propriétés classiques du cokrigage, à savoir l'interpolation exacte et l'additivité (qui permet de ramener l'estimation de toute fonction linéaire des variables à des estimations ponctuelles), le cokrigage universel possède les propriétés suivantes :

- **régularité du système**

L'existence et l'unicité de la solution sont garanties s'il y a indépendance linéaire des fonctions de base sur les données :

$$\left(\sum_{\ell} c_{\ell} f_{\ell}^{\ell}(\mathbf{x}_{\alpha}^i) = 0 \quad \forall i=1 \dots N \text{ et } \forall \alpha=1 \dots n_i \right) \Rightarrow c_{\ell} = 0 \quad \forall \ell.$$

- **cas de simplification du cokrigage universel (autokrigeabilité)**

Comme dans le cadre stationnaire, il existe deux situations remarquables où le cokrigage d'une variable se réduit à son krigeage. La première, évidente, est celle où toutes les variables Z_i sont indépendantes. La seconde est plus intéressante : dans le cas où les résidus sont en corrélation intrinsèque, on peut estimer chaque variable comme si elle était seule, à la triple condition que toutes les variables soient connues en tous les points de mesure (*homotopie*), que les dérivées $m_i(\mathbf{x})$ s'expriment à l'aide des mêmes fonctions de base, et qu'elles soient linéairement indépendantes.

V.2. Cokrigage intrinsèque

Il est possible d'élargir la théorie des fonctions aléatoires intrinsèques d'ordre k au cadre multivariable. On suppose que :

- chaque fonction aléatoire Z_i est intrinsèque d'ordre k_i , de fonctions de base $f_i^\ell(\mathbf{x})$ et de covariance généralisée $K_{ij}(\mathbf{h})$;
- la propriété intrinsèque s'étend aux accroissements généralisés croisés :

$$\forall i, j \in [1, N], \text{cov}[\sum_{\alpha} \lambda_{\alpha} Z_i(\mathbf{x}_{\alpha}^i), \sum_{\beta} \mu_{\beta} Z_j(\mathbf{x}_{\beta}^j)] = \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \mu_{\beta} K_{ij}(\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j)$$

$$\text{dès lors que } \sum_{\alpha} \lambda_{\alpha} f_i^\ell(\mathbf{x}_{\alpha}^i) = \sum_{\beta} \mu_{\beta} f_j^\ell(\mathbf{x}_{\beta}^j) = 0.$$

Le modèle linéaire de corégionalisation fait l'hypothèse que les covariances généralisées simples et croisées sont des combinaisons linéaires des mêmes covariances élémentaires $K_u(\mathbf{h})$:

$$\forall i, j \in [1, N], K_{ij}(\mathbf{h}) = \sum_{u=1}^S b_{ij}^u K_u(\mathbf{h})$$

avec $\mathbf{B}_u = (b_{ij}^u)_{i,j}$: matrice de corégionalisation, symétrique et de type positif.

Usuellement, on utilise des covariances généralisées polynomiales isotropes, auxquelles on ajoute un effet de pépite $\delta(\mathbf{h})$:

$$K_{ij}(\mathbf{h}) = b_{ij}^0 \delta(\mathbf{h}) + \sum_{u=1}^S b_{ij}^u (-1)^u |\mathbf{h}|^{2u-1} \text{ avec } b_{ij}^u = 0 \text{ lorsque } u-1 > \min(k_i, k_j).$$

La mise en œuvre de ce modèle nécessite d'abord de déterminer, séparément, les ordres $\{k_i, i = 1 \dots N\}$ des diverses fonctions aléatoires, puis de trouver les matrices de corégionalisation qui ajustent au mieux les covariances généralisées simples et croisées. Comme dans le cas monovarié, on a recours à des procédures automatiques qui utilisent des critères de moindres carrés pondérés et évitent les problèmes de biais rencontrés en cokrigage universel. Vu la complexité du modèle et le manque de robustesse de la procédure automatique, il est recommandé de vérifier la qualité de l'ajustement par validation croisée.

Une fois spécifié le modèle variographique multivariable, le cokrigage intrinsèque s'obtient en reprenant le système de cokrigage universel, mais en remplaçant les covariances usuelles par les covariances généralisées correspondantes. Ainsi, l'estimateur de la variable Z_{i_0} en un site $\mathbf{x}_0$ s'écrit

$$Z_{i_0}^*(\mathbf{x}_0) = \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i Z_i(\mathbf{x}_{\alpha}^i)$$

où les pondérateurs sont déterminés par le système

$$\begin{cases} \sum_{j=1}^N \sum_{\beta=1}^{n_j} \lambda_{\beta}^j K_{ij}(\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j) + \sum_{\ell} \mu_{\ell}^i f_i^{\ell}(\mathbf{x}_{\alpha}^i) = K_{i i_0}(\mathbf{x}_{\alpha}^i - \mathbf{x}_0) & \forall i = 1 \dots N, \forall \alpha = 1 \dots n_i \\ \sum_{\beta=1}^{n_i} \lambda_{\beta}^i f_i^{\ell}(\mathbf{x}_{\beta}^i) = \begin{cases} f_{i_0}^{\ell}(\mathbf{x}_0) & \text{si } i = i_0 \\ 0 & \text{sinon} \end{cases} & \forall \ell, \forall i = 1 \dots N \end{cases}$$

$$\text{La variance de cokrigage est : } \sigma_{CKI-k}^2(\mathbf{x}_0, i_0) = K_{i_0 i_0}(\mathbf{0}) - \sum_{\ell} \mu_{\ell}^{i_0} f_{i_0}^{\ell}(\mathbf{x}_0) - \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i K_{i i_0}(\mathbf{x}_{\alpha}^i - \mathbf{x}_0).$$

V.3. Dérive externe

Le (co)krigeage universel ou intrinsèque n'utilise que des dérivées polynomiales ou, dans de rares cas, trigonométriques, qui peuvent être inadaptées à la description du phénomène régionalisé. Nous allons présenter une méthode géostatistique multivariable qui consiste à considérer, pour mettre en œuvre l'estimation, une ou plusieurs variables auxiliaires connues sur l'ensemble du champ et donnant, en quelque sorte, la "forme" de la variable cible.

V.3.1. Un exemple introductif

En exploration pétrolière, on cherche à cartographier la profondeur d'horizons géologiques. On dispose en général de deux types de mesure :

- des mesures précises de la profondeur de l'horizon, faites sur des forages pétroliers. Leur nombre est cependant limité à cause du coût élevé des forages. Elles sont modélisées par une fonction aléatoire $Z(\mathbf{x})$;
- des mesures indirectes, très nombreuses mais peu précises, réalisées à partir de méthodes sismiques. Elles sont représentées par la variable régionalisée $\zeta(\mathbf{x})$.

$\zeta(\mathbf{x})$ est connue (presque) exhaustivement, et n'a donc pas besoin d'être randomisée. Elle ne donne que l'allure générale de la profondeur de l'horizon, mesurée par $Z(\mathbf{x})$. $\zeta(\mathbf{x})$ et $Z(\mathbf{x})$ se référant à la même grandeur physique, on peut raisonnablement supposer que l'espérance de $Z(\mathbf{x})$ est égale à $\zeta(\mathbf{x})$:

$$E[Z(\mathbf{x})] = \zeta(\mathbf{x}).$$

En fait, pour permettre plus de souplesse dans le modèle, on fait l'hypothèse que la relation entre l'espérance de $Z(\mathbf{x})$ et $\zeta(\mathbf{x})$ est affine, c'est-à-dire que l'espérance de $Z(\mathbf{x})$ est proportionnelle à $\zeta(\mathbf{x})$ à une constante additive près :

$$E[Z(\mathbf{x})] = a_0 + b_1 \zeta(\mathbf{x}).$$

$\zeta(\mathbf{x})$ est une variable explicative, ou encore une fonction de forme, qui décrit l'allure, ou encore la *tendance*, du phénomène, tandis que les mesures de $Z(\mathbf{x})$ donnent la valeur précise de la profondeur de l'horizon.

Un autre exemple pourrait être celui de la température au niveau du sol sur une zone géographique. On peut en effet supposer que la température $Z(\mathbf{x})$ est liée linéairement à l'altitude $\zeta(\mathbf{x})$, et utiliser cette dernière comme fonction de forme décrivant l'allure de $Z(\mathbf{x})$.

V.3.2. Krigage avec une dérive externe

Nous allons examiner le cas d'une fonction aléatoire $Z(\mathbf{x})$ dont on souhaite améliorer l'estimation en introduisant la fonction de forme déterministe $\zeta(\mathbf{x})$, ou dérive "externe".

Nous supposons que :

$$\begin{cases} E[Z(\mathbf{x})] = a_0 + b_1 \zeta(\mathbf{x}) & \forall \mathbf{x} \in \mathcal{D} \\ \text{cov}[Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})] = C(\mathbf{h}) & \forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{D} \end{cases}$$

où les coefficients a_0 et b_1 sont inconnus.

La condition de linéarité du krigage impose d'écrire l'estimation en un point quelconque $\mathbf{x}_0$ comme une combinaison linéaire pondérée des données disponibles :

$$Z^*(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha}).$$

La condition d'autorisation est automatiquement satisfaite, car les deux premiers moments de $Z(\mathbf{x})$ existent et sont finis. La condition de non biais s'écrit :

$$E[Z^*(\mathbf{x}_0)] = \sum_{\alpha=1}^n \lambda_{\alpha} E[Z(\mathbf{x}_{\alpha})] = a_0 \sum_{\alpha=1}^n \lambda_{\alpha} + b_1 \sum_{\alpha=1}^n \lambda_{\alpha} \zeta(\mathbf{x}_{\alpha}) = E[Z(\mathbf{x}_0)] = a_0 + b_1 \zeta(\mathbf{x}_0).$$

Pour garantir cette égalité quels que soient les coefficients a_0 et b_1 inconnus, on doit poser

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} = 1 \\ \sum_{\beta=1}^n \lambda_{\beta} \zeta(\mathbf{x}_{\beta}) = \zeta(\mathbf{x}_0) \end{cases}$$

Il reste à minimiser la variance de l'erreur d'estimation sous ces contraintes de non biais. L'explicitation du système de krigage fournit alors

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) - \mu_1 - \mu_2 \zeta(\mathbf{x}_{\alpha}) = C(\mathbf{x}_{\alpha} - \mathbf{x}_0) & \text{pour } \alpha = 1 \dots n \\ \sum_{\beta=1}^n \lambda_{\beta} = 1 \\ \sum_{\beta=1}^n \lambda_{\beta} \zeta(\mathbf{x}_{\beta}) = \zeta(\mathbf{x}_0) \end{cases}$$

On reconnaît le système de krigage universel classique, à ceci près que l'on a remplacé les fonctions de base f^j polynomiales par la fonction de forme $\zeta(\mathbf{x})$. Cette dernière doit être connue en tous les points de données de $Z(\mathbf{x})$, ainsi qu'au site à estimer.

La dérive externe étant unique, il est possible d'identifier les directions de l'espace où elle est approximativement constante, ce qui facilite l'analyse variographique (cf. chapitre 6).

V.3.3. Krigeage avec dérivées externes multiples

Nous allons à présent généraliser la méthode de la dérive externe en examinant le cas où il n'existe pas de fonction de covariance stationnaire, et en introduisant plusieurs fonctions de forme. On considère une fonction aléatoire Z intrinsèque d'ordre k , admettant une covariance généralisée K . Comme d'habitude, les fonctions de base (monômes de degré $\leq k$) définissant les combinaisons linéaires autorisées seront notées $f'(\mathbf{x})$ et l'on impose $f^0(\mathbf{x}) \equiv 1$.

Le krigeage avec dérivées externes consiste à ajouter au système de krigeage intrinsèque d'ordre k des conditions de non-biais supplémentaires, relatives à une ou plusieurs variables externes $\zeta_i(\mathbf{x})$, $i = 1 \dots N$, mesurées partout dans le domaine. En réalité, il suffit de connaître les ζ_i en tous les points de données de Z , ainsi qu'au site à estimer.

Le système de krigeage en un point $\mathbf{x}_0$ s'écrit :

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} K(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) - \sum_{\ell} \mu_{\ell} f'_{\ell}(\mathbf{x}_{\alpha}) - \sum_{i=1}^N v_i \zeta_i(\mathbf{x}_{\alpha}) = K(\mathbf{x}_{\alpha} - \mathbf{x}_0) & \forall \alpha = 1 \dots n \\ \sum_{\beta=1}^n \lambda_{\beta} f'_{\ell}(\mathbf{x}_{\beta}) = f'_{\ell}(\mathbf{x}_0) & \forall \ell \\ \sum_{\beta=1}^n \lambda_{\beta} \zeta_i(\mathbf{x}_{\beta}) = \zeta_i(\mathbf{x}_0) & \forall i = 1 \dots N \end{cases}$$

où les μ_{ℓ} et les v_i sont des multiplicateurs de Lagrange.

La dénomination de dérivées *externes* provient du fait que les conditions

$$\sum_{\alpha=1}^n \lambda_{\alpha} \zeta_i(\mathbf{x}_{\alpha}) = \zeta_i(\mathbf{x}_0) \quad \text{pour } i = 1 \dots N$$

sont des conditions de non-biais, indépendantes du choix de la covariance $K(\mathbf{h})$ et du degré k^8 . Au contraire, les équations

$$\sum_{\alpha=1}^n \lambda_{\alpha} f'_{\ell}(\mathbf{x}_{\alpha}) = f'_{\ell}(\mathbf{x}_0) \quad \forall \ell$$

correspondent à des conditions d'autorisation, et influent sur le choix de la covariance généralisée.

Comme pour l'estimation de la dérive — ou de ses coefficients — en krigeage universel, il est possible d'estimer les coefficients des dérivées externes, qui mesurent l'influence exercée par ces dernières. Mais de même qu'en krigeage universel, c'est l'analyse variographique (inférence de la covariance généralisée sous-jacente) qui s'avère l'étape la plus difficile dans la mise en œuvre de la technique des dérivées externes, car les combinaisons linéaires autorisées n'annulent pas les dérivées externes. Nous renvoyons au chapitre 6 sur le krigeage universel pour l'examen des problèmes d'analyse variographique.

⁸ Le krigeage universel est, en fin de compte, un cas particulier de krigeage avec dérivées externes, car les fonctions de base n'ont pas d'impact sur le choix de l'outil structural, qui est une covariance usuelle ou un variogramme, quel que soit le degré de la dérive polynomiale. En revanche, il n'en est pas de même du krigeage intrinsèque d'ordre k : la covariance généralisée est liée à l'ensemble de fonctions de base, qui forment une dérive *interne* (l'ordre k étant choisi en fonction du degré de cette dérive polynomiale de manière à l'éliminer du formalisme).

Résumé

Lorsque plusieurs variables régionalisées présentent des corrélations spatiales, on a souvent intérêt à les étudier et les estimer simultanément. Des outils multivariés (covariances ou variogrammes *simples et croisés*) permettent de décrire leur structure spatiale conjointe. L'analyse variographique multivariable est basée sur le **modèle linéaire de corégionalisation**, simple et suffisamment général pour la plupart des applications pratiques. L'ajustement des paramètres du modèle doit respecter certaines contraintes, et se fait à l'aide de procédures automatiques dès qu'il y a plus de deux variables.

Le **cokrigeage** permet d'estimer simultanément plusieurs variables régionalisées en tenant compte de leur structure *conjointe*. Cette technique est particulièrement avantageuse lorsque la variable cible est moins bien connue que des variables auxiliaires. Le cokrigeage possède les mêmes propriétés que le krigeage (interpolation exacte, lissage...) et peut être étendu à l'estimation de blocs. Sa principale limitation est l'augmentation des temps de calcul lorsque le nombre de variables est important.

L'**analyse krigeante** combine les méthodes géostatistiques et d'analyse des données en un même formalisme. Elle consiste à décomposer les fonctions aléatoires étudiées en *facteurs* indépendants, hiérarchisés en fonction de la quantité d'information qu'ils contiennent. Ces facteurs peuvent être estimés à partir des données par cokrigeage. Leurs valeurs sont une synthèse des données initiales, de la structure spatiale conjointe des variables régionalisées et de l'échelle d'observation.

Enfin, il est possible d'étendre les techniques multivariées à des corégionalisations non stationnaires (cokrigeage universel et intrinsèque). Une méthode intéressante est celle de la **dérive externe**, où une ou plusieurs variables connues exhaustivement sont utilisées comme fonctions de forme pour la variable d'intérêt.

ANNEXES

Annexe A :

rappels des notions usuelles de probabilité

I. Espace probabilisé

Une expérience aléatoire se modélise par la donnée d'un **espace probabilisé** défini par un triplet $(\Omega, \mathcal{A}, P)$:

- Ω désigne l'ensemble des résultats possibles ou *univers*.
- $\mathcal{A}$ est une tribu (σ -algèbre) sur Ω ; elle décrit l'ensemble des propositions (ou *événements*) que l'on peut énoncer sur les éléments de Ω .
- P est une fonction de probabilité (i.e. une mesure positive de poids total 1) définie sur $\mathcal{A}$.

On a ainsi substitué à l'expérience aléatoire réelle la collection de tous ses résultats possibles et des probabilités de réalisation des événements associés. Par exemple, pour le lancer d'un dé non pipé, on a $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{A}$ est l'ensemble des parties de Ω et P est la probabilité qui donne un poids égal à $1/6$ à chaque élément de Ω .

II. Variable aléatoire

Une **variable aléatoire** U est une grandeur qui à tout élément ω de Ω associe une valeur numérique — ou **réalisation** — $U(\omega) \in \mathcal{E}$. Désormais, nous prendrons $\mathcal{E} = \mathbb{R}$ (variable aléatoire continue) ou $\mathcal{E} = \mathbb{N}$ (variable aléatoire discrète).

III. Fonction de répartition et densité de probabilité

La **fonction de répartition** d'une variable aléatoire U est définie par

$$\forall u \in \mathcal{E} \text{ (R ou N)}, F(u) = P(U < u)$$

ou de manière rigoureuse par $F(u) = P(\{\omega \in \Omega / U(\omega) < u\})$.

C'est une fonction non décroissante, à valeurs dans l'intervalle $[0,1]$. On a : $F(-\infty) = 0$ et $F(+\infty) = 1$.

La **densité de probabilité** est égale à $P(U = u)$ pour une variable discrète. Dans le cas d'une variable absolument continue, c'est une fonction $f(u)$ positive telle que

$$F(u) = \int_{-\infty}^u f(t) dt.$$

En particulier, $\int_{-\infty}^{+\infty} f(t) dt = F(+\infty) = 1$. La densité $f(u)$ est proportionnelle à la probabilité que U soit dans l'intervalle infinitésimal $[u, u+du[$:

$$P(U \in [u, u+du[) = F(u+du) - F(u) \underset{du \rightarrow 0}{\sim} f(u) du.$$

La fonction de répartition et la densité de probabilité caractérisent la loi de probabilité de la variable aléatoire U .

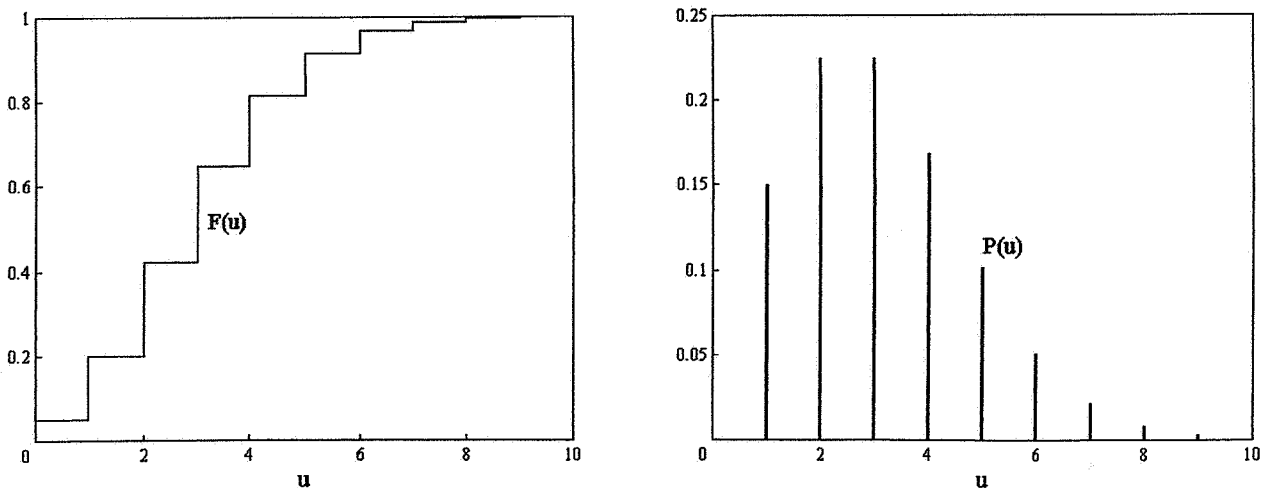


figure A.1 : fonction de répartition et densité de probabilité d'une variable aléatoire discrète (de Poisson)

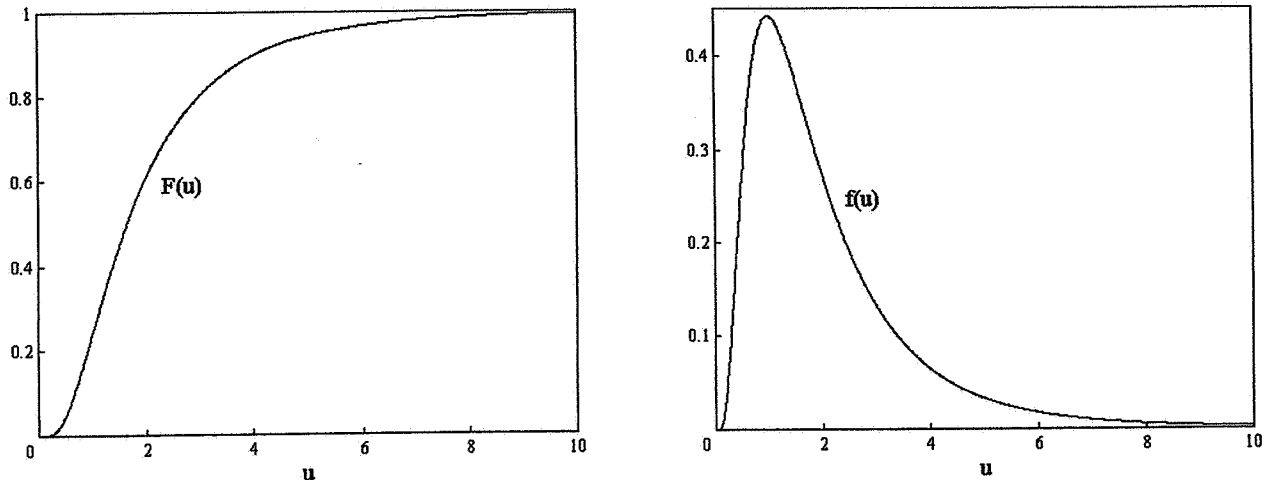


figure A.2 : fonction de répartition et densité de probabilité d'une variable aléatoire continue (lognormale)

IV. Espérance et moments d'une variable aléatoire

L'**espérance mathématique** désigne la "moyenne" autour de laquelle les valeurs de U se distribuent selon la loi de probabilité P . Elle est définie, pour une variable continue, par :

$$E(U) = \int_{\mathbb{R}} u f(u) du .$$

Si U est discrète, l'intégrale devient une somme sur les valeurs prises par U , pondérées par les probabilités correspondantes :

$$E(U) = \sum_i u_i f(u_i) = \sum_i u_i P(U = u_i) .$$

On appelle **moment d'ordre k** ($k \in \mathbb{N}$) la quantité

$$\mu_k = E(U^k)$$

et **moment centré d'ordre k**

$$m_k = E \{ [U - E(U)]^k \} .$$

Les moments donnent une description synthétique de la loi d'une variable aléatoire. Après le moment d'ordre 1, qui n'est autre que l'espérance, le plus important est le moment centré d'ordre 2, appelé **variance** et noté σ^2 :

$$m_2 = \sigma^2 = E\{[U - E(U)]^2\} = E(U^2) - [E(U)]^2 \geq 0.$$

La racine carrée de la variance est l'**écart-type**, noté σ . La variance et l'écart-type sont des indicateurs de la dispersion des valeurs de U autour de son espérance.

Deux indices sont usuellement calculés à partir des moments centrés :

$$a_3 = m_3 / \sqrt{m_2^3} \text{ appelé coefficient d'asymétrie}$$

$$a_4 = m_4 / m_2^2 \text{ appelé coefficient d'aplatissement.}$$

Ils donnent une idée de la forme de la densité de probabilité.

V. Variables aléatoires vectorielles

Un ensemble de variables aléatoires $(U_1, \dots, U_n)$ constitue une variable aléatoire vectorielle ou vecteur aléatoire. Un vecteur aléatoire peut être discret, continu, ou mixte (certaines de ses composantes peuvent être continues, d'autres discrètes). On définit sa **fonction de répartition** par :

$$F(u_1, \dots, u_n) = P(U_1 < u_1, \dots, U_n < u_n).$$

Dans le cas de variables absolument continues, on peut définir une **densité de probabilité conjointe** f dépendant de n arguments telle que :

$$\forall D \subset \mathbb{R}^n, P((U_1, \dots, U_n) \in D) = \int_D f(u_1, \dots, u_n) du_1 \dots du_n.$$

La densité conjointe est liée à la probabilité que le vecteur aléatoire prenne ses valeurs dans un domaine infinitésimal :

$$P(U_1 \in [u_1, u_1 + du_1[, \dots, U_n \in [u_n, u_n + du_n[) \underset{\substack{du_1 \rightarrow 0 \\ \vdots \\ du_n \rightarrow 0}}{\sim}} f(u_1, \dots, u_n) du_1 \dots du_n.$$

Les **densités marginales**, c'est-à-dire celles des composantes $U_1, \dots, U_n$, s'obtiennent par intégration par rapport aux arguments correspondant aux autres composantes :

$$f_{U_i}(u) = \int_{\mathbb{R}^{n-1}} f(u_1, \dots, u_{i-1}, u, u_{i+1}, \dots, u_n) du_1 \dots du_{i-1} du_{i+1} \dots du_n.$$

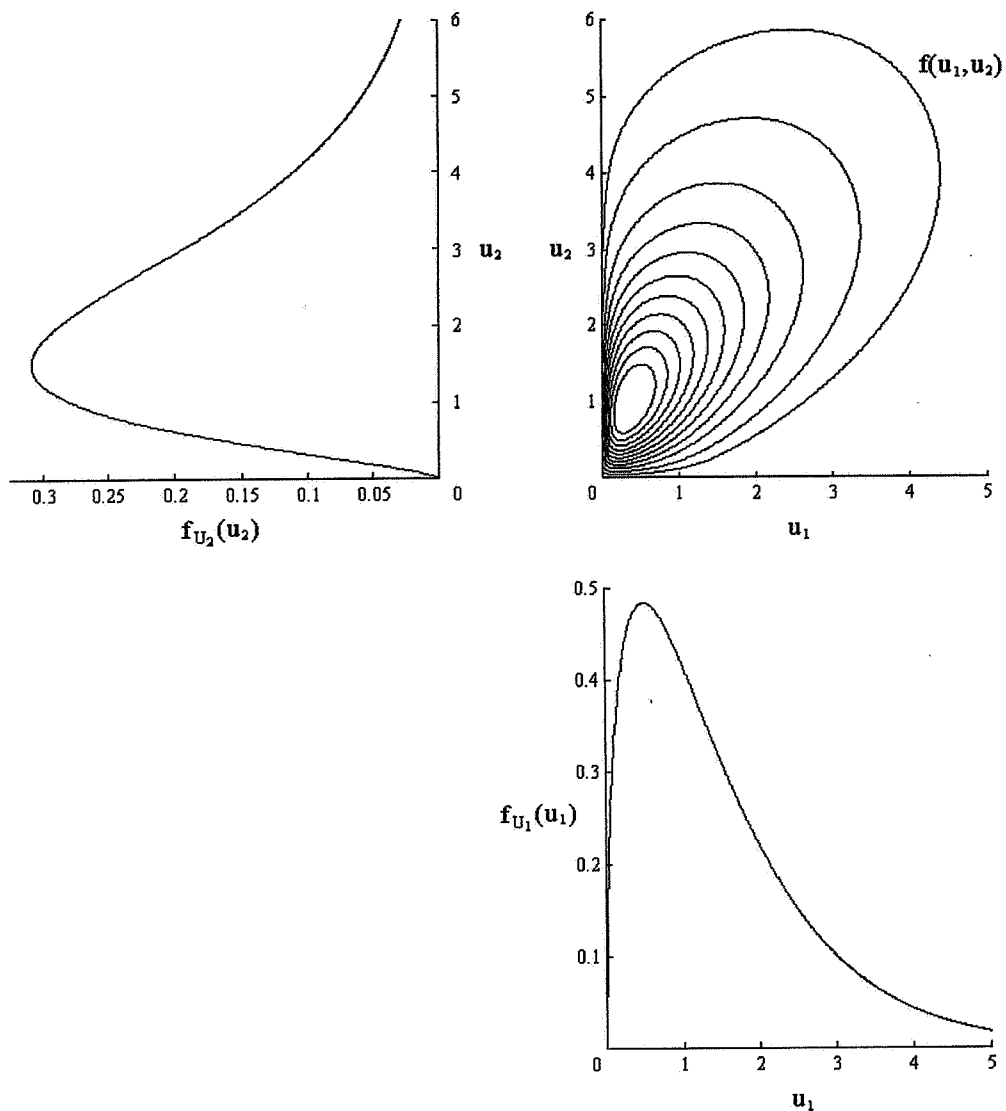


figure A.3 : un exemple de densité conjointe bivariable (représentée sous la forme de courbes de niveau) et les densités marginales associées

Les variables $U_1, \dots, U_n$ sont **indépendantes** si et seulement si la densité conjointe se factorise en le produit des densités marginales

$$f(u_1, \dots, u_n) = \prod_{i=1}^n f_{U_i}(u_i).$$

Un tel cas signifie qu'une information sur l'une des variables n'implique rien sur les autres.

Les deux premiers moments d'une variable vectorielle sont :

- le vecteur des espérances $(E(U_1), \dots, E(U_n))$
- la **matrice de variance-covariance**, qui décrit de manière synthétique les liaisons ou "interactions" entre les diverses composantes :

$$C = (\sigma_{ij})_{i,j}$$

où $\sigma_{ij} = \text{cov}(U_i, U_j) = E\{[U_i - E(U_i)][U_j - E(U_j)]\}$: covariance centrée entre U_i et U_j .

On a les propriétés suivantes :

- $\sigma_{ij} = E\{[U_i - E(U_i)][U_j - E(U_j)]\} = E(U_i U_j) - E(U_i)E(U_j)$.
- $\sigma_{ii} = \text{var}(U_i)$: la covariance entre une variable aléatoire et elle-même n'est autre que sa variance.
- $\text{cov}^2(U_i, U_j) \leq \text{var}(U_i) \text{var}(U_j)$: inégalité de Cauchy-Schwarz.
- Si $U_1, \dots, U_n$ sont indépendantes, alors $\forall i \neq j, \text{cov}(U_i, U_j) = 0$, c'est-à-dire que la matrice C est diagonale. En général, la réciproque est fausse.

On définit le *coefficient de corrélation linéaire* entre U_i et U_j par :

$$\rho_{ij} = \frac{\text{cov}(U_i, U_j)}{\sqrt{\text{var}(U_i) \text{var}(U_j)}}.$$

Ce coefficient, qui prend ses valeurs dans $[-1, 1]$, mesure l'importance de la liaison *linéaire* entre U_i et U_j . Si $\rho_{ij} = 0$, U_i et U_j n'ont pas de corrélation (cas des variables indépendantes) ; au contraire, si $\rho_{ij} = \pm 1$, elles sont linéairement liées et l'on peut écrire : $U_i = \alpha U_j + \beta$, avec α du signe de ρ_{ij} . Attention : une relation *non linéaire* existant entre U_i et U_j peut être invisible si l'on se limite au calcul de ρ_{ij} ; pour éviter des conclusions erronées, il convient toujours de visualiser le nuage de corrélation correspondant, i.e. le nuage des points formés par les réalisations disponibles du couple (U_i, U_j) .

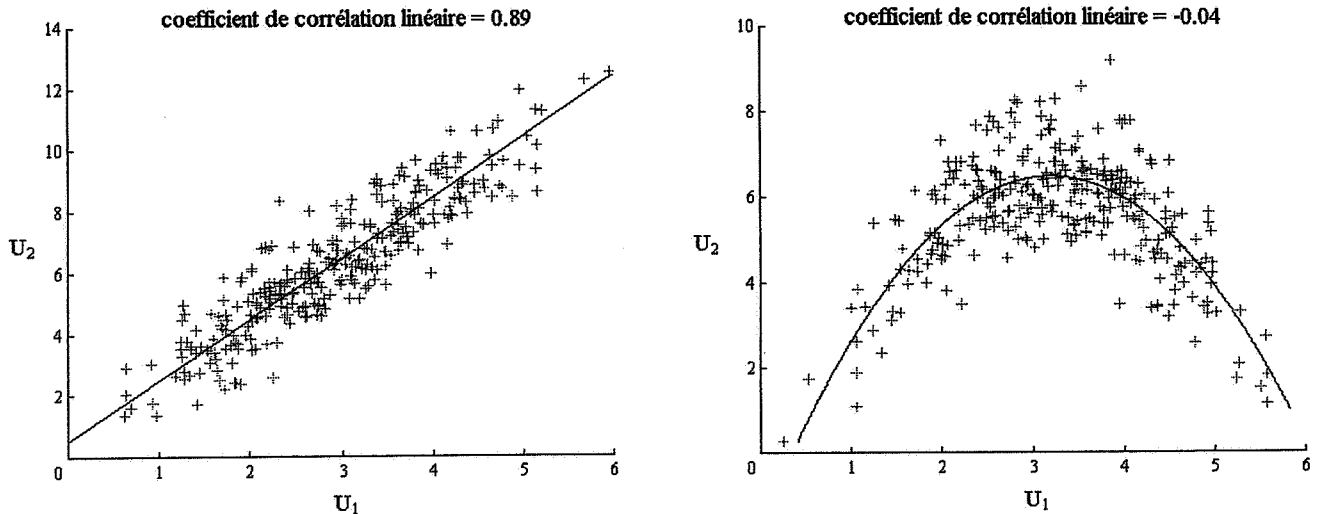


figure A.4 : exemples de nuages de corrélation entre deux variables aléatoires U_1 et U_2 ; dans le second cas, le coefficient de corrélation linéaire ne révèle pas la liaison (non linéaire) entre les variables

VI. Loi de probabilité conditionnelle

En général, la loi de probabilité d'une variable aléatoire V (appelée aussi loi *a priori*) est modifiée lorsque l'on dispose d'une information partielle sur l'expérience aléatoire, par exemple lorsqu'on connaît une réalisation d'une variable aléatoire U non indépendante de V .

Considérons un couple de variables aléatoires absolument continues (U, V) , de densité de probabilité conjointe $f(u, v)$ et de densités marginales $f_U(u)$ et $f_V(v)$. On appelle **densité conditionnelle** de V à $U \equiv u$ la fonction

$$f_V^u(v) = \frac{f(u, v)}{f_U(u)}.$$

On écrit parfois, de façon symbolique, $f_V^u(v) = f_V(v | U \equiv u)$.

On appelle **fonction de répartition conditionnelle** de V à $U \equiv u$ la fonction

$$F_V^u(v) = \int_{-\infty}^v f_V^u(t) dt, \text{ notée parfois } P(V < v | U \equiv u).$$

La densité *a priori* f_V de V est la moyenne des densités conditionnelles f_V^u pondérées par la densité de U :

$$f_V(v) = \int_{-\infty}^{+\infty} f_V^u(v) f_U(u) du.$$

Les densités conditionnelles fournissent un critère d'indépendance des variables U et V : elles sont indépendantes si et seulement si $f_V^u(v)$ ne dépend pas de u , auquel cas $f_V^u(v) = f_V(v)$. La donnée de la réalisation de U n'altère pas la loi de probabilité de V . La densité conjointe se factorise alors en le produit des densités marginales :

$$f(u, v) = f_U(u) f_V(v).$$

Les densités conditionnelles permettent de définir les **moments conditionnels** de V à $U \equiv u$. Par exemple, en ce qui concerne le moment d'ordre 1 (espérance) :

$$E(V|U \equiv u) = \int_{\mathbb{R}} v f_V^u(v) dv : \text{espérance conditionnelle de } V \text{ à } U \equiv u.$$

L'espérance conditionnelle de V à $U \equiv u$ ainsi définie est une valeur numérique, dont l'expression analytique dépend de u . En remplaçant formellement u par U dans cette expression, c'est-à-dire en restituant à U son caractère aléatoire, l'**espérance conditionnelle** devient une variable aléatoire, notée $E(V|U)$. Il s'agit de la variable aléatoire construite sur U qui est la plus "proche" de V , au sens où $E(V|U)$ est l'estimateur de V par une fonction de U qui minimise la variance de l'erreur. En particulier, cet estimateur a la même espérance que V :

$$E\{E(V|U)\} = E(V) \text{ (formule de l'espérance totale).}$$

VII. Fonction aléatoire

Une fonction aléatoire peut être vue comme un vecteur aléatoire ayant une infinité de composantes, soit $\{U(\mathbf{x}), \mathbf{x} \in \mathcal{D}\}$, où $\mathcal{D}$ représente le domaine de définition de la fonction aléatoire. Le passage d'un nombre fini de composantes (vecteur aléatoire) à un nombre non dénombrable (fonction aléatoire) soulève cependant quelques difficultés théoriques.

Une fonction aléatoire est caractérisée par sa **loi spatiale**, qui correspond à la donnée de toutes les lois de probabilité multivariées entre ses composantes, c'est-à-dire à l'ensemble des fonctions de répartition

$$F_{\mathbf{x}_1, \dots, \mathbf{x}_k}(u_1, \dots, u_k) = P(U(\mathbf{x}_1) < u_1, \dots, U(\mathbf{x}_k) < u_k)$$

pour tous les choix possibles de l'entier k , des valeurs $u_1, \dots, u_k$ et des points $\mathbf{x}_1, \dots, \mathbf{x}_k$ dans $\mathcal{D}$. Ces fonctions de répartition doivent satisfaire deux contraintes particulières :

- *contrainte de symétrie* : l'ordre dans lequel on choisit les valeurs $u_1, \dots, u_k$ et les points $\mathbf{x}_1, \dots, \mathbf{x}_k$ correspondants est indifférent, de sorte que $F_{\mathbf{x}_1, \dots, \mathbf{x}_k}(u_1, \dots, u_k)$ est invariante par permutation des indices $1, \dots, k$.

- *contrainte de compatibilité* : si l'on prend $u_k = +\infty$, l'inégalité $U(\mathbf{x}_k) < u_k$ est toujours vérifiée et est inutile dans la définition de $F_{\mathbf{x}_1, \dots, \mathbf{x}_k}(u_1, \dots, u_k)$. On doit donc avoir :

$$\begin{aligned} F_{\mathbf{x}_1, \dots, \mathbf{x}_{k-1}, \mathbf{x}_k}(u_1, \dots, u_{k-1}, +\infty) &= P(U(\mathbf{x}_1) < u_1, \dots, U(\mathbf{x}_{k-1}) < u_{k-1}, U(\mathbf{x}_k) < +\infty) \\ &= P(U(\mathbf{x}_1) < u_1, \dots, U(\mathbf{x}_{k-1}) < u_{k-1}) \\ &= F_{\mathbf{x}_1, \dots, \mathbf{x}_{k-1}}(u_1, \dots, u_{k-1}) \end{aligned}$$

On introduit aussi les **moments** d'une fonction aléatoire, qui décrivent synthétiquement sa loi spatiale. Le moment (non centré) d'ordre $(\alpha_1, \dots, \alpha_k)$ est défini par

$$\mu_{\alpha_1, \dots, \alpha_k}(\mathbf{x}_1, \dots, \mathbf{x}_k) = E[U(\mathbf{x}_1)^{\alpha_1} \dots U(\mathbf{x}_k)^{\alpha_k}]$$

et dépend des points $\mathbf{x}_1, \dots, \mathbf{x}_k$.

Les moments *centrés* sont calculés en considérant $U(\mathbf{x}) - E[U(\mathbf{x})]$ (fonction aléatoire d'espérance nulle en tout point $\mathbf{x}$) au lieu de $U(\mathbf{x})$. Contrairement à la loi spatiale, les moments d'une fonction aléatoire peuvent très bien ne pas exister. En géostatistique linéaire, on ne s'intéresse qu'aux deux premiers moments, à savoir l'espérance $E[U(\mathbf{x})]$ et la fonction de covariance

$$C(\mathbf{x}_1, \mathbf{x}_2) = E\{[U(\mathbf{x}_1) - E(U(\mathbf{x}_1))][U(\mathbf{x}_2) - E(U(\mathbf{x}_2))]\}.$$

Annexe B : l'ergodicité

Parallèlement à l'hypothèse de stationnarité, nous avons mentionné (cf. chapitre 2) la nécessité d'une hypothèse d'ergodicité pour pouvoir réaliser l'inférence statistique. Après avoir défini ce concept, nous analyserons ses implications sur l'inférence des deux premiers moments de la fonction aléatoire modélisant le phénomène régionalisé.

I. Définition

L'ergodicité est une propriété qui relie la loi spatiale d'une fonction aléatoire et la structure de ses réalisations dans l'espace. Elle consiste à dire que, en supposant vérifiée l'hypothèse de stationnarité, la moyenne spatiale converge vers l'espérance mathématique lorsqu'elle est calculée sur un domaine de plus en plus grand. Plus précisément, pour tout entier naturel k , tout système de points $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ et toute fonction φ de $\mathbb{R}^k$ dans $\mathbb{R}$ telle que $E\{\varphi[Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)]\} < \infty$, l'intégrale

$$\bar{\varphi}_{\Sigma}(\mathbf{x}_1, \dots, \mathbf{x}_k) = \frac{1}{|\Sigma|} \int_{\Sigma} \varphi(Z(\mathbf{x}_1 + \mathbf{h}), \dots, Z(\mathbf{x}_k + \mathbf{h})) d\mathbf{h}$$

converge presque sûrement (c'est-à-dire avec probabilité 1) vers $E\{\varphi[Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)]\}$ lorsque l'étendue du domaine d'intégration Σ devient infinie.

L'ergodicité sous-tend la stationnarité, faute de quoi $E\{\varphi[Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)]\}$ ne serait pas invariante par translation de l'ensemble des points $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$, alors que $\bar{\varphi}_{\Sigma}(\mathbf{x}_1, \dots, \mathbf{x}_k)$ l'est si Σ est infini ; l'énoncé précédent n'aurait alors plus de sens.

Cette hypothèse permet d'estimer l'espérance mathématique de la fonction aléatoire et ses autres moments, définis par des espérances, en remplaçant une moyenne sur l'ensemble des réalisations par la moyenne spatiale de la régionalisation sur un domaine Σ très étendu, en pratique, sur le champ $\mathcal{D}$ entier. Il est ainsi possible d'**inférer les moments de la loi spatiale (espérance, covariance, variogramme...) à partir d'une unique réalisation**. On voit que le recours à un modèle ergodique est nécessaire pour permettre l'inférence statistique.

Un exemple de processus ergodique est une fonction aléatoire stationnaire et purement péritique, pour laquelle toutes les valeurs sont indépendantes et identiquement distribuées. La propriété d'ergodicité (identité entre moyenne spatiale et espérance) découle alors de la loi forte des grands nombres. Cet exemple n'est toutefois qu'un cas particulier, d'ailleurs peu intéressant en théorie des variables régionalisées puisqu'il correspond à une absence de structuration spatiale. Il existe fort heureusement de nombreux processus ergodiques qui ne sont pas purement péritiques.

Sans entrer dans les détails, une interprétation que l'on peut donner de l'ergodicité est que les événements liés à $\{Z(\mathbf{x}_1 + \mathbf{h}), \dots, Z(\mathbf{x}_k + \mathbf{h})\}$ deviennent indépendants de ceux relatifs à $\{Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_k)\}$ quand la distance $|\mathbf{h}|$ tend vers l'infini. Ceci constitue une généralisation de la propriété d'indépendance. L'intégrale d'espace $\bar{\varphi}_\Sigma(\mathbf{x}_1, \dots, \mathbf{x}_k)$ mélange des quantités corrélées entre elles, mais sur un domaine suffisamment grand pour que les corrélations se compensent. On dit qu'il y a "brassage" ou "mélange" du processus.

L'ergodicité est une propriété stochastique qui n'a pas de "traduction" au niveau de la variable régionalisée. Il n'y a donc pas de contrôle *a priori* du bien-fondé de représenter une variable régionalisée par un processus ergodique. En revanche, nous allons voir qu'il est possible de valider l'hypothèse ergodique une fois spécifiés les paramètres du modèle probabiliste.

Dans le prochain paragraphe, nous allons nous intéresser au moment d'ordre 1, c'est-à-dire l'espérance mathématique ; dans ce cas, la vérification de l'hypothèse d'ergodicité fera appel à la notion de *portée intégrale* que nous allons définir au paragraphe II.2. L'étude des propriétés ergodiques du variogramme (moment d'ordre 2) est réservée au paragraphe III.

II. Moment d'ordre 1 (espérance mathématique)

II.1. Espérance et moyenne spatiale

On suppose que la fonction aléatoire $Z(\mathbf{x})$ qui représente la variable régionalisée étudiée est stationnaire d'ordre deux. On a alors

$$\begin{aligned} E[Z(\mathbf{x})] &= m \text{ constante dans l'espace} \\ \text{cov}[Z(\mathbf{x}), Z(\mathbf{y})] &= C(\mathbf{x} - \mathbf{y}) \text{ ne dépend que de } \mathbf{x} - \mathbf{y}. \end{aligned}$$

On définit la moyenne spatiale de Z sur un domaine Σ par

$$M^* = \frac{1}{|\Sigma|} \int_{\Sigma} Z(\mathbf{x}) d\mathbf{x}.$$

M^* est une variable aléatoire, dont on peut calculer les deux premiers moments (espérance et variance) :

$$\begin{aligned} E(M^*) &= m \\ \text{var}(M^*) &= \frac{1}{|\Sigma|^2} \int_{\Sigma} \int_{\Sigma} C(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} = \frac{1}{|\Sigma|^2} \iint 1_{\Sigma}(\mathbf{x}) 1_{\Sigma}(\mathbf{y}) C(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \\ &= \frac{1}{|\Sigma|^2} \iint 1_{\Sigma}(\mathbf{y}) 1_{\Sigma}(\mathbf{y} + \mathbf{h}) C(\mathbf{h}) d\mathbf{y} d\mathbf{h} \quad \text{où } \mathbf{h} = \mathbf{x} - \mathbf{y} \\ &= \frac{1}{|\Sigma|^2} \int C(\mathbf{h}) d\mathbf{h} \int 1_{\Sigma}(\mathbf{y}) 1_{\Sigma}(\mathbf{y} + \mathbf{h}) d\mathbf{y} = \frac{1}{|\Sigma|^2} \int C(\mathbf{h}) K(\mathbf{h}) d\mathbf{h} \end{aligned}$$

où $C(\mathbf{h})$ est la covariance centrée de Z ;

1_{Σ} est la fonction indicatrice du domaine Σ ;

$K(\mathbf{h})$ est le covariogramme géométrique de Σ , défini au chapitre 1.

L'ergodicité implique que, lorsque $|\Sigma|$ tend vers l'infini, la variance de M^* tend vers 0, ce qui assure la convergence de M^* vers son espérance m :

$$M^* = \frac{1}{|\Sigma|} \int_{\Sigma} Z(\mathbf{x}) d\mathbf{x} \xrightarrow{|\Sigma| \rightarrow +\infty} m \quad \text{sous hypothèse d'ergodicité.}$$

Asymptotiquement, M^* n'est plus aléatoire, mais presque sûrement égale à m . La convergence est également réalisée au niveau de la réalisation m^* de M^* , qui n'est autre que la moyenne spatiale de la variable régionalisée

$$m^* = \frac{1}{|\Sigma|} \int_{\Sigma} z(\mathbf{x}) d\mathbf{x} \xrightarrow{|\Sigma| \rightarrow +\infty} m.$$

L'espérance mathématique, qui, à première vue, n'est qu'un *paramètre* du modèle — c'est-à-dire un objet qui n'existe pas dans la réalité et dont la valeur pourrait être choisie conventionnellement —, prend alors une signification objective, puisqu'elle s'identifie à une *grandeur régionale* (la moyenne spatiale de la variable régionalisée sur Σ). Toutefois, il n'est pas possible en pratique de rendre Σ aussi grand que l'on veut : on est inévitablement limité par la taille du champ de la régionalisation, et le plus grand domaine que l'on puisse considérer est le champ $\mathcal{D}$ entier : faute de mieux, on prendra donc $\Sigma = \mathcal{D}$. On désire à présent savoir si $\mathcal{D}$ est assez grand pour que la moyenne spatiale m^* soit représentative de l'espérance mathématique m ; la réponse à cette question va faire intervenir le concept de *portée intégrale*.

La figure suivante présente une illustration de la propriété ergodique dans un espace à une dimension. On a simulé, le long d'une droite, une réalisation d'un processus gaussien d'espérance nulle et de covariance sphérique de portée et palier unité. On voit que la moyenne spatiale (ligne continue) fluctue autour de l'espérance mathématique, et s'en rapproche de plus en plus à mesure que la taille du domaine augmente.

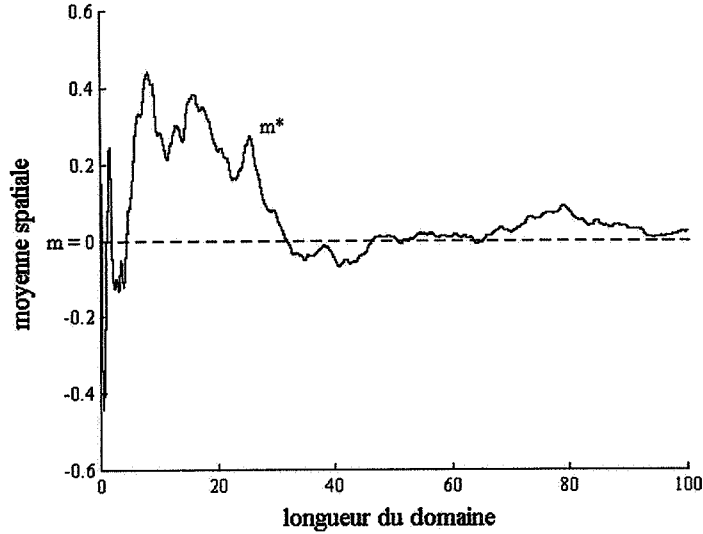


figure B.1 : comparaison entre la moyenne spatiale (courbe continue) et l'espérance mathématique (pointillés), en fonction de la taille du domaine

II.2. La portée intégrale

On établit que, lorsque $\mathcal{D}$ est très grand, la variance de M^* est égale à

$$\text{var}(M^*) \underset{|\mathcal{D}| \rightarrow +\infty}{\sim} \frac{1}{|\mathcal{D}|} \int C(\mathbf{h}) d\mathbf{h}$$

à condition que $0 < \int C(\mathbf{h}) d\mathbf{h} < +\infty$.

En effet, la covariance $C(\mathbf{h})$ est nulle ou quasi-nulle lorsque $|\mathbf{h}|$ est supérieur à la portée pratique a , d'où

$$\int C(\mathbf{h}) K(\mathbf{h}) d\mathbf{h} \approx \int_{|\mathbf{h}| < a} C(\mathbf{h}) K(\mathbf{h}) d\mathbf{h}$$

Or, lorsque la mesure du domaine $\mathcal{D}$ devient infinie, on peut écrire :

$$K(\mathbf{h}) = |\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}| \approx |\mathcal{D}| \text{ lorsque } |\mathbf{h}| < a.$$

Finalement, $\int C(\mathbf{h}) K(\mathbf{h}) d\mathbf{h} \approx |\mathcal{D}| \int C(\mathbf{h}) d\mathbf{h}$, ce qui justifie le résultat annoncé.

On définit la **portée intégrale** comme la valeur de l'intégrale de la fonction de covariance sur tout l'espace, et la portée intégrale *relative* en divisant la portée intégrale par la variance *a priori* :

$$A = \frac{1}{C(\mathbf{0})} \int C(\mathbf{h}) d\mathbf{h}.$$

La portée intégrale relative a une dimension d'espace (longueur dans $\mathbb{R}$, aire dans $\mathbb{R}^2$, volume dans $\mathbb{R}^3$). En la reportant dans la formule de la variance de M^* , il vient :

$$\text{var}(M^*) \underset{|\mathcal{D}| \rightarrow +\infty}{\sim} \frac{A}{|\mathcal{D}|} C(\mathbf{0}) = \frac{C(\mathbf{0})}{N} \quad \text{où } N = |\mathcal{D}|/A.$$

Si le champ $\mathcal{D}$ est suffisamment grand, la variance de M^* est identique à celle que l'on obtient en estimant m par la moyenne de N variables aléatoires indépendantes de variance $C(\mathbf{0})$. Plus N est grand, plus la variance de M^* est petite par rapport à la variance *a priori*, c'est-à-dire moins M^* est susceptible de fluctuer autour de son espérance m , laquelle prend une signification de plus en plus objective. Une fois définie la portée intégrale relative A , la vitesse de convergence ne dépend plus de la covariance, mais seulement du rapport $|\mathcal{D}|/A$. L'ergodicité n'est atteinte que si le champ est grand par rapport à la portée intégrale relative, qui représente une mesure de référence. Dans le cas contraire, rien ne garantit que l'espérance mathématique puisse s'identifier à la moyenne spatiale des valeurs.

Les pages suivantes donnent les valeurs des portées intégrales des modèles de covariance usuels (supposés isotropes et de palier unité). Dans la mesure du possible, on évitera d'utiliser des modèles de portée intégrale infinie, car ils ne permettent pas d'atteindre l'ergodicité, alors que celle-ci est requise lors de l'inférence statistique. Une condition nécessaire pour vérifier l'hypothèse ergodique est que la fonction de covariance soit intégrable.

	expression de la covariance isotrope	portée intégrale dans R	portée intégrale dans R ²	portée intégrale dans R ³
pépite	$C(r) = \begin{cases} 1 & \text{si } r=0 \\ 0 & \text{sinon} \end{cases}$	0	0	0
sphérique	$C(r) = \begin{cases} 1 - \frac{3r}{2a} + \frac{1}{2} \frac{r^3}{a^3} & \text{si } r \leq a \\ 0 & \text{sinon} \end{cases}$	$\frac{3}{4} \frac{a}{a}$	$\frac{\pi}{5} a^2$	$\frac{\pi}{6} a^3$
exponentiel	$C(r) = \exp(-r/a)$	2 a	$2\pi a^2$	$8\pi a^3$
cubique	$C(r) = \begin{cases} 1 - 7 \frac{r^2}{a^2} + \frac{35}{4} \frac{r^3}{a^3} - \frac{7}{2} \frac{r^5}{a^5} + \frac{3}{4} \frac{r^7}{a^7} & \text{si } r \leq a \\ 0 & \text{sinon} \end{cases}$	$\frac{35}{48} \frac{a}{a}$	$\frac{\pi}{6} a^2$	$\frac{7\pi}{60} a^3$
gaussien	$C(r) = \exp(-r^2/a^2)$	$\sqrt{\pi} a$	πa^2	$\pi \sqrt{\pi} a^3$
triangulaire	$C(r) = \begin{cases} 1 - \frac{r}{a} & \text{si } r \leq a \\ 0 & \text{sinon} \end{cases}$	a	—	—
quadratique	$C(r) = \begin{cases} 1 - 2 \frac{r}{a} + \frac{r^2}{a^2} & \text{si } r \leq a \\ 0 & \text{sinon} \end{cases}$	$\frac{2}{3} \frac{a}{a}$	$\frac{\pi}{6} a^2$	$\frac{2\pi}{15} a^3$

	expression de la covariance isotrope	portée intégrale dans R	portée intégrale dans R ²	portée intégrale dans R ³
cosinus amorti ($\alpha = 1$)	$C(r) = \exp(-br) \cos\left(\frac{2\pi r}{a}\right)$	$\frac{a^2 b}{a^2 b^2 + 4\pi^2}$	—	—
sinus cardinal	$C(r) = \frac{\sin(r/a)}{r/a}$	πa	∞	∞
J-Bessel	$C(r) = \left(\frac{r}{2a}\right)^{-\alpha} \Gamma(\alpha+1) J_{\alpha}\left(\frac{r}{a}\right)$	$\begin{cases} 2a \frac{\sqrt{\pi} \Gamma(\alpha+1)}{\Gamma(\alpha+1/2)} & \text{si } \alpha \geq \frac{1}{2} \\ \infty & \text{sinon} \end{cases}$	$\begin{cases} 4\pi\alpha a^2 & \text{si } \alpha > \frac{3}{2} \\ \infty & \text{sinon} \end{cases}$	$\begin{cases} 8\pi a^3 \sqrt{\pi} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha-1/2)} & \text{si } \alpha > \frac{5}{2} \\ \infty & \text{sinon} \end{cases}$
polynômes tronqués	$C(r) = \begin{cases} 1 - \frac{5r}{2a} + \frac{5r^3}{2a^3} - \frac{r^5}{a^5} & \text{si } r \leq a \\ 0 & \text{sinon} \end{cases}$	$\frac{5}{12}a$	$\frac{\pi}{21}a^2$	0
	$C(r) = \begin{cases} 1 - \frac{35r}{12a} + \frac{35r^3}{8a^3} - \frac{7r^5}{2a^5} + \frac{25r^7}{24a^7} & \text{si } r \leq a \\ 0 & \text{sinon} \end{cases}$	$\frac{35}{96}a$	$\frac{\pi}{27}a^2$	0
	$C(r) = \begin{cases} 1 - \frac{7r}{2a} + 7\frac{r^3}{a^3} - 7\frac{r^5}{a^5} + \frac{5r^7}{2a^7} & \text{si } r \leq a \\ 0 & \text{sinon} \end{cases}$	$\frac{7}{24}a$	$\frac{\pi}{45}a^2$	0
stable généralisé	$C(r) = [1 - \beta(r/a)^{\alpha}] \exp[-(r/a)^{\alpha}]$	$\frac{2(\alpha - \beta)}{\alpha} a \Gamma\left(\frac{\alpha+1}{\alpha}\right)$	$\frac{\pi(\alpha - 2\beta)}{\alpha} a^2 \Gamma\left(\frac{\alpha+2}{\alpha}\right)$	$\frac{4\pi(\alpha - 3\beta)}{3\alpha} a^3 \Gamma\left(\frac{\alpha+3}{\alpha}\right)$

	expression de la covariance isotrope	portée intégrale dans R	portée intégrale dans R ²	portée intégrale dans R ³
circulaire	$C(r) = \begin{cases} \frac{2}{\pi} \arccos\left(\frac{r}{a}\right) - \frac{2}{\pi} \frac{r}{a} \sqrt{1 - \frac{r^2}{a^2}} & \text{si } r \leq a \\ 0 & \text{sinon} \end{cases}$	$\frac{8}{3\pi} a$	$\frac{\pi}{4} a^2$	—
pentasphérique	$C(r) = \begin{cases} 1 - \frac{15}{8} \frac{r}{a} + \frac{5}{4} \frac{r^3}{a^3} - \frac{3}{8} \frac{r^5}{a^5} & \text{si } r \leq a \\ 0 & \text{sinon} \end{cases}$	$\frac{5}{8} a$	$\frac{\pi}{7} a^2$	$\frac{5\pi}{48} a^3$
stable	$C(r) = \exp[-(r/a)^\alpha]$	$2a \Gamma\left(\frac{\alpha+1}{\alpha}\right)$	$\pi a^2 \Gamma\left(\frac{\alpha+2}{\alpha}\right)$	$\frac{4}{3} \pi a^3 \Gamma\left(\frac{\alpha+3}{\alpha}\right)$
gamma	$C(r) = \frac{1}{(1+r/a)^\alpha}$	$\begin{cases} \frac{2a}{\alpha-1} & \text{si } \alpha > 1 \\ \infty & \text{sinon} \end{cases}$	$\begin{cases} \frac{2\pi a^2}{(\alpha-1)(\alpha-2)} & \text{si } \alpha > 2 \\ \infty & \text{sinon} \end{cases}$	$\begin{cases} \frac{8\pi a^3}{(\alpha-1)(\alpha-2)(\alpha-3)} & \text{si } \alpha > 3 \\ \infty & \text{sinon} \end{cases}$
Cauchy	$C(r) = \frac{1}{(1+r^2/a^2)^\alpha}$	$\begin{cases} \frac{\sqrt{\pi} a}{\alpha-1} & \text{si } \alpha > 1 \\ \infty & \text{sinon} \end{cases}$	$\begin{cases} \frac{\pi a^2}{(\alpha-1)(\alpha-2)} & \text{si } \alpha > 2 \\ \infty & \text{sinon} \end{cases}$	$\begin{cases} \frac{\pi \sqrt{\pi} a^3}{(\alpha-1)(\alpha-2)(\alpha-3)} & \text{si } \alpha > 3 \\ \infty & \text{sinon} \end{cases}$
K-Bessel	$C(r) = \frac{1}{2^{\alpha-1} \Gamma(\alpha)} (r/a)^\alpha K_\alpha(r/a)$	$2a \sqrt{\pi} \frac{\Gamma(\alpha+1/2)}{\Gamma(\alpha)}$	$4\pi \alpha a^2$	$8\pi a^3 \sqrt{\pi} \frac{\Gamma(\alpha+3/2)}{\Gamma(\alpha)}$
cosinus	$C(r) = \cos\left(\frac{2\pi r}{a}\right)$	∞	—	—

II.3. Calcul empirique de la portée intégrale

La valeur de la portée intégrale dépend du comportement à l'infini de la covariance choisie ; or ce comportement ne s'appuie pas sur les données expérimentales, dont la séparation est nécessairement limitée : la portée intégrale dépendra grandement du choix subjectif du modèle de covariance. Comment vérifier expérimentalement l'hypothèse d'ergodicité, ou comment calculer de manière moins subjective la portée intégrale ? Un test empirique est basé sur la notion de *variance de dispersion* exposée au chapitre 4.

Lorsque l'on considère un domaine V très étendu, la variance de la valeur moyenne sur V est inversement proportionnelle à la mesure de V :

$$\overline{C}(V, V) \approx \frac{A}{|V|} C(0) \quad (V \text{ grand}).$$

En conséquence, la variance de dispersion de V dans un domaine le contenant, par exemple le champ $\mathcal{D}$, peut s'écrire :

$$D^2(V|\mathcal{D}) \approx A C(0) \left(\frac{1}{|V|} - \frac{1}{|\mathcal{D}|} \right).$$

Cette variance de dispersion s'estime à partir des dispersions ponctuelles empiriques dans $\mathcal{D}$ et V à l'aide de la formule de Krige :

$$D^2(V|\mathcal{D}) = D^2(o|\mathcal{D}) - D^2(o|V) \approx \hat{s}^2(o|\mathcal{D}) - \hat{s}^2(o|V).$$

On peut ainsi vérifier si $D^2(V|\mathcal{D})$ est proportionnelle à $1/|V| - 1/|\mathcal{D}|$ lorsque V est très étendu. Si tel est le cas, le facteur de proportionnalité donne une valeur approchée de la portée intégrale, et par conséquent de la variance de la moyenne spatiale, sans recourir à un modèle variographique.

III. Le variogramme

Il convient de distinguer trois concepts : le variogramme *expérimental*, le variogramme *régional* et le variogramme *théorique*. Les propriétés du premier dépendent grandement de l'état de l'information disponible (nombre et disposition des données, forme de la distribution des valeurs, paramètres de calcul choisis...). Le second est le variogramme expérimental "complet", c'est-à-dire celui que l'on obtiendrait si l'on connaissait exhaustivement la variable régionalisée. Enfin, le troisième est un objet théorique, qui s'identifie au variogramme régional sous couvert de stationnarité et ergodicité.

Or, du fait que la régionalisation se répartit dans un champ inévitablement borné, l'ergodicité ne peut être atteinte exactement. Il existe toujours un écart entre le variogramme régional $\gamma_{\text{reg}}(\mathbf{h})$ et le variogramme théorique $\gamma(\mathbf{h})$; cet écart, appelé *fluctuation*, est mesuré par la variance de fluctuation :

$$D_{\text{fluctuation}}^2(\mathbf{h}) = E \{ [\gamma_{\text{reg}}(\mathbf{h}) - \gamma(\mathbf{h})]^2 \}$$

où le variogramme régional est considéré comme une variable aléatoire, dont l'espérance est précisément le variogramme théorique.

Le comportement de la variance de fluctuation a été étudié pour des processus aléatoires admettant une loi spatiale multigaussienne, qui est le seul cas pratique où l'on peut mener les calculs de manière explicite. On a alors :

$$D_{\text{fluctuation}}^2(\mathbf{h}) = \frac{1}{2|\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}|} \iint_{[\mathcal{D} \cap \mathcal{D}_{\mathbf{h}}] \times [\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}]} [\gamma(\mathbf{x} - \mathbf{y} + \mathbf{h}) + \gamma(\mathbf{x} - \mathbf{y} - \mathbf{h}) - 2\gamma(\mathbf{x} - \mathbf{y})]^2 d\mathbf{x} d\mathbf{y}.$$

Le résultat le plus important est que cette variance croît rapidement à mesure que la distance $|\mathbf{h}|$ considérée augmente, ce qui signifie en pratique que l'on peut tolérer un écart de plus en plus important entre le variogramme régional (ou encore le variogramme expérimental, qui en est une approximation) et le modèle de variogramme.

A titre d'illustration, la figure B.2 présente les variogrammes régionaux de dix réalisations d'une même fonction aléatoire; on apprécie que leur écart avec le variogramme théorique (un modèle linéaire + pépite) augmente avec la distance et devient considérable lorsque les distances sont proches du diamètre du champ (100).

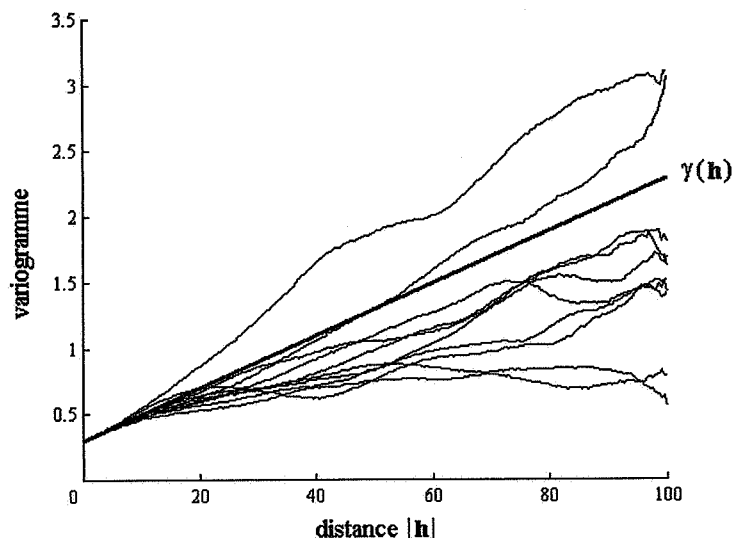


figure B.2 : variogramme théorique (trait épais)
et variogrammes régionaux de dix réalisations

Au contraire, pour les distances infiniment petites, la variance de fluctuation tend vers 0 et le variogramme régional est représentatif du variogramme théorique : ainsi, le comportement aux distances microscopiques, qui reflète les propriétés de régularité spatiale de la régionalisation, s'apparente à une *grandeur régionale*. Cette propriété a été baptisée “**microergodicité**”. Au niveau du variogramme expérimental, il est cependant nécessaire que l'échantillonnage soit suffisamment dense pour pouvoir connaître ce comportement microscopique.

Par exemple, si l'on visualise les variogrammes de la figure B.2 aux petites distances, on observe une adéquation presque parfaite entre les variogrammes régionaux et le variogramme théorique.

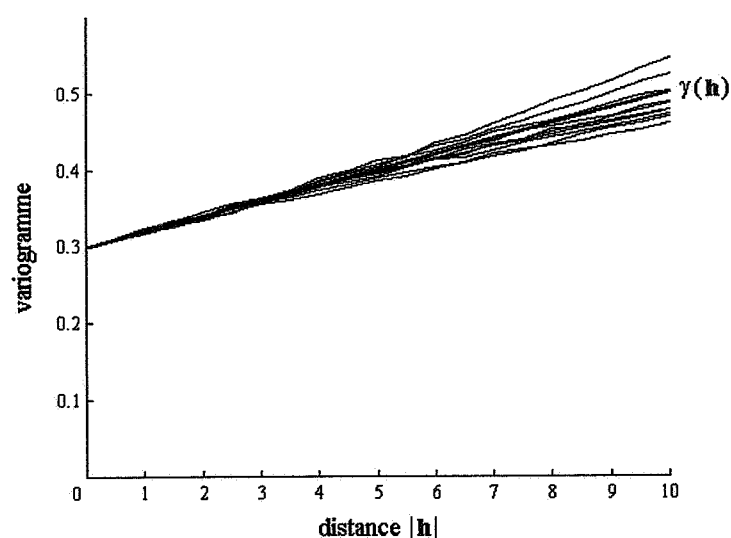


figure B.3 : illustration de la microergodicité

Annexe C : compléments sur le krigeage universel

l'évaluation optimale de la dérive et ses implications théoriques

I. Hypothèses et objectif

On rappelle le modèle de dichotomie utilisé en krigeage universel :

$$Z(\mathbf{x}) = Y(\mathbf{x}) + m(\mathbf{x})$$

où $Z(\mathbf{x})$ est la fonction aléatoire associée à la variable régionalisée étudiée ; elle est définie sur un champ $\mathcal{D} \subset \mathbb{R}^d$

$Y(\mathbf{x})$ est un résidu aléatoire

$m(\mathbf{x})$ est une dérive déterministe, qui peut se développer localement sous la forme

$$m(\mathbf{x}) = \sum_{\ell} a_{\ell} f^{\ell}(\mathbf{x})$$

avec des coefficients a_{ℓ} inconnus.

Les f^{ℓ} , appelées *fonctions de base*, sont le plus souvent des monômes des coordonnées du point $\mathbf{x}$. L'objectif de cette annexe est d'effectuer l'évaluation optimale de la dérive, à partir de la covariance ou du variogramme de Z supposé connu, et d'examiner les conséquences théoriques de cette estimation optimale. Cette analyse va en particulier soulever des difficultés liées au modèle même de dichotomie.

Deux approches successives vont être proposées pour évaluer la dérive : d'abord, une évaluation globale de la dérive, ensuite une évaluation optimale de chacun des coefficients inconnus a_{ℓ} .

II. Evaluation de la dérive dans son ensemble

Comme cela a été fait dans le cas du krigage ordinaire pour l'évaluation de l'espérance, on estime la dérive $m(\mathbf{x}_0)$ au point $\mathbf{x}_0$ par :

$$M^*(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha})$$

où $\{\mathbf{x}_{\alpha}, \alpha = 1 \dots n\}$ sont les points de mesure voisins de $\mathbf{x}_0$. La contrainte de linéarité est ainsi satisfaite.

Pour continuer, on doit distinguer deux cas, selon que le résidu est stationnaire du second ordre ou intrinsèque.

II.1. Résidu stationnaire d'ordre deux

- **autorisation** : pas de contrainte effective (toute combinaison linéaire est autorisée, car Z admet une espérance et une variance finies en tout point) ;
- **non-biais** : l'espérance de l'erreur d'estimation vaut :

$$\begin{aligned} E[M^*(\mathbf{x}_0) - m(\mathbf{x}_0)] &= \sum_{\alpha=1}^n \lambda_{\alpha} \sum_{\ell} a_{\ell} f^{\ell}(\mathbf{x}_{\alpha}) - \sum_{\ell} a_{\ell} f^{\ell}(\mathbf{x}_0) \\ &= \sum_{\ell} a_{\ell} \left[\sum_{\alpha=1}^n \lambda_{\alpha} f^{\ell}(\mathbf{x}_{\alpha}) - f^{\ell}(\mathbf{x}_0) \right] \end{aligned}$$

Pour que cette espérance soit nulle quels que soient les coefficients a_{ℓ} inconnus, on doit poser

$$\sum_{\alpha=1}^n \lambda_{\alpha} f^{\ell}(\mathbf{x}_{\alpha}) = f^{\ell}(\mathbf{x}_0) \quad \forall \ell.$$

- **optimalité** : la minimisation de la variance d'estimation sous les contraintes précédentes conduit au système

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + \sum_{\ell} \mu_{\ell} f^{\ell}(\mathbf{x}_{\alpha}) = 0 & \forall \alpha = 1 \dots n \\ \sum_{\beta=1}^n \lambda_{\beta} f^{\ell}(\mathbf{x}_{\beta}) = f^{\ell}(\mathbf{x}_0) & \forall \ell \end{cases}$$

soit, en écriture matricielle :

$$\begin{pmatrix} C(\mathbf{x}_\alpha - \mathbf{x}_\beta) & f^\ell(\mathbf{x}_\alpha) \\ f^\ell(\mathbf{x}_\beta) & 0 \end{pmatrix} \begin{pmatrix} \lambda_\beta \\ \mu_\ell \end{pmatrix} = \begin{pmatrix} 0 \\ f^\ell(\mathbf{x}_0) \end{pmatrix}$$

La variance d'estimation s'écrit :

$$\sigma_m^2(\mathbf{x}_0) = - \sum_{\ell} \mu_{\ell} f^{\ell}(\mathbf{x}_0).$$

Le premier membre du système est identique à celui du krigeage universel avec résidu stationnaire. Les conditions de régularité (existence et unicité de la solution) sont donc les mêmes.

L'évaluation de la dérive permet d'interpréter de manière intuitive la condition d'indépendance linéaire des fonctions de base sur les données, nécessaire pour que le système de krigeage précédent soit régulier. Ainsi, si toutes les données sont alignées dans $\mathbb{R}^2$ ou coplanaires dans $\mathbb{R}^3$, l'ajustement d'une dérive linéaire est théoriquement indéterminé. Plus généralement, si les sites d'observation sont sur une (hyper)courbe de degré k , on ne pourra pas introduire dans le modèle une dérive de degré supérieur ou égal à k .

II.2. Résidu intrinsèque strict

On suppose à présent que le résidu $Y(\mathbf{x})$ est intrinsèque strict. L'estimateur de la dérive est toujours de la forme

$$M^*(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha}).$$

La condition d'autorisation donne

$$\sum_{\alpha=1}^n \lambda_{\alpha} = 0$$

tandis que la condition de non-biais se développe en :

$$\underbrace{E\left[\sum_{\alpha=1}^n \lambda_{\alpha} Y(\mathbf{x}_{\alpha})\right]}_0 + \sum_{\alpha=1}^n \lambda_{\alpha} \sum_{\ell} a_{\ell} f^{\ell}(\mathbf{x}_{\alpha}) - \sum_{\ell} a_{\ell} f^{\ell}(\mathbf{x}_0) = 0$$

(combinaison linéaire autorisée)

soit

$$\sum_{\ell} a_{\ell} \left[\sum_{\alpha=1}^n \lambda_{\alpha} f^{\ell}(\mathbf{x}_{\alpha}) - f^{\ell}(\mathbf{x}_0) \right] = 0.$$

Sous la condition d'autorisation, étant donné que $f^0(\mathbf{x}) \equiv 1$, la condition de non-biais finale est :

$$\sum_{\ell \geq 1} a_{\ell} \left[\sum_{\alpha=1}^n \lambda_{\alpha} f^{\ell}(\mathbf{x}_{\alpha}) - f^{\ell}(\mathbf{x}_0) \right] - a_0 = 0.$$

Cette égalité ne peut être satisfaite quels que soient les coefficients a_{ℓ} inconnus, car la condition d'autorisation a isolé le coefficient constant a_0 de la dérive. En d'autres termes, **les contraintes d'autorisation et de non-biais sont incompatibles**, et l'estimation de la dérive est impossible. Une solution consisterait à renoncer à étudier la fonction Z elle-même, et à travailler sur les *accroissements* de Z , qui admettent des résidus stationnaires.

III. Evaluation optimale des coefficients de la dérive

Au lieu d'estimer globalement la dérive, on peut chercher à estimer séparément chacun de ses coefficients.

Pour tout ℓ , on estime a_{ℓ} en posant : $A_{\ell}^* = \sum_{\alpha=1}^n \lambda_{\alpha}^{\ell} Z(\mathbf{x}_{\alpha})$.

Lorsque le résidu est **stationnaire d'ordre deux**, le système de krigeage s'écrit sans problème. On aboutit, pour chaque indice ℓ , à un système analogue à celui de l'estimation de la dérive dans son ensemble :

$$(\forall \ell) \begin{cases} \sum_{\beta=1}^n \lambda_{\beta}^{\ell} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + \sum_m \mu_{\ell m} f^m(\mathbf{x}_{\alpha}) = 0 & \forall \alpha = 1 \dots n \\ \sum_{\beta=1}^n \lambda_{\beta}^{\ell} f^m(\mathbf{x}_{\beta}) = \delta_{\ell}^m & \forall m \end{cases}$$

où δ_{ℓ}^m représente l'indice de Kronecker (égal à 1 si $\ell = m$, 0 sinon)

$\mu_{\ell m}$ est un multiplicateur de Lagrange, lié à la covariance entre A_{ℓ}^* et A_m^* par :

$$\text{cov}(A_{\ell}^*, A_m^*) = -\mu_{\ell m}.$$

Cette relation montre au passage que la matrice des multiplicateurs de Lagrange ($\mu_{\ell m}$) est de type négatif (puisque'elle est l'opposée d'une matrice de variance-covariance). On notera aussi que la variance d'estimation de a_{ℓ} est $\sigma_{\ell}^2 = \text{var}(A_{\ell}^* - a_{\ell}) = -\mu_{\ell \ell}$.

Ainsi, lorsque le résidu est stationnaire d'ordre deux, on peut estimer séparément chaque coefficient de la dérive. Les systèmes matriciels à résoudre présentent tous le même premier membre que celui du krigeage universel ; les conditions de régularité (existence et unicité de la solution) sont donc les mêmes.

Remarque

Il est aussi possible d'estimer chaque coefficient par *identification*, à partir de l'estimation optimale globale de la dérive. En effet, les pondérateurs λ_α dans cette estimation sont des combinaisons linéaires des $f'(\mathbf{x}_0)$ qui apparaissent au second membre du système concerné, de sorte que l'on peut écrire :

$$M^*(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_\alpha Z(\mathbf{x}_\alpha) = \sum_{\alpha=1}^n \left[\sum_{\ell} \lambda_\alpha^\ell f'(\mathbf{x}_0) \right] Z(\mathbf{x}_\alpha) = \sum_{\ell} A_\ell^* f'(\mathbf{x}_0) \text{ avec } A_\ell^* = \sum_{\alpha=1}^n \lambda_\alpha^\ell Z(\mathbf{x}_\alpha).$$

On vérifie que A_ℓ^* est bien l'estimateur optimal du coefficient a_ℓ . L'inconvénient de cette méthode d'identification est qu'elle ne fournit pas directement la variance d'estimation des coefficients.

Lorsque le résidu est **intrinsèque strict**, on peut encore estimer les coefficients a_ℓ pour $\ell \geq 1$, tandis que l'estimation du coefficient a_0 est impossible (par incompatibilité entre les conditions d'autorisation et de non-biais). Ceci était prévisible, puisque le coefficient a_0 , qui représente le terme constant de la dérive, est annulé par les combinaisons linéaires autorisées, donc invisible dans le formalisme. En d'autres termes, travailler sur des combinaisons linéaires de poids total nul revient à travailler à une constante additive près, d'où une indétermination sur a_0 .

Plus précisément, pour $\ell \neq 0$, on obtient le système matriciel :

$$(\forall \ell \neq 0) \begin{pmatrix} \gamma(\mathbf{x}_\alpha - \mathbf{x}_\beta) & f^m(\mathbf{x}_\alpha) \\ f^m(\mathbf{x}_\beta) & 0 \end{pmatrix} \begin{pmatrix} \lambda_\beta^\ell \\ -\mu_{\ell m} \end{pmatrix} = \begin{pmatrix} 0 \\ \delta_\ell^m \end{pmatrix}$$

et $\text{cov}(A_\ell^*, A_m^*) = -\mu_{\ell m}$ pour $\ell \neq 0$ et $m \neq 0$.

La matrice du premier membre du système ne dépend pas de l'indice ℓ considéré. Par conséquent, on pourrait l'étendre à l'indice $\ell = 0$, ce qui fournirait un "estimateur" A_0^* . L'interprétation d'un tel estimateur nécessite d'avoir recours à un modèle de dichotomie avec **dérive aléatoire**. Plus précisément, en admettant la dichotomie $Z(\mathbf{x}) = Y(\mathbf{x}) + m(\mathbf{x})$, on définit une dérive aléatoire de la forme :

$$M(\mathbf{x}) = m(\mathbf{x}) + \int Y(\mathbf{x} + \mathbf{h}) p(\mathbf{h}) d\mathbf{h}$$

où p est une fonction de pondération de poids total 1. On impose de plus que p soit suffisamment régulière pour que $M(\mathbf{x})$ puisse se développer sous la forme

$$M(\mathbf{x}) \equiv \sum_{\ell} A_\ell f'(\mathbf{x})$$

où les A_ℓ sont des variables aléatoires.

Il est possible d'estimer tous les coefficients A_ℓ de la dérive aléatoire $M(\mathbf{x})$, y compris le coefficient A_0 . On constate alors que les estimateurs obtenus ne dépendent pas de p et sont identiques à ceux que l'on obtient en estimant les coefficients de la dérive déterministe $m(\mathbf{x})$, y compris en résolvant le système correspondant à l'indice $\ell = 0$.

En conséquence,

$$M^*(\mathbf{x}) = A_0^* + \sum_{\ell \geq 1} A_\ell^* f_\ell(\mathbf{x})$$

constitue un estimateur de la dérive aléatoire $M(\mathbf{x})$, qui diffère de la dérive déterministe $m(\mathbf{x})$ d'une moyenne pondérée présentant certaines propriétés de régularité. La forme de la fonction de pondération n'a pas besoin d'être spécifiée, car le résultat ne dépend pas de cette pondération ; en fait, seule la variance d'estimation de A_0 en dépend. Il est donc justifié de calculer "l'estimateur" A_0^* .

Toutes les difficultés ne sont pas résolues, puisque le résidu estimé

$$Y^*(\mathbf{x}) = Z(\mathbf{x}) - A_0^* - \sum_{\ell \geq 1} A_\ell^* f_\ell(\mathbf{x})$$

est une combinaison linéaire autorisée, donc admet une variance (et une covariance), en général non stationnaire, alors que le résidu vrai $Y(\mathbf{x})$, strictement intrinsèque, n'en admet pas. Ceci ne sera pas sans poser des problèmes lors de l'analyse variographique du résidu.

Ces complications montrent que le choix de la dichotomie en une partie déterministe et une partie aléatoire conduit à de sérieuses difficultés théoriques dans le cadre intrinsèque. Une interprétation possible est qu'un résidu intrinsèque strict peut s'écarter indéfiniment de son espérance (car il a une variance infinie) ; dans ce cas, la distinction dérive – résidu apparaît artificielle et ambiguë.

IV. Le variogramme du résidu estimé optimal

L'estimation optimale de la dérive permet, par différence avec les données concernant la variable brute Z , d'estimer le résidu $Y(\mathbf{x})$ aux points expérimentaux (il s'agit même de l'estimation optimale du résidu). L'objet de ce paragraphe est d'examiner les caractéristiques du variogramme du résidu ainsi estimé. Notons γ_R ce variogramme et γ le variogramme vrai :

$$\gamma(\mathbf{x}, \mathbf{y}) = \gamma(\mathbf{x} - \mathbf{y}) = \frac{1}{2} \text{var}[Z(\mathbf{x}) - Z(\mathbf{y})] = \frac{1}{2} \text{var}[Y(\mathbf{x}) - Y(\mathbf{y})]$$

$$\gamma_R(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \text{var}[Y^*(\mathbf{x}) - Y^*(\mathbf{y})].$$

Le problème d'indétermination du terme constant a_0 ne se pose pas ici, puisque l'on a besoin d'estimer seulement les *accroissements* du résidu, qui sont stationnaires d'ordre deux. Le variogramme γ_R est donc bien défini.

On établit que ce variogramme est de la forme :

$$\forall \mathbf{x}, \mathbf{y} \in \mathcal{Z}, \gamma_R(\mathbf{x}, \mathbf{y}) = \gamma(\mathbf{x}, \mathbf{y}) - \frac{1}{2} \sum_{\ell, s} \mu_{\ell s} [f^\ell(\mathbf{x}) - f^\ell(\mathbf{y})] [f^s(\mathbf{x}) - f^s(\mathbf{y})]$$

où les $\mu_{\ell s}$ sont les multiplicateurs de Lagrange qui apparaissent dans le système d'évaluation optimale de chacun des coefficients de la dérive, et $\mathcal{Z}$ le voisinage glissant utilisé dans l'estimation. En particulier, on a montré que la matrice des $\mu_{\ell m}$ est de type négatif (en tant que l'opposée d'une matrice de variance-covariance). Ceci implique l'inégalité :

$$\gamma_R(\mathbf{x}, \mathbf{y}) \leq \gamma(\mathbf{x}, \mathbf{y}).$$

Le variogramme du résidu estimé optimal est **biaisé** par rapport au variogramme vrai γ . De plus, même si ce dernier est invariant par translation ($\gamma(\mathbf{x}, \mathbf{y}) = \gamma(\mathbf{x} - \mathbf{y})$), il n'en est pas de même de γ_R car la différence entre γ et γ_R

$$\frac{1}{2} \sum_{\ell, s} \mu_{\ell s} [f^\ell(\mathbf{x}) - f^\ell(\mathbf{y})] [f^s(\mathbf{x}) - f^s(\mathbf{y})]$$

dépend en général des points $\mathbf{x}$ et $\mathbf{y}$ (et pas seulement de leur séparation). En pratique, on réalise l'inférence statistique du variogramme *moyen* du résidu estimé, lequel est toujours biaisé par rapport au variogramme vrai :

$$\bar{\gamma}_R(\mathbf{h}) = \frac{1}{|\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}|} \int_{\mathcal{D} \cap \mathcal{D}_{-\mathbf{h}}} \gamma_R(\mathbf{x} + \mathbf{h}, \mathbf{x}) d\mathbf{x}$$

où $\mathcal{D}$ représente le champ de la régionalisation, et $\mathcal{D}_{-\mathbf{h}}$ son translaté par le vecteur $-\mathbf{h}$.

V. L'invariance tensorielle

Dans ce paragraphe, on justifie l'utilisation de fonctions de base polynomiales (ou plus généralement appartenant à la famille des exponentielles-polynômes) plutôt que des fonctions quelconques.

Si l'on effectue une transformation *linéaire bijective* sur les fonctions f^ℓ , l'expression de la dérive devient :

$$m(\mathbf{x}) = \sum_{\ell} a_{\ell} f^{\ell}(\mathbf{x}) = \sum_{\ell} a'_{\ell} \varphi^{\ell}(\mathbf{x})$$

où les φ^{ℓ} sont les nouvelles fonctions de base ; les coefficients a'_{ℓ} se déduisent linéairement des a_{ℓ} .

Par linéarité des équations, on montre que les estimateurs optimaux des coefficients se transforment de la même façon que les vrais coefficients. Par suite, l'estimateur optimal de la dérive est inchangé :

$$m^*(\mathbf{x}) = \sum_{\ell} a_{\ell}^* f^{\ell}(\mathbf{x}) = \sum_{\ell} a_{\ell}^{*'} \varphi^{\ell}(\mathbf{x}) = m^*(\mathbf{x}).$$

Cette propriété présente un intérêt lorsque l'on doit diviser le champ $\mathcal{D}$ en plusieurs domaines $\mathcal{V}_i$ sur chacun desquels le modèle de dichotomie est admissible. Pour étudier le domaine $\mathcal{V}_i$, on place l'origine des coordonnées en un point $\mathbf{x}_i$ de ce domaine, ce qui revient à effectuer le changement de fonctions

$$\varphi'(\mathbf{x}) = f'(\mathbf{x} - \mathbf{x}_i).$$

Si l'espace engendré par les fonctions de base f' est invariant par translation, les nouvelles fonctions φ' s'expriment linéairement au moyen des f' . Dans ce cas, d'après l'invariance tensorielle, le changement d'origine n'affecte pas l'estimateur optimal de la dérive. C'est ce qui motive le choix des fonctions de base f' parmi la famille des exponentielles – polynômes, qui sont les seules fonctions à engendrer un espace vectoriel stable par translation. La dérive estimée est indépendante du choix du repère des coordonnées.

VI. Le théorème d'additivité

Le théorème d'additivité est ici donné sous sa forme la plus générale. Il permet de relier les estimateurs du krigeage simple, krigeage ordinaire et krigeage universel, et de comparer leur variance d'estimation.

VI.1. Résidu stationnaire d'ordre deux

L'estimateur de krigeage universel peut s'écrire :

$$Z_{\text{KU}}^*(\mathbf{x}_0) - M^*(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_{\alpha}^{\text{KS}} [Z(\mathbf{x}_{\alpha}) - M^*(\mathbf{x}_{\alpha})]$$

où les $\lambda_{\alpha}^{\text{KS}}$ représentent les poids de krigeage simple obtenus à l'aide de la covariance de Z , et M^* est l'estimation *optimale* de la dérive.

Cette égalité est identique à l'écriture du krigeage simple du résidu $Y(\mathbf{x}) = Z(\mathbf{x}) - m(\mathbf{x})$, où l'on a simplement remplacé les données concernant $Y(\mathbf{x})$ par leur estimation *optimale* :

$$Y^*(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_{\alpha}^{\text{KS}} Y^*(\mathbf{x}_{\alpha}).$$

En d'autres termes, on peut kriger le résidu à partir de ses valeurs estimées de manière *optimale* aux points de données (krigeage simple à moyenne nulle). Cette propriété n'est plus vraie si on utilise pour l'estimation de la dérive $m(\mathbf{x})$ un estimateur non optimal (moindres carrés, splines...).

Il existe aussi un théorème d'additivité pour les variances :

$$\sigma_{\text{KU}}^2(\mathbf{x}_0) = \sigma_{\text{KS}}^2(\mathbf{x}_0) + \text{var}[Z_{\text{KU}}^*(\mathbf{x}_0) - Z_{\text{KS}}^*(\mathbf{x}_0)]$$

Le terme

$$Z_{\Delta}(\mathbf{x}_0) = Z_{\text{KU}}^*(\mathbf{x}_0) - Z_{\text{KS}}^*(\mathbf{x}_0)$$

est parfois appelé "correction de dérive". Il s'agit d'une combinaison linéaire des données, en tant que différence des estimateurs du krigeage universel et du krigeage simple.

Au passage, on a montré l'inégalité :

$$\sigma_{KU}^2(\mathbf{x}_0) \geq \sigma_{KS}^2(\mathbf{x}_0).$$

VI.2. Résidu intrinsèque

Comme précédemment, on montre que :

$$Z_{KU}^*(\mathbf{x}_0) - \sum_{\ell \geq 1} A_{\ell}^*(\mathbf{x}_0) f'_{\ell}(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_{\alpha}^{KO} [Z(\mathbf{x}_{\alpha}) - \sum_{\ell \geq 1} A_{\ell}^*(\mathbf{x}_0) f'_{\ell}(\mathbf{x}_{\alpha})]$$

où les λ_{α}^{KO} représentent les poids de krigeage ordinaire obtenus à l'aide du variogramme de Z et les A_{ℓ}^* ($\ell \neq 0$) sont les estimateurs *optimaux* des coefficients de la dérive.

Cette égalité s'écrit encore :

$$Z_{KU}^*(\mathbf{x}_0) = Z_{KO}^*(\mathbf{x}_0) + \sum_{\ell \geq 1} A_{\ell}^*(\mathbf{x}_0) [f'_{\ell}(\mathbf{x}_0) - \sum_{\alpha=1}^n \lambda_{\alpha}^{KO} f'_{\ell}(\mathbf{x}_{\alpha})].$$

Dans cette dernière expression, le problème de l'indice $\ell = 0$ ne se pose plus, puisque le terme correspondant est nul. On peut donc étendre la somme à tous les indices et introduire l'estimateur A_0^* .

En ce qui concerne les variances, on a l'égalité :

$$\sigma_{KU}^2(\mathbf{x}_0) = \sigma_{KO}^2(\mathbf{x}_0) + \text{var}[Z_{KU}^*(\mathbf{x}_0) - Z_{KO}^*(\mathbf{x}_0)]$$

d'où l'on déduit en particulier $\sigma_{KU}^2(\mathbf{x}_0) \geq \sigma_{KO}^2(\mathbf{x}_0)$.

Bien qu'il établit un lien entre krigeage universel et krigeage simple (ou ordinaire), le théorème d'additivité requiert la résolution d'un système de krigeage universel pour pouvoir estimer la dérive (et par suite, le résidu) de manière *optimale*. En particulier, il serait erroné de réaliser un krigeage simple ou ordinaire du résidu estimé par une méthode quelconque.

Annexe D : compléments sur les fonctions aléatoires intrinsèques d'ordre k

Cette annexe rassemble plusieurs résultats complémentaires sur la théorie des fonctions aléatoires intrinsèques d'ordre k (cf. chapitre 6). Dans le premier paragraphe, on revient sur la définition initiale de ces fonctions aléatoires, qui sera ensuite conceptualisée de manière plus formelle ; on introduira alors les concepts de “représentation” et de “dérive”. On détaille ensuite des cas particuliers de représentations de fonctions aléatoires intrinsèques. Le troisième paragraphe fournit quelques propriétés des covariances généralisées, tandis que le quatrième paragraphe évoque la notion de fonction aléatoire intrinsèque générale. L'annexe se termine par une présentation du krigeage comme technique d'interpolation et une comparaison avec la théorie classique des splines.

I. Formalisation théorique des fonctions aléatoires intrinsèques d'ordre k

I.1. Notations

L'ensemble des combinaisons linéaires autorisées d'ordre k est noté Λ_k . Pour tout élément λ de Λ_k attribuant les poids λ_i aux sites $\mathbf{x}_i$, et pour tout vecteur $\mathbf{h}$ de l'espace $\mathbb{R}^d$, $\tau_{\mathbf{h}} \lambda$ désigne la combinaison linéaire translatée de λ par le vecteur $\mathbf{h}$, c'est-à-dire celle qui attribue les mêmes poids λ_i aux sites $\mathbf{x}_i + \mathbf{h}$. De manière explicite :

$$\text{si } \lambda = \{\lambda_i, \mathbf{x}_i\}_i, \text{ alors } \tau_{\mathbf{h}} \lambda = \{\lambda_i, \mathbf{x}_i + \mathbf{h}\}_i.$$

Rappelons que, par définition des fonctions de base f^{ℓ} , l'ensemble Λ_k des combinaisons linéaires autorisées d'ordre k est invariant par translation, de sorte que si $\lambda = \{\lambda_i, \mathbf{x}_i\}_i$ est un élément de Λ_k , il en est de même de $\tau_{\mathbf{h}} \lambda = \{\lambda_i, \mathbf{x}_i + \mathbf{h}\}_i$.

I.2. Première définition

On a convenu de définir, au chapitre 6, une fonction aléatoire intrinsèque d'ordre k comme une fonction aléatoire Z dont les combinaisons linéaires autorisées d'ordre k sont stationnaires, soit :

$$\forall \lambda = \{\lambda_i, \mathbf{x}_i\} \in \Lambda_k, \mathcal{Z}(\tau_{\mathbf{h}}\lambda) = \sum_i \lambda_i Z(\mathbf{x}_i + \mathbf{h}) \text{ est stationnaire d'ordre 2 en } \mathbf{h}$$

c'est-à-dire que l'espérance et la variance de $\mathcal{Z}(\tau_{\mathbf{h}}\lambda)$ existent et ne dépendent pas de $\mathbf{h}$.

I.3. Seconde définition

Il est possible de donner une seconde définition, plus formelle, en se référant à la fonction aléatoire $\mathcal{Z}$, qui est définie non plus sur l'espace géographique mais sur l'espace Λ_k des combinaisons linéaires autorisées d'ordre k . On établit que, si Z_1 est une fonction aléatoire intrinsèque d'ordre k (au sens de la définition précédente), alors pour toute fonction aléatoire Z_2 différant de Z_1 d'un polynôme de degré $\leq k$, et pour toute combinaison linéaire $\lambda = \{\lambda_i, \mathbf{x}_i\}_i$ autorisée d'ordre k , on a :

$$\sum_i \lambda_i Z_1(\mathbf{x}_i) \equiv \sum_i \lambda_i Z_2(\mathbf{x}_i)$$

ce qui s'écrit encore :

$$\forall \lambda \in \Lambda_k, \mathcal{Z}_1(\lambda) = \mathcal{Z}_2(\lambda).$$

Cette identité résulte du fait que les combinaisons linéaires autorisées d'ordre k annulent les polynômes de degré inférieur ou égal à k . Ainsi, Z_1 et Z_2 sont indiscernables si l'on se restreint à ne manipuler que de telles combinaisons linéaires. Cette propriété permet d'élargir la définition d'une fonction aléatoire intrinsèque d'ordre k à la **classe d'équivalence** des fonctions aléatoires qui diffèrent d'un polynôme (déterministe ou aléatoire) de degré $\leq k$ et dont l'un des représentants obéit à la première définition.

On montre (*théorème des représentations*) que cette seconde définition équivaut à définir une fonction aléatoire intrinsèque d'ordre k comme une application linéaire $\mathcal{Z}$ qui, à toute combinaison linéaire λ autorisée d'ordre k , associe une variable aléatoire $\mathcal{Z}(\lambda)$ telle que

$$\{\mathcal{Z}(\tau_{\mathbf{h}}\lambda), \mathbf{h} \in \mathbb{R}^d\} \text{ est une fonction aléatoire stationnaire d'ordre deux.}$$

Pour lever toute ambiguïté, nous allons modifier la définition donnée au chapitre 6 et dire qu'une fonction aléatoire $Z(\mathbf{x})$ non stationnaire est une *représentation* d'une fonction aléatoire $\mathcal{Z}$ intrinsèque d'ordre k si

$$\forall \lambda = \{\lambda_i, \mathbf{x}_i\} \in \Lambda_k, \mathcal{Z}(\lambda) = \sum_i \lambda_i Z(\mathbf{x}_i).$$

Ainsi, selon cette seconde définition, la variable régionalisée $z(\mathbf{x})$ est interprétée comme une réalisation d'une représentation $Z(\mathbf{x})$ d'une fonction aléatoire $\mathcal{Z}$ intrinsèque d'ordre k . Le formalisme des combinaisons linéaires autorisées d'ordre k ne permettra pas de distinguer une représentation d'une autre. Par suite, on ne pourra estimer que les grandeurs indépendantes de la représentation considérée ; de telles grandeurs sont appelées *caractéristiques intrinsèques*.

Par exemple, une "dérive" polynomiale de degré k , telle qu'elle a été définie en krigeage universel, n'est *pas* une caractéristique intrinsèque et chercher à l'estimer n'aura pas grand sens ; ceci est d'ailleurs sans importance, car une telle dérive est automatiquement annulée par les combinaisons linéaires autorisées d'ordre k et sera invisible dans le modèle. Le concept de fonction aléatoire intrinsèque d'ordre k va ainsi permettre d'esquiver les problèmes rencontrés en krigeage universel à cause de la dichotomie dérive – résidu.

I.4. Troisième définition

Les combinaisons linéaires autorisées d'ordre 0 (cadre intrinsèque classique) sont des accroissements usuels, lesquels annulent les constantes ; par extension, on peut dire que les combinaisons linéaires autorisées d'ordre k sont des *accroissements généralisés* qui annulent les polynômes de degré inférieur ou égal à k . Or, cette notion s'apparente à celle d'une **dérivation** d'ordre $(k+1)$. Intuitivement, on pourrait définir une fonction aléatoire intrinsèque d'ordre k comme un processus dont les dérivées d'ordre $(k+1)$ sont stationnaires d'ordre deux. Les coefficients du polynôme qui correspond à la différence entre deux représentations s'interprèteraient alors comme des constantes d'intégration.

Cette définition "intuitive" n'est pas dépourvue de fondement théorique. On démontre en effet que les primitives d'ordre $(k+1)$ d'une fonction aléatoire stationnaire d'ordre deux sont des représentations d'une même fonction aléatoire intrinsèque d'ordre k . Réciproquement, dans un espace à une ou plusieurs dimensions, si une fonction aléatoire $Z(\mathbf{x})$ est différentiable en moyenne quadratique, et si toutes ses dérivées¹ partielles d'ordre $(k+1)$ sont stationnaires, alors $Z(\mathbf{x})$ est une représentation d'une fonction aléatoire intrinsèque d'ordre k . Cette propriété caractérise d'ailleurs les fonctions aléatoires intrinsèques d'ordre k *différentiables* ; elle est plus générale qu'il n'y paraît à première vue, car on établit que, pourvu qu'elle soit continue en moyenne quadratique, une fonction aléatoire intrinsèque d'ordre k est la somme d'une fonction aléatoire stationnaire d'ordre deux et d'une fonction aléatoire intrinsèque d'ordre k différentiable (*théorème de décomposition*).

¹ Dans un espace à une dimension, le processus dérivé de $Z(\mathbf{x})$ est défini par

$$Z'(\mathbf{x}) = \lim_{|h| \rightarrow 0} \frac{Z(\mathbf{x} + h) - Z(\mathbf{x})}{h}$$

sous réserve que

$$\lim_{|h| \rightarrow 0} E \left\{ \left[\frac{Z(\mathbf{x} + h) - Z(\mathbf{x})}{h} - Z'(\mathbf{x}) \right]^2 \right\} = 0.$$

Dans l'espace $\mathbb{R}^d$, on définit de même les dérivées partielles d'une fonction aléatoire comme les processus dérivés (lorsqu'ils existent) de $Z(\mathbf{x})$ par rapport aux coordonnées du vecteur $\mathbf{x}$. Cette définition (dérivabilité en moyenne quadratique) n'implique pas nécessairement la dérivabilité des réalisations au sens usuel des fonctions.

I.5. Dérive d'une fonction aléatoire intrinsèque d'ordre k

Soient $\mathcal{Z}$ une fonction aléatoire intrinsèque d'ordre k et Z une de ses représentations. Un théorème établit que l'espérance de $Z(\mathbf{x})$, si elle existe, est un polynôme relativement aux coordonnées de $\mathbf{x}$ de degré au plus égal à $(k+1)$, de sorte que l'on a : $Z(\mathbf{x}) = Y(\mathbf{x}) + m(\mathbf{x})$, où $E[Y(\mathbf{x})] = 0$ et $m(\mathbf{x})$ est polynomiale. Cette décomposition diffère de la dichotomie du krigeage universel, car en général $Y(\mathbf{x})$ n'est ni stationnaire ni intrinsèque.

On pourrait imaginer appeler dérive de $\mathcal{Z}$ le polynôme $m(\mathbf{x})$, mais l'inconvénient d'une telle définition est qu'elle est tributaire de la représentation choisie. Plus précisément, dans $m(\mathbf{x})$, seul le terme de degré $(k+1)$ est une *caractéristique intrinsèque* ; c'est celui-ci que l'on définit comme la **dérive** de $\mathcal{Z}$. Les termes de degrés inférieurs dépendent de la représentation, mais ils sont invisibles dans le formalisme puisque, par définition, toute combinaison linéaire λ autorisée d'ordre k annule les polynômes de degré $\leq k$. Ainsi, l'espérance de $\mathcal{Z}(\lambda)$ est caractérisée uniquement par la dérive.

Par ailleurs, les combinaisons linéaires autorisées d'ordre k contiennent celles d'ordres supérieurs. Par conséquent, une fonction aléatoire intrinsèque d'ordre k , stationnaire sur les combinaisons d'ordre k , est également stationnaire sur celles d'ordres plus élevés : c'est donc, en particulier, une fonction aléatoire intrinsèque d'ordre $k+1$, $k+2$, etc. Ce résultat généralise le fait qu'une fonction aléatoire stationnaire vérifie l'hypothèse intrinsèque (la réciproque est, bien entendu, fausse).

Mieux encore : toute fonction aléatoire intrinsèque d'ordre k est une fonction aléatoire intrinsèque d'ordre $k+1$ sans dérive, ce qui implique que toute combinaison linéaire autorisée d'ordre $k+1$ est d'espérance nulle :

$$\forall \lambda \in \Lambda_{k+1}, E[\mathcal{Z}(\lambda)] = 0.$$

Par suite, lorsque l'on étudiera une fonction aléatoire intrinsèque d'ordre k , on admettra toujours que les combinaisons linéaires autorisées d'ordre k ont une espérance nulle (absence de dérive). Si tel n'est pas le cas, il suffira de passer à l'ordre $k+1$. En particulier, lors de l'étude des fonctions aléatoires intrinsèques (d'ordre 0), nous avons toujours supposé l'absence de dérive ; cette restriction est finalement peu contraignante puisqu'une fonction aléatoire intrinsèque (d'ordre 0) avec dérive n'est autre qu'une fonction aléatoire intrinsèque d'ordre 1 sans dérive.

II. Cas particuliers de représentations

II.1. Représentations internes

Parmi les représentations d'une même fonction aléatoire intrinsèque, certaines s'écrivent comme des combinaisons linéaires autorisées d'ordre k . Elles admettent donc, en tout point du champ, une espérance et une variance finies. De même, elles possèdent une fonction de covariance $C(\mathbf{x}, \mathbf{y})$, en générale non stationnaire, c'est-à-dire dépendant des positions des sites $\mathbf{x}$ et $\mathbf{y}$ et pas seulement de leur séparation. De telles représentations sont dites *internes*.

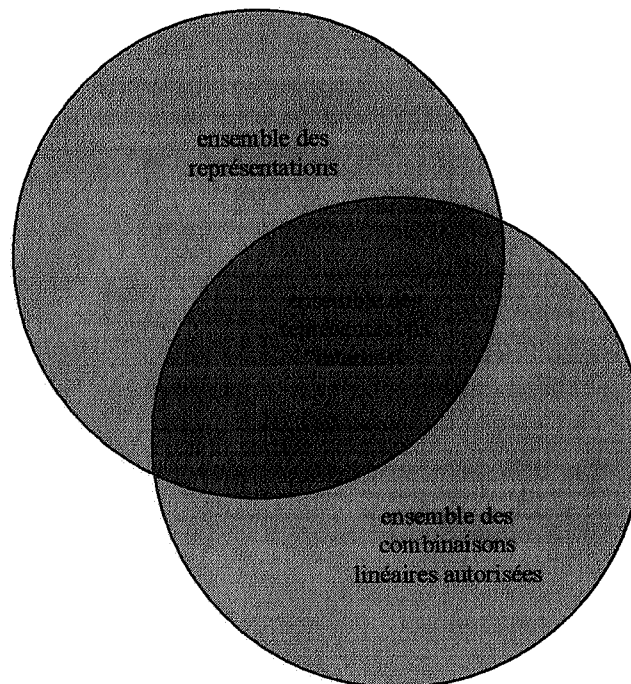


figure D.1 : représentations internes d'une même fonction aléatoire intrinsèque d'ordre k

Par exemple, si $\mathcal{Z}$ est une fonction aléatoire intrinsèque usuelle (d'ordre 0) et $Z(\mathbf{x})$ est une de ses représentations, toutes les autres représentations sont définies par $Z(\mathbf{x}) + A_0$, où A_0 est une constante déterministe ou aléatoire, éventuellement corrélée avec $Z(\mathbf{x})$. Ainsi, si l'on choisit $A_0 = -Z(\mathbf{x}_0)$ avec $\mathbf{x}_0$ site arbitraire fixé, la représentation

$$Y(\mathbf{x}) = Z(\mathbf{x}) - Z(\mathbf{x}_0)$$

est un accroissement, donc une combinaison linéaire autorisée d'ordre 0.

Remarque

Le recours à cette représentation interne a permis d'établir la règle de calcul des variances de combinaisons linéaires autorisées d'ordre 0, en remplaçant la covariance (inexistante) par l'opposé du variogramme, cf. chapitre 2.

Plus généralement, soient $\mathcal{Z}$ une fonction aléatoire intrinsèque d'ordre k et $Z(\mathbf{x})$ l'une de ses représentations. On définit une fonction aléatoire

$$Y(\mathbf{x}) = Z(\mathbf{x}) - \sum_{\alpha} \lambda_{\alpha}(\mathbf{x}) Z(\mathbf{x}_{\alpha})$$

où $\{\mathbf{x}_{\alpha}\}$ sont des sites fixés une fois pour toutes, en nombre suffisant

$\{\lambda_{\alpha}\}$ sont des scalaires, dépendant du point $\mathbf{x}$, vérifiant le système d'équations linéaires :

$$\forall \ell, \sum_{\alpha} \lambda_{\alpha} f^{\ell}(\mathbf{x}_{\alpha}) = f^{\ell}(\mathbf{x}).$$

Ces équations coïncident avec les conditions d'autorisation et de non-biais du krigeage universel avec une dérive de degré k , où l'on estime la valeur en $\mathbf{x}$ à partir des valeurs aux points $\mathbf{x}_{\alpha}$. $Y(\mathbf{x})$ représente alors l'erreur de krigeage universel, qui est une combinaison linéaire autorisée d'ordre k . D'autre part, les coefficients λ_{α} sont des combinaisons linéaires des $f^{\ell}(\mathbf{x})$, qui sont des monômes de degré inférieur ou égal à k ; par suite, les valeurs $Z(\mathbf{x}_{\alpha})$ étant fixées, $Y(\mathbf{x})$ ne diffère de $Z(\mathbf{x})$ que d'un polynôme aléatoire de degré $\leq k$ et constitue donc une représentation de la fonction aléatoire intrinsèque $\mathcal{Z}$.

On peut construire des représentations internes en considérant, non plus des combinaisons linéaires discrètes, mais des mesures à support continu :

$$Y(\mathbf{x}) = Z(\mathbf{x}) - \int \lambda_{\mathbf{x}}(d\mathbf{u}) Z(\mathbf{u}) \quad \text{avec} \quad \forall \ell, \int \lambda_{\mathbf{x}}(d\mathbf{u}) f^{\ell}(\mathbf{u}) = f^{\ell}(\mathbf{x}).$$

La covariance d'une représentation interne est liée à la covariance généralisée par la formule :

$$\forall \mathbf{x}, \mathbf{y} \in \mathcal{D}, C(\mathbf{x}, \mathbf{y}) = K(\mathbf{x} - \mathbf{y}) + \sum_{\ell} a_{\ell}(\mathbf{y}) f^{\ell}(\mathbf{x}) + \sum_{\ell} a_{\ell}(\mathbf{x}) f^{\ell}(\mathbf{y})$$

où les a_{ℓ} sont des fonctions continues telles que l'expression ci-dessus soit de type positif.

II.2. Représentations localement stationnaires

Parmi les représentations internes d'une fonction aléatoire intrinsèque, il arrive que certaines admettent une covariance qui, dans une région de l'espace, ne dépend que de la séparation entre les points d'appui. On définit une représentation *localement stationnaire* sur un domaine D borné comme une représentation qui admet sur D une covariance stationnaire, et coïncide donc, sur D , avec une fonction aléatoire stationnaire. Cette définition rejoint celle de la quasi-stationnarité introduite au chapitre 2 ; le domaine D sera soit le champ entier de la régionalisation, soit le voisinage glissant employé pour l'estimation locale.

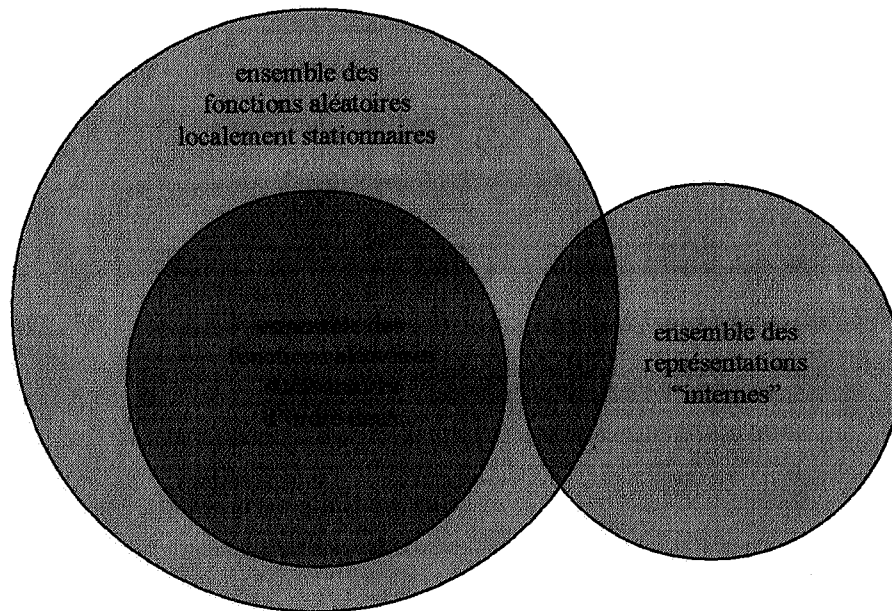


figure D.2 : représentations internes et localement stationnaires d'une même fonction aléatoire intrinsèque d'ordre k

Par exemple, soient une fonction aléatoire intrinsèque d'ordre 0 unidimensionnelle, de variogramme linéaire $\gamma(h) = \omega |h|$, et $Z(x)$ une de ses représentations. La fonction aléatoire définie par

$$Y(x) = Z(x) - \frac{1}{2}[Z(-a) + Z(a)]$$

est une représentation localement stationnaire sur l'intervalle $[-a, a]$, bien qu'elle ne soit pas stationnaire sur $\mathbb{R}$. Elle admet en effet, sur $[-a, a]$, la covariance

$$\forall x, y \in [-a, a], C(x, y) = \omega[a - |x - y|].$$

Une explication "intuitive" est qu'un variogramme linéaire peut être vu comme un schéma triangulaire de portée et palier arbitrairement grands ; dès lors, il n'est pas étonnant de trouver des représentations qui coïncident avec un processus stationnaire dont la covariance est ce schéma triangulaire. Ce résultat peut par ailleurs être élargi à l'espace à plusieurs dimensions.

Pour toute fonction aléatoire intrinsèque d'ordre 0 de variogramme $\gamma(h)$, la représentation interne

$$Y(\mathbf{x}) = Z(\mathbf{x}) - \int \lambda_{\mathbf{x}}(d\mathbf{u}) Z(\mathbf{u}) .$$

est localement stationnaire sur D si la mesure $\lambda_{\mathbf{x}}$, à support dans D , est solution du système :

$$\begin{cases} \int \lambda_{\mathbf{x}}(d\mathbf{u}) \gamma(\mathbf{u} - \mathbf{v}) = \mu \quad \forall \mathbf{v} \in D \\ \int \lambda_{\mathbf{x}}(d\mathbf{u}) = 1 \end{cases}$$

où μ est un scalaire. En effet, on peut développer la covariance entre deux valeurs $Y(\mathbf{x}_1)$ et $Y(\mathbf{x}_2)$ à l'aide du variogramme :

$$\begin{aligned} C(\mathbf{x}_1, \mathbf{x}_2) &= -\gamma(\mathbf{x}_1 - \mathbf{x}_2) + \int \lambda_{\mathbf{x}_1}(d\mathbf{u}) \gamma(\mathbf{u} - \mathbf{x}_2) + \int \lambda_{\mathbf{x}_2}(d\mathbf{v}) \gamma(\mathbf{x}_1 - \mathbf{v}) - \iint \lambda_{\mathbf{x}_1}(d\mathbf{u}) \lambda_{\mathbf{x}_2}(d\mathbf{v}) \gamma(\mathbf{u} - \mathbf{v}) \\ &= -\gamma(\mathbf{x}_1 - \mathbf{x}_2) + 2\mu - \mu \int \lambda_{\mathbf{x}_2}(d\mathbf{v}) = -\gamma(\mathbf{x}_1 - \mathbf{x}_2) + \mu \\ &= -\gamma(\mathbf{x}_1 - \mathbf{x}_2) + \mu \text{ ne dépend que de } \mathbf{x}_1 - \mathbf{x}_2. \end{aligned}$$

Si elle existe, une telle représentation interne est appelée *représentation privilégiée* dans D .

Plus généralement, toute fonction aléatoire intrinsèque d'ordre k dont une covariance généralisée est formée de covariances usuelles, covariances puissances ou splines, admet des représentations localement stationnaires. Il n'est pas certain toutefois qu'il en soit ainsi de n'importe quelle fonction aléatoire intrinsèque.

D'après le paragraphe précédent, la covariance stationnaire est de la forme :

$$\forall \mathbf{x}, \mathbf{y} \in D, C(\mathbf{x} - \mathbf{y}) = K(\mathbf{x} - \mathbf{y}) + \sum_{\ell} a_{\ell}(\mathbf{y}) f^{\ell}(\mathbf{x}) + \sum_{\ell} a_{\ell}(\mathbf{x}) f^{\ell}(\mathbf{y})$$

avec des fonctions a_{ℓ} continues. Or, sur le domaine D , $C(\mathbf{x} - \mathbf{y})$ est une covariance généralisée admissible de la fonction aléatoire intrinsèque, ce qui montre qu'elle ne diffère de $K(\mathbf{x} - \mathbf{y})$ que par un polynôme pair de degré inférieur ou égal à $2k$.

Le tableau 6.0 donne la forme des covariances stationnaires associées aux modèles usuels de covariances généralisées (du type polynomial + spline) ; A est un scalaire positif et D est le diamètre du domaine de validité de ces formules. On notera que l'expression des covariances stationnaires dépend explicitement du paramètre D .

$k \leq 2$	$K(\mathbf{h}) = -b_0 \mathbf{h} + b_s \mathbf{h} ^2 \ln \mathbf{h} + b_1 \mathbf{h} ^3 - b_2 \mathbf{h} ^5$
dans R	$C(\mathbf{h}) = A + b_0 (D/2 - \mathbf{h}) + b_s (D^2/8 - [3/2 + \ln(D/4)] \mathbf{h} ^2 + \mathbf{h} ^2 \ln \mathbf{h}) + b_1 (\mathbf{h} ^3 - 3D \mathbf{h} ^2/2 + D^3/4) - b_2 (\mathbf{h} ^5 - 5D \mathbf{h} ^4/2 + 5D^3 \mathbf{h} ^2/2 - D^5/2)$
dans R^2	$C(\mathbf{h}) = A + b_0 (\pi D/4 - \mathbf{h}) + b_s (D^2/4 - [1 + \ln(D/2)] \mathbf{h} ^2 + \mathbf{h} ^2 \ln \mathbf{h}) + b_1 (\mathbf{h} ^3 - 9\pi \mathbf{h} ^2/16 + 3\pi D^3/16) - b_2 (\mathbf{h} ^5 - 225\pi D \mathbf{h} ^4/256 + 75\pi D^3 \mathbf{h} ^2/64 - 15\pi D^5/32)$
dans R^3	$C(\mathbf{h}) = A + b_0 (D - \mathbf{h}) + b_s (3D^2/8 - [11/6 + \ln(D/4)] \mathbf{h} ^2 + \mathbf{h} ^2 \ln \mathbf{h}) + b_1 (\mathbf{h} ^3 - 2D \mathbf{h} ^2 + D^3) - b_2 (\mathbf{h} ^5 - 3D \mathbf{h} ^4 + 5D^3 \mathbf{h} ^2 - 3D^5)$

tableau D.1 : covariances localement stationnaires associées aux covariances généralisées usuelles

La notion de représentation localement stationnaire présente un intérêt en théorie de la simulation géostatistique, puisqu'elle permet de simuler un processus non stationnaire à partir de méthodes applicables au cadre stationnaire. Egalement, elle permet de remplacer, dans le système de krigeage, la covariance généralisée (fonction de type positif conditionnel) par une covariance usuelle (fonction de type positif stricte), ce qui est avantageux pour la résolution numérique de ce système.

III. Les covariances généralisées

III.1. Caractérisation des covariances généralisées

Les covariances généralisées d'une fonction aléatoire intrinsèque d'ordre k sont définies à un polynôme pair de degré inférieur à $2k$ près. Elles satisfont les conditions suivantes :

- 1) parité : $K(\mathbf{h}) = K(-\mathbf{h})$
- 2) pour toute combinaison linéaire autorisée d'ordre k , $\lambda = \{\lambda_i, \mathbf{x}_i\}_i$, on a :

$$\text{var}[\mathcal{Z}(\lambda)] = \text{var}\left[\sum_i \lambda_i Z(\mathbf{x}_i)\right] = \sum_i \sum_j \lambda_i \lambda_j K(\mathbf{x}_i - \mathbf{x}_j).$$

Cette relation permet de calculer non seulement les variances, mais aussi les covariances de combinaisons linéaires ; plus précisément, pour deux combinaisons linéaires autorisées d'ordre k, $\lambda = \{\lambda_i, \mathbf{x}_i\}_i$ et $\mu = \{\mu_j, \mathbf{y}_j\}_j$, on a :

$$\text{cov}[\mathcal{Z}(\lambda), \mathcal{Z}(\mu)] = \text{cov}\left[\sum_i \lambda_i Z(\mathbf{x}_i), \sum_j \mu_j Z(\mathbf{y}_j)\right] = \sum_i \sum_j \lambda_i \mu_j K(\mathbf{x}_i - \mathbf{y}_j).$$

Pour une fonction aléatoire intrinsèque d'ordre k sans dérive, les égalités précédentes s'écrivent :

$$\begin{cases} \forall \lambda \in \Lambda_k, E[\mathcal{Z}(\lambda)^2] = E\left\{\left[\sum_i \lambda_i Z(\mathbf{x}_i)\right]^2\right\} = \sum_i \sum_j \lambda_i \lambda_j K(\mathbf{x}_i - \mathbf{x}_j) \\ \forall \lambda, \mu \in \Lambda_k, E[\mathcal{Z}(\lambda) \mathcal{Z}(\mu)] = E\left\{\left[\sum_i \lambda_i Z(\mathbf{x}_i)\right]\left[\sum_j \mu_j Z(\mathbf{y}_j)\right]\right\} = \sum_i \sum_j \lambda_i \mu_j K(\mathbf{x}_i - \mathbf{y}_j) \end{cases}$$

Une propriété caractéristique des covariances généralisées est que ce sont des fonctions paires de type positif conditionnel à l'ordre k, c'est-à-dire que pour toute combinaison linéaire autorisée d'ordre k, $\lambda = \{\lambda_i, \mathbf{x}_i\}_i$, on a :

$$\sum_i \sum_j \lambda_i \lambda_j K(\mathbf{x}_i - \mathbf{x}_j) \geq 0.$$

III.2. Propriétés spectrales

Une condition nécessaire et suffisante pour qu'une fonction $K(\mathbf{h})$ continue soit de type positif conditionnel à l'ordre k (donc une covariance généralisée) est que $(-1)^{k+1} \Delta^{k+1}(K(\mathbf{h}))$, où Δ^{k+1} désigne le laplacien itéré (k+1) fois au sens des distributions, ait pour transformée de Fourier une mesure positive $\chi(\mathbf{du})$ telle que :

$$1) \quad \chi(\mathbf{du}) \text{ n'a pas d'atome à l'origine, i.e. } \chi(\{\mathbf{0}\}) = 0$$

$$2) \quad \int_{\mathbb{R}^d} \frac{\chi(\mathbf{du})}{(1 + 4\pi^2 |\mathbf{u}|^2)^{k+1}} < +\infty$$

$\chi(\mathbf{du})$ est appelée la *mesure spectrale* associée à $K(\mathbf{h})$ (à l'ordre k).

La covariance généralisée admet une représentation intégrale de la forme :

$$K(\mathbf{h}) = \int_{\mathbb{R}^d} \frac{\cos(2\pi \langle \mathbf{u}, \mathbf{h} \rangle) - P_k(2\pi \langle \mathbf{u}, \mathbf{h} \rangle)}{(4\pi^2 |\mathbf{u}|^2)^{k+1}} \chi(\mathbf{du}) + Q_k(\mathbf{h})$$

où Q_k est un polynôme pair de degré $2k+2$ ($2k$ en l'absence de dérive) et P_k est le développement limité du cosinus à l'ordre $2k$ (ce terme est introduit pour assurer la convergence de l'intégrale au voisinage de $\mathbf{u} = \mathbf{0}$) :

$$\forall t \in \mathbb{R}, P_k(t) = 1 - t^2/2! + t^4/4! + \dots + (-1)^k t^{2k}/(2k!).$$

Cette propriété est la généralisation du théorème de Bochner selon lequel une fonction continue est de type positif (donc une covariance) si et seulement si elle est la transformée de Fourier d'une mesure positive sommable. Elle permet d'établir les résultats suivants :

- on peut trouver deux réels a et b positifs tels que, pour tout $\mathbf{h}$, on ait : $|K(\mathbf{h})| \leq a + b |\mathbf{h}|^{2k+2}$
- en l'absence de dérive, $K(\mathbf{h}) / |\mathbf{h}|^{2k+2}$ tend vers 0 lorsque $|\mathbf{h}|$ tend vers l'infini.

III.3. Changement d'ordre

Toute covariance généralisée d'ordre k est admissible aux ordres supérieurs. En particulier, les covariances usuelles sont également des variogrammes (au signe près) et des covariances généralisées.

La réciproque n'est vraie que sous certaines conditions. Ainsi, une fonction aléatoire intrinsèque usuelle (d'ordre 0) est définie à une constante additive près, qui peut être déterministe ou aléatoire ; elle admet des représentations stationnaires si son variogramme est borné. De la même manière, on établit qu'une fonction aléatoire intrinsèque d'ordre k admet des représentations qui sont celles d'une fonction aléatoire intrinsèque d'ordre $(k-1)$ lorsque sa covariance généralisée est telle que $K(\mathbf{h}) / |\mathbf{h}|^{2k}$ est borné ; dans ce cas, on obtient une covariance généralisée d'ordre $k-1$ en annulant les termes de degré $2k$ dans l'expression de $K(\mathbf{h})$.

IV. Fonctions aléatoires intrinsèques générales

En toute généralité, les fonctions de base sont issues de la famille des exponentielles — polynômes, c'est-à-dire qu'elles sont de la forme :

$$f^\ell(\mathbf{x}) = \sum_q P_q(\mathbf{x}) \exp \langle \mathbf{u}_q | \mathbf{x} \rangle$$

où P_q sont des polynômes, $\mathbf{u}_q$ sont des vecteurs (éventuellement complexes) et $\langle | \rangle$ désigne le produit scalaire usuel. Cette condition vient du fait que les combinaisons linéaires autorisées doivent former un ensemble invariant par translation, de façon à ce que l'hypothèse de stationnarité qui porte sur ces combinaisons linéaires ait un sens. Explicitement, on doit avoir

$$\left(\sum_i \lambda_i f^\ell(\mathbf{x}_i) = 0 \quad \forall \ell \right) \Rightarrow \left(\sum_i \lambda_i f^\ell(\mathbf{x}_i + \mathbf{h}) = 0 \quad \forall \ell \text{ et } \forall \mathbf{h} \right)$$

ce qui implique que les translatées $f^\ell(\mathbf{x}_i + \mathbf{h})$ s'expriment linéairement en fonction des $f^\ell(\mathbf{x}_i)$:

$$\forall \ell, \forall \mathbf{h}, f^\ell(\mathbf{x} + \mathbf{h}) = \sum_m \alpha_{\ell m}(\mathbf{h}) f^m(\mathbf{x}).$$

Les fonctions aléatoires intrinsèques d'ordre k correspondent au cas où les $f^\ell(\mathbf{x})$ sont des monômes des coordonnées. Lorsque les fonctions de base sont plus générales, les fonctions aléatoires correspondantes sont appelées *fonctions aléatoires intrinsèques généralisées*. Par exemple, on peut choisir des fonctions de base trigonométriques, dans le but de décrire des phénomènes à tendance périodique (en particulier, les phénomènes temporels). Le krigeage intrinsèque correspondant est appelé *krigeage trigonométrique*.

V. Le krigeage comme méthode d'interpolation

La propriété d'interpolation exacte des données apparente le krigeage ponctuel à une technique de cartographie. Cependant, tel que nous avons présenté le krigeage, chaque point à estimer nécessite la résolution d'un système linéaire d'équations. Or, beaucoup de méthodes de cartographie (moindres carrés, splines...) ne se construisent pas en pondérant les données voisines du point à estimer, mais en ajustant une fonction de forme prédéfinie à l'ensemble des données.

L'objectif de ce paragraphe est d'examiner dans quelle mesure le krigeage peut être considéré comme une fonction d'interpolation et de le comparer à d'autres techniques, en particulier celle des splines.

V.1. Présentation duale du krigeage

Considérons le système de krigeage intrinsèque d'ordre k d'un point $\mathbf{x}$:

$$\begin{pmatrix} K(\mathbf{x}_\alpha - \mathbf{x}_\beta) & f^\ell(\mathbf{x}_\alpha) \\ f^\ell(\mathbf{x}_\beta) & 0 \end{pmatrix} \begin{pmatrix} \lambda_\beta \\ \mu_\ell \end{pmatrix} = \begin{pmatrix} K(\mathbf{x} - \mathbf{x}_\alpha) \\ f^\ell(\mathbf{x}) \end{pmatrix}$$

Les solutions λ_β et μ_ℓ de ce système sont des combinaisons linéaires des termes $K(\mathbf{x} - \mathbf{x}_\alpha)$ et $f^\ell(\mathbf{x})$ qui apparaissent au second membre.

Par suite, l'estimateur

$$z^*(\mathbf{x}) = \sum_{\beta=1}^n \lambda_\beta z(\mathbf{x}_\beta)$$

est aussi une combinaison linéaire des $K(\mathbf{x} - \mathbf{x}_\alpha)$ et $f^\ell(\mathbf{x})$, et on peut écrire :

$$z^*(\mathbf{x}) = \sum_{\alpha=1}^n b_\alpha K(\mathbf{x} - \mathbf{x}_\alpha) + \sum_{\ell} a_\ell^* f^\ell(\mathbf{x}).$$

Cette expression, écrite à l'aide de la variable régionalisée et non de la fonction aléatoire, constitue la **formulation duale du krigeage**. Elle peut servir de définition au krigeage, qui apparaît alors dépourvu de connotation probabiliste ; en particulier, la notion de variance d'estimation est abandonnée.

Il reste à spécifier les coefficients b_α et a_ℓ^* . En remplaçant, dans la formulation duale, les fonctions $K(\mathbf{x} - \mathbf{x}_\alpha)$ et $f^\ell(\mathbf{x})$ par leurs expressions issues du système direct, et en procédant par identification, on obtient le système dont b_α et a_ℓ^* sont solutions :

$$\begin{cases} \sum_{\beta=1}^n b_\beta K(\mathbf{x}_\alpha - \mathbf{x}_\beta) + \sum_{\ell} a_\ell^* f^\ell(\mathbf{x}_\alpha) = z(\mathbf{x}_\alpha) & \forall \alpha = 1 \dots n \\ \sum_{\beta=1}^n b_\beta f^\ell(\mathbf{x}_\beta) = 0 & \forall \ell \end{cases}$$

soit :

$$\begin{pmatrix} K(\mathbf{x}_\alpha - \mathbf{x}_\beta) & f^\ell(\mathbf{x}_\alpha) \\ f^\ell(\mathbf{x}_\beta) & 0 \end{pmatrix} \begin{pmatrix} b_\beta \\ a_\ell^* \end{pmatrix} = \begin{pmatrix} z(\mathbf{x}_\alpha) \\ 0 \end{pmatrix}$$

Les premières équations de ce système correspondent à une contrainte d'**interpolation exacte des données** :

$$\forall \alpha = 1 \dots n, z^*(\mathbf{x}_\alpha) = \sum_{\beta=1}^n b_\beta K(\mathbf{x}_\alpha - \mathbf{x}_\beta) + \sum_{\ell} a_\ell^* f^\ell(\mathbf{x}_\alpha) = z(\mathbf{x}_\alpha).$$

Le système dual présente le même membre de gauche que le système direct, donc les mêmes propriétés de régularité. Ainsi, ce système étant supposé régulier, les coefficients b_α et a_ℓ^* sont entièrement caractérisés par les équations ci-dessus.

On notera par ailleurs que ces coefficients ne dépendent pas du point $\mathbf{x}$ à estimer, qui n'intervient pas dans le système. Le krigeage dual apparaît ainsi sous la forme d'une fonction interpolatrice, combinaison linéaire de la fonction de covariance généralisée K et des fonctions de base f^ℓ . Ces dernières étant des fonctions indéfiniment dérivables, z^* présentera les mêmes propriétés de continuité que la fonction K .

Une conséquence notable est la propriété de **lissage** du krigeage, car les modèles usuels de covariance généralisée n'ont de singularités analytiques qu'à l'origine et au voisinage de la portée (si celle-ci existe) et sont partout ailleurs réguliers. Par exemple, si la covariance est un effet de pépite pur, un schéma exponentiel ou un modèle puissance, les seuls points singuliers de z^* sont les sites de données, alors que la régionalisation est discontinue dans le premier cas, non dérivable dans le second cas, et continue jusqu'à un certain ordre dans le troisième cas.

Un autre point à souligner est que, si K est une covariance usuelle, les coefficients a_ℓ^* coïncident avec les estimateurs optimaux des coefficients de la dérive dans le modèle de krigeage universel (cf. annexe C).

L'intérêt de la formulation duale du krigeage est qu'il suffit de résoudre un seul système pour obtenir les coefficients b_α et a_i^* , et en déduire l'estimation de tous les points du champ, alors que l'approche directe nécessite de résoudre autant de systèmes qu'il y a de points à estimer. En contrepartie, on renonce au calcul des variances d'estimation. On dispose de plus d'une équation implicite des courbes de niveau, en posant $z^*(x) = z_0$. Quelques complications apparaissent lorsqu'on travaille en voisinage glissant : la fonction interpolatrice change suivant le voisinage utilisé et peut créer des discontinuités sur la carte finale (ces problèmes ne se posent pas en voisinage unique, mais l'inférence de la fonction de covariance généralisée peut ne pas être acceptable à l'échelle du champ).

V.2. Equivalence entre splines et krigeage

La formulation duale du krigeage ne se réfère pas au concept de fonction aléatoire, ce qui la situe parmi les techniques d'interpolation déterministes. Dans ce paragraphe, on cherche à comparer le krigeage au formalisme des splines. Sous leur forme la plus générale, les splines d'interpolation sont des fonctions minimisant l'intégrale d'espace du carré d'un opérateur linéaire, tout en respectant les valeurs aux points de données.

On démontre que toute spline d'interpolation s'identifie à un krigeage intrinsèque, et réciproquement qu'à tout krigeage correspond une spline. Mais il n'est pas facile d'explicitier l'opérateur linéaire de la spline associée à une covariance généralisée donnée. Les splines utilisées habituellement (splines cubiques, laplaciennes...) ne correspondent qu'à une classe restreinte de covariances généralisées.

Exemples

- dans R , la covariance généralisée puissance impaire $K_p(r) = (-1)^{p+1} |r|^{2p+1}$ correspond à la minimisation du carré de l'opérateur différentiel d^{p+1}/dr^{p+1} .
- les splines laplaciennes de lissage, présentées lors de l'estimation de la dérive en krigeage universel (cf. chapitre 6), coïncident, sauf aux points de données, avec un krigeage à l'aide de la covariance généralisée $K(h) = |h|^2 \ln(|h|) + \rho \delta$, où δ représente un effet de pépité de palier 1, d'où d'ailleurs le nom de "spline" donné à la covariance généralisée $|h|^2 \ln(|h|)$.

L'équivalence entre splines et krigeage conduit à la conclusion suivante : utiliser une certaine classe de fonctions splines comme interpolateur revient à effectuer un krigeage avec des fonctions de base et une covariance généralisée arbitrairement choisies. Le propre des méthodes géostatistiques est de faire précéder l'interpolation par une analyse variographique. L'ordre k et la covariance généralisée choisis décrivent la structure spatiale du phénomène régionalisé et suppriment l'arbitraire de l'interpolateur.

Bien qu'équivalents au niveau théorique, les formalismes des splines et du krigeage répondent à des préoccupations différentes : l'objectif des fonctions splines est de donner une carte "esthétique" (i.e. lisse, régulière) quelles que soient les données expérimentales ; au contraire, le krigeage se soucie de la précision de l'estimation.

Annexe E : introduction aux problèmes de filtrage

Le krigeage universel constitue une première approche, élémentaire, de ce type de problème. En admettant la dichotomie

$$Z(\mathbf{x}) = Y(\mathbf{x}) + m(\mathbf{x})$$

on peut interpréter la dérive $m(\mathbf{x})$ comme la variable d'intérêt ("signal") et le résidu $Y(\mathbf{x})$ comme une perturbation ("bruit"). Les techniques d'estimation de la dérive (krigeage, splines de lissage, moindres carrés...) répondent alors au problème du filtrage du "bruit" $Y(\mathbf{x})$.

Il est intéressant d'élargir cette notion de filtrage lorsque la variable d'intérêt ne peut se réduire à une dérive polynomiale. Nous allons nous placer dans le cadre très général des fonctions aléatoires intrinsèques d'ordre k , dont on souhaite filtrer une composante indésirable. Le problème du filtrage de composantes est également abordé lors de l'examen des techniques multivariées (chapitre 7 et annexe G).

I. Filtrage d'erreurs

I.1. Cas d'une erreur non systématique

On suppose que les mesures sont entachées d'erreurs :

$$Z(\mathbf{x}_\alpha) = Y(\mathbf{x}_\alpha) + \varepsilon(\mathbf{x}_\alpha)$$

où ε est une fonction aléatoire d'espérance nulle, *indépendante de Y* et telle que :

$$\text{cov}[\varepsilon(\mathbf{x}_\alpha), \varepsilon(\mathbf{x}_\beta)] = C_\varepsilon(\mathbf{x}_\alpha, \mathbf{x}_\beta).$$

On se place dans le cas où Y est une fonction aléatoire intrinsèque d'ordre k admettant une fonction de covariance généralisée $K_Y(\mathbf{h})$.

Alors Z admet elle aussi une covariance généralisée, somme de K_Y et C_ε :

$$K_Z(\mathbf{x}, \mathbf{x} + \mathbf{h}) = K_Y(\mathbf{h}) + C_\varepsilon(\mathbf{x}, \mathbf{x} + \mathbf{h}).$$

La variable d'intérêt n'est pas $Z(\mathbf{x})$, variable "bruitée", mais $Y(\mathbf{x})$. Pour estimer $Y(\mathbf{x})$ en un point quelconque $\mathbf{x}_0$ à partir des valeurs mesurées $\{Z(\mathbf{x}_\alpha), \alpha = 1 \dots n\}$, on pose :

$$Y^*(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_\alpha Z(\mathbf{x}_\alpha).$$

Les conditions d'autorisation s'écrivent :

$$\sum_{\beta=1}^n \lambda_\beta f^\ell(\mathbf{x}_\beta) = f^\ell(\mathbf{x}_0) \quad \forall \ell$$

où les f^ℓ désignent une famille complète de monômes de degré inférieur ou égal à k .

Du fait que l'erreur ε est d'espérance nulle, la contrainte de non biais est toujours vérifiée.

En écrivant

$$\text{var}[Y(\mathbf{x}_0) - Y^*(\mathbf{x}_0)] = \text{var}[Y(\mathbf{x}_0) - \sum_{\alpha=1}^n \lambda_\alpha Y(\mathbf{x}_\alpha)] + \text{var}[\sum_{\alpha=1}^n \lambda_\alpha \varepsilon(\mathbf{x}_\alpha)]$$

et en minimisant cette expression sous les conditions d'autorisation, on aboutit au système de krigeage :

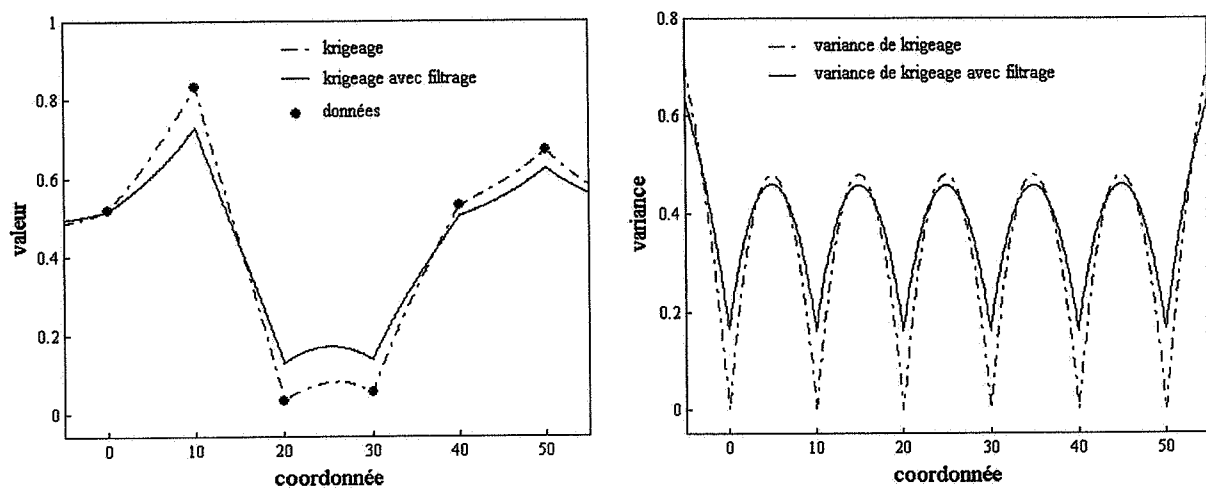
$$\begin{cases} \sum_{\beta=1}^n \lambda_\beta K_Z(\mathbf{x}_\alpha, \mathbf{x}_\beta) + \sum_{\ell} \mu_\ell f^\ell(\mathbf{x}_\alpha) = K_Y(\mathbf{x}_\alpha - \mathbf{x}_0) & \forall \alpha = 1 \dots n \\ \sum_{\beta=1}^n \lambda_\beta f^\ell(\mathbf{x}_\beta) = f^\ell(\mathbf{x}_0) & \forall \ell \end{cases}$$

et à la variance d'estimation:

$$\sigma_{K_{IK-Y}}^2(\mathbf{x}_0) = K_Y(\mathbf{0}) - \sum_{\alpha=1}^n \lambda_\alpha K_Y(\mathbf{x}_\alpha - \mathbf{x}_0) - \sum_{\ell} \mu_\ell f^\ell(\mathbf{x}_0).$$

Le système ci-dessus diffère du système de krigeage intrinsèque de Z par le second membre (où K_Y remplace K_Z). De même, par rapport à la variance du krigeage intrinsèque de Z , la variance d'estimation de Y fait intervenir la covariance généralisée de Y au lieu de celle de Z .

La figure E.1 donne une illustration du krigeage avec filtrage d'erreurs sur un exemple unidimensionnel ; les erreurs de mesure ont une portée de 15m et une amplitude de 0.2, tandis que le modèle variographique sous-jacent a une portée de 30m et une amplitude de 0.8.



*figure E.1 : comparaison entre le krigeage standard et avec filtrage d'erreurs de mesure
à gauche : les estimations ; à droite : les variances d'estimation*

I.2. Cas particulier d'un bruit blanc

L'inférence de la covariance généralisée K_Y n'est possible que si l'on a identifié la fonction de covariance C_ε des erreurs. Un cas particulier très simple est celui où les erreurs forment un **bruit blanc** (effet de pépité pur).

On a alors

$$\text{cov}[\varepsilon(\mathbf{x}_\alpha), \varepsilon(\mathbf{x}_\beta)] = C_\varepsilon(\mathbf{x}_\alpha, \mathbf{x}_\beta) = C_0 \delta(\mathbf{x}_\alpha - \mathbf{x}_\beta) = \begin{cases} C_0 & \text{si } \mathbf{x}_\alpha = \mathbf{x}_\beta \\ 0 & \text{sinon} \end{cases}$$

Dans ce cas, la covariance généralisée de Z n'est autre que

$$K_Z(\mathbf{h}) = K_Y(\mathbf{h}) + C_0 \delta(\mathbf{h}).$$

Il suffit alors de réaliser l'analyse variographique de la variable bruitée Z , et d'identifier la composante pépitique dans la modélisation comme la covariance des erreurs de mesure. S'il existe d'autres causes responsables d'un effet de pépité (microstructure ou déstructuration partielle de la variable non bruitée Y), il convient de n'attribuer qu'une *partie* de l'effet de pépité à la présence de ε ; la détermination de C_0 requiert alors des informations quantitatives ou qualitatives sur l'amplitude de l'erreur de mesure. La covariance généralisée de Y se déduit par différence entre K_Z et $C_0 \delta(\mathbf{h})$.

On notera que le krigeage avec filtrage d'erreur n'est pas un interpolateur exact. Il est facile de montrer que le filtrage d'un bruit blanc ne diffère du krigeage de la variable bruitée Z qu'aux points expérimentaux, le second y prenant la valeur expérimentale, ce qui n'est pas le cas du premier. Ils donneront donc la même carte lorsqu'aucun point estimé n'est confondu avec un point de données. En revanche, la variance de krigeage de Z et la variance de filtrage ne sont pas égales (en tout point non échantillonné, elles diffèrent de la variance C_0 des erreurs de mesure), ce qui montre que le même estimateur n'a pas la même précision selon l'interprétation qu'on lui donne.

La figure E.2 reprend l'exemple de la figure E.1, mais en attribuant aux erreurs de mesure une covariance pépétique d'amplitude 0.2, au lieu d'une covariance de portée 15m.

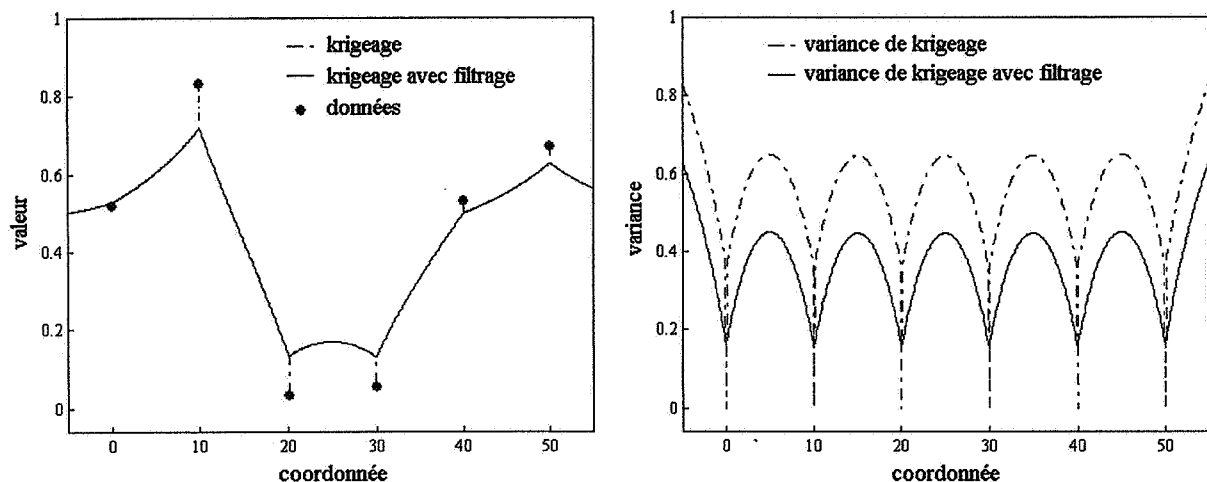


figure E.2 : comparaison entre le krigeage standard et avec filtrage d'un bruit blanc
à gauche : les estimations ; à droite : les variances d'estimation

On peut généraliser cette approche au cas où l'amplitude C_0 de l'erreur commise dépend du point de mesure considéré. En particulier, si on dispose de plusieurs ensembles de données de qualités différentes (mesurés par des procédés différents ou avec une précision différente), il est avantageux de considérer une amplitude C_0 propre à chaque ensemble. Par exemple, un ensemble de mesures très précises peut permettre d'inférer directement $K_Y(\mathbf{h})$, tandis que les covariances des autres ensembles de données présenteront un effet de pépité plus important ; la différence avec $K_Y(\mathbf{h})$ sera alors identifiée comme la valeur de C_0 propre à l'ensemble de mesures considéré. Un autre cas particulier est celui où l'on connaît un intervalle de précision pour chaque mesure, par exemple $\varepsilon(\mathbf{x}_\alpha) \in [-\delta_\alpha, \delta_\alpha]$ avec une probabilité de 95% ; en supposant que l'erreur ε suit une loi gaussienne, on peut associer à cet intervalle une variance égale à $C_0(\mathbf{x}_\alpha) = (\delta_\alpha/2)^2$.

I.3. Cas d'une erreur systématique

On suppose toujours valide la décomposition :

$$Z(\mathbf{x}_\alpha) = Y(\mathbf{x}_\alpha) + \varepsilon(\mathbf{x}_\alpha)$$

où $Y(\mathbf{x})$ est une fonction aléatoire intrinsèque d'ordre k , de covariance généralisée $K_Y(\mathbf{h})$
 ε est une erreur aléatoire d'espérance m_ε **inconnue**, *indépendante de Y* et telle que :

$$\text{cov}[\varepsilon(\mathbf{x}_\alpha), \varepsilon(\mathbf{x}_\beta)] = C_\varepsilon(\mathbf{x}_\alpha, \mathbf{x}_\beta).$$

Si l'on essaie de réaliser un krigeage de $Y(\mathbf{x}_0)$ en posant

$$Y^*(\mathbf{x}_0) = \sum_{\alpha=1}^n \lambda_\alpha Z(\mathbf{x}_\alpha)$$

on s'aperçoit que les conditions d'autorisation et de non-biais ne peuvent être remplies simultanément. La première impose en effet que la somme des poids soit égale à 1 (pour que l'erreur d'estimation soit une combinaison linéaire autorisée), tandis que la seconde exige que cette somme soit nulle. Sans information supplémentaire, on ne peut réaliser le krigeage de la composante sous-jacente Y .

II. Filtrage d'une dérive temporelle

Le filtrage d'une dérive est envisageable dans le modèle de krigeage universel (cf. chapitre 6) ; il suffit par exemple d'évaluer la variable régionalisée et d'ôter une estimation de la dérive, ce qui revient à estimer directement le résidu sous-jacent. Des problèmes d'indétermination se posent lorsque le résidu est intrinsèque ; les difficultés se concentrent en fait au niveau du terme constant, que l'on ne peut attribuer de manière univoque ni à la dérive ni au résidu (annexe C). Dans le modèle des fonctions aléatoires intrinsèques d'ordre k ou intrinsèques générales (cf. chapitre 6 et annexe D), le problème du filtrage d'une dérive est ambigu, car on renonce précisément à séparer la régionalisation en une partie correspondant à une dérive et une partie résiduelle.

Une exception est constituée par les variables régionalisées spatio-temporelles, représentées par des fonctions aléatoires $Z(\mathbf{x}, t)$ qui dépendent à la fois de la position spatiale $\mathbf{x}$ et de l'instant t . Sous la condition que les composantes spatiale et temporelle se séparent, il sera possible de filtrer cette dernière composante et obtenir une estimation débarrassée des fluctuations temporelles.

Un premier modèle est celui d'une variable où la composante temporelle se réduit à une dérive périodique de moyenne nulle :

$$Z(\mathbf{x}, t) = Z_k(\mathbf{x}) + m_\omega(t)$$

où $Z(\mathbf{x}, t)$ est une représentation d'une fonction aléatoire intrinsèque générale (cf. annexe D), c'est-à-dire une fonction aléatoire stationnaire sur les combinaisons linéaires annulant à la fois les monômes $f'(\mathbf{x})$ de degré $\leq k$ et les fonctions trigonométriques $\sin(\omega t)$ et $\cos(\omega t)$;

$Z_k(\mathbf{x})$ est une représentation d'une fonction aléatoire intrinsèque d'ordre k , qui ne possède qu'une dérive spatiale polynomiale de degré k ;

$m_\omega(t)$ est une dérive temporelle périodique : $m_\omega(t) = a \sin(\omega t) + b \cos(\omega t)$, avec des coefficients a et b indéterminés.

On note K une covariance généralisée de la variable $Z(\mathbf{x}, t)$ avec dérivées spatiale et temporelle. On filtre la dérive temporelle en annulant les termes relatifs à cette dérive dans le second membre du système de krigeage intrinsèque, soit :

$$\begin{cases} \sum_{\beta=1}^n \lambda_\beta K(\mathbf{x}_\alpha - \mathbf{x}_\beta) + \sum_{\ell} \mu_\ell f'(\mathbf{x}_\alpha) + v_1 \sin(\omega t_\alpha) + v_2 \cos(\omega t_\alpha) = K(\mathbf{x}_\alpha - \mathbf{x}_0) & \forall \alpha = 1 \dots n \\ \sum_{\beta=1}^n \lambda_\beta f'(\mathbf{x}_\beta) = f'(\mathbf{x}_0) & \forall \ell \\ \sum_{\beta=1}^n \lambda_\beta \sin(\omega t_\beta) = 0 \\ \sum_{\beta=1}^n \lambda_\beta \cos(\omega t_\beta) = 0 \end{cases}$$

où μ_ℓ , v_1 et v_2 sont des multiplicateurs de Lagrange. Résoudre ce système revient à estimer la représentation intemporelle $Z_k(\mathbf{x})$.

Un modèle plus recherché est celui où la variable est la somme de deux représentations de fonctions aléatoires intrinsèques générales *indépendantes* — l'une spatiale associée aux monômes de degré $\leq k$, l'autre temporelle associée aux fonctions trigonométriques $\sin(\omega t)$ et $\cos(\omega t)$:

$$Z(\mathbf{x}, t) = Z_k(\mathbf{x}) + Z_\omega(t).$$

La covariance généralisée de $Z(\mathbf{x}, t)$ est la somme des covariances généralisées des composantes $Z_k(\mathbf{x})$ et $Z_\omega(t)$:

$$K(\mathbf{x}_\alpha - \mathbf{x}_\beta, t_\alpha - t_\beta) = K_1(\mathbf{x}_\alpha - \mathbf{x}_\beta) + K_2(t_\alpha - t_\beta).$$

L'inférence de K à partir des données brutes $\{Z(\mathbf{x}_\alpha, t_\alpha), \alpha = 1 \dots n\}$ permet, par séparation des variables $\mathbf{x}$ et t , d'en déduire K_1 et K_2 .

On filtre la composante temporelle $Z_\omega(t)$ en annulant du second membre du système de krigeage tout ce qui concerne cette composante :

$$\begin{cases} \sum_{\beta=1}^n \lambda_\beta K(\mathbf{x}_\alpha - \mathbf{x}_\beta, t_\alpha - t_\beta) + \sum_{\ell} \mu_\ell f'(\mathbf{x}_\alpha) + v_1 \sin(\omega t_\alpha) + v_2 \cos(\omega t_\alpha) = K_1(\mathbf{x}_\alpha - \mathbf{x}_0) & \forall \alpha = 1 \dots n \\ \sum_{\beta=1}^n \lambda_\beta f'(\mathbf{x}_\beta) = f'(\mathbf{x}_0) & \forall \ell \\ \sum_{\beta=1}^n \lambda_\beta \sin(\omega t_\beta) = 0 \\ \sum_{\beta=1}^n \lambda_\beta \cos(\omega t_\beta) = 0 \end{cases}$$

Annexe F :

une méthode d'analyse des données :

l'analyse en composantes principales

I. Introduction

Les méthodes d'analyse de données ne font pas appel à des modèles probabilistes (ni variable aléatoire, ni fonction aléatoire), mais utilisent la **statistique descriptive** ; par ailleurs, elles ne tiennent pas compte de la position spatiale des observations. Elles ne font donc pas partie des méthodes géostatistiques proprement dites, mais sont néanmoins utiles dans le cadre de l'étude exploratoire de données multivariées. L'analyse en composantes principales est une technique d'*analyse factorielle* (au même titre que l'analyse factorielle des correspondances ou l'analyse des correspondances multiples) ; elle fournit des représentations synthétiques d'un tableau de données.

Nous nous plaçons désormais dans le cas où l'on dispose d'un ensemble de n échantillons $\{ \mathbf{x}_1, \dots, \mathbf{x}_n \}$ sur lesquels ont été mesurées N variables $\{ z_1, \dots, z_N \}$. Ces mesures peuvent être représentées par l'intermédiaire de la matrice suivante, de taille $n \times N$, appelée matrice des données et notée $\mathbf{Z}$:

$$\mathbf{Z} = \begin{pmatrix} z_1(\mathbf{x}_1) & \cdots & z_i(\mathbf{x}_1) & \cdots & z_N(\mathbf{x}_1) \\ \vdots & & & & \vdots \\ z_1(\mathbf{x}_\alpha) & & z_i(\mathbf{x}_\alpha) & & z_N(\mathbf{x}_\alpha) \\ \vdots & & & & \vdots \\ z_1(\mathbf{x}_n) & \cdots & z_i(\mathbf{x}_n) & \cdots & z_N(\mathbf{x}_n) \end{pmatrix} = [z_i(\mathbf{x}_\alpha)]_{\substack{\alpha=1..n \\ i=1..N}}$$

Géométriquement, on peut représenter la matrice des données de deux manières : soit n échantillons dans $\mathbf{R}^N$ (chaque coordonnée est associée à une variable), soit N variables dans $\mathbf{R}^n$ (chaque coordonnée est associée à un échantillon). Ces représentations sont respectivement appelées le **nuage des échantillons** (ou **nuage des individus**) et le **nuage des variables**.

Dans le paragraphe qui suit, nous allons définir des variables synthétiques, baptisées *facteurs*, qui vont faciliter la visualisation de ces nuages, via leurs projections dans des sous-espaces de faible dimension.

II. Construction de facteurs synthétiques orthogonaux

II.1. Principe

L'analyse en composantes principales cherche à construire, à partir des variables initiales z_i ($i = 1 \dots N$), d'autres variables y_p ($p = 1 \dots N$) statistiquement non corrélées, c'est-à-dire telles que

$$\text{cov}(y_p, y_q) = \frac{1}{n} \sum_{\alpha=1}^n y_p(\mathbf{x}_\alpha) y_q(\mathbf{x}_\alpha) = 0 \quad \text{si } p \neq q.$$

Ces quantités non corrélées sont appelées "**facteurs**". On souhaite de plus que les facteurs soient des combinaisons linéaires des variables de départ et qu'ils aient une moyenne nulle sur l'ensemble des échantillons.

II.2. Définition des facteurs

Soient $\mathbf{Z}$ la matrice $n \times N$ des données **centrées**¹ et $\mathbf{V}$ la matrice $N \times N$ de variance-covariance associée :

$$\mathbf{V} = (\sigma_{ij})_{i,j=1 \dots N} = \frac{1}{n} \mathbf{Z}^t \mathbf{Z}.$$

Soit $\mathbf{Y}$ la matrice $n \times N$ contenant les valeurs des facteurs y_p ($p = 1 \dots N$) pour chacun des n échantillons. Les facteurs étant non corrélés, leur matrice de variance-covariance, notée $\mathbf{D}$, est diagonale :

$$\mathbf{D} = \frac{1}{n} \mathbf{Y}^t \mathbf{Y} = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_N \end{pmatrix}$$

avec $d_p = \text{var}(y_p) \quad \forall p \in [1, N]$.

¹ c'est-à-dire que l'on remplace chaque variable par la variable centrée, obtenue en lui soustrayant sa valeur moyenne sur l'ensemble des échantillons.

Comme la transformation des variables en facteurs est supposée linéaire, on cherche une matrice $\mathbf{Q}$, orthogonale de taille $N \times N$, telle que :

$$\mathbf{Y} = \mathbf{Z} \mathbf{Q} \quad \text{avec} \quad \mathbf{Q}^t \mathbf{Q} = \mathbf{I}$$

où $\mathbf{I}$ désigne la matrice identité d'ordre N .

En multipliant cette équation à gauche par $\frac{1}{n} \mathbf{Y}^t$, il vient :

$$\frac{1}{n} \mathbf{Y}^t \mathbf{Y} = \frac{1}{n} \mathbf{Y}^t \mathbf{Z} \mathbf{Q} = \frac{1}{n} (\mathbf{Z} \mathbf{Q})^t \mathbf{Z} \mathbf{Q} = \mathbf{Q}^t \frac{1}{n} \mathbf{Z}^t \mathbf{Z} \mathbf{Q} = \mathbf{Q}^t \mathbf{V} \mathbf{Q}.$$

Soit : $\mathbf{D} = \mathbf{Q}^t \mathbf{V} \mathbf{Q}$, ou encore $\mathbf{V} = \mathbf{Q} \mathbf{D} \mathbf{Q}^t$ puisque $\mathbf{Q}$ est orthogonale. On voit que $\mathbf{Q}$ est une matrice de vecteurs propres orthonormés de $\mathbf{V}$; les valeurs propres, positives ou nulles, ne sont autres que les variances d_p des facteurs y_p .

L'analyse en composantes principales consiste donc à diagonaliser la matrice de variance-covariance $\mathbf{V}$:

$$\mathbf{V} = \mathbf{Q} \mathbf{D} \mathbf{Q}^t \quad \text{avec} \quad \mathbf{Q}^t \mathbf{Q} = \mathbf{I}$$

et à définir les facteurs par : $\mathbf{Y} = \mathbf{Z} \mathbf{Q}$.

Chaque facteur est associé à une valeur propre de $\mathbf{V}$, égale à sa variance statistique. De plus, les facteurs sont *centrés*, c'est-à-dire de moyennes nulles, en tant que combinaisons linéaires de variables centrées.

II.3. Hiérarchisation des facteurs

Désormais, on range les valeurs propres de $\mathbf{V}$ par ordre décroissant : $d_1 \geq d_2 \geq \dots \geq d_N$, i.e. on numérote les facteurs par ordre décroissant des valeurs propres associées.

Les variances des facteurs décomposent la variance totale des données, puisque la trace de la matrice $\mathbf{V}$ est conservée par le changement de base orthonormée :

$$\text{trace}(\mathbf{V}) = \sum_{i=1}^N \sigma_{ii} = \sum_{p=1}^N d_p : \text{variance totale des données.}$$

On montre que cette décomposition est *optimale* : y_1 est la combinaison linéaire des variables z_i centrées de variance maximale ; de même, y_2 est la combinaison linéaire des variables z_i centrées, non corrélée avec y_1 , de variance maximale, etc. Ainsi, **les facteurs extraient successivement une part maximale de la variance des données.**

Les valeurs propres de V indiquent la part de la variance totale associée aux différents facteurs et permettent de les hiérarchiser. Le rapport

$$\frac{\text{variance du facteur}}{\text{variance totale}} = \frac{d_p}{\text{trace}(V)}$$

est une mesure adimensionnée (entre 0 et 1) de la quantité d'information contenue dans le facteur y_p . Plus elle est grande, plus y_p constitue une variable pertinente pour différencier les échantillons.

Les valeurs propres de V sont positives (ce sont les variances des facteurs). Les deux valeurs extrêmes que peut prendre la première valeur propre d_1 sont les suivantes :

- $d_1 = \text{trace}(V)$, auquel cas $d_2 = d_3 = \dots = d_N = 0$: la matrice V est de rang 1, ainsi que la matrice Z ; toutes les variables de départ se déduisent l'une de l'autre par une transformation linéaire.
- $d_1 = \text{trace}(V)/N$, d'où $d_2 = d_3 = \dots = d_N = \text{trace}(V)/N$: le nuage des échantillons est sphérique ; il n'y a pas de direction privilégiée. Ceci se produit lorsque les variables de départ sont mutuellement non corrélées.

L'analyse en composantes principales permet de *résumer* l'information multivariable contenue dans les données, en ne retenant que les premiers facteurs (ceux ayant les variances les plus fortes), par exemple ceux dont la somme des variances représente 80% de la variance totale.

III. Le nuage des échantillons

III.1. Deux critères de représentation équivalents

Le nuage des échantillons $\{z_1(\mathbf{x}_\alpha), \dots, z_N(\mathbf{x}_\alpha)\}_{\alpha=1 \dots n}$ est défini dans un espace de dimension N (l'espace des variables), et est impossible à visualiser dès que $N > 3$. Pour obtenir une représentation synthétique des données, l'idée est de **projeter** ce nuage dans un sous-espace de faible dimension (en pratique, un plan), en choisissant le sous-espace le plus "proche" du nuage, à savoir celui qui minimise son *moment d'inertie*. On montre que ce **critère de proximité** est équivalent au **critère de fidélité** suivant : le sous-espace cherché est celui qui déforme le moins possible les distances entre les points du nuage par l'opération de projection. Nous allons voir que l'analyse en composantes principales répond à ce problème.

III.2. Représentation du nuage des échantillons

Les vecteurs propres de V étant numérotés par ordre décroissant des valeurs propres associées, on établit que le sous-espace de dimension k ($k \in [1, N]$) qui minimise le moment d'inertie du nuage des échantillons est précisément celui engendré par les k premiers vecteurs propres de V .

On appelle *p-ième axe principal d'inertie*, ou encore *p-ième axe factoriel*, la droite engendrée par le p -ième vecteur propre de V , noté q_p (p -ième colonne de Q). Cet axe est relié au p -ième facteur y_p par la relation

$$y_p = Z q_p.$$

En d'autres termes, le facteur y_p est le vecteur de R^n dont les composantes sont les coordonnées des échantillons sur le p -ième axe factoriel.

Ainsi donc, **en projetant le nuage des échantillons dans les sous-espaces engendrés par les premiers axes factoriels, on obtient les visualisations les plus proches et les plus fidèles possibles de ce nuage dans des sous-espaces de faible dimension.** Nous donnons ci-dessous un exemple de représentation d'un nuage d'échantillons dans le premier plan factoriel. Les observations sont les moyennes annuelles de sept composantes du champ magnétique terrestre (*déclinaison D , inclinaison I , composante horizontale H , composante nord X , composante est Y , composante verticale Z , intensité du champ F*), enregistrées en 169 observatoires au cours de l'année 1988.

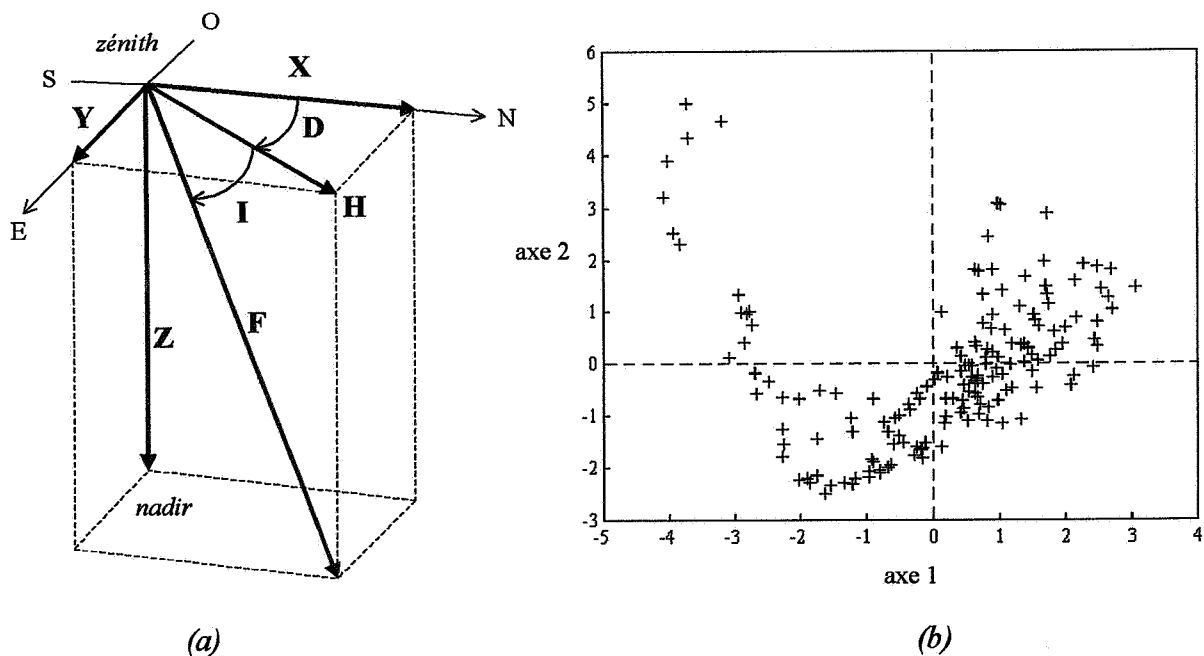


figure F.1 : application de l'analyse en composantes principales à des données magnétiques
schéma explicatif des variables mesurées (a) et représentation du nuage des échantillons (b)

L'examen du plan factoriel formé par les deux premiers axes principaux permet de voir l'allure générale du nuage, parfois sans chercher à identifier les échantillons lorsque leur nombre est élevé. Il est en particulier intéressant de détecter des zones de forte ou de faible densité, qui correspondent à des dispersions non homogènes de la population. A ce titre, l'exemple précédent est instructif, car une vingtaine d'échantillons forme une branche qui se démarque du reste du nuage (pour information, un examen des données révèle qu'il s'agit des observatoires magnétiques situés dans l'hémisphère austral).

La variance du facteur correspondant à un axe factoriel peut se décomposer échantillon par échantillon : on appelle **contribution relative** de l'échantillon α ($\alpha \in [1, n]$) à l'inertie du p -ième axe factoriel, la quantité :

$$CTR_p(\alpha) = \frac{y_{p,\alpha}^2}{n d_p}$$

où $y_{p,\alpha}$, α -ième composante de y_p , représente la coordonnée sur l'axe de l'échantillon α . Etant donné que le facteur y_p est centré et de variance d_p , on a la relation

$$\sum_{\alpha=1}^n CTR_p(\alpha) = 1.$$

On peut aussi définir la contribution d'un ensemble d'échantillons comme la somme des contributions relatives de chacun de ses échantillons.

La contribution mesure le rôle joué par l'échantillon, ou l'ensemble d'échantillons, dans la définition du p -ième axe factoriel. Ce sont les échantillons à forte contribution qui influent sur l'interprétation de l'axe. L'examen de ces contributions permet de repérer les échantillons les plus "atypiques" du nuage, qui sont souvent ceux qui jouent le plus grand rôle dans la détermination des premiers axes factoriels (on pourra éventuellement mettre en évidence des observations aberrantes).

III.3. Réduction des variables de départ (standardisation)

Il est généralement préférable de **standardiser les variables initiales** (c'est-à-dire leur soustraire leur moyenne et les diviser par leur écart-type), surtout si ces variables sont exprimées dans des unités différentes : les variables standardisées sont adimensionnées, de moyennes nulles et de variances unités, et ainsi ramenées à une échelle comparable.

L'analyse en composantes principales repose alors sur la matrice de corrélation $\mathbf{R}$ au lieu de la matrice de variance-covariance $\mathbf{V}$. Les éléments diagonaux de la matrice $\mathbf{R}$ sont les variances des variables standardisées, c'est-à-dire 1. Une valeur propre de $\mathbf{R}$ inférieure à 1 signifie que le facteur associé a une variance inférieure à celle des variables standardisées ; en quelque sorte, il ne discrimine pas les échantillons autant que les variables initiales, et peut par conséquent sembler d'un intérêt secondaire.

Ceci conduit à un premier critère possible pour choisir le nombre d'axes factoriels que l'on va retenir pour expliquer et interpréter les données : on retient les axes pour lesquels les valeurs propres associées sont supérieures à 1 (*critère de Kaiser*). Mais d'autres critères sont envisageables, basés sur l'analyse de l'histogramme des valeurs propres de $\mathbf{R}$: règle du point d'inflexion, règle du coude...

IV. Le nuage des variables

IV.1. Corrélation des variables avec les facteurs

Désormais, nous travaillons avec des données standardisées pour ramener toutes les variables à une échelle unique :

$$\forall i \in [1, N], \forall \alpha \in [1, n], \tilde{z}_i(\mathbf{x}_\alpha) = \frac{z_i(\mathbf{x}_\alpha) - m_i}{\sqrt{\sigma_{ii}}}$$

où $m_i = \frac{1}{n} \sum_{\alpha=1}^n z_i(\mathbf{x}_\alpha)$ est la moyenne statistique de la variable z_i

$\sigma_{ii} = \frac{1}{n} \sum_{\alpha=1}^n [z_i(\mathbf{x}_\alpha) - m_i]^2$ est la variance statistique de z_i .

L'analyse en composantes principales consiste à diagonaliser la matrice de corrélation, soit :

$$\mathbf{R} = \tilde{\mathbf{Q}} \tilde{\mathbf{D}} \tilde{\mathbf{Q}}^t = \mathbf{A}^t \mathbf{A} \text{ avec } \tilde{\mathbf{Q}}^t \tilde{\mathbf{Q}} = \mathbf{I} \text{ et } \mathbf{A} = \sqrt{\tilde{\mathbf{D}}} \tilde{\mathbf{Q}}^t.$$

Les facteurs sont définis par $\mathbf{Y} = \tilde{\mathbf{Z}} \tilde{\mathbf{Q}}$, ce qui s'écrit aussi $\tilde{\mathbf{Z}} = \mathbf{Y} \tilde{\mathbf{Q}}^t$.

Le vecteur $\mathbf{a}_i$, i -ème colonne de $\mathbf{A}$, contient les coefficients de corrélation entre la variable $\tilde{z}_i$ et les différents facteurs.

En effet, $\forall i, p \in [1, N]$

$$\text{corr}(\tilde{z}_i, y_p) = \frac{\text{cov}(\tilde{z}_i, y_p)}{\sqrt{1} \sqrt{\tilde{d}_p}} = \frac{1}{\sqrt{\tilde{d}_p}} \text{cov}\left(\sum_{j=1}^N \tilde{q}_{ij} y_j, y_p\right) = \frac{1}{\sqrt{\tilde{d}_p}} \sum_{j=1}^N \tilde{q}_{ij} \underbrace{\text{cov}(y_j, y_p)}_{\substack{0 \text{ si } j \neq p \\ \tilde{d}_p \text{ sinon}}} = \tilde{q}_{ip} \sqrt{\tilde{d}_p} = a_{pi}.$$

De plus, le produit scalaire des vecteurs $\mathbf{a}_i$ et $\mathbf{a}_j$ est égal au coefficient de corrélation entre les variables z_i et z_j correspondantes :

$$\mathbf{a}_i^t \mathbf{a}_j = \rho_{ij}$$

puisque par construction $\mathbf{A}^t \mathbf{A} = \mathbf{R}$. En particulier, $\mathbf{a}_i^t \mathbf{a}_i = \rho_{ii} = 1$, ce qui montre que la longueur des vecteurs $\mathbf{a}_i$ est égale à l'unité.

IV.2. Représentation du nuage des variables

La matrice A peut servir à représenter géométriquement le nuage des variables. Chaque variable z_i , associée au vecteur $\mathbf{a}_i$ de longueur unité, est représentée par un point sur la sphère unité de $\mathbf{R}^N$. Le cosinus entre les vecteurs $\mathbf{a}_i$ et $\mathbf{a}_j$ est égal au coefficient de corrélation ρ_{ij} entre les variables z_i et z_j .

Pour faciliter la visualisation, on projette la position des variables sur les plans engendrés par les premiers axes factoriels. Les graphiques que l'on obtient sont appelés "cercles des corrélations", car les variables sont représentées par des points à l'intérieur d'un cercle unité (projection de la sphère). En règle générale, la plus ou moins grande proximité – ou au contraire l'éloignement – des projections des variables indique leur plus ou moins forte corrélation – ou antagonisme. Il faut toutefois prendre garde que la proximité à l'intérieur d'un cercle n'implique pas nécessairement une proximité sur la sphère, sauf si les projections sont situées près de la circonférence du cercle ; dans le cas contraire, pour éviter des conclusions erronées, il convient d'examiner les projections du nuage des variables sur plusieurs plans factoriels.

Nous donnons ci-dessous une illustration sur les données géomagnétiques précédemment introduites.

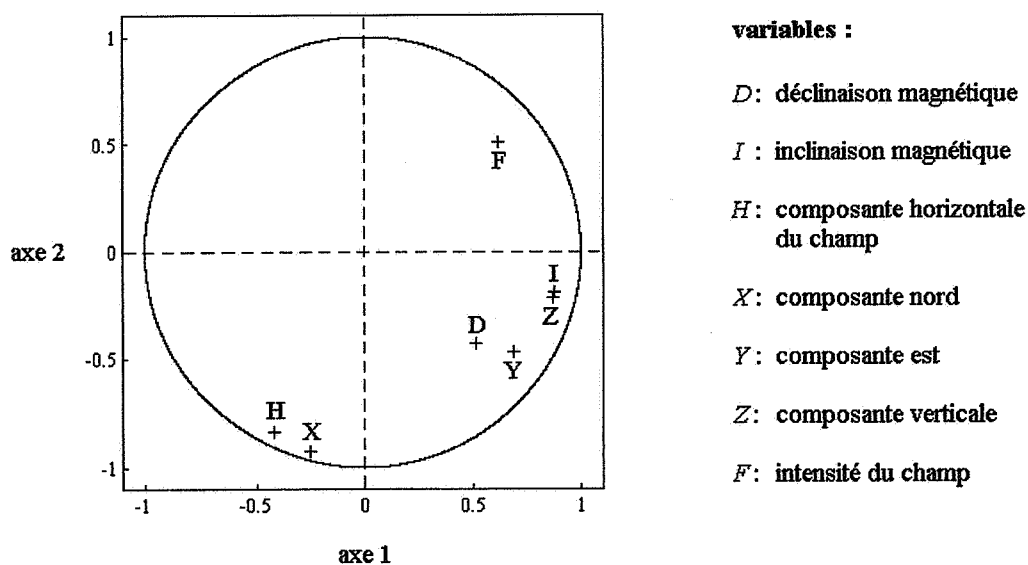


figure F.2 : nuage des variables sur des données géomagnétiques

Les corrélations entre les variables originales et les facteurs permettent parfois de donner à ces derniers une interprétation.

V. Éléments supplémentaires

V.1. Échantillons supplémentaires

On peut considérer certains échantillons comme des données supplémentaires : bien que ne participant pas à la détermination des axes principaux du nuage des échantillons, on les projette sur les plans déterminés par ces axes. Ils peuvent en particulier servir à l'interprétation des axes. On parle d'*échantillons illustratifs*, par opposition aux échantillons ayant participé à la construction des axes, lesquels sont appelés *échantillons actifs*. On peut citer deux applications de cette pratique : d'une part, le cas d'un ou plusieurs échantillons ayant une très (trop) forte contribution à l'inertie de l'un des premiers axes factoriels ; d'autre part, le cas d'échantillons n'appartenant pas au champ étudié, mais à un domaine voisin, et que l'on souhaite comparer aux échantillons actifs.

V.2. Variables supplémentaires

Il est utile d'enrichir l'interprétation des facteurs à l'aide de variables supplémentaires ou illustratives. Ainsi, une variable quantitative dont on connaît les valeurs sur les n échantillons, peut être représentée dans R^n sans qu'elle participe à la détermination des axes principaux : elle sera projetée sur les plans engendrés par ces axes construits à partir des variables *actives*. L'existence de variables supplémentaires très corrélées aux facteurs permet de donner un sens à des facteurs difficiles à interpréter à l'aide des variables actives.

VI. Conclusion

L'analyse en composantes principales permet de décrire des observations multivariées et d'analyser les relations existant entre les variables mesurées sur ces observations. L'objectif essentiel d'une telle analyse est de hiérarchiser l'information multivariable sur des *facteurs*, mutuellement orthogonaux et classés en fonction de la quantité d'information qu'ils contiennent. Ainsi, les premiers facteurs résument l'information contenue dans les données. En ce qui concerne la visualisation des nuages des échantillons et des variables sur des plans factoriels, un graphique est certes plus parlant qu'un tableau de chiffres, mais il convient de faire attention aux lectures rapides conduisant à des interprétations erronées.

Une limitation de l'analyse en composantes principales est de n'étudier que les relations linéaires entre variables (des relations fonctionnelles peuvent être invisibles dans l'analyse). Une autre limitation, plus fondamentale en théorie des variables régionalisées, est qu'elle ne tient pas compte de la position des observations lorsqu'elles sont situées dans l'espace géographique.

Annexe G : l'analyse krigéante

L'analyse krigéante permet de pratiquer l'**analyse des données** lorsque les observations sont localisées dans l'espace géographique, à partir d'une interprétation du modèle linéaire de corégionalisation. Elle s'articule autour de trois étapes : analyse variographique multivariable à l'aide du modèle linéaire de corégionalisation ; décomposition des fonctions aléatoires en *facteurs* indépendants, synthétisant l'information contenue dans les variables et dans leur structure conjointe ; estimation de ces facteurs par cokrigéage.

Nous allons détailler les deux dernières étapes, en divisant la présentation en trois parties : le cas d'une seule variable ; celui de plusieurs variables dont les covariances ou variogrammes simples et croisés sont proportionnels entre eux (*corrélation intrinsèque*) ; enfin, le cas de plusieurs variables décrites par un modèle linéaire de corégionalisation. Cette dernière partie sera une synthèse des deux précédentes.

I. Cas d'une seule variable : le modèle linéaire de régionalisation

I.1. Décomposition de la variable

Souvent, on peut distinguer dans la covariance ou le variogramme expérimental(e) des ruptures de pentes, qui indiquent un changement de structuration à partir d'une certaine échelle. Le modèle variographique associé est un schéma gigogne, formé par la superposition de plusieurs structures élémentaires. Considérons par exemple une fonction aléatoire $Z(\mathbf{x})$ stationnaire d'ordre deux¹, de moyenne m et de fonction de covariance gigogne

$$C(\mathbf{h}) = \sum_{u=1}^S b_u \rho_u(\mathbf{h})$$

où $\forall u \in [1, S]$, $b_u > 0$ et $\rho_u(\mathbf{h})$ est une covariance qui vaut 1 en $\mathbf{h} = \mathbf{0}$.

¹ Le même modèle peut s'appliquer à une fonction aléatoire *localement* stationnaire, c'est-à-dire dont l'espérance $m(\mathbf{x})$ est approximativement constante dans tout voisinage local du champ.

Alors, $Z(\mathbf{x})$ se décompose en une somme de *composantes spatiales* stationnaires X_u , $u = 1 \dots S$, indépendantes (ou plus exactement spatialement non corrélées), de moyennes nulles et de covariances respectives $b_u \rho_u(\mathbf{h})$:

$$Z(\mathbf{x}) = \sum_{u=1}^S X_u(\mathbf{x}) + m$$

avec : $\forall u \in [1, S], \forall \mathbf{x}, E[X_u(\mathbf{x})] = 0$

$$\forall u, v \in [1, S], \forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{D}, \text{cov}[X_u(\mathbf{x} + \mathbf{h}), X_v(\mathbf{x})] = \begin{cases} b_u \rho_u(\mathbf{h}) & \text{si } u = v \\ 0 & \text{sinon} \end{cases}$$

D'autres modèles peuvent être élaborés, par exemple en superposant une fonction aléatoire strictement intrinsèque $X_S(\mathbf{x})$ et $S-1$ fonctions aléatoires stationnaires d'ordre deux $X_u(\mathbf{x})$ de moyennes nulles, non corrélées entre elles et non corrélées avec les *accroissements* de $X_S(\mathbf{x})$:

$$Z(\mathbf{x}) = \sum_{u=1}^{S-1} X_u(\mathbf{x}) + X_S(\mathbf{x}).$$

Le variogramme correspondant est la somme de $S-1$ structures stationnaires et d'une structure intrinsèque stricte :

$$\gamma(\mathbf{h}) = \sum_{u=1}^{S-1} b_u g_u(\mathbf{h}) + b_S g_S(\mathbf{h}).$$

Le modèle linéaire de régionalisation permet ainsi d'**interpréter une covariance ou un variogramme gigogne** en décomposant la fonction aléatoire $Z(\mathbf{x})$ en plusieurs composantes *orthogonales* (i.e. spatialement non corrélées ou à accroissements spatialement non corrélés) agissant à différentes échelles.

I.2. Estimation des composantes spatiales : krigéage avec filtrage de composantes

I.2.1. Krigéage des composantes stationnaires et intrinsèque

Dans le cas du **modèle intrinsèque** précédent, où $Z(\mathbf{x})$ est la superposition de $S-1$ composantes stationnaires et d'une composante strictement intrinsèque, on peut chercher à estimer les valeurs des composantes spatiales $X_u(\mathbf{x})$ à partir des données $\{Z(\mathbf{x}_\alpha), \alpha = 1 \dots n\}$. Le moyen le plus simple est de réaliser un krigéage ordinaire de $X_u(\mathbf{x})$.

On posera :

$$X_u^*(\mathbf{x}) = \sum_{\alpha} \lambda_{\alpha}^u Z(\mathbf{x}_{\alpha})$$

pour aboutir au système :

$$\begin{cases} \sum_{\beta} \lambda_{\beta}^u \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) - \mu_u = b_u g_u(\mathbf{x}_{\alpha} - \mathbf{x}) & \text{pour } \alpha = 1 \dots n \\ \sum_{\beta} \lambda_{\beta}^u = \begin{cases} 0 & \text{si } u = 1 \dots S-1 \text{ (} X_u \text{ stationnaire)} \\ 1 & \text{si } u = S \text{ (} X_S \text{ intrinsèque stricte)} \end{cases} \end{cases}$$

La différence avec le système de krigeage ordinaire réside dans le second membre : seule apparaît la structure g_u relative à la composante $X_u(\mathbf{x})$ et, dans le cas où $X_u(\mathbf{x})$ est stationnaire, la condition de non-biais est modifiée².

A titre indicatif, la variance de krigeage associée est :

$$\sigma_u^2(\mathbf{x}) = \begin{cases} b_u + b_u \sum_{\alpha} \lambda_{\alpha}^u g_u(\mathbf{x}_{\alpha} - \mathbf{x}) & \text{pour } u = 1 \dots S-1 \\ \sum_{v=1}^{S-1} b_v - \mu_S + b_S \sum_{\alpha} \lambda_{\alpha}^S g_S(\mathbf{x}_{\alpha} - \mathbf{x}) & \text{pour } u = S \end{cases}$$

Il convient de remarquer que cette variance n'est jamais nulle, même en un point de données. Cela traduit le fait qu'il subsiste toujours une incertitude dans l'estimation puisque les données portent sur Z , pas sur les X_u .

Par linéarité des équations, on vérifie que la somme des composantes estimées est égale au krigeage ordinaire de la variable de départ :

$$\sum_{u=1}^{S-1} X_u^*(\mathbf{x}) + X_S^*(\mathbf{x}) = Z^*(\mathbf{x}).$$

Dans le **cadre stationnaire** d'ordre deux, on peut remplacer le variogramme par la covariance comme outil structural. Le système de krigeage d'une composante stationnaire $X_u(\mathbf{x})$ est alors :

$$\begin{cases} \sum_{\beta} \lambda_{\beta}^u C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + \mu_u = b_u \rho_u(\mathbf{x}_{\alpha} - \mathbf{x}) & \text{pour } \alpha = 1 \dots n \\ \sum_{\beta} \lambda_{\beta}^u = 0 \end{cases}$$

² On est limité à *une seule* composante intrinsèque stricte : le krigeage est impossible dans le cas où il existerait plusieurs composantes strictement intrinsèques (pas de condition d'autorisation possible). Ceci se comprend si l'on se souvient qu'un processus intrinsèque n'est défini qu'à une constante additive près.

I.2.2. Krigeage de la moyenne locale

Dans le cadre stationnaire, on est intéressé par une estimation de la moyenne m . Cette estimation a déjà été menée lors de la présentation du krigeage ordinaire. Nous allons nous placer dans le cas plus général d'une fonction aléatoire *localement* stationnaire et chercher à estimer la moyenne locale $m(\mathbf{x})$:

$$Z(\mathbf{x}) = \sum_{u=1}^S X_u(\mathbf{x}) + m(\mathbf{x}).$$

L'estimateur de la moyenne locale en un point $\mathbf{x}$ quelconque est une combinaison linéaire pondérée des données avoisinantes :

$$M^*(\mathbf{x}) = \sum_{\alpha} \lambda_{\alpha}^m Z(\mathbf{x}_{\alpha}).$$

L'espérance de l'erreur d'estimation vaut :

$$E[M^*(\mathbf{x}) - m(\mathbf{x})] = \sum_{\alpha} \lambda_{\alpha}^m m(\mathbf{x}_{\alpha}) - m(\mathbf{x}).$$

L'hypothèse de stationnarité locale permet d'identifier $m(\mathbf{x}_{\alpha})$ à $m(\mathbf{x})$, d'où :

$$E[M^*(\mathbf{x}) - m(\mathbf{x})] = \left(\sum_{\alpha} \lambda_{\alpha}^m - 1 \right) m(\mathbf{x}).$$

Pour assurer la nullité de cette espérance quelle que soit la moyenne locale $m(\mathbf{x})$, on doit poser

$$\sum_{\alpha} \lambda_{\alpha}^m = 1.$$

Il ne reste qu'à minimiser la variance de l'erreur sous cette contrainte (en introduisant un multiplicateur de Lagrange μ). Tous calculs faits, on obtient le système :

$$\begin{cases} \sum_{\beta} \lambda_{\beta}^m C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + \mu = 0 & \text{pour } \alpha = 1 \dots n \\ \sum_{\beta} \lambda_{\beta}^m = 1 \end{cases}$$

La variance de krigeage est égale à $\sigma_m^2(\mathbf{x}) = -\mu$.

On retrouve le système de krigeage ordinaire de la moyenne dans le cas stationnaire d'ordre deux, avec un sens élargi (krigeage d'une moyenne *locale*). On prendra garde à ne pas confondre l'estimation de la moyenne locale par krigeage avec le calcul de la moyenne arithmétique des données dans le voisinage de krigeage, laquelle attribue un poids égal à tous les échantillons quelle que soit leur configuration géométrique.

On vérifie que la somme des estimations des composantes stationnaires et de la moyenne locale redonne l'estimation par krigeage ordinaire de la variable de départ :

$$Z^*(\mathbf{x}) = \sum_{u=1}^S X_u^*(\mathbf{x}) + M^*(\mathbf{x}).$$

Les composantes d'un modèle gigogne sont des variables synthétiques qui représentent différents types de structuration spatiale et dont la superposition coïncide avec la variable régionalisée. On estime une composante ou une somme de composantes en annulant, dans le second membre du système de krigeage ordinaire, les termes non relatifs à ces composantes, et en modifiant éventuellement la condition sur la somme des pondérateurs. Par exemple, on peut éliminer ou "filtrer" l'effet de pépite, assimilé à un bruit blanc indésirable, et obtenir une estimation de la somme des autres composantes. Il convient de noter que le krigeage avec filtrage n'est pas un interpolateur exact³ car on estime une composante spatiale sous-jacente et non la variable de départ.

II. Décomposition de plusieurs fonctions aléatoires en corrélation intrinsèque

II.1. Décomposition des variables

II.1.1. Cas stationnaire du second ordre

Considérons N fonctions aléatoires conjointement **stationnaires** $\mathbf{Z}(\mathbf{x}) = [Z_1(\mathbf{x}), \dots, Z_N(\mathbf{x})]$, de moyennes respectives $\mathbf{m} = [m_1, \dots, m_N]$ et de covariances de la forme :

$$\mathbf{C}(\mathbf{h}) = \mathbf{V} \rho_0(\mathbf{h})$$

où $\mathbf{C}(\mathbf{h}) = [C_{ij}(\mathbf{h})]_{i,j}$ est la matrice des covariances simples et croisées de $\mathbf{Z}(\mathbf{x})$

$\mathbf{V} = (\sigma_{ij})_{i,j} = [C_{ij}(\mathbf{0})]_{i,j}$ est la matrice de variance-covariance de $\mathbf{Z}(\mathbf{x})$

$\rho_0(\mathbf{h})$ est une covariance telle que $\rho_0(\mathbf{0}) = 1$.

³ L'interpolation exacte peut d'ailleurs être importune lorsque le modèle variographique présente un effet de pépite, et que la grille d'estimation rencontre le réseau d'échantillonnage en quelques points. En filtrant la composante pépitique, on évite que la carte des valeurs estimées présente des discontinuités en ces points. Dans ce cas, le "krigeage" ne sera plus un interpolateur exact des données.

Nous pouvons faire une analyse en composantes principales en diagonalisant la matrice de variance-covariance V (cf. annexe F). Soient Q une matrice orthogonale de vecteurs propres de V et D la matrice diagonale de valeurs propres associées :

$$V = Q D Q^t = A^t A \text{ avec } Q^t Q = I, D \text{ matrice diagonale et } A = \sqrt{D} Q^t.$$

Alors la relation⁴

$$Y(x) = [Z(x) - m] \times A^{-1}$$

définit un ensemble de *facteurs spatiaux* $Y(x) = [Y_1(x), \dots, Y_N(x)]$ tels que :

$$\begin{cases} \forall p \in [1, N], \forall x \in \mathcal{D}, E[Y_p(x)] = 0 \\ \forall p, q \in [1, N], \forall x \in \mathcal{D}, \text{cov}[Y_p(x), Y_q(x)] = \begin{cases} 0 & \text{si } p \neq q \\ 1 & \text{si } p = q \end{cases} \end{cases}$$

Or, en tant que combinaisons linéaires de variables en corrélation intrinsèque, les facteurs sont également en corrélation intrinsèque, c'est-à-dire :

$$\forall p, q \in [1, N], \forall x, x+h \in \mathcal{D}, \text{cov}[Y_p(x+h), Y_q(x)] \text{ est proportionnel à } \rho_0(h).$$

Par suite :

- 1) la nullité en $h = 0$ pour $p \neq q$ entraîne la propriété fondamentale :

$$\forall p, q \in [1, N], p \neq q, \forall x, x+h \in \mathcal{D}, \text{cov}[Y_p(x+h), Y_q(x)] = 0.$$

Ainsi, dans le cas de la corrélation intrinsèque, les facteurs obtenus par analyse en composantes principales sont *spatialement non corrélés*. Nous verrons que cette propriété disparaît dans le cas général.

- 2) la valeur unité en $h = 0$ pour $p = q$ permet d'écrire :

$$\forall p \in [1, N], \forall x, x+h \in \mathcal{D}, \text{cov}[Y_p(x+h), Y_p(x)] = \rho_0(h).$$

Chaque facteur spatial est associé à une valeur propre positive de V qui mesure en quelque sorte la "quantité" d'information contenue dans ce facteur. Dans le cas où cette valeur est nulle, le facteur correspondant est nul ; cela signifie que les variables de départ $Z(x)$ sont liées linéairement. L'examen des valeurs propres permet de classer les facteurs par ordre d'importance.

⁴ L'utilisation de la matrice A^{-1} au lieu de Q sert à normer les facteurs, i.e. à assurer qu'ils aient une variance unité.

Il est souvent préférable de travailler sur les variables standardisées, c'est-à-dire centrées et réduites (qui sont encore en corrélation intrinsèque)

$$\forall i \in [1, N], \tilde{Z}_i(\mathbf{x}) = \frac{Z_i(\mathbf{x}) - m_i}{\sqrt{\sigma_{ii}}}.$$

Lorsque les variables $\{Z_i(\mathbf{x}), i = 1 \dots N\}$ sont exprimées dans des unités différentes, la standardisation permet de les ramener à une échelle unique : les variables deviennent adimensionnées, de moyennes nulles et de variances unitaires. L'analyse en composantes principales revient alors à diagonaliser la matrice de corrélation

$$\mathbf{R} = \left(\frac{\sigma_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}} \right)_{i,j} = (\rho_{ij})_{i,j}$$

au lieu de la matrice de variance-covariance $\mathbf{V}$.

II.1.2. Cas intrinsèque

La décomposition précédente peut être généralisée au **cadre intrinsèque**. Les fonctions aléatoires $\mathbf{Z}(\mathbf{x}) = [Z_1(\mathbf{x}), \dots, Z_N(\mathbf{x})]$ sont en corrélation intrinsèque lorsque leurs *variogrammes* simples et croisés sont proportionnels à un même variogramme :

$$\Gamma(\mathbf{h}) = \mathbf{P} g_0(\mathbf{h})$$

où $\Gamma(\mathbf{h}) = [\gamma_{ij}(\mathbf{h})]_{i,j}$ est la matrice des variogrammes simples et croisés de $\mathbf{Z}(\mathbf{x})$

$\mathbf{P} = (p_{ij})_{i,j}$ est une matrice symétrique appelée matrice des "paliers"⁵

$g_0(\mathbf{h})$ est un variogramme de base, de palier unité s'il est stationnaire.

Pour chaque paire d'indices (i,j) , on définit le *coefficient de codispersion* comme le coefficient de corrélation des accroissements $Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{x})$ et $Z_j(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{x})$:

$$cc_{ij}(\mathbf{h}) = \frac{\gamma_{ij}(\mathbf{h})}{\sqrt{\gamma_{ii}(\mathbf{h}) \gamma_{jj}(\mathbf{h})}} \in [-1, 1].$$

Sous l'hypothèse de corrélation intrinsèque, cc_{ij} ne dépend pas de $\mathbf{h}$. Le calcul des coefficients de codispersion et la vérification de leur invariance en fonction de $\mathbf{h}$ constituent donc un test du modèle de corrélation intrinsèque.

⁵ Cette dénomination n'est pas adaptée lorsque g_0 est un variogramme sans palier, par exemple un modèle puissance, auquel cas $\mathbf{P}$ a plutôt le sens d'une matrice de pentes ou de facteurs d'échelle.

Dans le cadre stationnaire du second ordre⁶, en supposant que $g_0(\mathbf{h})$ est de palier unité et que les covariances s'annulent lorsque $|\mathbf{h}|$ tend vers l'infini (condition nécessaire d'ergodicité), on a :

$$\mathbf{P} = \lim_{|\mathbf{h}| \rightarrow +\infty} \Gamma(\mathbf{h}) = \mathbf{C}(\mathbf{0}) = \mathbf{V}$$

de sorte qu'en cas de stationnarité $cc_{ij}(\mathbf{h})$ coïncide avec le coefficient de corrélation ρ_{ij} entre Z_i et Z_j .

Ainsi, en cas de stationnarité d'ordre deux, $\mathbf{P}$ est la matrice de variance-covariance de $\mathbf{Z}(\mathbf{x})$. Nous définirons donc les facteurs spatiaux $\mathbf{Y}(\mathbf{x}) = [Y_1(\mathbf{x}), \dots, Y_N(\mathbf{x})]$ de manière similaire au cas stationnaire, à savoir :

$$\mathbf{Y}(\mathbf{x}) = \mathbf{Z}(\mathbf{x}) \times \mathbf{A}^{-1}$$

avec $\mathbf{P} = \mathbf{Q} \mathbf{D} \mathbf{Q}^t = \mathbf{A}^t \mathbf{A}$, $\mathbf{Q}^t \mathbf{Q} = \mathbf{I}$, $\mathbf{D}$ matrice diagonale et $\mathbf{A} = \sqrt{\mathbf{D}} \mathbf{Q}^t$.

En général, les facteurs ainsi définis ne sont pas stationnaires, mais intrinsèques stricts (en tant que combinaisons linéaires de fonctions aléatoires intrinsèques). On démontre que les *accroissements* de ces facteurs sont spatialement non corrélés, à condition que les covariances croisées d'accroissements soient paires⁷.

On a alors :

$$\begin{cases} \forall p \in [1, N], \forall \mathbf{x}, \forall \mathbf{h}, E[Y_p(\mathbf{x} + \mathbf{h}) - Y_p(\mathbf{x})] = 0 \\ \forall p \in [1, N], \forall \mathbf{x}, \forall \mathbf{h}, \text{var}[Y_p(\mathbf{x} + \mathbf{h}) - Y_p(\mathbf{x})] = 2g_0(\mathbf{h}) \\ \forall p, q, p \neq q, \forall \mathbf{h}_1, \mathbf{h}_2, \forall \mathbf{x}, \mathbf{y}, \text{cov}[Y_p(\mathbf{x} + \mathbf{h}_1) - Y_p(\mathbf{x}), Y_q(\mathbf{y} + \mathbf{h}_2) - Y_q(\mathbf{y})] = 0 \end{cases}$$

Les facteurs sont *orthogonaux* (leurs accroissements sont spatialement non corrélés) : ils ne contiennent pas d'information redondante. Les premiers facteurs, ceux associés aux plus grandes valeurs propres de $\mathbf{D}$, constituent un résumé optimal de l'information multivariable. Ce résumé est d'autant plus exhaustif que la somme des valeurs propres concernées est proche de la trace de la matrice $\mathbf{D}$.

⁶ Dans ce cas, les variogrammes croisés sont reliés aux fonctions de covariance croisées par la relation

$$\Gamma(\mathbf{h}) = \mathbf{C}(\mathbf{0}) - [\mathbf{C}(\mathbf{h}) + \mathbf{C}(-\mathbf{h})] / 2$$

qui montre que les variogrammes croisés ne retiennent que le terme pair des covariances croisées. En particulier, l'hypothèse de corrélation intrinsèque sur les variogrammes est moins forte que celle sur les covariances, et ne garantit pas, dans le cas stationnaire, la non corrélation spatiale des facteurs définis précédemment. Il faut faire l'hypothèse supplémentaire que les covariances croisées sont paires, auquel cas la donnée des variogrammes croisés équivaut à celle des covariances croisées.

⁷ c'est-à-dire très précisément : $\forall i, j \in [1, N], \forall \mathbf{x}, \mathbf{x} + \mathbf{h}_1, \mathbf{x} + \mathbf{h}_2 \in \mathcal{D}$,

$$\text{cov}[Z_i(\mathbf{x} + \mathbf{h}_1) - Z_i(\mathbf{x}), Z_j(\mathbf{x} + \mathbf{h}_2) - Z_j(\mathbf{x})] = \text{cov}[Z_j(\mathbf{x} + \mathbf{h}_1) - Z_j(\mathbf{x}), Z_i(\mathbf{x} + \mathbf{h}_2) - Z_i(\mathbf{x})].$$

II.2. Estimation des facteurs spatiaux

Si le point $\mathbf{x}$ est un des points de données où toutes les variables sont connues, alors $\mathbf{Y}(\mathbf{x})$ sera évalué par l'une des relations de définition des facteurs, celle relative à l'hypothèse de stationnarité retenue. En toute rigueur, ce calcul suppose, dans le cadre stationnaire, les moyennes de $\mathbf{Z}(\mathbf{x})$ connues et non estimées.

Si $\mathbf{x}$ ne coïncide pas avec un point de données, on estimera $\mathbf{Y}(\mathbf{x})$ par cokrigage à partir des données relatives à $\mathbf{Z}(\mathbf{x})$, ou, ce qui revient au même, à partir des valeurs de $\mathbf{Y}$ aux points de données. Or, ces derniers étant en corrélation intrinsèque, leur cokrigage se réduit à un krigeage séparé (en cas d'homotopie), ce qui simplifie considérablement les calculs.

Par exemple, dans le cadre stationnaire d'ordre deux :

- *cas général (moyennes inconnues, possible hétérotopie)*

On estime $Y_p(\mathbf{x})$, p -ième composante de $\mathbf{Y}(\mathbf{x})$, par :

$$Y_p^*(\mathbf{x}) = \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i Z_i(\mathbf{x}_{\alpha}^i).$$

Les pondérateurs sont déterminés par le système

$$\begin{cases} \sum_{j=1}^N \sum_{\beta=1}^{n_j} \lambda_{\beta}^j C_{ij}(\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j) + \mu_i = a_{pi} \rho_0(\mathbf{x}_{\alpha}^i - \mathbf{x}) & \text{pour } i = 1 \dots N, \alpha = 1 \dots n_i \\ \sum_{\beta=1}^{n_i} \lambda_{\beta}^i = 0 & \text{pour } i = 1 \dots N \end{cases}$$

où a_{pi} est le terme générique de la matrice $\mathbf{A}$.

- *homotopie (m_i connues)*

L'estimation se simplifie en :

$$Y_p^*(\mathbf{x}) = \sum_{\alpha=1}^n \lambda_{\alpha} Y_p(\mathbf{x}_{\alpha})$$

où les poids sont solutions du système

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} \rho_0(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + \mu = \rho_0(\mathbf{x}_{\alpha} - \mathbf{x}) & \text{pour } \alpha = 1 \dots n \\ \sum_{\beta=1}^n \lambda_{\beta} = 0 \end{cases}$$

III. Le modèle linéaire de corégionalisation

III.1. Décomposition des variables

Le modèle linéaire de corégionalisation de N fonctions aléatoires $\mathbf{Z}(\mathbf{x}) = [Z_1(\mathbf{x}), \dots, Z_N(\mathbf{x})]$ stationnaires d'ordre deux de moyennes $\mathbf{m} = [m_1, \dots, m_N]$ suppose que toutes les covariances sont des combinaisons linéaires des mêmes structures de base :

$$\mathbf{C}(\mathbf{h}) = \sum_{u=1}^S \mathbf{B}_u \rho_u(\mathbf{h})$$

où $\mathbf{C}(\mathbf{h}) = [C_{ij}(\mathbf{h})]_{i,j}$ est la matrice des covariances simples et croisées de $\mathbf{Z}(\mathbf{x})$

$\mathbf{B}_u = (b_{ij}^u)_{i,j}$, $u = 1 \dots S$, sont les matrices de corégionalisation, symétriques de type positif
 $\rho_u(\mathbf{h})$, $u = 1 \dots S$, sont des fonctions de covariance qui valent 1 en $\mathbf{h} = \mathbf{0}$

On interprète ce modèle en décomposant $\mathbf{Z}(\mathbf{x})$ en un ensemble de composantes spatiales $\mathbf{X}_u$, $u = 1 \dots S$:

$$\mathbf{Z}(\mathbf{x}) - \mathbf{m} = \sum_{u=1}^S \mathbf{X}_u(\mathbf{x})$$

telles que

- 1) lorsque $u \neq v$, $\mathbf{X}_u = [X_1^u(\mathbf{x}), \dots, X_N^u(\mathbf{x})]$ et $\mathbf{X}_v = [X_1^v(\mathbf{x}), \dots, X_N^v(\mathbf{x})]$ sont spatialement non corrélés :

$$\forall u, v \in [1, S], u \neq v, \forall i, j \in [1, N], \forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{D}, \text{cov}[X_i^u(\mathbf{x} + \mathbf{h}), X_j^v(\mathbf{x})] = 0$$

- 2) $\mathbf{X}_u = [X_1^u(\mathbf{x}), \dots, X_N^u(\mathbf{x})]$ a pour matrice de covariances simples et croisées $\mathbf{B}_u \rho_u(\mathbf{h})$:

$$\forall u \in [1, S], \forall i, j \in [1, N], \forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{D}, \text{cov}[X_i^u(\mathbf{x} + \mathbf{h}), X_j^u(\mathbf{x})] = b_{ij}^u \rho_u(\mathbf{h}).$$

De même que nous avons défini des coefficients de codispersion dans le modèle de corrélation intrinsèque, nous pouvons définir des *coefficients de corrélation régionalisée* pour chaque indice u :

$$\rho_{ij}^u = \frac{b_{ij}^u}{\sqrt{b_{ii}^u b_{jj}^u}}$$

et on a : $-1 \leq \rho_{ij}^u \leq 1$, puisque $|b_{ij}^u| \leq \sqrt{b_{ii}^u b_{jj}^u}$ (la matrice $\mathbf{B}_u$ étant de type positif).

Le coefficient de corrélation régionalisée mesure le lien ou "interaction" entre les composantes relatives à la structure de base ρ_u . Cette structure est représentative d'une certaine échelle de variation (donnée par exemple par sa portée pratique ou sa portée intégrale).

Pour chaque indice u , les composantes de $\mathbf{X}_u$ sont en corrélation intrinsèque, peuvent être décomposées en *facteurs* spatialement non corrélés. Le modèle complet de corégionalisation s'écrit :

$$\mathbf{Z}(\mathbf{x}) - \mathbf{m} = \sum_{u=1}^S \mathbf{Y}_u(\mathbf{x}) \times \mathbf{A}_u$$

où $\forall u \in [1, S], \mathbf{Y}_u(\mathbf{x}) = [Y_1^u(\mathbf{x}), \dots, Y_N^u(\mathbf{x})]$

$\forall u \in [1, S], \forall p \in [1, N], \forall \mathbf{x}, E[Y_p^u(\mathbf{x})] = 0$

$\forall u, v \in [1, S], \forall p, q \in [1, N], \forall \mathbf{x}, \forall \mathbf{h}, \text{cov}[Y_p^u(\mathbf{x} + \mathbf{h}), Y_q^v(\mathbf{x})] = \begin{cases} 0 & \text{si } u \neq v \text{ ou } p \neq q \\ \rho_u(\mathbf{h}) & \text{sinon} \end{cases}$

$\forall u \in [1, S], \mathbf{B}_u = \mathbf{A}_u^t \mathbf{A}_u, \mathbf{A}_u = \sqrt{\mathbf{D}_u} \mathbf{Q}_u^t, \mathbf{Q}_u^t \mathbf{Q}_u = \mathbf{I}$ et $\mathbf{D}_u$ matrice diagonale.

La même démarche s'applique au cas de plusieurs fonctions aléatoires **intrinsèques** ; le modèle de corégionalisation fait l'hypothèse que leurs *variogrammes* simples et croisés sont des combinaisons linéaires des mêmes structures de base :

$$\Gamma(\mathbf{h}) = \sum_{u=1}^S \mathbf{B}_u g_u(\mathbf{h})$$

où $\Gamma(\mathbf{h}) = [\gamma_{ij}(\mathbf{h})]_{i,j}$ est la matrice des variogrammes simples et croisés de $\mathbf{Z}(\mathbf{x})$

$\mathbf{B}_u = (b_{ij}^u)_{i,j}, u = 1 \dots S$, sont les *matrices de corégionalisation*, de type positif

$g_u(\mathbf{h}), u = 1 \dots S$, sont des variogrammes de base, dont nous supposons les k premiers bornés et les $S-k$ derniers non bornés ($k \in [0, S]$).

Les *coefficients de corrélation régionalisée* sont toujours définis pour chaque indice u par :

$$\rho_{ij}^u = \frac{b_{ij}^u}{\sqrt{b_{ii}^u b_{jj}^u}} \in [-1, 1].$$

On interprète ce modèle en décomposant dans un premier temps les variables $\mathbf{Z}(\mathbf{x})$ en un ensemble de composantes $\mathbf{X}_u, u = 1 \dots S$:

$$\mathbf{Z}(\mathbf{x}) = \sum_{u=1}^S \mathbf{X}_u(\mathbf{x}).$$

Les composantes de $\mathbf{X}_u$, dont les variogrammes sont proportionnels à $g_u(\mathbf{h})$, peuvent être décomposées en facteurs spatiaux. On obtient finalement :

$$\mathbf{Z}(\mathbf{x}) = \sum_{u=1}^S \mathbf{Y}_u(\mathbf{x}) \times \mathbf{A}_u$$

où $\forall u \in [1, S]$, $\mathbf{Y}_u(\mathbf{x}) = [Y_1^u(\mathbf{x}), \dots, Y_N^u(\mathbf{x})]$.

Les facteurs stationnaires $\{Y_p^u(\mathbf{x}), u = 1 \dots k, p = 1 \dots N\}$ et les *accroissements* des facteurs non stationnaires $\{Y_p^u(\mathbf{x}), u = k + 1 \dots S, p = 1 \dots N\}$ sont d'espérance nulle et spatialement non corrélés⁸.

Mathématiquement, en distinguant les facteurs stationnaires ($u = 1 \dots k$) des facteurs intrinsèques stricts ($u = k + 1 \dots S$), on a :

$$\forall p \in [1, N], \forall u \in [1, k], \forall \mathbf{x}, E[Y_p^u(\mathbf{x})] = 0$$

$$\forall p \in [1, N], \forall u \in [k + 1, S], \forall \mathbf{x}, \forall \mathbf{h}, E[Y_p^u(\mathbf{x} + \mathbf{h}) - Y_p^u(\mathbf{x})] = 0$$

et

$$\forall p, q \in [1, N], \forall u, v \in [1, k], u \neq v \text{ ou } p \neq q, \forall \mathbf{x}, \forall \mathbf{h}, \text{cov}[Y_p^u(\mathbf{x} + \mathbf{h}), Y_q^v(\mathbf{x})] = 0$$

$$\forall p, q \in [1, N], \forall u, v \in [k + 1, S], u \neq v \text{ ou } p \neq q, \forall \mathbf{x}, \mathbf{y}, \forall \mathbf{h}_1, \mathbf{h}_2,$$

$$\text{cov}[Y_p^u(\mathbf{x} + \mathbf{h}_1) - Y_p^u(\mathbf{x}), Y_q^v(\mathbf{y} + \mathbf{h}_2) - Y_q^v(\mathbf{y})] = 0$$

$$\forall p, q \in [1, N], \forall u \in [1, k], \forall v \in [k + 1, S], \forall \mathbf{x}, \mathbf{y}, \forall \mathbf{h}, \text{cov}[Y_p^u(\mathbf{x}), Y_q^v(\mathbf{y} + \mathbf{h}) - Y_q^v(\mathbf{y})] = 0$$

Les facteurs ainsi définis sont spatialement *orthogonaux*. Chaque facteur est associé à une valeur propre de la matrice de corégionalisation correspondante, qui indique son importance relative, ce qui permet de **hiérarchiser les facteurs**. Toutefois, il faut prendre garde au fait que les matrices associées à des structures non stationnaires n'ont pas le même sens que les autres (ce sont des matrices de *pentés*, et non des matrices de *paliers*) ; dans la hiérarchie, les facteurs non stationnaires doivent être privilégiés par rapport aux facteurs stationnaires.

⁸ à condition de faire l'hypothèse supplémentaire que les covariances croisées des composantes stationnaires sont paires, ainsi que les covariances croisées des *accroissements* des composantes non stationnaires.

III.2. Estimation des facteurs spatiaux

Hormis le cas de la corrélation intrinsèque, si le point $\mathbf{x}$ à estimer fait partie des points de données, on obtient un système linéaire indéterminé (trop d'inconnues par rapport au nombre d'équations). Donc, quel que soit le point $\mathbf{x}$, on estimera la valeur des facteurs spatiaux par cokrigage à partir des données multivariées disponibles.

L'estimateur du facteur $Y_p^u(\mathbf{x})$, p -ième composante de $\mathbf{Y}_u(\mathbf{x})$, sera de la forme :

$$Y_p^u(\mathbf{x}) = \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i Z_i(\mathbf{x}_{\alpha}^i).$$

Dans le cas où toutes les variables sont stationnaires, de moyennes inconnues, tous les facteurs sont stationnaires, et l'on peut utiliser les fonctions de covariance dans le système de cokrigage. Il vient :

$$\begin{cases} \sum_{j=1}^N \sum_{\beta=1}^{n_j} \lambda_{\beta}^j C_{ij}(\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j) + \mu_i = a_{pi}^u \rho_u(\mathbf{x}_{\alpha}^i - \mathbf{x}) & \text{pour } i = 1 \dots N, \alpha = 1 \dots n_i \\ \sum_{\beta=1}^{n_i} \lambda_{\beta}^i = 0 & \text{pour } i = 1 \dots N \end{cases}$$

où a_{pi}^u désigne le terme générique de la matrice $\mathbf{A}_u$.

Dans le cas plus général où les variables vérifient l'hypothèse intrinsèque conjointe avec une seule structure strictement intrinsèque⁹, il est aussi possible d'estimer par cokrigage les différents facteurs spatiaux. Le système de cokrigage final est :

$$\begin{cases} \sum_{j=1}^N \sum_{\beta=1}^{n_j} \lambda_{\beta}^j \gamma_{ij}(\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j) - \mu_i = a_{pi}^u g_u(\mathbf{x}_{\alpha}^i - \mathbf{x}) & \text{pour } i = 1 \dots N, \alpha = 1 \dots n_i \\ \sum_{i=1}^N a_{qi}^u \sum_{\beta=1}^{n_i} \lambda_{\beta}^i = \begin{cases} 0 & \text{pour un facteur spatial stationnaire} \\ \delta_p^q & \text{pour un facteur spatial intrinsèque strict} \end{cases} & \text{pour } q = 1 \dots N \end{cases}$$

où δ_p^q est le symbole de Kronecker (égal à 1 si $p = q$, 0 sinon).

⁹ On ne peut réaliser le cokrigage des facteurs s'il existe plus d'une structure non stationnaire (pas de condition d'autorisation possible).

IV. Différence entre l'analyse krigéante et l'analyse en composantes principales

On se place dans le cadre du modèle linéaire de corégionalisation. Sous l'hypothèse de stationnarité d'ordre deux conjointe, la matrice de variance-covariance V de $Z(\mathbf{x})$ est la somme des matrices de corégionalisation B_u , que l'on choisisse le modèle sur les covariances ou sur les variogrammes.

L'analyse krigéante revient à étudier *séparément* chaque matrice de corégionalisation B_u . Les facteurs correspondants¹⁰ sont spatialement non corrélés :

$$\text{cov}[Y_p^u(\mathbf{x} + \mathbf{h}), Y_q^u(\mathbf{x})] = 0 \text{ si } p \neq q.$$

En revanche, l'analyse en composantes principales, menée sur la matrice de variance-covariance V , ne fournit en général pas des facteurs *spatialement* non corrélés :

$$\text{cov}[Y_p(\mathbf{x} + \mathbf{h}), Y_q(\mathbf{x})] \neq 0.$$

L'analyse en composantes principales ne tient pas compte de la position des données, ni par conséquent de la variation de leur structuration spatiale en fonction de la distance. Les covariances ou variogrammes croisés entre les facteurs ne sont pas, en général, identiquement nuls, ce qui indique qu'il existe des redondances d'information entre ces facteurs. En particulier, ceux-ci ne constituent pas des variables autokrigeables, comme on le pense parfois à tort. De plus, la matrice de variance-covariance n'a de signification que sous l'hypothèse de stationnarité d'ordre deux, alors que l'on peut encore définir les matrices de corégionalisation dans le cadre intrinsèque strict.

¹⁰ Les facteurs spatiaux sont obtenus en décomposant chaque matrice B_u en valeurs et vecteurs propres :

$$Q_u^t B_u Q_u = D_u \text{ avec } Q_u^t Q_u = I \text{ et } D_u \text{ diagonale.}$$

Comme dans l'analyse en composantes principales classique, il convient au préalable de **réduire les variables de départ** pour les ramener à une échelle unique et les rendre comparables. Ceci revient à rechercher des facteurs orthogonaux par rapport à une métrique M différente de l'identité, ou encore résoudre le problème de valeurs propres généralisé

$$Q_u^t B_u Q_u = D_u \text{ avec } Q_u^t M Q_u = I$$

soit

$$B_u Q_u = M Q_u D_u \text{ avec } Q_u^t M Q_u = I.$$

Si les variables $\{Z_i, i = 1 \dots N\}$ sont stationnaires, de variances respectives $\sigma_{i,i}^2$, on choisira la métrique $M = \text{diag}(\sigma_{i,i}^2)$, ce qui revient à réduire chaque Z_i par son écart-type $\sigma_{i,i}$. Dans le cas où il existe une structure intrinsèque stricte g_S dans le modèle, on peut réduire Z_i par la racine carrée du terme diagonal b_{ii}^S de la matrice B_S correspondante ; ce terme mesure le "degré d'infinité" de l'écart-type de Z_i et constitue donc un facteur de normation convenable.

Annexe H : compléments sur les modèles variographiques multivariables

I. Le modèle bilinéaire de corégionalisation

Le modèle linéaire de corégionalisation (*cf.* chapitre 7) implique que les covariances croisées soient des fonctions paires. Le modèle bilinéaire de corégionalisation, présenté ci-dessous, autorise des covariances croisées non paires, ce qui permet en particulier de modéliser les effets de retard. Il est d'un grand intérêt pour l'étude de phénomènes temporels où les variables évoluent de manière asynchrone.

I.1. Le modèle linéaire complexe de corégionalisation

Considérons des fonctions aléatoires $\mathbf{Z}(\mathbf{x}) = [Z_1(\mathbf{x}), \dots, Z_N(\mathbf{x})]$ conjointement stationnaires, de moyennes $\mathbf{m} = [m_1, \dots, m_N]$, et à valeurs complexes. Le modèle linéaire de corégionalisation suppose que leurs covariances simples et croisées¹ sont des combinaisons linéaires des mêmes structures de base :

$$\mathbf{C}(\mathbf{h}) = \sum_{u=1}^S \mathbf{B}_u \rho_u(\mathbf{h})$$

avec $\mathbf{C}(\mathbf{h}) = [C_{ij}(\mathbf{h})]_{i,j}$: matrice des covariances simples et croisées

$\mathbf{B}_u = (b_{ij}^u)_{i,j}$: matrice complexe de corégionalisation

$\rho_u(\mathbf{h})$: fonction de covariance complexe.

Pour que le modèle soit autorisé, il suffit que toutes les matrices $\mathbf{B}_u$ soient hermitiennes, i.e. $\mathbf{B}_u = \overline{\mathbf{B}_u}^t$, et de type positif.

¹ La covariance entre deux fonctions aléatoires complexes Z_i et Z_j conjointement stationnaires est définie par :

$$C_{ij}(\mathbf{h}) = E[Z_i(\mathbf{x} + \mathbf{h}) \overline{Z_j(\mathbf{x})}] - E[Z_i(\mathbf{x} + \mathbf{h})] E[\overline{Z_j(\mathbf{x})}]$$

où la barre indique l'opération de conjugaison complexe.

On interprète ce modèle en décomposant les variables en un ensemble de facteurs spatialement orthogonaux (cf. chapitre 7 et annexe G) :

$$\mathbf{Z}(\mathbf{x}) - \mathbf{m} = \sum_{u=1}^S \mathbf{Y}_u(\mathbf{x}) \times \mathbf{A}_u$$

avec $\forall u \in [1, S], \mathbf{Y}_u(\mathbf{x}) = [Y_1^u(\mathbf{x}), \dots, Y_N^u(\mathbf{x})]$

$$\forall u \in [1, S], \forall p \in [1, N], \forall \mathbf{x}, E[Y_p^u(\mathbf{x})] = 0$$

$$\forall u, v \in [1, S], \forall p, q \in [1, N], \forall \mathbf{x}, \forall \mathbf{h}, \text{cov}[Y_p^u(\mathbf{x} + \mathbf{h}), Y_q^v(\mathbf{x})] = \begin{cases} 0 & \text{si } u \neq v \text{ ou } p \neq q \\ \rho_u(\mathbf{h}) & \text{sinon} \end{cases}$$

$$\forall u \in [1, S], \mathbf{B}_u = \mathbf{A}_u^t \overline{\mathbf{A}_u}.$$

En introduisant les parties réelles (indexées par ⁽¹⁾) et imaginaires (indexées par ⁽²⁾), et en notant i le complexe dont le carré vaut -1 , il vient :

$$\mathbf{Z}(\mathbf{x}) - \mathbf{m} = \sum_{u=1}^S \{ \mathbf{Y}_u^{(1)}(\mathbf{x}) \times \mathbf{A}_u^{(1)} - \mathbf{Y}_u^{(2)}(\mathbf{x}) \times \mathbf{A}_u^{(2)} \} + i \sum_{u=1}^S \{ \mathbf{Y}_u^{(1)}(\mathbf{x}) \times \mathbf{A}_u^{(2)} + \mathbf{Y}_u^{(2)}(\mathbf{x}) \times \mathbf{A}_u^{(1)} \}.$$

1.2. Le modèle bilinéaire de corégionalisation

Prenons la partie réelle de la relation précédente :

$$\mathbf{Z}^{(1)}(\mathbf{x}) - \mathbf{m}^{(1)} = \sum_{u=1}^S \{ \mathbf{Y}_u^{(1)}(\mathbf{x}) \times \mathbf{A}_u^{(1)} - \mathbf{Y}_u^{(2)}(\mathbf{x}) \times \mathbf{A}_u^{(2)} \}.$$

Ce modèle apparaît comme la différence de deux modèles linéaires, et a de ce fait été baptisé *modèle bilinéaire de corégionalisation*. Les seules covariances non nulles sont celles entre deux composantes de mêmes indices :

$$\forall p \in [1, N], \forall u \in [1, S], \forall \mathbf{x}, \mathbf{x} + \mathbf{h}, \begin{cases} \text{cov}[Y_p^{u(1)}(\mathbf{x} + \mathbf{h}), Y_p^{u(1)}(\mathbf{x})] = C_{11}^u(\mathbf{h}) \\ \text{cov}[Y_p^{u(1)}(\mathbf{x} + \mathbf{h}), Y_p^{u(2)}(\mathbf{x})] = C_{12}^u(\mathbf{h}) \\ \text{cov}[Y_p^{u(2)}(\mathbf{x} + \mathbf{h}), Y_p^{u(2)}(\mathbf{x})] = C_{22}^u(\mathbf{h}) \end{cases}$$

On a la relation suivante avec le modèle linéaire complexe de corégionalisation :

$$\forall u \in [1, S], \rho_u(\mathbf{h}) = [C_{11}^u(\mathbf{h}) + C_{22}^u(\mathbf{h})] + i[C_{12}^u(-\mathbf{h}) - C_{12}^u(\mathbf{h})].$$

Pour limiter le nombre de paramètres, on pose :

$$C_{11}^u(\mathbf{h}) = C_{22}^u(\mathbf{h}) = \rho_u^{(1)}(\mathbf{h}) / 2$$

$$C_{12}^u(-\mathbf{h}) = -C_{12}^u(\mathbf{h}) = \rho_u^{(2)}(\mathbf{h}) / 2$$

ce qui implique que $\rho_u^{(1)}(\mathbf{h})$ est une fonction paire, et $\rho_u^{(2)}(\mathbf{h})$ une fonction impaire.

Sous ces hypothèses, la matrice des covariances simples et croisées de $\mathbf{Z}^{(1)}(\mathbf{x})$ s'écrit :

$$\mathbf{C}(\mathbf{h}) = \sum_{u=1}^S [\mathbf{B}_u^{(1)} \rho_u^{(1)}(\mathbf{h}) - \mathbf{B}_u^{(2)} \rho_u^{(2)}(\mathbf{h})] / 2$$

avec $\mathbf{B}_u = \mathbf{B}_u^{(1)} + i \mathbf{B}_u^{(2)}$: matrice complexe de corégionalisation (hermitienne, de type positif)

$\rho_u(\mathbf{h}) = \rho_u^{(1)}(\mathbf{h}) + i \rho_u^{(2)}(\mathbf{h})$: fonction de covariance complexe.

Les conditions pour que le modèle soit autorisé sont finalement :

- $\mathbf{B}_u = \mathbf{B}_u^{(1)} + i \mathbf{B}_u^{(2)}$: matrice hermitienne de type positif
- $\rho_u(\mathbf{h}) = \rho_u^{(1)}(\mathbf{h}) + i \rho_u^{(2)}(\mathbf{h})$: covariance complexe
- $\rho_u^{(1)}(\mathbf{h})$ fonction paire et $\rho_u^{(2)}(\mathbf{h})$ fonction impaire.

Les covariances simples sont paires, car $\mathbf{B}_u^{(2)}$ est antisymétrique et ses termes diagonaux sont nuls. En revanche, les covariances croisées ne le sont pas nécessairement. Par conséquent, ce modèle est apte à décrire des corégionalisations où existent des effets de retard (covariance croisée atteignant son maximum pour un vecteur de séparation non nul). Les termes $\rho_u^{(1)}(\mathbf{h})$ modélisent la partie paire des covariances croisées, et $\rho_u^{(2)}(\mathbf{h})$ la partie impaire.

I.3. Exemples

- 1) Dans l'espace unidimensionnel (par exemple, l'axe temporel), considérons la covariance complexe périodique $\rho(h) = \cos(ah) + i \sin(ah)$. Le modèle bilinéaire associé s'écrit :

$$\mathbf{C}(h) = [\mathbf{B}_u^{(1)} \cos(ah) - \mathbf{B}_u^{(2)} \sin(ah)] / 2.$$

Les covariances simples et croisées sont des sinusoides ; elles permettent de décrire des variables temporelles périodiques, de même période, mais présentant des déphasages entre elles.

On peut sophistiquer ce modèle en introduisant un terme d'amortissement :

$$\rho(\mathbf{h}) = e^{-b|\mathbf{h}|^2} [\cos(a \cdot \mathbf{h}) + i \sin(a \cdot \mathbf{h})]$$

ce qui conduit à des covariances simples et croisées périodiques amorties :

$$\mathbf{C}(\mathbf{h}) = e^{-b|\mathbf{h}|^2} [\mathbf{B}_u^{(1)} \cos(a \cdot \mathbf{h}) - \mathbf{B}_u^{(2)} \sin(a \cdot \mathbf{h})] / 2 .$$

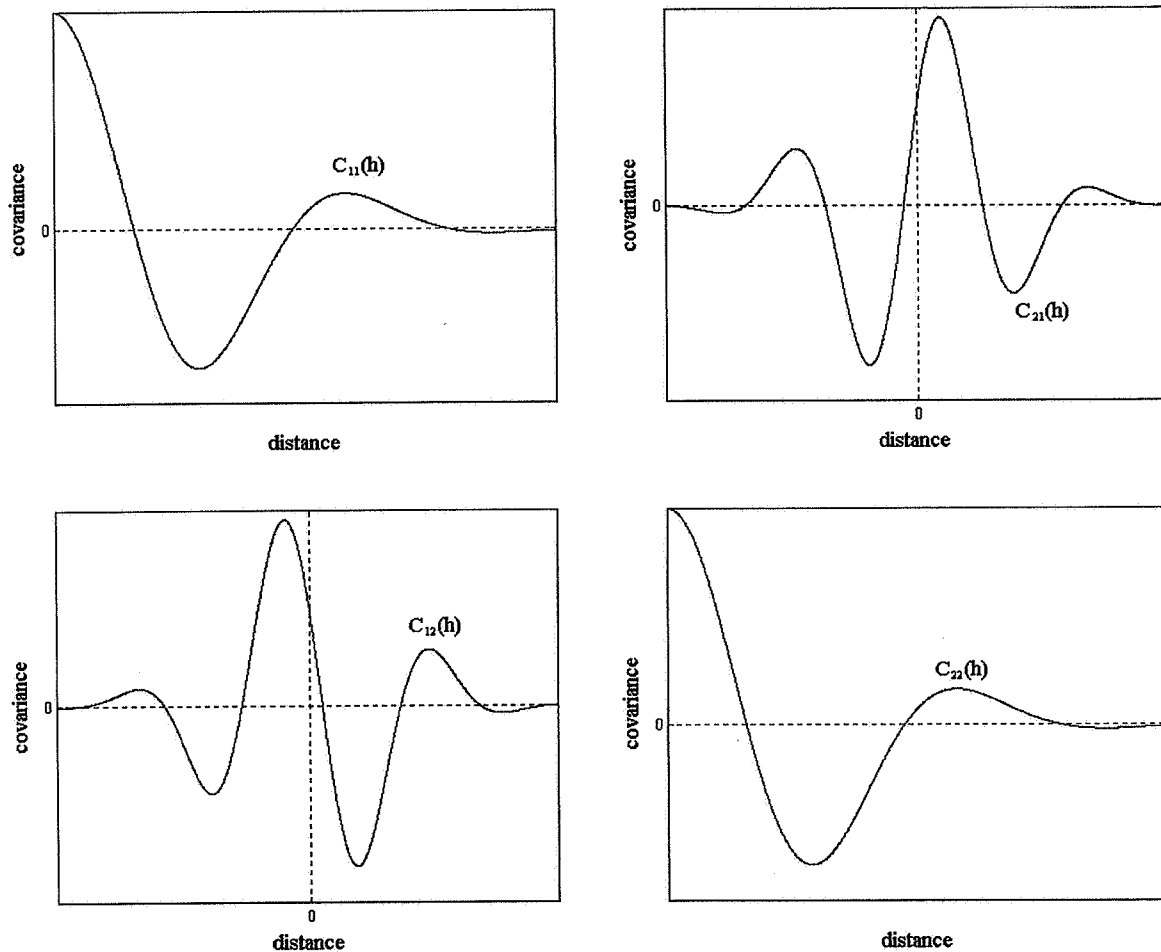


figure H.1 : modèle bilinéaire de corégionalisation de deux variables unidimensionnelles

2) Le modèle précédent se généralise à l'espace à plusieurs dimensions en posant :

$$\rho(\mathbf{h}) = e^{-b|\mathbf{h}|^2} [\cos(\langle \mathbf{a} | \mathbf{h} \rangle) + i \sin(\langle \mathbf{a} | \mathbf{h} \rangle)]$$

avec $\mathbf{a} \in \mathbb{R}^d$.

Le modèle bilinéaire associé s'écrit :

$$\mathbf{C}(\mathbf{h}) = e^{-b|\mathbf{h}|^2} [\mathbf{B}_u^{(1)} \cos(\langle \mathbf{a} | \mathbf{h} \rangle) - \mathbf{B}_u^{(2)} \sin(\langle \mathbf{a} | \mathbf{h} \rangle)] / 2 .$$

II. Fonction aléatoire et ses dérivées partielles

Une fonction aléatoire Z définie sur $\mathbb{R}^d$ est différentiable en moyenne quadratique au point $\mathbf{x}$ si ses dérivées partielles en ce point existent, i.e. s'il existe un vecteur aléatoire $\{Z_i(\mathbf{x}), i = 1 \dots d\}$ tel que :

$$\lim_{|\mathbf{h}| \rightarrow 0} E \left\{ [Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x}) - \sum_{i=1}^d h_i Z_i(\mathbf{x})]^2 / |\mathbf{h}|^2 \right\} = 0.$$

Le vecteur aléatoire $\{Z_i(\mathbf{x}), i = 1 \dots d\}$ est appelé **gradient** de $Z(\mathbf{x})$ au point $\mathbf{x}$. La fonction aléatoire est dérivable en moyenne quadratique si elle est différentiable en tout point².

Soit une fonction aléatoire $Z(\mathbf{x})$ **stationnaire d'ordre deux**. On montre que $Z(\mathbf{x})$ est dérivable en moyenne quadratique si sa covariance $C(\mathbf{h})$ est deux fois différentiable en $\mathbf{h} = \mathbf{0}$, auquel cas elle l'est en tout $\mathbf{h}$. La dérivée partielle $Z_i(\mathbf{x})$ par rapport à la i -ème coordonnée

$$Z_i(\mathbf{x}) = \frac{\partial Z(\mathbf{x})}{\partial x_i}$$

est alors une fonction aléatoire stationnaire d'ordre deux.

En écrivant la dérivée comme un accroissement infinitésimal, on peut expliciter les covariances du modèle à partir de celle de $Z(\mathbf{x})$. La covariance croisée entre $Z(\mathbf{x})$ et $Z_i(\mathbf{x})$ est impaire (donc nulle à l'origine) et égale à :

$$\text{cov}[Z_i(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})] = -\text{cov}[Z(\mathbf{x} + \mathbf{h}), Z_i(\mathbf{x})] = \frac{\partial C(\mathbf{h})}{\partial h_i}$$

et la covariance simple de $Z_i(\mathbf{x})$ vaut : $\text{cov}[Z_i(\mathbf{x} + \mathbf{h}), Z_i(\mathbf{x})] = -\frac{\partial^2 C(\mathbf{h})}{\partial h_i^2}$.

Cette situation échappe au modèle linéaire de corégionalisation, car la covariance croisée est une fonction impaire. On peut également expliciter la covariance croisée entre deux dérivées partielles $Z_i(\mathbf{x})$ et $Z_j(\mathbf{x})$, qui est une fonction paire :

$$\text{cov}[Z_i(\mathbf{x} + \mathbf{h}), Z_j(\mathbf{x})] = -\frac{\partial^2 C(\mathbf{h})}{\partial h_i \partial h_j}.$$

² En toute rigueur, cette définition ne garantit pas que les *réalisations* de $Z(\mathbf{x})$ soient dérivables au sens usuel des fonctions : en fait, cette dernière propriété ne peut être établie, même si l'on connaît la loi spatiale entière de la fonction aléatoire. Si toutefois, $Z(\mathbf{x})$ est telle que toutes ses réalisations sont des fonctions dérivables (on dira qu'elle est presque sûrement différentiable), et si elle est dérivable en moyenne quadratique, alors les dérivées partielles, au sens de la différentiabilité presque sûre et en moyenne quadratique, coïncident : la définition donnée au niveau de la fonction aléatoire rejoint donc celle que l'on attend en pratique, au niveau des réalisations.

Le modèle multivariable dérive entièrement de la covariance de $Z(\mathbf{x})$, laquelle peut être ajustée grâce au modèle linéaire de régionalisation. Le fait que $C(\mathbf{h})$ doive être deux fois dérivable est restrictif, et interdit *a priori* l'usage de schémas linéaires à l'origine, qui sont les modèles les plus fréquemment employés. En fait, on peut contourner cette difficulté en supposant que la dérivation n'est pas exactement celle des valeurs ponctuelles de $Z(\mathbf{x})$, mais celle de valeurs moyennes sur un petit volume ; la covariance de $Z(\mathbf{x})$ est alors régularisée au voisinage de l'origine et devient deux fois dérivable. Cette approche revient à considérer la dérivation non pas au sens des fonctions, mais au sens des distributions.

Ces résultats se généralisent au cas d'une fonction aléatoire $Z(\mathbf{x})$ **intrinsèque d'ordre k** , en remplaçant la covariance usuelle par une covariance généralisée ; les $Z_i(\mathbf{x})$ sont alors des fonctions aléatoires intrinsèques d'ordre $k - 1$. Ce cadre est beaucoup plus flexible, car il permet de traiter des fonctions aléatoires non stationnaires.

Citons deux applications de ce modèle :

- estimation directe³ des composantes du gradient à partir de mesures concernant $Z(\mathbf{x})$;
- si de surcroît l'on dispose de mesures concernant le gradient, le modèle variographique multivariable permet d'estimer par *cokrigage* à la fois les valeurs de la variable et de ses dérivées partielles. Par exemple, en hydrogéologie, on cherche à estimer un champ de transmissivité à partir de mesures concernant la charge hydraulique et la transmissivité. Dans ce cas, il est préférable de réaliser un ajustement *simultané* des covariances expérimentales de la variable et de ses dérivées, afin de ne pas privilégier la première par rapport aux dernières.

Dans certains domaines d'application (géothermie, météorologie, hydrogéologie, etc.), les variables étudiées sont reliées par des équations différentielles ou aux dérivées partielles. Avant de mettre en œuvre les techniques géostatistiques d'estimation, il est indispensable de s'assurer que le modèle variographique multivariable est compatible avec ces équations. Le paragraphe suivant considère une équation particulière, l'équation de Laplace, qui intervient notamment en hydrogéologie.

³ On peut également utiliser la propriété d'additivité du krigage : le krigage de la i -ème dérivée partielle de Z est égal à la i -ème dérivée partielle du krigage de Z . Toutefois, cette approche, à première vue plus simple, nécessite d'évaluer les dérivées partielles à l'aide de différences finies : le résultat dépendra de la discrétisation choisie. Au contraire, l'évaluation directe des dérivées partielles est plus précise, car indépendante de la maille de la grille sur laquelle elle est effectuée.

III. Fonction aléatoire et son laplacien

On considère un couple de fonctions aléatoires (Y, Z) reliées par l'équation de Laplace :

$$Y(\mathbf{x}) = \Delta Z(\mathbf{x}).$$

Le formalisme des fonctions aléatoires stationnaires d'ordre deux est insuffisant pour étudier cette équation, car il restreint considérablement l'ensemble des solutions. En effet, si $Y(\mathbf{x})$ est stationnaire, $Z(\mathbf{x})$ ne le sera pas nécessairement et pourra présenter une "dérive" linéaire ou quadratique (annulée par l'opérateur laplacien). Aussi est-il préférable de se placer dans le cadre plus général des fonctions aléatoires intrinsèques d'ordre k .

Si Y est une fonction aléatoire intrinsèque d'ordre k , on démontre qu'il existe une unique fonction aléatoire Z intrinsèque d'ordre $k+2$, deux fois dérivable, telle que $Y = \Delta Z$. L'unicité de la solution vient du fait qu'une fonction aléatoire intrinsèque d'ordre k est indéterminée à un polynôme de degré k près et contient en fait une *famille* de fonctions aléatoires.

Notons $K_Y(\mathbf{h})$ une covariance généralisée de Y , $K_Z(\mathbf{h})$ une covariance de Z , et $K_{YZ}(\mathbf{h})$ et $K_{ZY}(\mathbf{h})$ des covariances croisées de Y et Z . Ces covariances sont reliées par les formules :

$$\begin{cases} K_Y(\mathbf{h}) = \Delta^2 K_Z(\mathbf{h}) \\ K_{YZ}(\mathbf{h}) = K_{ZY}(\mathbf{h}) = \Delta K_Z(\mathbf{h}) \end{cases}$$

L'explicitation de K_Z et K_{YZ} à partir de K_Y se résout grâce aux propriétés spectrales des fonctions de covariance.

Exemples

- Dans $\mathbb{R}^3$, si K_Y est une covariance généralisée linéaire de pente ω , alors :

$$\begin{cases} K_Y(\mathbf{h}) = -\omega |\mathbf{h}| \\ K_{YZ}(\mathbf{h}) = -\omega |\mathbf{h}|^3 / 12 \\ K_Z(\mathbf{h}) = -\omega |\mathbf{h}|^5 / 360 \end{cases}$$

- Dans $\mathbb{R}^3$, si K_Y est une covariance exponentielle, de palier C et portée pratique $3a$, alors :

$$\begin{cases} K_Y(\mathbf{h}) = C e^{-|\mathbf{h}|/a} \\ K_{YZ}(\mathbf{h}) = C a^2 (1 + e^{-|\mathbf{h}|/a} + 2 a e^{-|\mathbf{h}|/a} / |\mathbf{h}| - 2a / |\mathbf{h}|) \\ K_Z(\mathbf{h}) = C a^4 (3 - |\mathbf{h}|/a + |\mathbf{h}|^2 / 6a^2 + e^{-|\mathbf{h}|/a} + 4 a e^{-|\mathbf{h}|/a} / |\mathbf{h}| - 4a / |\mathbf{h}|) \end{cases}$$

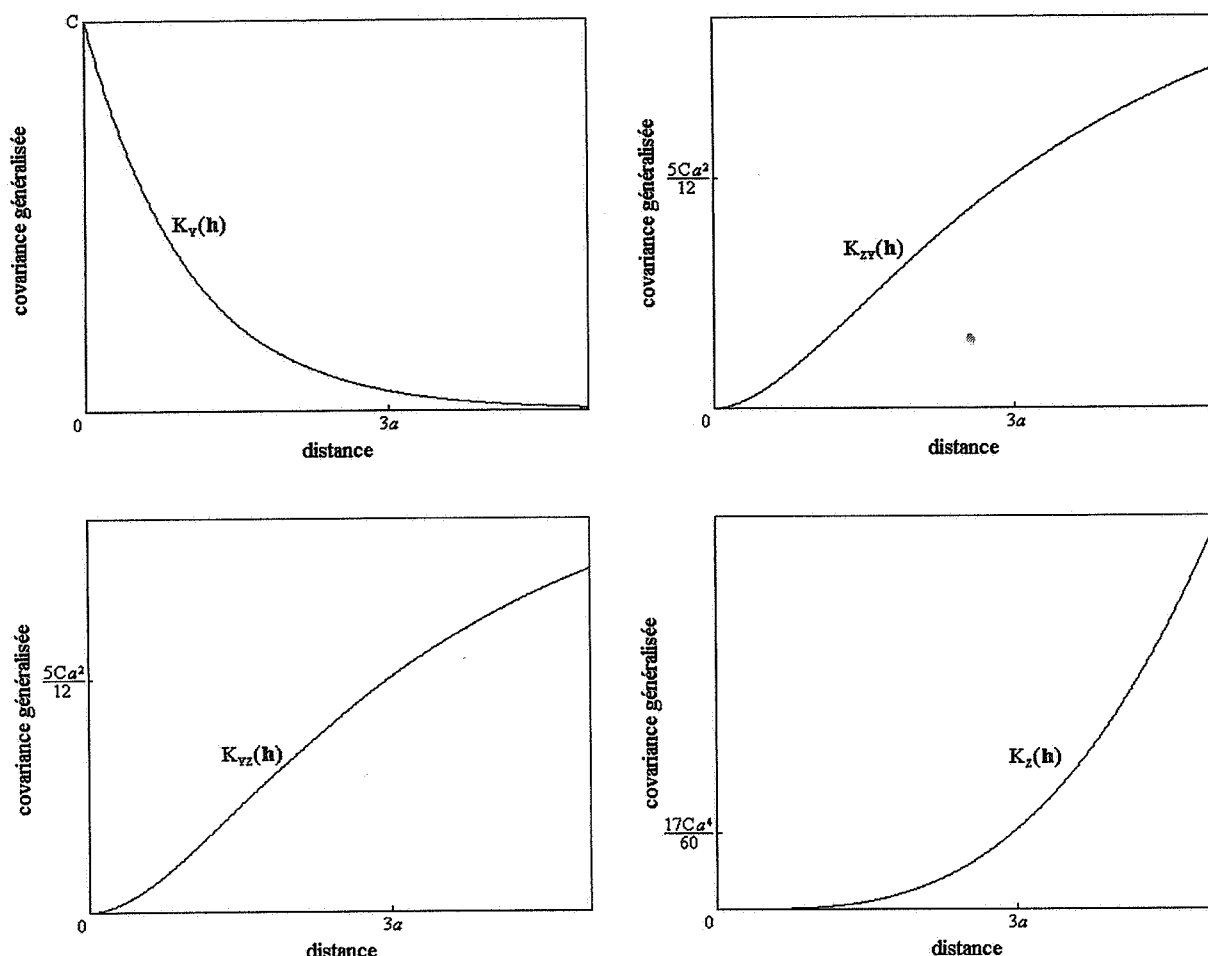


figure H.2 : exemple de modèle variographique d'une variable et de son laplacien

On détermine ainsi des triplets de covariances (K_Y , K_{YZ} , K_Z) qui sont théoriquement admissibles. Pour la modélisation, il est conseillé de réaliser l'ajustement simultané des covariances expérimentales des deux variables.

Une fois spécifié le modèle variographique bivariable, on peut résoudre les problèmes d'estimation par cokrigage intrinsèque, à partir des données disponibles de Y et Z . Le cokrigage fournit un estimateur précis et sans biais des deux variables Y et Z . En outre, par linéarité des systèmes d'équations, les estimations obtenues vérifient l'équation de Laplace ; on peut également ajouter des conditions aux limites linéaires⁴ (e.g. Dirichlet, Neumann, surfaces libres, conditions de périodicité).

⁴ Il suffit pour cela d'ajouter des données correspondant à ces conditions aux limites. Si cette procédure introduit de nouvelles variables (par exemple, les conditions de Neumann concernent le flux de Z et introduisent la dérivée de Z dans une direction), il est nécessaire de réaliser un modèle variographique multivariable incorporant ces variables. L'approche par cokrigage a l'avantage d'être flexible, car elle ne requiert pas que les conditions aux limites soient connues sur toute la frontière du domaine, ni même qu'elles soient situées sur la frontière : en fait, il n'y a pas de restriction sur le nombre ou la localisation des conditions aux limites.

INDEX
ET
BIBLIOGRAPHIE

Index des mots clés

A

abondance 19, 26, 162
 additive, variable..... 14, 60, 214
 additivité..... 190, 214, 234, 251, 283
 analyse en composantes principales
 63, 261, 365, 380, 388
 analyse krigéante..... 269, 294, 375
 analyse structurale, *voir* analyse
 variographique
 analyse variographique
 41, 99, 223, 240, 252, 268, 305, 309, 389
 anisotropie 16, 23, 32, 43, 79, 103, 117,
 123, 128, 176, 187, 212, 254
 anisotropie géométrique 23, 103
 anisotropie zonale 104, 129
 autokrigeabilité 292, 305, 388

B

biais conditionnel.....
 111, 175, 192, 195, 197, 198, 220, 221

C

caractéristique intrinsèque 347
 carte variographique.....
 31, 103, 108, 124, 128, 274
 champ 14, 20, 261
 champ, problèmes de délimitation
 19, 113, 162, 166
 changement de support..... 14, 140
 coefficient de codispersion..... 381, 384
 coefficient de corrélation.... 53, 112, 216, 274,
 285, 286, 288-291, 318, 382

coefficient de corrélation régionalisée..... 384
 cokrigeage 278
 cokrigeage intrinsèque 306
 cokrigeage ordinaire 278, 283
 cokrigeage simple..... 282
 cokrigeage universel..... 304
 combinaisons linéaires autorisées
 60, 78, 173, 185, 243, 345
 comportement à l'origine..... 23, 43, 76, 81,
 87, 110, 121, 125, 141, 142, 206, 246
 composantes spatiales..... 299, 376, 384, 386
 composite 67, 102
 condition d'universalité 181
 configuration de krigéage 40, 173, 284
 corégionalisation 261
 corrélation intrinsèque 293, 296, 379
 corrélogramme 53, 55
 covariance 53, 55, 60, 75, 140, 318
 covariance croisée 262, 266
 covariance expérimentale..... 84
 covariance généralisée 246, 253, 353
 covariance non centrée 86, 167
 covariogramme expérimental..... 29, 34, 41
 covariogramme géométrique.. 24, 91, 102, 167
 covariogramme transitif.....
 21, 31, 53, 76, 143, 167
 croix de sondage..... 125

D

dérive 57, 62, 87, 228, 229,
 237, 335, 347, 348, 363
 dérive aléatoire 339
 dérive externe 228, 229, 307
 dichotomie..... 228, 347, 359
 donnée robuste 111

E

écart-type53
 échantillonnage aléatoire stratifié
 35, 41, 156, 160
 échantillonnage aléatoire uniforme
 36, 154, 160
 échantillonnage irrégulier
 36, 39, 41, 145, 152, 187
 échantillonnage préférentiel.....
 38, 69, 107, 132, 145, 152
 échantillonnage régulier ...26, 33, 41, 157, 160
 échelle de travail
 15, 56, 58, 89, 100, 101,
 110, 120, 207, 227, 229
 effet de bord..... 19, 114, 162, 167
 effet d'écran.....187
 effet d'écran inverse..... 188, 214
 effet de pépite
 23, 59, 81, 88, 101, 121, 126, 129,
 143, 147, 180, 185, 187, 194, 206,
 210, 224, 253, 294, 300, 324, 361
 effet de relais..... 188, 214
 effet de retard..... 263, 267, 389, 391
 effet de support14, 67, 140, 216, 221
 effet de trou.....76, 87, 94, 150, 211
 effet d'information219, 221
 effet proportionnel.....
 59, 71, 72, 73, 130, 132, 203
 ergodicité54, 76, 167, 262, 323
 erreurs de localisation.....222
 erreurs de mesure81, 88, 101, 302, 359
 erreur standardisée 111, 127, 131
 espérance mathématique.....
52, 60, 177, 181, 183,
 230, 315, 318, 323, 324
 estimation16
 estimation globale16, 26, 33, 139, 152
 estimation locale16, 40, 171, 227, 278
 estimation par plus proche voisin197
 étude exploratoire.....62
 évaluation de la dérive.....237, 336
 évaluation de la moyenne 183, 378
 évaluation par moyenne mobile.....238
 exponentielles – polynômes.....
 229, 243, 304, 342

F

facteurs spatiaux
294, 297, 298, 380, 382, 386, 387
 famille complète..... 243
 filtrage..... 359
 filtrage de composantes 376
 fluctuation 115, 332
 fonction aléatoire..... 48, 320
 fonction aléatoire intrinsèque..... 57, 77
 fonction aléatoire intrinsèque d'ordre k.....
345, 359, 394
 fonction aléatoire intrinsèque généralisée.....
 355, 364
 fonction aléatoire stationnaire d'ordre deux
 55, 75
 fonction de type négatif conditionnel ... 78, 87
 fonction de type positif22, 43, 76, 77
 fonction de type positif conditionnel. 246, 354
 fonctions de base 229, 243, 304, 335
 formule de Krige151, 156, 331

G

géostatistique intrinsèque,
 voir géostatistique probabiliste
 géostatistique multivariable 261
 géostatistique probabiliste.....47, 163, 167
 géostatistique transitive.....19, 162, 167
 grandeur régionale
15, 53, 152, 167, 183, 300, 325
 graphique quantiles contre quantiles 70

H

hétérotopie.....261, 278, 287, 383
 histogramme.....
 31, 33, 62, 64, 69, 114, 119, 195, 201
 homotopie261, 281, 293, 383
 hypothèse intrinsèque 57, 77, 185, 232
 hypothèse intrinsèque conjointe 265
 hypothèse intrinsèque d'ordre k 245, 345
 hypothèse quasi-intrinsèque..... 58

I

indépendance linéaire sur les données
 235, 251, 337
 inégalité de Cauchy-Schwarz
 21, 53, 75, 140, 263, 266, 318
 inférence statistique..... 54, 115, 253
 interpolateur.....356
 interpolation exacte.... 189, 234, 251, 283, 357
 intervalle de confiance33, 171
 invariance tensorielle341
 inverse des distances 189, 199
 isotropie, *voir* anisotropie

K

krigeage171
 krigeage aléatoire222
 krigeage factoriel, *voir* analyse krigeante
 krigeage intrinsèque d'ordre k... 243, 250, 256
 krigeage ordinaire 40, 181, 185, 193, 199, 202
 krigeage par bloc..... 214, 236, 251
 krigeage simple..... 177, 193, 204, 213
 krigeage transitif40, 228
 krigeage trigonométrique356
 krigeage universel 228, 255, 335

L

largeur de bande..... 83, 109, 129
 lissage.191, 195, 197, 201, 217, 234, 251, 357
 lognormalité.....70, 71, 85, 135, 315
 log-translation.....85
 loi spatiale.....51, 75, 320

M

madogramme135
 matrice de corégionalisation
 269, 271, 297, 301, 384, 385, 388
 matrice de type positif.....266, 270
 médiane64, 152
 méthode des moments84, 268

modèle de transition..... 79, 88
 modèle linéaire de corégionalisation
269, 277, 297, 375, 384, 389
 modèle linéaire de régionalisation..77, 99, 375
 modèles de variogrammes..... 87
 modélisation..... 29, 39, 43, 109, 115,
 126, 129, 253, 271, 276
 moindres carrés 237, 359
 moment 51, 60, 75, 315, 321, 323
 mouvement brownien 78, 97

N

non-biais.....28, 35, 37, 38, 112,
 131, 153, 173, 238, 279
 nuage de corrélation
 62, 111, 127, 131, 196, 218, 274, 319
 nuage de corrélation différée..... 62, 66, 72
 nuée variographique 107, 108, 122

O

objectivité..... 15, 19, 153, 198, 300, 325
 optimalité 174

P

palier 79, 87, 110, 168, 208
 phénomène régionalisé 13
 pondérateurs négatifs..... 188
 pondération des données
43, 68, 85, 107, 137, 145, 152
 portée23, 43, 76, 79, 89, 141, 175, 207
 portée intégrale..... 143, 326
 présentation duale du krigeage..... 356
 principe de composition des termes de
 ligne, de section et de tranche 157
 principe de composition des variances
 élémentaires..... 158
 principe d'économie15, 51, 229
 principe de réalisme..... 15
 propriété d'orthogonalité190, 234, 251
 pseudo-variogramme croisé 267

Q

quartiles 64, 107
quasi-stationnarité .. 58, 71, 114, 175, 193, 351

R

reconstruction opératoire 15, 50, 153
régionalisation, *voir* variable régionalisée
régularisation 14, 98, 102, 139, 140, 172
régularité du système de cokrigage .. 283, 305
régularité du système de krigeage
..... 189, 235, 251
relation d'additivité, *voir* formule de Krige
représentation 346, 349
résidu 228, 304, 335

S

semi-variogramme, *voir* variogramme
splines 358, 359
splines de lissage 239
stationnarité 54, 227, 304, 323
stationnarité de second ordre
..... 55, 87, 114, 167, 177, 181, 230
stationnarité de second ordre conjointe 262
stationnarité locale, *voir* quasi-stationnarité
stationnarité stricte 54
statistique classique 47, 155, 160, 163
statistique descriptive 161, 163, 365
structure 48, 49, 51, 75, 187
structures gigognes 24, 99, 296, 375
support
..... 14, 67, 98, 102, 140, 154, 160, 163, 214
surfaces d'influence 39, 42

T

taux de sondage 164
teneur de coupure 217
théorème d'additivité 192, 213, 342
théorème de Bochner 77, 264, 355
tolérances 42, 83, 108, 128

V

valeur aberrante 15, 62, 64, 107, 152, 252
validation croisée 111, 127, 130, 175,
..... 200, 252, 259, 288, 292
variable aléatoire
..... 27, 35, 38, 39, 41, 47, 49, 52, 313
variable régionalisée 13
variables standardisées 301, 370, 381, 388
variance (*a priori*)
..... 52, 55, 56, 60, 75, 78, 79,
..... 86, 140, 147, 156, 180, 301, 316
variance de cokrigage 280, 282, 305, 306
variance de dispersion .. 56, 146, 155, 156, 331
variance de krigeage 171, 178, 182, 183,
..... 186, 232, 234, 236, 250
variance d'estimation globale
..... 28, 29, 33, 36, 37, 38, 139, 145,
..... 153, 155, 156, 158, 160, 161, 163
variance d'extension 144, 150, 158
variogramme
..... 53, 55, 57, 62, 77, 141, 144, 332
variogramme croisé 265
variogramme d'ordre 1, *voir* madogramme
variogramme expérimental
..... 83, 107, 118, 122, 129, 268, 275, 332
variogramme local 132
variogramme logarithmique 135
variogramme pondéré 85, 107, 137
variogramme régional 115, 332
variogramme régularisée 98, 141
variogrammes relatifs 85, 107, 133
variographie, *voir* analyse variographique
voisinage de krigeage
..... 110, 174, 192, 193, 199, 202, 220, 286
voisinage glissant 175
voisinage unique 174

Z

zone d'influence 86

Bibliographie

La bibliographie est divisée en deux parties : la première contient des ouvrages de portée générale sur la géostatistique linéaire, la seconde des ouvrages plus spécialisés. La plupart des références proviennent du Centre de Géostatistique de l'Ecole des Mines de Paris¹.

Ouvrages généraux

- [1] Arnaud M. et Emery X., *Estimation et interpolation spatiale, méthodes déterministes et méthodes géostatistiques*, Hermès Science Publications, Paris, 2000, 221 p.
- [2] Chauvet P., *Aide-mémoire de géostatistique linéaire*, Presses de l'Ecole des Mines de Paris, 1999, 370 p.
- [3] Chilès J.P. et Delfiner P., *Geostatistics: Modeling Spatial Uncertainty*, Wiley, New York, 1999, 696 p.
- [4] Cressie N. A. C., *Statistics for spatial data*, John Wiley & Sons, Inc. New York, 1993.
- [5] Delfiner P., *Basic introduction to geostatistics*, Centre de géostatistique, Ecole des Mines de Paris, 1979, 125 p [C-78].
- [6] Isaaks E. H. et Srivastava R. M., *An Introduction to applied geostatistics*, Oxford University Press, New York, 1989, 561 p.
- [7] Journel A. G. et Huijbregts C. J., *Mining geostatistics*, Academic Press, London, 1978, 600 p.
- [8] Matheron G., *Les variables régionalisées et leur estimation*, Une application de la théorie des fonctions aléatoires aux sciences de la nature, Masson, Paris, 1965, 305 p.
- [9] Matheron G., *La théorie des variables régionalisées et ses applications*, Les cahiers du centre de morphologie mathématique de Fontainebleau, Fasc. 5, Ecole des Mines de Paris, 1970, 212 p.

¹ Ecole des Mines de Paris, Centre de Géostatistique, 35, rue Saint Honoré, 77 305 Fontainebleau Cedex, France

- [10] Matheron G., *The theory of regionalized variables and its applications*, Les cahiers du centre de morphologie mathématique de Fontainebleau, Fasc. 5, Ecole des Mines de Paris, 1971, 212p.
- [11] Matheron G., *Estimer et choisir*, Les cahiers du centre de morphologie mathématique de Fontainebleau, Fascicule 7, Ecole des Mines de Paris, 1978, 175 p.
- [12] Matheron G., *Estimating and choosing*, Springer-Verlag, Berlin, 1989.
- [13] Rivoirard J., *Concepts et méthodes de la géostatistique*, Centre de géostatistique, Ecole des Mines de Paris, 1995, 95 p. [C-158].
- [14] Wackernagel H., *Cours de géostatistique multivariable*, Centre de géostatistique, Ecole des Mines de Paris, 1993, 80 p. [C-146].
- [15] Wackernagel H., *Multivariate geostatistics : an introduction with applications*, Springer-Verlag, Berlin, 1998, 291 p.

Ouvrages spécialisés

Géostatistique transitive

- [16] Bez N, *Statistiques individuelles et géostatistique transitive en écologie halieutique*, Thèse de docteur en géostatistique, Centre de géostatistique, Ecole des Mines de Paris, 1997, 276 p.

Krigeage

- [17] Rivoirard J., *Le comportement des poids de krigeage*, Thèse de docteur-ingénieur en sciences et techniques minières, Centre de géostatistique, Ecole des Mines de Paris, 1984, 72 p. [L-803].
- [18] Rivoirard J., *Two key parameters when choosing the kriging neighborhood*, in *Mathematical Geology*, Vol. 19, No. 8, 1987, p.851-858.

Modèles non stationnaires

- [19] Chilès J.-P., *Géostatistique des phénomènes non stationnaires (dans le plan)*, Thèse de docteur-ingénieur, Université de Nancy I, 1977, 152 p.
- [20] Delfiner P., *The intrinsic model of order k*, Ecole d'été 1982, Centre de géostatistique et de morphologie mathématique de Fontainebleau, Ecole des Mines de Paris, 1982, 157 p. [C-97].

- [21] Delfiner P. et Matheron G., *Les fonctions aléatoires intrinsèques d'ordre k*, Centre de géostatistique et de morphologie mathématique, Fontainebleau, Ecole des Mines de Paris, 1980, 36 p [C-84].
- [22] Dubrule O., *Krigeage et splines en cartographie automatique, Application à des exemples pétroliers*, Thèse de docteur-ingénieur, Centre de géostatistique, Ecole des Mines de Paris, 1981, 150 p.
- [23] Matheron G., *La théorie des fonctions aléatoires intrinsèques généralisées*, Centre de géostatistique, Ecole des Mines de Paris, note géostatistique n°117, 1971, 64 p.

Géostatistique multivariable

- [24] Goulard M. et Voltz M., *Linear coregionalization model: tools for estimation and choice of cross-variogram matrix*, Mathematical Geology, Vol. 24, 1992, p. 269-286.
- [25] Matheron G., *Pour une analyse krigeante des données régionalisées*, Centre de géostatistique et de morphologie mathématique, Ecole des Mines de Paris, 1982, 22 p. [N-732].
- [26] Sandjivy L., *Analyse krigeante des données de prospection géochimique*, Thèse de docteur-ingénieur en sciences et techniques minières, Centre de géostatistique, Ecole des Mines de Paris, 1987, 164 p.

Autres

- [27] Alfaro Sironvalle M., *Etude de la robustesse des simulations de fonctions aléatoires*, Thèse de docteur-ingénieur en sciences et techniques minières, Centre de géostatistique, Ecole des Mines de Paris, 1979, 161 p.
- [28] Armstrong M. et Delfiner P., *Towards a more robust variogram : a case study on coal*, Centre de géostatistique, Ecole des Mines de Paris, 1980, 38 p [N-671].
- [29] Dong A., *Estimation géostatistique des phénomènes régis par des équations aux dérivées partielles*, Thèse de doctorat, Centre de géostatistique, Ecole des Mines de Paris, 1990, 262 p.
- [30] Lantuéjoul C., *Cours de sélectivité*, Centre de géostatistique, Ecole des Mines de Paris, 1990, 72 p. [C-140].
- [31] Matheron G., *Effet proportionnel et lognormalité ou : le retour du serpent de mer*, Centre de géostatistique, Ecole des Mines de Paris, note n°124, 1974, 43 p.
- [32] Séguret S., *Géostatistique des phénomènes à tendance périodique (dans l'espace-temps)*, Thèse de doctorat, Centre de géostatistique, Ecole des Mines de Paris, 1991, 254 p.