



Modélisation géologique 3D et caractérisation des incertitudes par la méthode du champ de potentiel

Christophe Aug

► To cite this version:

Christophe Aug. Modélisation géologique 3D et caractérisation des incertitudes par la méthode du champ de potentiel. Sciences de la Terre. École Nationale Supérieure des Mines de Paris, 2004. Français. NNT : 2004ENMP1254 . pastel-00001077

HAL Id: pastel-00001077

<https://pastel.hal.science/pastel-00001077>

Submitted on 6 Aug 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THESE

pour obtenir le grade de

Docteur de l'Ecole des Mines de Paris

Spécialité « **Géostatistique** »

présentée et soutenue publiquement

par

Christophe Aug

le 13 décembre 2004

MODELISATION GEOLOGIQUE 3D ET CARACTERISATION DES INCERTITUDES PAR LA METHODE DU CHAMP DE POTENTIEL

Directeur de thèse : Jean-Paul Chilès

Jury

MM. Jean-Pierre	Brun	Président
Jean-Laurent	Mallet	Rapporteur
Aurèle	Parriaux	Rapporteur
Jean-Paul	Chilès	Examineur
Gabriel	Courrioux	Examineur
Christian	Lajaunie	Examineur

À mes parents

«[...] l'objet visé [...] n'est pas une somme d'éléments qu'il faudrait d'abord isoler et analyser, mais un ensemble, c'est-à-dire une forme, une structure : l'élément ne préexiste pas à l'ensemble, il n'est ni plus immédiat ni plus ancien, ce ne sont pas les éléments qui déterminent l'ensemble, mais l'ensemble qui détermine les éléments : la connaissance du tout et de ses lois, de l'ensemble et de sa structure, ne saurait être déduite de la connaissance séparée des parties qui le composent [...]»

Georges Perec (La Vie mode d'emploi)

«Setting out all the sample bottles on a long table, Dr. O'Geery provides the camera with a summation of his findings, casually gesturing to various groupings while he sips coffee from a Garfield mug. "What we have here is a nice banquet of igneous, sedimentary, and metamorphic sample [...]"»

Mark Z. Danielewski (House of Leaves)

«Ainsi, même en moi, quelque chose a traversé ces saisons sans croître ni s'abolir, l'alluvionnement des heures a réservé certains espaces-témoins, et tandis que je déambule, cherchant la raison de moi-même, dans ce terrain vague que je suis devenu, tâtonnant sur d'énormes masses de dépôt, tout d'un coup je trébuche au bord d'une faille au fond de laquelle le sol d'antan est resté nu, mesurant alors l'épaisseur de cette matière qu'il faut que je sonde et tamise, afin de retrouver des assises et des fondations.»

Michel Butor (L'emploi du temps)

«Je sais que je ne sais pas ce que je ne sais pas ; j'envie ceux qui sauront davantage, mais je sais qu'ils auront tout comme moi à mesurer, peser, déduire et se méfier des déductions produites, faire dans le faux la part du vrai et tenir compte dans le vrai de l'éternelle admixtion du faux.»

Marguerite Yourcenar (L'OEuvre au Noir)

Et voici venu le temps des remerciements...

Je vais tâcher d'être bref, efficace et de n'oublier personne.

Je tiens tout d'abord à exprimer ma plus profonde gratitude à Jean-Paul Chilès qui m'a confié ce sujet et m'a fait confiance tout au long de ces trois années. Il a su, par un subtil mélange, me laisser travailler de façon autonome tout en faisant régulièrement le point avec moi. Je le remercie aussi pour sa disponibilité et son partage de ses connaissances, très larges, de notre domaine d'étude.

Mes remerciements vont également aux autres membres du jury, en particulier aux rapporteurs, Messieurs Mallet et Parriaux, qui ont volontiers accepté de juger mon mémoire, messieurs Courrioux et Lajaunie (que je reciterai) ainsi qu'à Monsieur Brun qui a assuré la présidence de ce jury.

Un grand merci à tous les membres de l'équipe CDG du BRGM qui m'ont très sympathiquement accueilli en leur sein lors de mes visites à Orléans ou lors de mes appels à l'aide pour l'"Editeur". Je pense en particulier à Antonio, Philippe, Bernard, Guillaume, Elisabeth. Enfin, ce travail ne serait rien sans le soutien régulier de Gaby, qu'il soit ici très chaleureusement remercié.

Sans leur aide une grande partie de ces travaux n'aurait pas été possible : François "Fifi" Geffroy et Didier "Goupil" Renard pour toutes les questions informatiques, les secrétaires Nathalie et Isabelle, ainsi que Françoise à la bibliothèque du CG.

Je remercie aussi les autres chercheurs du Centre de Géostatistique : Hélène, Jacques, Hans, Chantal, Gaëlle, Pierre, Jean-Philippe, Nicolas, Serge, et plus spécialement les deux Christian pour leurs aides ponctuelles : "Cachou" et "Lantu".

Une petite pensée, à tous les thésards de géostat, passés (... , Laurent, David, Julien, Simon, Eric, Marta, Sarah, Xavier) et actuels (futurs ?) : Caroline ma collègue de bureau, Mireille, Pierre, Mathieu, Martial, Marie... Sans oublier mes collègues des autres centres, surtout ceux du côté obscurs (CMM) : Nicolas, Thomas, Romain, Gab, Costin, Raffi, Mathilde... et de ma famille (geophy) Alex, François et leurs prédécesseurs.

Un coucou à l'ensemble de mes amis de l'IST (spéciale dédicace pour Dr. Fred) et de mes études antérieures, ainsi qu'aux demeures de HOL.com.

Un merci aussi à tous ceux qui ont crû en moi, qui m'ont encouragé au premier rang desquels mes parents, Mamy alias Petie, ma soeur Helyette et son mari (et un bisou pour le petit Aurélien), mes "beaux-parents". Un gros bisou pour Judith : je t'aime.

Table des matières

1	Introduction	1
1.1	La carte géologique " classique "	1
1.2	La carte géologique numérique 3D et la modélisation	2
1.3	Les défis de la modélisation géologique	3
1.4	Organisation générale	4
2	Les méthodes de modélisation géologique	7
2.1	De nombreuses applications	7
2.2	Les techniques probabilistes et déterministes de modélisation géologique	8
2.2.1	L'interpolation	8
2.2.2	Les simulations stochastiques	9
2.2.3	Les modèles génétiques	10
2.2.4	Les modèles mixtes	10
2.3	La modélisation géométrique	11
2.3.1	Quand la CAO inspire...	11
2.3.2	Les deux approches : surfacique et volumique	11
2.4	Les méthodes surfaciques	12
2.4.1	Les surfaces d'élévation	12
2.4.2	Les surfaces paramétriques	13
2.4.3	Les surfaces triangulées et l'approche de <i>Gocad</i>	13
2.5	Méthodes volumiques et surfaces implicites	15
2.5.1	Définition	15
2.5.2	L'approche d' <i>EarthVision</i>	16
2.5.3	Visualisation	16
3	La méthode des potentiels dans l'<i>Éditeur Géologique</i>	19
3.1	Le cadre de la géostatistique multivariable	19
3.1.1	Variogrammes simples et croisés	19
3.1.2	Le cokrigeage simple	20
3.1.3	Le cokrigeage universel	20
3.2	Rappels sur la notion de champ et de gradient	22
3.2.1	Définitions	22
3.2.2	Propriétés	23

3.3	Présentation de la méthode des potentiels	23
3.3.1	Originalité	23
3.3.2	Les hypothèses	23
3.3.3	Les principes	24
3.4	L'interpolation par cokrigeage	25
3.4.1	Les données	25
3.4.2	La résolution du système	25
3.4.3	Remarques sur le cokrigeage dual	26
3.4.4	Quelques précisions	26
3.5	La gestion de plusieurs formations	27
3.5.1	La pile stratigraphique et les règles géologiques	27
3.5.2	L'utilisation de plusieurs champs de potentiel et la détermination des formations	28
3.6	Non-stationnarité et discontinuités par fonctions de dérive	31
3.6.1	Prise en compte de la non-stationnarité	31
3.6.2	Prise en compte des failles	31
3.7	L'environnement <i>Open CAS.CADE</i>	32
4	Le modèle de covariance	35
4.1	Utilisation des dérivées	35
4.1.1	Le cas de la méthode des potentiels	35
4.1.2	Différentiabilité d'une fonction aléatoire	36
4.1.3	Dérivée directionnelle d'une fonction aléatoire	36
4.1.4	Les modèles admissibles pour la covariance du potentiel	36
4.2	Le modèle heuristique	37
4.2.1	Le modèle heuristique initial	37
4.2.2	Rationalisation du choix	37
4.3	La détermination du modèle de covariance à partir des données d'orientation	38
4.3.1	Lien entre covariance du potentiel et covariance du gradient	38
4.3.2	Choix du schéma cubique	39
4.3.3	Calcul des composantes dans le cas du modèle cubique	39
4.4	Des données d'orientation aux données de gradient	41
4.4.1	La nature des données d'orientation	41
4.4.2	Le passage aux données de gradient	42
4.4.3	Le calcul pratique des variogrammes des composantes de gradient	43
4.5	Quelques éléments sur la modélisation : anisotropie et ajustement automatique	43
4.5.1	L'anisotropie zonale	43
4.5.2	Un programme d'ajustement automatique	45
5	Variographie des données d'orientation - Exemples des foliations méta- morphiques	47
5.1	Le Velay	47
5.1.1	Présentation du jeu de données	47
5.1.2	Variogrammes expérimentaux	52

5.1.3	L'ajustement des variogrammes	53
5.1.4	Modèle de la covariance du potentiel	58
5.2	Le Limousin	58
5.2.1	Présentation du jeu de données	58
5.2.2	Variogrammes expérimentaux	60
5.2.3	L'ajustement des variogrammes	66
5.2.4	Modèle de la covariance du potentiel	72
5.3	Les Cévennes	72
5.3.1	Présentation du jeu de données	72
5.3.2	Variogrammes expérimentaux	73
5.3.3	L'ajustement des variogrammes	73
5.3.4	Modèle de la covariance du potentiel	76
5.4	Les Alpes Lepontines	77
5.4.1	Présentation du jeu de données	77
5.4.2	Variogrammes expérimentaux	80
5.4.3	L'ajustement des variogrammes	80
5.4.4	Modèle de la covariance du potentiel	83
5.5	Synthèse de la variographie des quatre jeux de données	83
6	La détermination des incertitudes sur la position des interfaces	87
6.1	Impact du modèle de covariance sur l'interpolation dans le cas du Limousin	87
6.1.1	Présentation	87
6.1.2	Comparaison des interpolations	88
6.2	Détermination des incertitudes sur la position des interfaces	88
6.2.1	Une approche empirique	88
6.2.2	Retour à la méthode du potentiel : variance d'estimation	91
6.2.3	Probabilité d'appartenance à une formation	91
6.2.4	Probabilité d'appartenance à une interface - Définition du potentiel réduit	92
6.2.5	Cartographie du potentiel réduit dans le cas du Limousin	92
6.3	Ecart-types en profondeur et effet de pépite sur les données de potentiel . .	97
6.3.1	Détermination des écart-types en profondeur	97
6.3.2	L'effet de pépite sur les données de potentiel	97
7	Les incertitudes sur les paramètres du modèle par l'approche bayésienne	101
7.1	Rappels sur l'estimation par approche bayésienne dans le cas de champs aléatoires gaussiens	101
7.1.1	L'approche bayésienne en géostatistique : une rencontre relativement récente	101
7.1.2	La formule de Bayes	102
7.1.3	Application au cas d'un champ aléatoire gaussien	102
7.1.4	Les lois <i>a priori</i>	103
7.1.5	Lois <i>a posteriori</i> des paramètres de la covariance	104
7.2	Application au cas du Limousin	106

7.2.1	Les paramètres et l'algorithme utilisés	106
7.2.2	Composante verticale du gradient	107
7.2.3	Composante horizontale du gradient	109
7.2.4	Conclusions	112
8	Illustration du problème des vecteurs normés par l'exemple du Limousin	115
8.1	Rappel de l'hypothèse	115
8.1.1	Utilisation des gradients	115
8.1.2	Le module du gradient	115
8.2	Simulations des vecteurs normés et non-normés : l'exemple du Limousin . .	116
8.2.1	Simulations non conditionnelles du potentiel	116
8.2.2	Calcul des gradients	117
8.2.3	Les variogrammes	119
8.2.4	Conséquences	119
9	Le traitement des sondages incomplets - Aspects théoriques	123
9.1	Les données d'inégalité	123
9.1.1	Les types de données	123
9.1.2	Les inégalités	124
9.1.3	Bornes	125
9.1.4	L'importance de la validation des contraintes	126
9.2	Les méthodes utilisant des données d'inégalité	126
9.2.1	Vérification <i>a posteriori</i>	127
9.2.2	Conditions sur les poids	127
9.2.3	Méthodes non-linéaires	127
9.2.4	Conditions sur les estimations	127
9.2.5	Les inégalités en tant que données	128
9.3	La génération des gaussiennes tronquées et les chaînes de Markov	128
9.3.1	Les gaussiennes tronquées	129
9.3.2	La méthode d'acceptation et rejets	129
9.3.3	Définition et propriétés des chaînes de Markov	130
9.3.4	La probabilité de transition de Geman et l'échantillonneur de Gibbs	131
9.4	L'algorithme de validation des contraintes	133
9.4.1	Résumé des différentes étapes de l'algorithme	133
9.4.2	La prise en compte des inégalités dans la méthode des potentiels	134
10	Le traitement des sondages incomplets - Applications	137
10.1	Le cas de la ville de Lyon : Etudes préliminaires	137
10.1.1	Présentation des données	137
10.1.2	Analyse exploratoire	139
10.1.3	Analyse structurale	140
10.2	Le cas de la ville de Lyon : Résultats de la validation des contraintes	144
10.2.1	Convergence	144
10.2.2	Krigeage simple	149

10.2.3	Krigeage ordinaire	150
10.2.4	Comparaison avec et sans validation	153
10.2.5	Fonctions de dérive et de failles	155
10.2.6	Validation croisée	161
10.3	Méthode des potentiels et données d'inégalité : étude sur des données syn- thétiques	164
10.3.1	Un forage, une inégalité	164
10.3.2	Deux sondages, deux inégalités	165
10.3.3	Le cas de lacunes	166
11	Conclusions et perspectives	171
11.1	Un outil de modélisation géologique 3D : l' <i>Editeur Géologique</i>	171
11.2	Les apports principaux de la thèse	171
11.3	Perspectives	173
	Annexes	177
A	Matrice de covariance utilisée dans la méthode du potentiel	177
A.1	Les gradients	177
A.2	Les incréments de potentiel	178
A.3	Les interactions gradients / interfaces	179
A.4	Les conditions d'universalité	179
A.5	Les termes de failles	180
A.6	Le système de cokrigeage	180
B	Calcul du modèle de variogramme du gradient de potentiel	183
B.1	Détermination du lien entre covariance du potentiel et covariance du gradient	183
B.2	Le cas du modèle cubique	184
C	Rappels sur la minimisation d'une forme quadratique ou quelconque	185
D	Simulations	187
D.1	Les simulations conditionnelles	187
D.1.1	Généralités	187
D.1.2	Le krigeage conditionnant	188
D.2	Génération d'une variable aléatoire uniforme par la méthode de congruence	188
D.3	Approximation polynomiale de la fonction de répartition gaussienne	189
D.4	Simulation de variables aléatoires gaussiennes	189
	Bibliographie	193

Table des figures

2.1	Exemple de simulation avec un modèle de type objet : lentilles sableuses. D'après Benito García-Morales, 2003. Image GOCAD.	9
2.2	Exemple de simulation avec un modèle de type pixel : chenal dans un réservoir (5 faciès par la méthode des pluri-gaussiennes.)	10
2.3	Exemple de simulation par un modèle mixte stochastique et génétique : chenaux méandriformes. D'après Lopez, 2003.	11
2.4	Surface d'élévation.	13
2.5	Surface triangulée.	14
2.6	Exemple de modélisation par surfaces implicites : phases de déformation dans les Alpes centrales suisses. D'après Maxelon, 2004.	16
2.7	Schéma de principe de l'outil <i>EarthVision</i> . D'après Ledez, 2003.	17
2.8	Exemples de polygones de degré 3 à 6 créés par l'algorithme de <i>Marching Cubes</i>	17
3.1	Schéma du principe de la méthode d'interpolation de l' <i>Editeur</i> . Avant interpolation.	24
3.2	Schéma du principe de la méthode d'interpolation de l' <i>Editeur</i> . Après interpolation.	24
3.3	Exemple de pile stratigraphique.	27
3.4	Coupe géologique : relations entre les champs de potentiel. On dispose de 7 couches géologiques de A à G, les interfaces étant matérialisées par des lignes en couleur. La topographie est la ligne supérieure en noir.	29
3.5	Détermination de la formation dans le cas d'un seul champ de potentiel.	29
3.6	Traitement des relations et détermination de la formation pour 2 interfaces sécantes avec une relation de type "erod".	30
3.7	Traitement des relations et détermination de la formation pour 2 interfaces sécantes avec une relation de type "onlap".	30
3.8	Exemple de modèle d'influence pour les failles infinies. Coupe transversale en haut, coupe longitudinale en bas.	33
3.9	Exemple de modèle d'influence pour les failles finies. Coupe transversale en haut, coupe longitudinale en bas.	33
3.10	Zones d'influence des failles vues de haut, faille infinie à gauche, faille finie à droite.	33
4.1	Modèle cubique de portée $a = 600m$ et de palier $C_0 = 1$	40

4.2	Modèles de variogrammes de portées 1000m, 5000m, 10000m, 20000m, et 30000m pour les composantes du gradient. En rouge, le schéma sphérique de $\mathbb{R}^5$ correspondant à une composante du gradient dans la direction perpendiculaire et en noir, le schéma à effet de trou correspondant à une composante du gradient dans la direction parallèle.	41
4.3	Données d'orientation : azimuth et pendage.	42
4.4	Projections d'un vecteur normal sur les trois axes de coordonnées.	42
4.5	Exemple de calcul pour le variogramme de G_x suivant la direction perpendiculaire à l'axe (Ox).	44
5.1	Plan de position des données de foliation de la région Velay.	48
5.2	Photographies de granites. La foliation est marquée par l'orientation préférentielle de la biotite (S) et par l'étiement des enclaves micacées (m). Privas, marge méridionale du dôme du Velay. D'après Ledru <i>et al.</i> , 2001.	49
5.3	Histogrammes des données de foliation du Velay.	50
5.4	Histogrammes et diagrammes quantile contre quantile (distribution expérimentale vs distribution uniforme) des composantes du gradient de la région Velay.	51
5.5	Variogrammes des données de gradient pour la région Velay. Le variogramme γ_z de G_z est en noir, les variogrammes de G_x sont dans des teintes bleues (bleu pour $\gamma_{x\perp}$ et cyan pour $\gamma_{x\parallel}$), les variogrammes de G_y sont dans des teintes rouges (rouge pour $\gamma_{y\perp}$ et magenta pour $\gamma_{y\parallel}$).	52
5.6	Velay : Variogrammes des composantes du gradient à courte échelle : détection d'une éventuelle microstructure.	54
5.7	Velay : Ajustement du variogramme de la composante verticale du gradient.	55
5.8	Velay : Ajustement des variogrammes des composantes horizontales du gradient.	56
5.9	Velay : Ajustement automatique des variogrammes des composantes du gradient.	57
5.10	Plan de position des données de foliation de la région Limousin.	59
5.11	Histogrammes des données de foliation du Limousin.	60
5.12	Histogrammes et diagrammes quantile contre quantile (distribution expérimentale vs distribution uniforme) des composantes du gradient de la région Limousin.	61
5.13	Variogrammes des données de gradient pour la région Limousin. Le variogramme γ_z de G_z est en noir, les variogrammes de G_x sont dans des teintes bleues (bleu pour $\gamma_{x\perp}$ et cyan pour $\gamma_{x\parallel}$), les variogrammes de G_y sont dans des teintes rouges (rouge pour $\gamma_{y\perp}$ et magenta pour $\gamma_{y\parallel}$).	62
5.14	Limousin : Variogrammes des composantes du gradient à courte échelle : détection d'une éventuelle microstructure.	63
5.15	Limousin. Variogrammes des données de la composante x du gradient pour différentes projections.	64
5.16	Limousin. Variogrammes des données de la composante y du gradient pour différentes projections.	65
5.17	Limousin : Ajustement du variogramme de la composante verticale du gradient.	67
5.18	Limousin : Premier ajustement des variogrammes horizontaux des composantes du gradient.	68

5.19	Limousin : Second ajustement des variogrammes des composantes horizontales du gradient.	69
5.20	Limousin : Ajustement automatique des variogrammes des composantes du gradient.	71
5.21	Plan de position des données de foliation de la région des Cévennes.	74
5.22	Histogrammes des données de foliation des Cévennes.	74
5.23	Histogrammes et diagrammes quantile contre quantile (distribution expérimentale vs distribution gaussienne) des composantes du gradient de la région Cévennes.	75
5.24	Variogrammes des données de gradient de la région Cévennes. Le variogramme γ_z de G_z est en noir, les variogrammes de G_x sont dans des teintes bleues (bleu pour $\gamma_{x\perp}$ et cyan pour $\gamma_{x\parallel}$), les variogrammes de G_y sont dans des teintes rouges (rouge pour $\gamma_{y\perp}$ et magenta pour $\gamma_{y\parallel}$).	76
5.25	Cévennes : Ajustement du variogramme de la composante verticale du gradient.	77
5.26	Cévennes : Ajustement des variogrammes des composantes horizontales du gradient.	78
5.27	Plan de position des données de foliation de la région des Alpes centrales suisses.	79
5.28	Histogrammes des données de foliation des Alpes centrales suisses.	79
5.29	Histogrammes et diagrammes quantile contre quantile (distribution expérimentale vs distribution gaussienne) des composantes du gradient de la région des Alpes centrales suisses.	81
5.30	Variogrammes des données de gradient de la région des Alpes centrales suisses. Le variogramme γ_z de G_z est en noir, les variogrammes de G_x sont dans des teintes bleues (bleu pour $\gamma_{x\perp}$ et cyan pour $\gamma_{x\parallel}$), les variogrammes de G_y sont dans des teintes rouges (rouge pour $\gamma_{y\perp}$ et magenta pour $\gamma_{y\parallel}$).	82
5.31	Alpes centrales suisses : Ajustement du variogramme de la composante verticale du gradient.	83
5.32	Alpes centrales suisses : Ajustement des variogrammes des composantes horizontales du gradient.	84
6.1	Plan de position des données de la région Limousin. Croix noires : données de foliation. Disques rouges : données d'interface.	89
6.2	Comparaison de l'interpolation du toit de "uig" (Limousin) pour deux modèles différents. En noir, modèle "conventionnel". En rouge, modèle ajusté.	89
6.3	Comparaison des isopotentielles issues du champ de potentiel servant à modéliser le toit de "uig" (Limousin) pour deux modèles différents, conventionnel et ajusté.	90
6.4	Comparaison de l'interpolation du toit de "uig" (Limousin) pour deux modèles différents dans une coupe. En noir, modèle "conventionnel". En rouge, modèle ajusté.	90
6.5	Limousin - modèle "par défaut" : carte du potentiel réduit.	93
6.6	Limousin - modèle ajusté : carte du potentiel réduit.	93
6.7	Limousin, Coupe 1 - modèle "par défaut" : carte du potentiel réduit.	94
6.8	Limousin, Coupe 1 - modèle ajusté : carte du potentiel réduit.	94
6.9	Limousin, Coupe 2 - modèle ajusté : carte du potentiel réduit.	95
6.10	Limousin, Coupe A - modèle ajusté : carte du potentiel réduit.	95
6.11	Influence de l'effet de pépite du gradient sur les incertitudes.	96
6.12	Limousin, Coupe A - modèle ajusté : carte de l'écart-type en profondeur. (Le rapport entre longueur et profondeur n'est pas respecté.)	97

6.13	Limousin, modèle ajusté : bloc 3D de l'écart-type en profondeur.	98
6.14	Limousin. Courbe de la moyenne des écart-types sur l'interface de la Coupe 1 en fonction de l'effet de pépite (échelle semi-logarithmique).	98
7.1	Limousin : Histogrammes de la distribution <i>a posteriori</i> des paramètres pour la composante verticale.	108
7.2	Limousin : Variogramme expérimental (en noir) et variogramme modélisé (en rouge) en utilisant les modes <i>a posteriori</i> des paramètres de la composante verticale du gradient.	109
7.3	Limousin : Histogrammes de la distribution <i>a posteriori</i> des paramètres pour la composante horizontale G_x	110
7.4	Limousin : Variogramme expérimental (en noir) et variogramme modélisé (en rouge) en utilisant les modes <i>a posteriori</i> des paramètres de la composante horizontale G_x	111
7.5	Limousin : Nuages de corrélation entre différents paramètres en utilisant les échantillons de la densité <i>a posteriori</i> . (Cercle bleu : G_z - Triangle rouge : G_x .)	113
8.1	Cartes du potentiel pour deux simulations.	116
8.2	Histogrammes du potentiel pour deux simulations.	117
8.3	Variogrammes de la simulation 1. Le modèle de variogramme est en rouge, les variogrammes des gradients simulés non normés (complétés) sont en vert et normés en bleu.	120
8.4	Variogrammes de la simulation 2. Le modèle de variogramme est en rouge, les variogrammes des gradients simulés non normés (complétés) sont en vert et normés en bleu.	121
9.1	Contraintes d'inégalité dans les forages	124
9.2	Coupe d'un forage pétrolier - Trace d'une interface en fonction de l'information utilisée	126
9.3	Majoration de la densité de Gauss par loi uniforme pour la partie centrale de la gaussienne.	130
9.4	Comparaison de la validation des contraintes dans l'approche multicouche à gauche, et dans la méthode des potentiels à droite.	134
10.1	Carte de localisation des données - Légende : + rouge : données exactes, $\square$ bleu : données d'inégalité.	138
10.2	Zoom de la carte de localisation des données avec photo aérienne en fond de carte.	138
10.3	Histogramme des données dures.	139
10.4	Diagramme quantile contre quantile (distribution expérimentale vs distribution gaussienne), à gauche et diagramme de corrélation différée ($h = 1500m$), à droite.	141
10.5	Nuages de corrélation entre les données d'altitude et les coordonnées géographiques X et Y	141
10.6	Variogramme omnidirectionnel expérimental pour les 271 données dures, en haut sans tenir compte des failles et en bas en tenant compte des failles. C'est celui-ci qui sera utilisé dans la suite de l'étude.	142

10.7	Carte variographique, à gauche et variogrammes directionnels ($45^\circ N$ et $135^\circ N$) expérimentaux pour les 271 données dures, à droite.	143
10.8	Variogramme omnidirectionnel modélisé pour les 271 données dures	144
10.9	Forage A005 - Courbe de l'altitude du toit du socle en fonction du numéro d'itération, à gauche, et histogramme des valeurs obtenues pour 10000 itérations, à droite.	145
10.10	Forage A005 - Moyenne et variance de l'altitude de l'interface séries sédimentaires/socle en fonction du numéro d'itérations.	146
10.11	Forage A224 - Moyenne et variance de l'altitude de l'interface séries sédimentaires/socle en fonction du numéro d'itérations.	146
10.12	Courbe du paramètre $\sqrt{R}$ en fonction du numéro d'itérations et courbes de $\widehat{var}(Z)$ (en noir) et var_I (en rouge) également en fonction du numéro d'itérations.	148
10.13	Histogrammes après validation des contraintes (krigeage simple $m = 210$).	149
10.14	Nuage de corrélation entre les données totales après validation des contraintes par krigeage simple (KS) pour $m = 210$ et pour $m = 0$, à gauche et carte de localisation, à droite. En bleu les points qui s'éloignent le plus de la première bissectrice.	150
10.15	Histogramme des données souples pour le krigeage ordinaire et nuage de corrélation entre les données souples après validation des contraintes par krigeage simple pour $m = 210$ et par krigeage ordinaire.	151
10.16	Carte de localisation de toutes les données et valeur des variances d'estimation. La légende fournit des intervalles pour les valeurs des variances. En noir, les données dures (variance nulle). Les trois autres couleurs concernent les données souples.	152
10.17	Carte interpolée à partir des données dures et souples (KO).	152
10.18	Carte de différence entre les cartes utilisant toutes les données, l'une avec l'approche de type KS, l'autre le KO.	153
10.19	Histogrammes des données souples sans validation des contraintes.	154
10.20	Histogramme des erreurs standardisées, à gauche et nuage de corrélation entre les valeurs obtenues avec et sans validation, à droite.	155
10.21	Carte interpolée à partir des seules données dures.	156
10.22	Carte des écarts entre les altitudes estimées sans validation (Figure 10.21) et avec validation (Figure 10.17).	156
10.23	Carte de différence entre les cartes utilisant toutes les données, l'une avec l'approche de type KU et dérive linéaire, l'autre avec le KO.	157
10.24	Carte de localisation des zones de failles. En bleu, les points qui s'éloignent le plus de la première bissectrice à la Figure 10.25.	158
10.25	Nuage de corrélation entre une validation des données d'inégalité sans et avec fonctions de failles. En bleu, les points qui s'éloignent le plus de la première bissectrice.	159
10.26	Carte de différence entre les cartes utilisant toutes les données, l'une sans les fonctions de failles, l'autre avec.	159
10.27	Carte interpolée utilisant toutes les données (zoom sur les failles au sud), (a) sans les fonctions de failles, (b) avec les fonctions de failles. Les croix noires correspondent aux données dures et les blanches aux données souples.	160

10.28	Validation croisée : carte de localisation des données - Croix noires : données d'égalité ; Cercles rouges : données d'inégalité.	161
10.29	Validation croisée : histogrammes des données dures et souples.	162
10.30	Validation croisée : cartes de localisation et nuages de corrélation entre les données simulées et réelles sans fonctions de failles (en haut) et après ajout de fonctions de failles, en bas.	163
10.31	Tracés de l'interface entre deux couches géologiques.	165
10.32	Détermination de 2 interfaces géologiques.	166
10.33	Détermination des interfaces entre 4 couches géologiques sans validation des contraintes.	168
10.34	Détermination des interfaces entre 4 couches géologiques avec validation des contraintes.	169
A.1	Relations entre points sur des interfaces différents.	178
B.1	Configuration de calcul.	184
D.1	Schéma illustrant l'estimation d'une variable 1D à partir de quelques points de données en haut, simulation conditionnelle en bas.	187

Liste des tableaux

5.1	Statistiques élémentaires des données de gradient pour le Velay.	50
5.2	Portées et paliers des structures de base du modèle de variogramme pour les données du Velay.	53
5.3	Effets de pépité et paliers des variogrammes des composantes du gradient pour les données du Velay.	53
5.4	Portées et paliers des structures de base du modèle de variogramme pour les données du Velay obtenus par le programme d'ajustement automatique (avec valeur du χ^2).	58
5.5	Effets de pépité et paliers des variogrammes des composantes du gradient pour les données du Velay obtenus par le programme d'ajustement automatique.	58
5.6	Statistiques élémentaires des données de gradient pour le Limousin.	59
5.7	Portées et paliers des structures de base du modèle de variogramme pour les données du Limousin.	66
5.8	Effets de pépité et paliers des variogrammes des composantes du gradient pour les données du Limousin.	66
5.9	Portées et paliers des structures de base du nouveau modèle pour les variogrammes de G_x et G_y pour les données du Limousin.	67
5.10	Effets de pépité et paliers du nouveau modèle pour les variogrammes de G_x et G_y pour les données du Limousin.	67
5.11	Portées et paliers des structures de base du modèle de variogramme pour les données du Limousin obtenus par le programme d'ajustement automatique (avec valeur du χ^2).	70
5.12	Effets de pépité et paliers des variogrammes des composantes du gradient pour les données du Limousin obtenus par le programme d'ajustement automatique.	70
5.13	Portées et paliers des structures de base du modèle de variogramme pour les données du Limousin projetées sur des axes après rotation de 30°	72
5.14	Effets de pépité et paliers des variogrammes des composantes du gradient pour les données du Limousin projetées sur des axes après rotation de 30°	72
5.15	Statistiques élémentaires des données de gradient pour les Cévennes.	73
5.16	Portées et paliers des structures de base du modèle de variogramme pour les données des Cévennes.	76

5.17	Effets de pépite et paliers des variogrammes des composantes du gradient pour les données des Cévennes.	77
5.18	Statistiques élémentaires des données de gradient pour les Alpes Lépointines.	80
5.19	Portées et paliers des structures de base du modèle de variogramme pour les données des Alpes Lépointines.	80
5.20	Effets de pépite et paliers des variogrammes des composantes du gradient pour les données des Alpes Lépointines.	82
6.1	Limousin, Coupe 1. Statistiques élémentaires des écart-types en profondeur du toit de “uig” en fonction de l’effet de pépite sur le potentiel.	98
7.1	Limousin : Statistiques <i>a posteriori</i> du modèle de variogramme pour les données du gradient vertical.	107
7.2	Limousin : Statistiques <i>a posteriori</i> du modèle de variogramme pour les données de la composante horizontale G_x du gradient.	111
8.1	Rappel des paramètres du modèle de covariance (gauche) et des statistiques élémentaires des données de gradient (droite) pour le Limousin.	118
8.2	Statistiques élémentaires des données de gradient simulées (simulation 1).	118
8.3	Statistiques élémentaires des données de gradient simulées (simulation 2).	118
10.1	Statistiques élémentaires des données dures (altitude du toit du socle en m).	139
10.2	Caractéristiques du modèle retenu	143
10.3	Statistiques élémentaires des données souples (en m).	151
10.4	Statistiques élémentaires de la variance d’estimation des données souples (en m^2).	151
10.5	Statistiques élémentaires des données souples sans et avec validation des contraintes (en m).	154
10.6	Statistiques élémentaires pour la validation croisée (en m)	164

Chapitre 1

Introduction

Pendant longtemps, l'Homme s'est contenté de se représenter des réalités tridimensionnelles (objets, bâtiments, lieux géographiques...) sur des supports bidimensionnels : croquis, plans, cartes ; ceci pour des raisons pratiques, car les volumes (maquette ou globe par exemple) sont peu maniables, donc difficilement transportables. Parmi ces documents, les cartes sont certainement ceux qui portent la charge émotionnelle la plus importante, car elles sont destinées à nous décrire la Terre, à nous représenter le monde. Elles possèdent donc à la fois une destinée utilitaire et conceptuelle. A notre époque, les cartes ne sont plus ornées d'animaux fantastiques et ne possèdent plus de *Terra incognita*, car toutes les zones du globe sont plus ou moins bien connues. Il reste cependant des parties de notre planète qui sont moins bien appréhendées ; il s'agit, outre les fonds océaniques, de tout ce qui se trouve sous nos pieds : le sous-sol. Si les plus anciennes cartes géographiques remontent à l'Antiquité, les premières cartes géologiques ne datent en France que du XIX^e siècle, même si les préoccupations à ce sujet sont un peu antérieures.

Dans notre pays, le Bureau de Recherches Géologiques et Minières (BRGM) est chargé de dresser la carte géologique pour l'ensemble du territoire national. Actuellement toute la surface de ce territoire est couverte par des cartes géologiques 2D et la carte du futur sera 3D. Ainsi de nombreuses recherches sont menées actuellement au BRGM sur la réalité virtuelle, la visualisation 3D, le calcul parallèle, les méthodologies de modélisation 3D etc. Ce travail de thèse, réalisé au Centre de Géostatistique de l'École des Mines de Paris, en partenariat avec le BRGM, s'inscrit donc dans le cadre d'un de ces programmes sur la modélisation géométrique 3D au service de la géologie. Il revêt essentiellement des aspects méthodologiques, en participant au développement d'un outil appelé *Éditeur Géologique*, par l'incorporation de techniques géostatistiques. Cette thèse se situe donc en amont du travail du géologue, qui lui, va être utilisateur de l'outil de modélisation.

1.1 La carte géologique " classique "

La principale difficulté que l'on rencontre lorsqu'on tente de se représenter le sous-sol est le caractère partiel de la connaissance géologique de la surface de la Terre et du sous-sol. Elle nécessite la collecte de données en nombre forcément limité qu'il faut ensuite exploiter.

Ces données sont de natures diverses : qualitatives et quantitatives, à échantillonnage plus ou moins fin et souvent affectées de “bruit” etc. On distingue les croquis et observations sur le terrain, les prélèvements d'échantillons en affleurement ou en carottes, les examens de lames minces de ces échantillons ainsi que des fossiles présents, les diagraphies, les mesures pétrophysiques sur carottes, les mesures géophysiques (sismique, gravimétrie, géomagnétisme, méthodes électriques...) etc. On voit que ces informations, parfois subjectives, sont prises à des échelles différentes, ce qui ajoute à la complexité du problème. De plus, la confiance accordée à ces données, parfois incomplètes, n'est pas la même (données précises mais localisées, données moins précises mais connues de façon plus continue).

A partir des données recueillies sur le terrain, il est possible de se représenter le sol et le sous-sol à l'aide de cartes ou de coupes. Il est nécessaire, pour fabriquer la carte, d'y inclure certaines hypothèses et concepts des Sciences de la Terre ; elle est toujours le résultat d'une interprétation. La carte, accompagnée d'une notice explicative, est une synthèse sur une surface à deux dimensions des informations géologiques connues sur une aire donnée. Elle correspond à la projection sur un plan de l'intersection des formations géologiques et de la surface topographique. Elle fournit un inventaire des types de roches en fournissant leur localisation précise à l'affleurement (ou sous la couverture végétale) et indique les traits structuraux. Ces informations sont indiquées par des plages de couleurs ainsi que des symboles. Par ce biais, la carte permet de connaître l'âge des terrains, ainsi que la lithologie. Le fond topographique reste visible, ainsi, dans les zones de relief, en interprétant les données de la carte et en extrapolant, est-il possible d'obtenir une connaissance préliminaire de l'extension des formations jusqu'à une certaine profondeur.

Les cartes géologiques sont publiées à des échelles différentes ; en France, le BRGM édite les plus détaillées au 1 : 50 000, la carte au 1 : 1 000 000 couvrant tout le pays. Il faut préciser aussi que les cartes ne sont jamais définitives et évoluent au cours du temps en fonction de l'accroissement de la connaissance. Elles reflètent donc le savoir des géosciences à un moment donné.

1.2 La carte géologique numérique 3D et la modélisation

La troisième dimension est accessible en carte grâce à des isolignes et des palettes de couleurs, et surtout par l'intermédiaire de coupes. Néanmoins, on reste dans un environnement bidimensionnel. Ce dernier est inadapté à représenter pleinement les relations complexes qui lient les surfaces géologiques 3D et leur évolution temporelle. La carte ne montre que quelques aspects du modèle qui n'existe peut-être complètement - et encore ? - que dans le cerveau du géologue.

Ces dernières années, le développement des techniques informatiques a dès lors permis de se représenter le monde souterrain dans un environnement tridimensionnel et d'intégrer des données de natures différentes. On parle alors de modélisation, le modèle géologique étant la représentation informatique de la zone du sous-sol étudiée. Il est composé d'un modèle numérique de terrain (MNT) qui est une grille donnant l'altitude en fonction de la latitude et de la longitude et qui fournit une connaissance en 3D de la topographie. La carte géologique étant l'intersection entre le sous-sol et la surface de la Terre, si on plaque

cette carte sur un MNT on obtient la partie supérieure d'un bloc 3D.

La modélisation 3D appliquée à la cartographie consiste à créer l'intégralité de ce cube constitué par des volumes limités par des surfaces dont la géométrie est la conséquence de l'histoire géologique. Ces volumes et ces surfaces sont les représentations des objets géologiques qui composent le modèle.

A la base, on trouve la formation (ou couche) géologique, qui est homogène d'un point de vue lithologique ou historique. Deux dépôts contemporains en contact mais de nature différente ne sont pas deux formations géologiques identiques. Par ailleurs, deux couches de grès - par exemple - n'ayant pas été déposées en même temps sont également des objets distincts.

Les interfaces (ou horizons) sont les surfaces limites séparant les formations. Les failles, qui sont les résultats de déformation cassante, constituent une catégorie particulière. On distingue la faille normale engendrée par une phase d'extension, la faille inverse témoignant d'une compression et le décrochement qui résulte d'un coulissage horizontal. Au cours de leur histoire, les formations sont aussi soumises à des déformations souples créant des plis (anticlinaux, synclinaux...), sont érodées etc.

La géométrie et la topologie sont donc complexes et sont régies par des lois spécifiques dont il faut tenir compte lors de la modélisation. Enfin, le modèle que l'on a déterminé doit être géologiquement réaliste.

1.3 Les défis de la modélisation géologique

Bien que de nombreux progrès aient été réalisés dans le domaine de la modélisation géologique, il reste de multiples problèmes en suspens si l'on souhaite créer les outils les plus fiables possibles et de construire les modèles les plus réalistes. Avant de résumer les qualités d'un bon modelleur, précisons avant toute chose qu'il est certainement illusoire de vouloir vérifier tous ces critères. Il s'agit surtout de trouver un compromis pour une solution optimale.

Le modelleur doit avant tout être flexible et permettre une grande interactivité avec la possibilité de mise à jour facile des modèles. La visualisation est essentielle à chaque étape pour corriger les éventuelles erreurs au cours de la modélisation. Toutes les fonctionnalités doivent être accessibles à l'utilisateur, contrairement à un outil de type boîte noire, afin de maîtriser tous les processus internes. Il faut cependant trouver un compromis avec l'utilisateur de base, qui n'est pas forcément formé aux techniques mathématiques. Tout en gardant un accès possible aux paramètres pour une utilisation plus experte, les étapes délicates peuvent faire l'objet d'une automatisation.

Il faut ensuite pousser l'intégration des différents types de données au plus haut point en gardant à l'esprit l'objectif général de l'étude. Les données qualitatives comme les interprétations du géologue doivent faire l'objet d'un formalisme précis, afin d'être manipulées comme n'importe quelle donnée quantitative.

Des aspects informatiques entrent aussi en jeu, comme le temps de calcul et l'espace mémoire nécessaire.

Un des critères les plus importants est la robustesse du modelleur, c'est-à-dire sa capacité

à fournir des résultats pertinents, y compris dans des cas complexes, tout en respectant les données. Enfin, il convient de décrire et de quantifier, si possible, les incertitudes associées à tous les résultats obtenus.

Tous ces points mettent en avant des problèmes spécifiques, qu'il a fallu tenter de résoudre tout d'abord lors de la mise au point de l'outil *Editeur Géologique* depuis bientôt une dizaine d'années pour les premiers développements théoriques mais également dans le cadre de cette thèse. Ainsi le problème des incertitudes nous a-t-il guidé tout au long de ces travaux.

1.4 Organisation générale

Le prochain chapitre (chapitre 2) présente les différentes méthodes de modélisation géologique et l'accent est évidemment mis sur les surfaces implicites, type de surfaces présent dans l'*Editeur*.

Après un rapide rappel concernant le cadre multivariable, la modélisation géologique par la méthode du potentiel telle qu'elle est implémentée dans le modèleur du BRGM fait l'objet du chapitre 3. Cette méthode de modélisation considère les interfaces comme des iso-surfaces d'un champ de potentiel défini en tout point de l'espace.

Le chapitre 4 est quant à lui consacré au modèle de covariance du champ de potentiel et à sa détermination à partir des données d'orientation. Une chaîne de traitement a été implémentée dans l'*Editeur*, afin d'obtenir les paramètres du modèle de covariance à partir de ces données. Les variogrammes expérimentaux des données de gradient sont calculés puis ajustés avec une méthode de moindres carrés non linéaires : la méthode de Levenberg-Marquardt.

Le chapitre 5, quant à lui regroupe les applications variographiques de différents jeux de données. Ces jeux recouvrent trois zones françaises échantillonnant de manière assez dense le toit d'un ensemble de formations métamorphiques : Velay, Limousin et Cévennes ainsi qu'une zone dans les Alpes suisses.

Ensuite, les chapitres 6 et 7 sont consacrés aux incertitudes, respectivement sur la position des interfaces en profondeur et sur les paramètres du modèle, cette dernière étape s'effectuant à l'aide d'une approche bayésienne. Pour le premier point, après le calcul des variances d'estimation attachées aux incréments de potentiel, la cartographie des incertitudes a été obtenue grâce au calcul de ce que l'on a appelé "potentiel réduit". Quant à l'approche bayésienne, les paramètres - dérive, portée, palier et effet de pépité - ont été simulés selon une distribution *a posteriori* qui permet de prendre en compte une information *a priori* ainsi que l'impact des données par l'intermédiaire d'une fonction de vraisemblance. Au chapitre 8 est étudié, sur des simulations non conditionnelles du potentiel recouvrant le cas du Limousin, l'impact de la normation des vecteurs d'orientation sur les variogrammes. Il a fallu vérifier cette hypothèse, car on assimile, au cours du processus d'interpolation, orientations et gradients.

Enfin, un autre type d'incertitudes est abordé : celui des données souples. En effet, les passes incomplètes des sondages sont présentes en grand nombre et sont souvent inutilisables. Par contre, il est possible d'exploiter la relation d'ordre "le point situé à la fin du

forage est au-dessus de la surface que je désire modéliser” en simulant une valeur respectant à la fois cette relation et la structure spatiale des données. Les principes des techniques permettant de modéliser une interface géologique à l’aide de données exactes et de données d’inégalités sont expliquées au chapitre 9, et des applications de cette prise en compte sont exposées au chapitre 10.

Chapitre 2

Les méthodes de modélisation géologique

La modélisation géologique consiste à recréer deux éléments d'une scène géologique : la géométrie et la géologie proprement dite à savoir les propriétés des couches la constituant. Ces propriétés sont la lithologie, l'âge, la porosité, la perméabilité ou toute autre propriété physique, mécanique, chimique etc. Le choix d'une méthode parmi un nombre pléthorique de techniques est donc guidé en premier lieu par les applications et donc les utilisateurs (géologues tectoniciens, ingénieurs réservoirs...) ¹.

2.1 De nombreuses applications

En effet, un sédimentologue qui désire connaître l'évolution spatio-temporelle d'un environnement donné n'a pas les mêmes besoins qu'un tectonicien travaillant sur différentes phases de déformation au niveau d'un massif montagneux ou qu'un ingénieur réservoir d'une compagnie pétrolière s'intéressant aux écoulements de fluides dans une formation bien définie. Ces différentes applications conduisent aussi à des modèles réalisés à des résolutions bien différentes (plurikilométrique pour une étude à l'échelle d'un bassin sédimentaire ou d'une chaîne de montagnes, décamétrique pour un réservoir pétrolier, métrique pour une unité génétique, centimétrique à l'échelle granulométrique).

En sciences de la Terre, les entreprises pétrolières ont été parmi les premières à développer, pour des raisons économiques, des outils 3D. L'imagerie plus réaliste et plus précise fournie par la sismique ou la modélisation de réservoir en 3D contribue à la diminution des incertitudes inhérentes à l'exploration du sous-sol.

De nombreux modeleurs existent et continuent d'être développés. Par la visualisation en 3D et l'apport de données d'origines diverses, ils contribuent à mieux connaître le sous-sol de notre planète. Cependant, la connaissance scientifique en tant que telle n'est pas la seule concernée ; il existe de nombreux domaines où l'utilisation de la modélisation et de la visualisation 3D est prometteuse. Ces domaines sont le génie civil et la géotechnique (grands

¹En ce qui nous concerne, en cartographie 3D, on se contente d'affecter aux volumes construits l'âge des terrains définis dans une pile chrono-stratigraphique.

chantiers urbains, percement de tunnels, études d'impact pour les ouvrages d'art...), l'exploitation des ressources minières et pétrolières (modélisation de réservoir par exemple), l'environnement et la gestion des risques (stockage souterrain, migration des polluants, comportement mécanique des terrains à l'aplomb de mines désaffectées...). En cartographie géologique, entre autre, l'usage de l'outil informatique permet une large diffusion des informations et de les partager à tous les intervenants des géosciences. Enfin, bien sûr, les modèles numériques 3D ne servent pas seulement à la visualisation, mais sont aussi intégrés dans des calculs (écoulement, mécanique etc.).

Il ne faut cependant pas oublier que le modèle restera toujours une interprétation, certes cohérente avec les données disponibles et certaines hypothèses, mais une simplification quand même de la réalité complexe sous-jacente.

Ces différences d'objectifs permettent déjà de distinguer les modèles statiques tentant de décrire la réalité à un moment donné et les modèles dynamiques qui visent à reconstituer les processus géologiques qui ont permis au cours du temps d'aboutir à cette réalité.

Nous allons voir que si l'on choisit une approche probabiliste, il est possible d'aborder la modélisation de deux manières : l'interpolation et les simulations. L'interpolation va nous servir à construire surfaces et volumes pour des modèles statiques à partir des données disponibles. Quant aux simulations, elles fournissent différentes réalisations de fonctions aléatoires ayant même loi spatiale. Elles peuvent être combinées à des méthodes génétiques afin d'expliquer l'évolution des structures géologiques par l'intermédiaire de différents *scenarii*.

2.2 Les techniques probabilistes et déterministes de modélisation géologique

Cette section ne se veut pas un inventaire exhaustif des techniques de modélisation géologiques, mais présente différentes pistes permettant de reconstruire une scène géologique.

2.2.1 L'interpolation

L'interpolation consiste à estimer une grandeur continue à partir d'un ensemble discret de cette grandeur. Elle conduit cependant inévitablement à des approximations. Les techniques d'interpolation vont donc nous permettre de reconstituer les formations géologiques à partir des données disponibles. Un des interpolateurs les plus simples à utiliser est l'inverse des distances séparant les points de l'espace. Quant au krigeage ponctuel, il va consister à estimer un point en pondérant les données voisines. C'est un estimateur linéaire, sans biais et à variance d'erreur minimale. Le krigeage, en tant qu'interpolateur exact, s'apparente aux techniques habituelles de cartographie comme les splines ou les moindres carrés. Ces dernières méthodes consistent à ajuster une fonction de forme prédéfinie à l'ensemble des données.

Le résultat du krigeage est une image lissée de la réalité. Si l'on veut reproduire les principales caractéristiques de dispersion de la variable étudiée, il faut faire appel à des techniques de simulations.

2.2.2 Les simulations stochastiques

Bien que la géologie d'un environnement donné soit intrinsèquement déterministe, puisqu'il est généré par des processus physiques connus, il est néanmoins possible pour le modéliser d'utiliser un cadre probabiliste. L'incertitude de notre connaissance de la réalité géologique est due à l'insuffisance de l'information disponible et à la complexité de cette réalité. On considère alors le dépôt sédimentaire par exemple comme la réalisation d'un processus aléatoire défini par sa loi spatio-temporelle et grâce aux simulations stochastiques, on obtient alors plusieurs réalisations d'un même objet. La connaissance du processus génétique à l'origine de la formation géologique n'est pas obligatoire. Contrairement aux modèles génétiques que nous verrons plus tard (2.2.3), il est nécessaire ici de bien connaître les dépôts, c'est-à-dire le résultat du processus, pour l'inférence du modèle statistique. Ces techniques sont, de plus, souples d'utilisation et offrent des possibilités pour le conditionnement des modèles. D'une manière générale, ils sont souvent peu réalistes et n'autorisent pas de compréhension précise des phénomènes géologiques.

Une première catégorie de modèles probabilistes est celle des **modèles objets**² où un objet est décrit par sa position, définie par un point aléatoire dans l'espace et ses propriétés comme la forme, la taille ou l'orientation. Ils fournissent des images intuitives, quoique souvent simplistes des formations géologiques et le conditionnement aux données peut s'avérer difficile. La Figure 2.1 est un exemple synthétique qui illustre des lentilles de sables dans un réservoir pétrolier (Benito García-Morales, 2003).

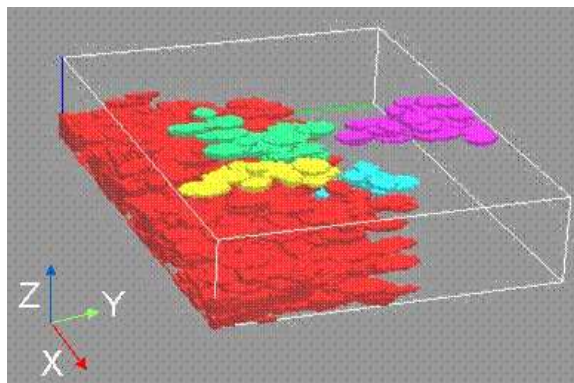


FIG. 2.1 – Exemple de simulation avec un modèle de type objet : lentilles sableuses. D'après Benito García-Morales, 2003. Image GOCAD.

L'autre grande classe est celle des **modèles pixels**. Ils utilisent une grille, un pixel étant une cellule élémentaire, et sont facilement contraints par les données. En revanche dans de nombreux cas, ils ne donnent pas une image géologiquement réaliste des formations modélisées. La Figure 2.2 présente des faciès obtenus par la méthode des plurigaussiennes seuillées (Armstrong *et al.*, 2003) dans le cadre d'un réservoir pétrolier synthétique.

²On parle aussi de modèles booléens.

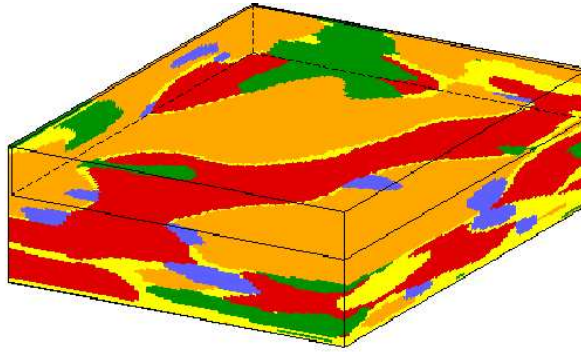


FIG. 2.2 – Exemple de simulation avec un modèle de type pixel : chenal dans un réservoir (5 faciès par la méthode des pluri-gaussiennes.)

2.2.3 Les modèles génétiques

Contrairement aux modèles précédents, les modèles génétiques prennent en compte la physique des phénomènes et donc nécessitent la connaissance des processus géologiques mis en œuvre dans la Nature lors de la formation des dépôts. Ces modèles dits aussi déterministes permettent de construire des formations géologiques (nature et topologie) à partir de paramètres initiaux.

Un grand nombre de modèles existe en fonction du type d'environnement sédimentaire étudié. Lopez (2003), qui recense les modèles existants pour les dépôts fluviaux, distingue plusieurs catégories en fonction de leur complexité. Les modèles de remplissage de bassin à l'échelle décakilométrique sont les plus simples, un grand nombre de facteurs étant constants. Les modèles géologiques, déjà plus complexes, utilisent des paramètres comme le taux d'aggradation ou la compaction. Enfin, les modèles hydrauliques et géomorphologiques utilisent les équations principales linéarisées de l'hydraulique.

Cette approche génétique vise à créer des modèles complexes qui permettent donc de bien décrire et expliquer les phénomènes en jeu. Ils sont donc réalistes, mais *a contrario* le conditionnement à des données s'avère difficile à réaliser.

2.2.4 Les modèles mixtes

Les deux approches, génétique et stochastique, sont un peu antagonistes quant à leurs qualités et défauts : complexes mais non contraintes, conditionnement possible mais simplistes. C'est pourquoi des méthodes mixtes ont été développées en combinant ces deux façons de voir. Lopez (2003) utilise un modèle mixte adapté aux réservoirs chenalisés méandriformes (Figure 2.3). Il utilise à la fois les équations d'hydrodynamique fluviale et les simulations stochastiques.

Plus modestement ce travail s'attachera à des problèmes d'interpolation des formations et non à la genèse des dépôts.

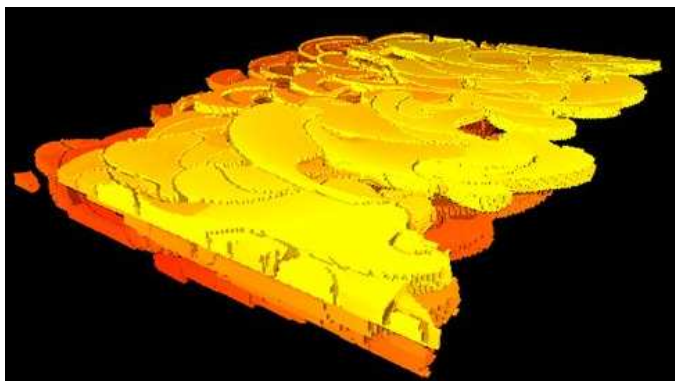


FIG. 2.3 – Exemple de simulation par un modèle mixte stochastique et génétique : chenaux méandriformes. D'après Lopez, 2003.

2.3 La modélisation géométrique

On ne s'intéresse dans cette section qu'au côté géométrique de la modélisation, autrement dit à la façon de reconstruire un objet à partir de volumes ou de surfaces.

2.3.1 Quand la CAO inspire...

La CAO ou conception assistée par ordinateur consiste à concevoir des objets complexes à l'aide de l'outil informatique. La géométrie de l'objet que l'on veut créer est alors très bien connue, ce qui n'est pas le cas en modélisation géologique où l'on souhaite représenter un objet déjà existant. Or l'information géologique est hétérogène, plus ou moins bien échantillonnée et connue avec des précisions variables (Mallet, 2002). Dans ces conditions, la réalité ne peut être qu'approchée. De plus, cette géométrie est le résultat d'une histoire géologique précise. De nombreuses techniques plus ou moins inspirées de celles de la CAO ou de l'imagerie médicale sont appliquées en modélisation géologique.

2.3.2 Les deux approches : surfacique et volumique

On distingue les méthodes qui consistent à construire des surfaces à partir de données présentes sur celles-ci et d'autres qui créent des volumes. Un volume est soit défini par l'espace à l'intérieur de surfaces limites cousues entre elles, soit par une variable définie à 3D qui régit la géologie de façon explicite ou implicite (cf. section 2.5).

Approche volumique.

La première idée qui peut venir à l'esprit est d'assembler des volumes élémentaires, un peu à la manière d'un jeu de construction. Cette méthode de grille volumique correspond au modèle pixel vu dans le cadre des modèles de simulation (2.2.2). Elle permet de construire un volume en utilisant des petits cubes élémentaires (voxels) dans une grille 3D, auxquels on assigne une propriété. Cette étape est effectuée par des techniques d'interpolation comme le krigeage et les voxels ayant la même propriété appartiennent aux mêmes

volumes. Le défaut majeur de cette méthode, outre la nécessité de requérir un espace mémoire important, est la discrétisation qui s'avère impropre à certaines applications comme les rais sismiques. Cette manière de faire est rare en modélisation géologique, on peut citer néanmoins des applications dans le cadre d'inversion gravimétrique (Guillen *et al.*, 2004) ou magnétique avec de larges mailles. Une autre manière explicite de construire des volumes est celle des diagrammes de Voronoï. En fait plusieurs aspects sont combinés ici puisqu'il s'agit d'une méthode de type grille volumique avec des surfaces limites, puisque les voxels utilisés sont des polyèdres (Nullans (1998), Courrioux *et al.* (2001)).

L'approche implicite, vue son importance dans le cadre de la méthode des champs de potentiel, fait l'objet d'une section particulière (2.5).

Approche surfacique.

La création de volumes proprement dits peut se faire en fermant des surfaces obtenues à l'aide des méthodes décrites en 2.4. Les surfaces sont alors les limites de l'objet géologiques. Le problème principal intervient à la couture pour des géométries complexes surtout lorsque les surfaces ont été obtenues indépendamment les unes des autres.

Il existe cependant encore d'autres méthodes. L'une d'elle consiste à créer le volume à partir de coupes en s'inspirant de ce qui se fait en imagerie médicale (Boissonnat, 1988). Elle nécessite néanmoins de bien connaître les interfaces, afin de bien "coller" les sections entre elles. En géologie, dans de nombreux cas, elle s'avère difficile à mettre en œuvre.

La suite de ce chapitre va consister à préciser les différents types de surfaces rencontrées en modélisation ainsi qu'à exposer l'approche volumique par surfaces implicites.

2.4 Les méthodes surfaciques

Il existe deux grandes classes de représentation de surfaces : implicites et paramétriques. Les surfaces les plus simples sont les surfaces d'élévation, mais elles ne peuvent pas être utilisées pour une modélisation purement 3D.

2.4.1 Les surfaces d'élévation

Ce sont des surfaces (Figure 2.4) où la cote z est exprimée en fonction des deux autres coordonnées spatiales x et y : $z = f(x, y)$. Le MNT est par exemple une surface d'élévation. On utilise ces surfaces pour modéliser les interfaces de couches sédimentaires subhorizontales dans une optique multicouche. Les méthodes géostatistiques classiques sont parfaitement adaptées à ce type d'approche (Blanchin et Chilès, 1993). Bien que la visualisation soit tridimensionnelle, la modélisation s'effectue à 2D. Comme elle porte habituellement sur plusieurs interfaces, on parle aussi parfois de modélisation 2,5D. Comme à un point (x, y) , il ne correspond pour une interface donnée qu'une seule valeur de z , certains objets géologiques ne peuvent être modélisés : surfaces replissées, diapirs, lentilles, intrusions granitiques etc.

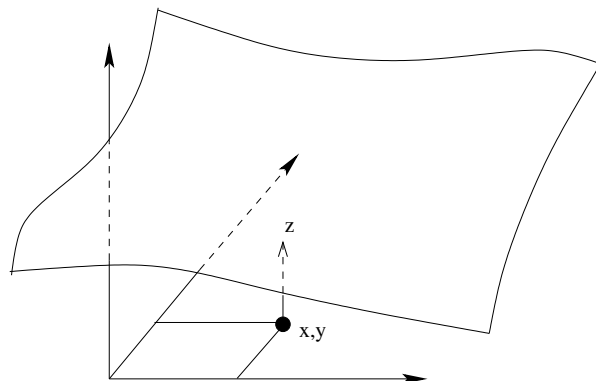


FIG. 2.4 – Surface d'élévation.

2.4.2 Les surfaces paramétriques

Ce sont les surfaces par excellence de la CAO : splines, Bézier... On en trouve des illustrations par exemple dans Auerbach et Schaeben (1990) ou dans le logiciel *Geosurf* au sein de l'outil RML (Reservoir Modeling Line) de l'Institut Français du Pétrole.

Elles possèdent un repère propre en 2D et s'expriment en associant un point 3D (x, y, z) de l'espace à un point (u, v) dans le repère de la surface : $F(u, v) = (x, y, z)$. On a donc : $x = f_1(u, v)$, $y = f_2(u, v)$, $z = f_3(u, v)$ avec f_1 , f_2 et f_3 des fonctions quelconques de u et v .

Par exemple, l'équation paramétrique d'une sphère de centre $C(c_x, c_y, c_z)$ et de rayon R est donnée par :

$$\begin{cases} x = c_x + R \cos \theta \cos \phi \\ y = c_y + R \cos \theta \sin \phi \\ z = c_z + R \sin \theta \end{cases}$$

Signalons que ce sont souvent des fonctions polynomiales qui sont utilisées pour f_1 , f_2 , f_3 . Cette formulation est donc explicite par opposition à celle, implicite, utilisée dans l'*Editeur*. L'attrait de ces surfaces repose principalement sur la relative simplicité de calcul de la courbure ou de la tangente et sur le fait qu'elles soient esthétiquement belles. Mais elles respectent difficilement les contraintes imposées par les données géologiques. En effet, le paramétrage s'avère parfois difficile à trouver et les polynômes ne sont pas adaptés à la modélisation de surfaces discontinues (présence de failles), mais on peut néanmoins utiliser des surfaces continues par morceaux.

2.4.3 Les surfaces triangulées et l'approche de *Gocad*

En géologie, les surfaces ne sont pas connues en continu, car l'information est toujours partielle. Il y a donc nécessité de reconstruire (et de visualiser) la surface à partir d'un nuage de points en utilisant des techniques d'interpolation et de triangulation.

Par définition (Frey et Borouchaki, 1996), une triangulation est "un ensemble d'entités décrivant un graphe, appelé graphe d'adjacence, dont les noeuds sont les triangles et

les arêtes représentent les relations de voisinage des paires de triangles adjacents.” Les surfaces triangulées sont donc créées en reliant les points de l’espace par des triangles (Figure 2.5).

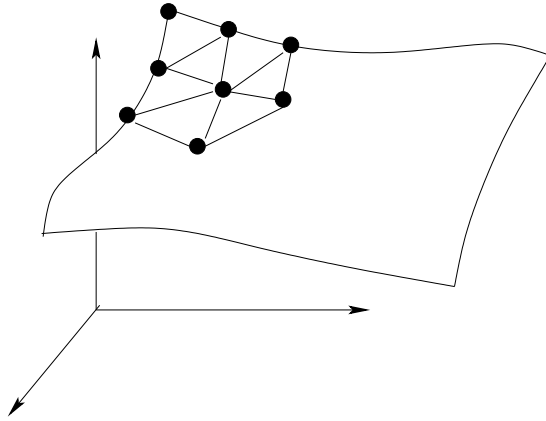


FIG. 2.5 – Surface triangulée.

Cette approche ressemble aux techniques d’éléments finis utilisées par exemple pour résoudre des problèmes comportant des équations différentielles. L’objectif est de trouver la bonne triangulation et le lissage adéquat pour le respect au mieux des données.

Il est intéressant ici de se pencher sur une approche originale. Depuis 1989, Jean-Laurent Mallet et son équipe de l’École Nationale Supérieure de Géologie de Nancy développent le modelleur *Gocad*. La méthode DSI (Discrete Smooth Interpolation) implémentée dans cet outil entre dans la catégorie des modèles déformables (Mallet, 2002). Une surface triangulée initiale s’ajuste aux données au cours d’un processus itératif utilisant la minimisation d’un critère de rugosité au sens des moindres carrés. La surface initiale possède une triangulation indépendante des données et à la fin du processus la méthode garantit une solution lisse. De plus, le DSI est parfaitement bien adapté au traitement de surfaces localement interrompues donc à la modélisation de zones très faillées. Cette approche puissante peut s’avérer néanmoins complexe à mettre en oeuvre dans des cas particulièrement ardu (Ledez, 2003).

De nombreux algorithmes de triangulation basés ou non sur la triangulation de Delaunay existent. De plus, lorsque l’on traite de reconstruction d’objets à partir de données discrètes, la cohérence topologique des modèles construits doit être assurée. Une solution est d’utiliser des techniques de partition cellulaire de type cartes généralisées ou G-Maps (Schneider, 2002). Halb wachs *et al.* (1996) les utilisent pour la représentation d’un réseau de failles.

Nous ne rentrerons pas dans le détail de ces méthodes mais nous allons nous attarder un peu plus sur la définition et la visualisation des surfaces implicites.

2.5 Méthodes volumiques et surfaces implicites

C'est ce type d'approche qui est mis en œuvre dans l'*Editeur*.

Notons que dans ce cadre des surfaces implicites l'interpolation peut se faire à l'aide de fonctions RBF (Radial Basis Functions) dont la formulation est identique à celle du krigeage dual (Turk et O'Brien, 2002), (Cowan *et al.*, 2003). On pourra aussi se reporter utilement à la thèse de Ledez (2003) traitant de la modélisation d'objets naturels par formulation implicite.

2.5.1 Définition

La géologie est décrite par une variable $f(x, y, z)$ de la manière suivante : une formation correspond à l'ensemble des points pour lesquels f est compris entre deux bornes f_{min} et f_{max} . Une interface est alors une surface implicite $\mathcal{S}$, c'est-à-dire définie comme l'ensemble des points (x, y, z) vérifiant $f(x, y, z) = V$ où V est une constante³. V est appelée une isovaleur du champ et est un concept familier des géoscientifiques (lignes de niveaux, isopaques, isopièzes... en 2D, isobares en 3D). Si on reprend l'exemple de la sphère présentée plus haut, sa définition implicite est la suivante :

$$(x - c_x)^2 + (y - c_y)^2 + (z - c_z)^2 - R^2 = 0$$

Il s'agit donc de trouver une expression calculable en tout point d'une fonction (dite implicite) inconnue $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ qui est donc un champ scalaire. On voit donc que deux étapes sont nécessaires pour construire une surface :

1. Trouver une expression de f de telle sorte que cette fonction champ soit calculable en tout point, c'est la phase d'interpolation du champ.
2. Rendre explicites les surfaces isovaleurs de l'interpolation ainsi que les volumes, c'est la phase de représentation.

Ce travail de thèse est centré sur la première étape, la seconde est explicitée au point 2.5.3. Pour chaque valeur de V , la fonction implicite f partitionne l'espace en deux sous-domaines : "intérieur" et "extérieur", ce qui permet de définir des objets de frontière la surface.

Si on compare cette approche à l'approche multicouche, on constate que z n'est plus la grandeur étudiée, mais la troisième coordonnée de l'espace et les structures complexes sont désormais modélisables. La Figure 2.6 qui présente un modèle de nappes dans les Alpes suisses⁴ a été réalisé par Maxelon avec l'*Editeur Géologique*, modeleur exploitant des surfaces implicites.

La méthode des potentiels qui est l'approche mise en œuvre dans l'*Editeur* est décrite au chapitre 3.

³ $\mathcal{S}$ est aussi appelée iso-surface de f .

⁴Pour toute information concernant la géologie de cet exemple, on pourra se reporter à Maxelon (2004).

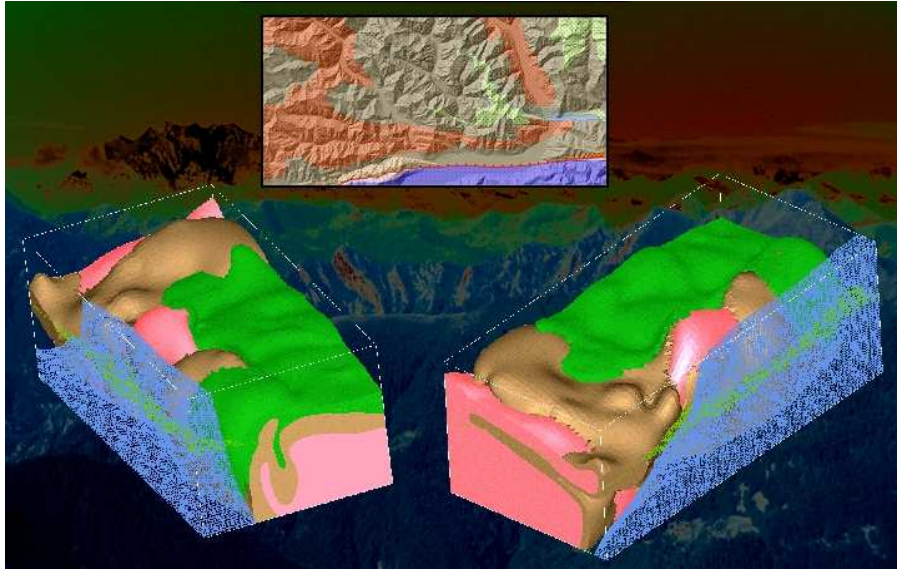


FIG. 2.6 – Exemple de modélisation par surfaces implicites : phases de déformation dans les Alpes centrales suisses. D'après Maxelon, 2004.

2.5.2 L'approche d'*Earth Vision*

Earth Vision, développé par la société *Dynamic Graphics*, est un logiciel de modélisation géologique qui utilise l'approche implicite. Il fonctionne de la manière suivante (Figure 2.7) :

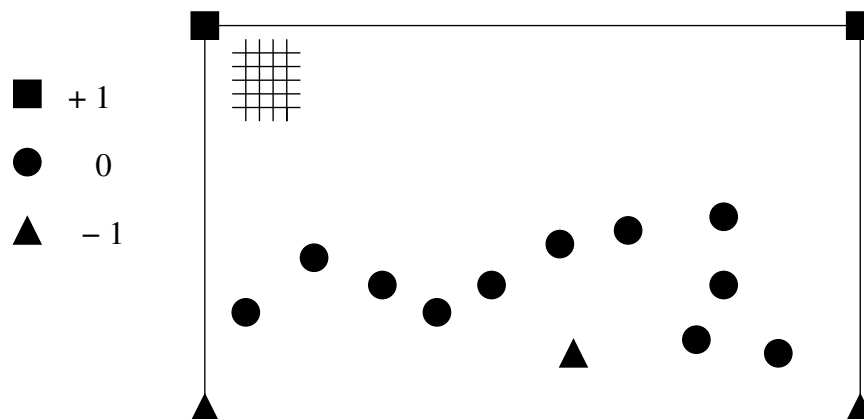
- Une grille cartésienne 3D, dans laquelle est définie une propriété, est construite autour des points de données.
- Ces points sont affectés de la valeur 0.
- Un sens doit être assigné à la propriété afin de distinguer l'intérieur de l'extérieur de l'objet. Cela se fait par l'intermédiaire de nœuds de contrôle, par exemple +1 à l'extérieur et -1 à l'intérieur.
- La propriété est interpolée dans la grille.

Cette méthode innovante et flexible a cependant des limites, car pour des géologies complexes, le nombre de nœuds de contrôle devient rapidement important. Enfin, la reconstruction de zones complexes (changements brusques de la courbure) n'est pas limitée par le nombre de données, mais par la taille de la grille.

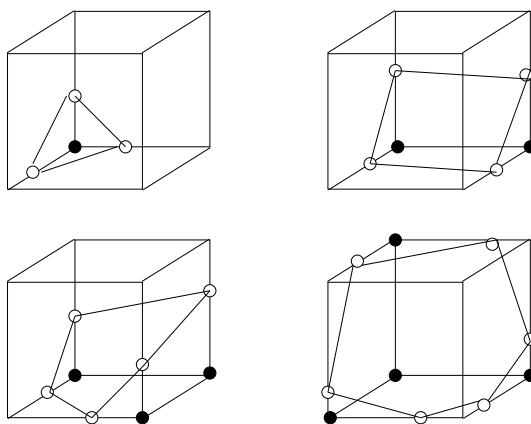
2.5.3 Visualisation

Contrairement aux surfaces paramétriques qui sont explicites, donc bien adaptées à la visualisation, pour les surfaces implicites il y a nécessité de représenter les surfaces isovaleurs. Donc, à partir du moment où le champ est interpolé, la difficulté est de visualiser l'isosurface considérée. Différentes méthodes existent comme le lancer de rayon ou les systèmes de particules, mais nous allons décrire ici la technique utilisée dans l'*Éditeur* et qui consiste à expliciter l'isosurface qui nous intéresse.

L'algorithme de triangulation *Marching Cubes* (Lorensen et Cline, 1987) construit une re-

FIG. 2.7 – Schéma de principe de l'outil *EarthVision*. D'après Ledez, 2003.

présentation polygonale de la surface à partir de voxels. Il est basé sur une décomposition en cellules hexaédriques faisant intervenir des triangles, quadrilatères, pentagones et hexagones (Fig. 2.8). De nombreuses améliorations ont été depuis ajoutées à l'algorithme initial. Il fonctionne ainsi : il évalue dans un premier temps la fonction implicite en tous nœuds d'un maillage lâche. L'éventualité de rencontre de la surface avec un cube initial est ensuite déterminée. Si c'est le cas, un polygone, intersection de la surface et du cube, est construit. On procède ensuite à un suivi de la surface en examinant les cubes voisins qui l'intersectent. La résolution du maillage est plus fine là où cela s'avère nécessaire comme lors de brusques changements de courbure. Toutes les configurations du problème (détermination de l'arête intersectée par la surface) et toutes les solutions associées (détermination des points à relier pour tracer le polygone *ad hoc*) sont répertoriées de façon optimisée dans deux tableaux. Ce qui en fait un algorithme simple et rapide.

FIG. 2.8 – Exemples de polygones de degré 3 à 6 créés par l'algorithme de *Marching Cubes*.

Chapitre 3

La méthode des potentiels dans *l'Éditeur Géologique*

L'*Éditeur Géologique*, développé par le BRGM dans l'environnement *CAS.CADE* de Matra Datavision et dans le cadre du projet de recherche "Géomodèle 3D", est un logiciel de construction automatique de modèles géologiques volumiques. L'idée est de modéliser les interfaces entre couches, frontières des objets volumiques, en utilisant le principe de construction des surfaces implicites en deux temps : d'abord, l'interpolation d'un champ scalaire assimilé à un champ de potentiel, puis la représentation des isopotentielles correspondant aux interfaces. En ce qui concerne la visualisation, elle peut se faire en 3D ou par la création de coupes dans le volume.

3.1 Le cadre de la géostatistique multivariable

La méthode des potentiels s'inscrit dans le cadre multivariable (Wackernagel, 2003), car différents types de données sont exploités conjointement afin de déterminer l'expression mathématique d'une surface géologique à modéliser. Le premier type de données sont des points, situés sur l'interface, et considérés comme ayant même potentiel ; on travaillera en fait sur des incréments de celui-ci. Les autres données sont angulaires et correspondent aux normales aux plans de stratification, de foliation etc., on les considèrera comme des gradients du potentiel. La méthode proprement dite est expliquée à la section 3.3 et suivantes, ce paragraphe-ci vise à rappeler quelques notions de géostatistique multivariable.

3.1.1 Variogrammes simples et croisés

Soit $\mathcal{D}$ le champ d'étude sur lequel sont définies N variables régionalisées $z_i(\mathbf{x})$. Celles-ci sont considérées comme des réalisations de fonctions aléatoires $Z_i(\mathbf{x})$. Si on se place dans l'hypothèse de stationnarité du second ordre conjointe, on a :

$$\left\{ \begin{array}{ll} E[Z_i(\mathbf{x})] = m_i & \forall i = 1, \dots, N, \quad \forall \mathbf{x} \in \mathcal{D} \\ Cov\{Z_i(\mathbf{x}), Z_j(\mathbf{x} + \mathbf{h})\} = E\{[Z_i(\mathbf{x} + \mathbf{h}) - m_i][Z_j(\mathbf{x}) - m_j]\} = C_{ij}(\mathbf{h}) & \forall i, j = 1, \dots, N, \quad \forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{D} \end{array} \right. \quad (3.1)$$

Ce qui définit les covariances simples ou directes $C_{ii}(\mathbf{h})$ et croisées $C_{ij}(\mathbf{h})$ $i \neq j$.

Dans le cadre de l'hypothèse intrinsèque conjointe, on définit cette fois les variogrammes simples $\gamma_{ii}(\mathbf{h})$ et croisés $\gamma_{ij}(\mathbf{h})$ $i \neq j$:

$$\begin{cases} E[Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{x})] = 0 & \forall i = 1, \dots, N, \quad \forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{D} \\ Cov\{Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{x}), Z_j(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{x})\} = 2\gamma_{ij}(\mathbf{h}) & \forall i, j = 1, \dots, N, \quad \forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathcal{D} \end{cases} \quad (3.2)$$

On en déduit une autre expression pour les variogrammes :

$$\gamma_{ij}(\mathbf{h}) = \frac{1}{2} E\{[Z_i(\mathbf{x} + \mathbf{h}) - Z_i(\mathbf{x})][Z_j(\mathbf{x} + \mathbf{h}) - Z_j(\mathbf{x})]\} \quad (3.3)$$

3.1.2 Le cokrigage simple

Le cokrigage consiste à estimer la valeur d'une variable à partir des mesures de cette variable ainsi que de celles des variables additionnelles. Il s'agit donc d'une extension naturelle du krigeage au cas multivariable.

Soit $Z_i(\mathbf{x})$ une fonction aléatoire. Pour commencer, nous allons nous placer dans le cas où les moyennes m_i sont connues¹ et travailler sur les variables centrées $Z_i - m_i$.

Ecrivons l'estimation d'une variable comme une combinaison linéaire de toutes les variables :

$$Z_{i_0}^*(\mathbf{x}_0) - m_{i_0} = \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i [Z_i(\mathbf{x}_{\alpha}^i) - m_i] \quad (3.4)$$

La condition de non-biais étant automatiquement satisfaite, les poids de cokrigage doivent être choisis tels que la variance de l'erreur d'estimation soit minimale. Ceci mène au système de cokrigage :

$$\sum_{j=1}^N \sum_{\beta=1}^{n_j} \lambda_{\beta}^j C_{ij}(\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j) = C_{ii_0}(\mathbf{x}_{\alpha}^i - \mathbf{x}_0) \quad \forall i = 1, \dots, N, \quad \forall \alpha = 1, \dots, n_i \quad (3.5)$$

Les propriétés du krigeage comme l'additivité ou l'interpolation exacte restent valables. Dans le cas de figure qui nous intéresse, on dispose de deux types de variables qui ne sont pas mesurées aux mêmes sites d'échantillonnage, on parle alors d'hétérotopie.

3.1.3 Le cokrigage universel

Plaçons-nous maintenant dans le cadre d'un phénomène avec dérive (modèle non-stationnaire). La dérive déterministe $m_i(\mathbf{x})$ est définie comme une combinaison linéaire de fonctions de base.

$$m_i(\mathbf{x}) = \sum_l a_l^i f_l^i(\mathbf{x}) \quad (3.6)$$

Deux cas se présentent à nous :

¹On parle de cokrigage simple.

- soit les fonctions de base sont indépendantes pour les deux variables ;
- soit elles sont liées.

Supposons dans un premier temps que les fonctions de dérive sont indépendantes. On utilise dans notre système des fonctions de base polynomiales (monômes de degré $\leq k$) avec $f_i^0(\mathbf{x}) \equiv 1 \quad \forall i = 1, \dots, N$, les fonctions f_i^l étant fixées et les coefficients a_i^l inconnus. On considère la fonction aléatoire $Z_i(\mathbf{x})$ comme étant la superposition de la dérive $m_i(\mathbf{x})$ et d'un résidu $Y_i(\mathbf{x})$:

$$\forall i = 1, \dots, N, \quad Z_i(\mathbf{x}) = Y_i(\mathbf{x}) + m_i(\mathbf{x}) \quad \text{avec} \quad m_i(\mathbf{x}) = \sum_l a_i^l f_i^l(\mathbf{x}) \quad (3.7)$$

L'hypothèse de stationnarité² d'ordre 2 (conjointe) se fait évidemment sur les résidus $Y_i(\mathbf{x})$. Les fonctions aléatoires $Y_i(\mathbf{x})$ et donc $Z_i(\mathbf{x})$ admettent des fonctions de covariance simples et croisées. L'espérance de $Y_i(\mathbf{x})$ étant inconnue, on l'inclut dans la dérive et la covariance de $Y_i(\mathbf{x})$ est égale à la covariance de $Z_i(\mathbf{x})$.

Remarque : Cette dichotomie entre une dérive et un résidu peut être interprétée comme la séparation d'une composante régionale, régulière et d'une composante résiduelle, erratique. L'estimateur de cokrigage universel s'écrit :

$$Z_{i_0}^*(\mathbf{x}_0) = \sum_{i=1}^N \sum_{\alpha=1}^{n_i} \lambda_{\alpha}^i Z_i(\mathbf{x}_{\alpha}^i) \quad (3.8)$$

Les pondérateurs de cokrigage sont choisis tels que l'estimation de l'erreur soit sans biais $\forall a_i^l$ et que la condition d'optimalité soit vérifiée. On aboutit ainsi au système suivant :

$$\begin{cases} \sum_{j=1}^N \sum_{\beta=1}^{n_j} \lambda_{\beta}^j C_{ij}(\mathbf{x}_{\alpha}^i - \mathbf{x}_{\beta}^j) + \sum_l \mu_l^i f_i^l(\mathbf{x}_{\alpha}^i) = C_{ii_0}(\mathbf{x}_{\alpha}^i - \mathbf{x}_0) & \forall i = 1, \dots, N, \quad \forall \alpha = 1, \dots, n_i \\ \sum_{\beta=1}^{n_i} \lambda_{\beta}^i f_i^l(\mathbf{x}_{\beta}^i) = \delta_{i_0}^i f_{i_0}^l(\mathbf{x}_0) & \forall l, \quad \forall i = 1, \dots, N \end{cases} \quad (3.9)$$

Avec $\delta_{i_0}^i$ valant 1 si $i = i_0$, 0 sinon.

Or, comme on le verra par la suite dans notre cas d'étude, la variable principale et la variable additionnelle sont liées. Il en est donc de même pour les fonctions de dérive.

Prenons l'exemple, à une dimension, d'une variable principale Z_1 et d'une variable additionnelle Z_2 telle que : $Z_2(x) = Z_1'(x)$. Les fonctions de base de la deuxième variable sont aussi des dérivées des fonctions de la première variable :

$$f_2^l(x) = \frac{d}{dx} f_1^l(x) \quad (3.10)$$

Par conséquent, les pondérateurs de cokrigage doivent être tels que, pour toute fonction de dérive f_1^l :

$$f_1^l(x_0) = \sum_{\beta=1}^{n_1} \lambda_{\beta}^1 f_1^l(x_{\beta}^1) + \sum_{\beta=1}^{n_2} \lambda_{\beta}^2 \frac{d}{dx} f_1^l(x_{\beta}^2) \quad (3.11)$$

²On parlera de covariance sous-jacente stationnaire.

On en déduit que dans cet exemple particulier, le système est de la forme :

$$\left\{ \begin{array}{l} \sum_{\beta=1}^{n_1} \lambda_{\beta}^1 C_{11} (x_{\alpha}^1 - x_{\beta}^1) + \sum_{\beta=1}^{n_2} \lambda_{\beta}^2 C_{12} (x_{\alpha}^1 - x_{\beta}^2) + \sum_l \mu_l^1 f_1^l (x_{\alpha}^1) = C_{11} (x_{\alpha}^1 - x_0) \\ \forall \alpha = 1, \dots, n_1 \\ \sum_{\beta=1}^{n_1} \lambda_{\beta}^1 C_{21} (x_{\alpha}^2 - x_{\beta}^1) + \sum_{\beta=1}^{n_2} \lambda_{\beta}^2 C_{22} (x_{\alpha}^2 - x_{\beta}^2) + \sum_l \mu_l^2 \frac{d}{dx} f_1^l (x_{\alpha}^2) = C_{21} (x_{\alpha}^2 - x_0) \\ \forall \alpha = 1, \dots, n_2 \\ \sum_{\beta=1}^{n_1} \lambda_{\beta}^1 f_1^l (x_{\beta}^1) + \sum_{\beta=1}^{n_2} \lambda_{\beta}^2 \frac{d}{dx} f_1^l (x_{\beta}^2) = f_1^l (x_0) \quad \forall l \end{array} \right. \quad (3.12)$$

3.2 Rappels sur la notion de champ et de gradient

3.2.1 Définitions

Une notion importante en physique est celle de champ (scalaire), comme par exemple le champ de potentiel électrostatique ou le champ de température. La définition formelle d'un champ est la suivante : le champ d'une grandeur P dans un domaine $\mathcal{D}$ de l'espace à un instant t est défini par la donnée de $P(M, t)$ en tout point M de $\mathcal{D}$.

Dans notre cas de figure, on considère le champ comme permanent (ou stationnaire), c'est-à-dire que sa valeur en chaque point ne dépend pas du temps t : $P(M, t) = P(M)$. De plus, il doit être différentiable puisque nous sommes amenés à travailler avec des dérivées du potentiel. Nous allons assimiler par la suite la notion de champ en tant que domaine comprenant l'ensemble des points où les valeurs du potentiel sont considérées et en tant qu'ensemble des valeurs prises en chaque point du domaine. De même, on ne fera pas la distinction entre la grandeur scalaire et la fonction qui à un point de l'espace associe cette grandeur. Par convention, nous appellerons par la suite cette grandeur "potentiel", bien qu'il n'existe pas réellement un potentiel au sens géologique du terme. Le nom vient du fait que P représente le potentiel du champ vectoriel **grad** P .

Définissons donc maintenant l'opérateur gradient qui sera abondamment utilisé ultérieurement. Il permet d'exprimer la variation dP de $P(M)$ lors d'un déplacement élémentaire $d\mathbf{l}$ de M . On a :

$$dP = \mathbf{grad}P(M) \cdot d\mathbf{l} \quad (3.13)$$

On notera désormais **grad** P par la lettre **G**. Dans un repère cartésien, les coordonnées de **G** sont :

$${}^t\mathbf{G} = (G^x, G^y, G^z) = \left(\frac{\partial P}{\partial x}, \frac{\partial P}{\partial y}, \frac{\partial P}{\partial z} \right) \quad (3.14)$$

L'opérateur gradient transforme donc le champ de potentiel (scalaire) en un champ de gradient (vectoriel).

Enfin, une ligne de champ d'un champ vectoriel est une courbe tangente en tout point au champ.

3.2.2 Propriétés

Dans la méthode des potentiels nous allons tirer parti du formalisme de cette notion de champ. Ainsi, l'ensemble des points M tels que $P(M) = V$ où V est une constante définit une surface. De plus, si on considère un point M sur une telle surface, le vecteur $\mathbf{G}(M)$ est normal à cette surface et est dirigé dans le sens des $\mathbf{G}$ croissants. Nous choisirons d'orienter arbitrairement le potentiel vers le haut, c'est-à-dire des terrains les plus anciens aux plus récents dans le cas d'une série normale. Si on fait une analogie avec la topographie, $P(M)$ est à comparer à l'altitude du point M et les lignes de champ de $\mathbf{G}(M)$ s'apparentent aux lignes de plus grande pente, orthogonales aux lignes $P(M) = V$ qui correspondent dans ce cas aux lignes de niveau.

3.3 Présentation de la méthode des potentiels

3.3.1 Originalité

L'intérêt principal de la méthode implémentée dans l'*Éditeur* est de construire des surfaces à partir de quelques points connus sur une ou plusieurs surfaces et quand des données additionnelles d'orientation sont disponibles. Alors que les méthodes classiques imposent que ces contraintes d'orientation soient aussi des points de contraintes de la surface, la méthode des potentiels n'a pas cette exigence. Or, cette obligation se rencontre peu dans la Nature. En effet les données d'orientation ne sont pas nécessairement situées sur les interfaces, mais échantillonnent l'anisotropie principale de la formation (stratification, foliation, plan de clivage).

3.3.2 Les hypothèses

Nous supposons que l'interface à modéliser - ou les interfaces si plusieurs sont représentées par le même champ de potentiel - appartient à une famille de surfaces subparallèles suivant à peu de choses près les données structurales. Cette hypothèse est vérifiée la plupart du temps sur le terrain aussi bien pour des terrains sédimentaires, plutoniques ou métamorphiques. En effet, dans le cas des bassins sédimentaires, il existe un parallélisme général entre les plans de stratification et les limites lithologiques. De même, dans le cas de granites par exemple il y a une similarité entre la forme de la frontière de l'intrusion granitique et les structures internes ; celles-ci servent d'ailleurs à déterminer la forme du pluton. Le parallélisme est également fort pour les roches métamorphiques entre les principales interfaces lithologiques et les foliations métamorphiques.

Enfin nous supposons que les données d'orientation puissent être transformées en données de gradient. Pour cela, la donnée de la polarité des structures est nécessaire. La polarité indique le sens (intérieur/extérieur) de la formation par rapport à la donnée d'orientation. Le module du gradient est constant et fixé arbitrairement à 1 par construction. La validité de cette dernière hypothèse est discutée au chapitre 8.

3.3.3 Les principes

On choisit de représenter les limites géologiques par des fonctions implicites. Les fondements de la méthode des potentiels ont été élaborés par Lajaunie *et al.* (1997).

Un point dans l'espace 3D est défini par le triplet (x, y, z) de ses coordonnées. On dispose de deux types de données (Figure 3.1) : des points de passage sur les interfaces géologiques et des données ponctuelles d'orientation qui sont des vecteurs normaux aux lignes isovaleurs (donnée : azimuth, angle, polarité). Le principe de cette technique est d'interpréter les points d'une interface comme appartenant à une isopotentielle et d'assimiler les directions de pendage à des gradients, c'est-à-dire des dérivées du potentiel. Notre objectif est donc d'interpoler tridimensionnellement le potentiel (Figure 3.2). On va effectuer ainsi un cokri-

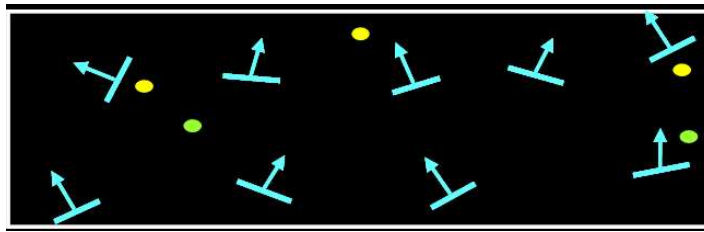


FIG. 3.1 – Schéma du principe de la méthode d'interpolation de l'*Editeur*. Avant interpolation.

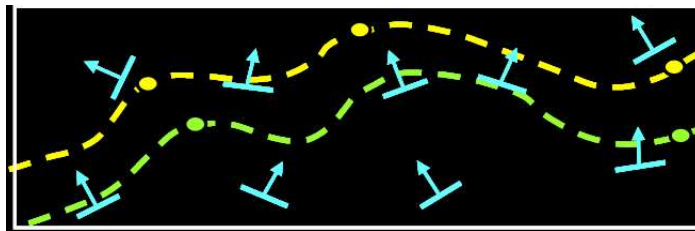


FIG. 3.2 – Schéma du principe de la méthode d'interpolation de l'*Editeur*. Après interpolation.

geage universel du potentiel (cf. section suivante) à partir de ces deux types de données que nous appellerons souvent par la suite données d'interface et données de gradient. D'autant plus que l'intérêt de la méthode repose sur l'utilisation de ces gradients qui apportent une information structurale à la modélisation, qui sont nombreuses et faciles à mesurer sur le terrain. Ce sont elles qui apportent une contribution non nulle au résultat comme on va le voir par la suite.

A partir du moment où le potentiel est connu, il suffit de se représenter l'isopotentielle correspondant à l'interface. En définitive, nous avons déterminé des surfaces qui passent par les points d'interface et respectent globalement le champ directionnel.

3.4 L'interpolation par cokrigeage

3.4.1 Les données

Tout d'abord, considérons le potentiel comme une réalisation d'une fonction aléatoire différentiable Z . Tous les points appartenant à une interface sont à un même potentiel, dont la valeur est inconnue. Les données numériques qu'on peut en déduire sont des accroissements de potentiel entre deux points de l'interface, tous de valeur nulle. Les données étant soit des incréments, soit des gradients de potentiel, celui-ci ne peut être connu qu'à une constante près. On estimera donc le potentiel à une constante près, ou, ce qui revient au même, l'accroissement de potentiel $Z(x) - Z(x_0)$ par rapport à une origine arbitraire x_0 . Ce point que l'on considère comme le point de "référence", sert d'origine au potentiel. L'incrément de potentiel entre deux points situés sur une même interface est donc nul.

Soit I une interface et J un jeu de n points sur cette interface et soit m le nombre de données d'orientation pour ce potentiel. Un point de l'interface sera noté $\mathbf{x}_j$ ($1 \leq j \leq n$). Nous utilisons donc $n - 1$ incréments de potentiel linéairement indépendants et tous égaux à 0. En terme de fonctions aléatoires, ces incréments sont de la forme :

$$Z(\mathbf{x}_j) - Z(\mathbf{x}_{j'}) = 0$$

par exemple :

$$Z(\mathbf{x}_j) - Z(\mathbf{x}_{j-1}) = 0 \quad j = 2, \dots, n \quad (3.15)$$

Si plusieurs interfaces sont modélisées avec le même champ de potentiel, le jeu de données J est l'union des jeux élémentaires relatifs aux différentes interfaces. De même, comme les données d'orientation sont considérées comme une réalisation de la fonction aléatoire gradient de Z , on a :

$$\frac{\partial Z}{\partial x}(\mathbf{x}_i) = G_i^x, \quad \frac{\partial Z}{\partial y}(\mathbf{x}_i) = G_i^y, \quad \frac{\partial Z}{\partial z}(\mathbf{x}_i) = G_i^z \quad (3.16)$$

avec $\mathbf{x}_i$ le point où le gradient est mesuré et G_i^x, G_i^y, G_i^z les composantes de ce gradient.

3.4.2 La résolution du système

Rappelons que l'objectif est de tracer des isosurfaces de potentiel correspondant à des toits ou à des bases de formations géologiques. Ce problème est résolu en interpolant la valeur du potentiel en tout point de l'espace et en traçant la surface isopotentielle qui nous intéresse. Le domaine interpolé est évidemment discrétisé, le potentiel est calculé aux noeuds d'une grille qui peut être resserrée si nécessaire.

Exprimons donc l'incrément du potentiel en fonction des données disponibles :

$$[Z(\mathbf{x}) - Z(\mathbf{x}_0)]^* = \sum_i (\lambda_i G_i^x + \mu_i G_i^y + \nu_i G_i^z) + \sum_j \tau_j [Z(\mathbf{x}_j) - Z(\mathbf{x}_{j-1})] \quad (3.17)$$

Même si le dernier terme de cette expression est nul, il intervient néanmoins dans le système de krigeage, car les pondérateurs λ_i, μ_i, ν_i sont différents de ceux calculés sur les

données de gradient seules.

A partir de ces données et d'un modèle de covariance particulier dont on reparlera abondamment au chapitre 4, on forme la matrice de covariance de notre système de cokrigage. On parle de cokrigage, car nous sommes dans un cas multivariable avec une variable primaire, celle que l'on cherche à estimer autrement dit le potentiel ou son incrément, et une variable secondaire qui sont les données d'orientation assimilées au gradient du potentiel. Pour comprendre l'architecture de la matrice de covariance on se référera utilement à l'annexe A. Cette annexe montre aussi le système de cokrigage universel dual tel qu'il est exploité dans le code, et qui permet d'exprimer l'estimation du potentiel en un point sous la forme :

$$\begin{aligned} Z^*(x) = & A_0 + \sum_i (a_i K_{ZG^x}(\mathbf{x} - \mathbf{x}_i) + b_i K_{ZG^y}(\mathbf{x} - \mathbf{x}_i) + c_i K_{ZG^z}(\mathbf{x} - \mathbf{x}_i)) \\ & + \sum_j d_j (K_Z(\mathbf{x} - \mathbf{x}_j) - K_Z(\mathbf{x} - \mathbf{x}_{j-1})) + \sum_l e_l f_l(\mathbf{x}) \end{aligned} \quad (3.18)$$

A_0 est une constante qui représente $Z^*(x_0)$ et peut être prise égale à 0.

$K_Z(\mathbf{x} - \mathbf{x}_j) = Cov(Z(\mathbf{x}), Z(\mathbf{x}_j))$ et $K_{ZG^x}(\mathbf{x} - \mathbf{x}_i) = Cov(Z(\mathbf{x}), Z'_x(\mathbf{x}_i))$ sont les fonctions de covariance simple du potentiel d'une part, et croisée du potentiel et de sa dérivée partielle selon x , d'autre part.

3.4.3 Remarques sur le cokrigage dual

Le cokrigage dual a été préféré à la formulation habituelle du cokrigage, car il est plus adapté à la cartographie, l'équation d'estimation donnant au moins sous forme implicite l'équation des lignes de niveau (Galli *et al.*, 1984). Je rappelle qu'une ligne de niveau Z_0 est le lieu des points $\mathbf{x}$ tels que : $Z^*(x) = Z_0$. Le traitement du problème par l'approche duale est plus rapide, car une seule résolution du système est entreprise. En effet, on travaille en voisinage unique afin d'éviter les artefacts de changement de voisinage. On calcule donc ces coefficients une fois pour toute, l'estimation revenant alors à un simple produit scalaire. Le gain est à la fois en terme de quantité de calculs et en mémoire utilisée. Par contre, on ne dispose pas de la variance de cokrigage.

La méthode de résolution mise en oeuvre ne procède pas par la recherche de pivot maximal, mais par la prise du pivot sur la diagonale, ce qui ne pose aucun problème si la matrice est définie positive. Or la matrice de cokrigage ne l'est pas en présence de fonctions de dérive. D'où une résolution en deux temps où l'on résout d'abord le système sans les conditions d'universalité, puis un système de dimension égale au nombre de conditions d'universalité et dont le premier membre est une matrice définie positive (Chilès et Delfiner, 1999).

3.4.4 Quelques précisions

Il faut noter que dans l'algorithme utilisé dans le logiciel *Editeur Géologique*, les données sont centrées normées, afin de résoudre d'éventuels problèmes de stabilité numérique dus à des valeurs très étalées. Les coordonnées de toutes les données sont donc comprises entre 0 et 1. Ainsi $X(i)$ est remplacé par $\frac{X(i) - X_c}{c} + \frac{1}{2}$ avec $X_c = \frac{X_{min} + X_{max}}{2}$ et

$c = 2 * \text{Max}(X_{max} - X_c, Y_{max} - Y_c, Z_{max} - Z_c)$. Les portées du modèle de covariance sont aussi divisées par c .

Lors du calcul d'interpolation, il est possible de réduire le temps de calcul (mais aussi la précision!) par l'intermédiaire d'un rayon de simplification. Celui-ci permet le regroupement en un seul point de tous les points situés dans un disque de rayon R , les valeurs concernées étant moyennées.

3.5 La gestion de plusieurs formations

Nous nous sommes intéressés jusqu'à présent essentiellement à la géométrie d'une interface. Conjointement à l'interpolation des données, des éléments spécifiques à la géologie ont été mis en place. Ceci permet de rendre le modèle réaliste et topologiquement cohérent, en particulier lorsque plusieurs formations doivent être modélisées et plus précisément de trancher le cas où les interfaces de deux couches sont sécantes (Courrioux et Calcagno, 2001).

3.5.1 La pile stratigraphique et les règles géologiques

La pile stratigraphique indique la chronologie des dépôts et obéit au "principe de superposition". S'il n'y a pas d'inversion due à la tectonique, les terrains les plus anciens sont en bas de la pile, les plus récents en haut. Lorsque les couches sont sub-horizontales, un seul champ de potentiel suffit à les modéliser. Les toits et murs des différentes formations sont alors des isopotentiellelles particulières de ce champ.

Par contre, quand cette situation n'est plus de mise, il faut utiliser autant de champs de potentiel qu'il existe de groupes d'interfaces sécantes. On définit une série comme un ensemble de couches (formations) concordantes et on utilise un champ de potentiel par série. La Figure 3.3 montre une pile stratigraphique avec trois séries F1 la plus ancienne, F2, et F3 la plus récente. La série F1 contient trois formations (F1-1, F1-2, F1-3), F2 deux (F2-1 et F2-2) et F3 une seule (F3). Dans ce dernier cas, série et formation sont confondues.

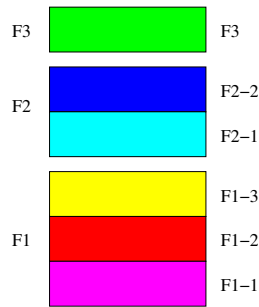


FIG. 3.3 – Exemple de pile stratigraphique.

Il reste à imposer des règles spécifiques à la géologie afin de fixer les relations entre séries et résoudre ainsi les problèmes de contacts entre elles.

Généralement, une transgression (avancée de la mer) se traduit par la discordance nette de

terrains “transgressifs” sur les couches antérieures basculées ou plissées puis érodées avant l’arrivée de la mer. On définit donc deux règles : érosion (dénommée “erod”) et dépôt (“onlap”). Elles peuvent apparaître simplistes au regard des phénomènes géologiques existants, mais il est néanmoins possible de résoudre avec elles toutes les situations rencontrées. La surface dénommée “erod” est la dernière interface avant la transgression et fusionne en fait deux réalités. Elle est à la fois le toit - dont on n’a aucune information - de la dernière couche déposée puis érodée et la surface d’érosion de l’ensemble des couches précédentes. On peut voir la pile comme une succession d’interfaces de type “erod” encadrant des interfaces de type “onlap”. Une surface de type “erod” est toujours prioritaire sur une surface “onlap”. De même, entre deux surfaces “erod”, la plus récente est prioritaire. La première et la dernière interface de la pile sont “erod” pour des raisons algorithmiques.

Le paragraphe suivant va illustrer sur des cas simples l’emploi de ces éléments.

3.5.2 L’utilisation de plusieurs champs de potentiel et la détermination des formations

Sur la Figure 3.4, les six interfaces représentées appartiennent à 4 champs de potentiel différents. La topographie correspond à l’interface noire. Par commodité et comme l’objectif n’est pas d’effectuer une interprétation géologique des modèles créés, toutes les interfaces seront nommées d’après le nom de la formation “sous-jacente” ; on les considérera toutes comme des toits, même lorsqu’il y a amalgame avec une surface d’érosion. On peut dire que les toits de A et B appartiennent au même champ de potentiel. Il en est de même pour les toits de D et E. Les 2 autres potentiels sont les horizons C et F.

Le toit de C (interface mauve) est “erod” et les toits de D et E (interfaces rose et rouge) sont en “onlap” sur lui. Les toits de A et B (interfaces vertes) sont aussi qualifiées “d’onlap”. Le toit de F (interface jaune) est aussi “erod”, mais comme il est plus récent que l’horizon C, ce dernier s’arrête sur lui (principe de superposition : “ce qui recoupe est plus récent que ce qui est recoupé”).

Regardons maintenant comment il est possible de déterminer dans quelle formation géologique se trouve un point de l’espace. Prenons le cas d’un seul champ de potentiel. Il suffit alors de comparer la valeur du potentiel au point considéré et les isovaleurs des différentes interfaces. A la Figure 3.5, le point M appartient à la formation 3, car le potentiel en ce point $P(M)$ est tel que : $v_2 < P(M) < v_3$ avec $V3$ valeur du potentiel en tout point du toit de la formation 3 et $V2$ valeur du potentiel en tout point du mur de la formation 3. Prenons maintenant le cas de deux interfaces sécantes (Figures 3.6 et 3.7). L’interpolation nécessite l’utilisation de deux champs de potentiel distincts et des règles spécifiques. A la Figure 3.6, la formation f2 est “erod” et on cherche à déterminer dans quelle formation se trouve un point M présent dans chaque compartiment créé par les deux interfaces. Soit P_1 et P_2 les potentiels relatifs aux deux interfaces et v_1 et v_2 les valeurs des isopotentielle correspondantes. On commence par comparer $P_2(M)$ et v_2 . Si $P_2(M) > v_2$ alors M appartient à la formation f3. Sinon, on effectue la comparaison entre $P_1(M)$ et v_1 . Ainsi, si $P_1(M) > v_1$ alors M est dans f2, sinon il est dans f1.

On fait de même pour la Figure 3.7 dans le cas où f2 est “onlap”. Par contre, on commence ici par comparer $P_1(M)$ et v_1 . Si $P_1(M) < v_1$ alors M appartient à la formation f1. Sinon,

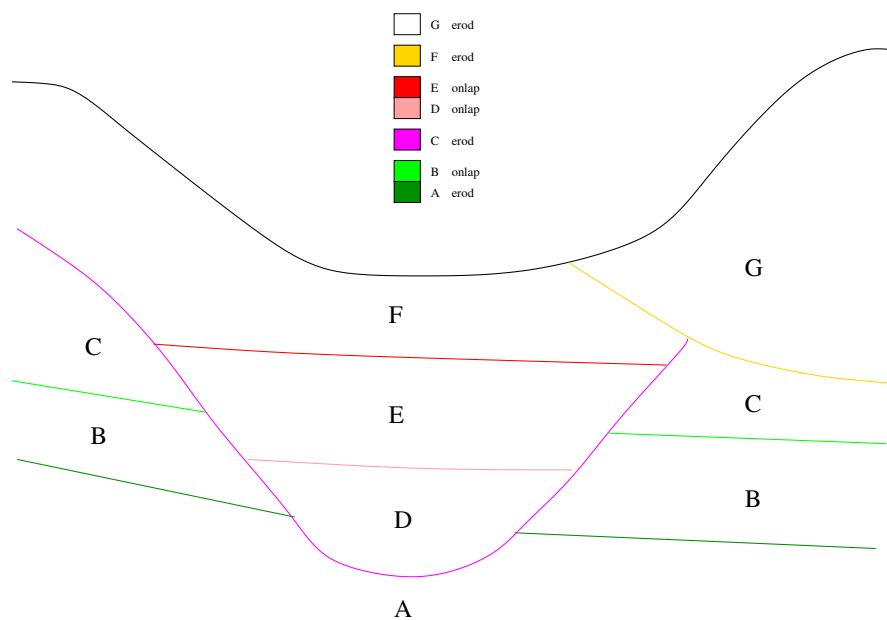


FIG. 3.4 – Coupe géologique : relations entre les champs de potentiel. On dispose de 7 couches géologiques de A à G, les interfaces étant matérialisées par des lignes en couleur. La topographie est la ligne supérieure en noir.

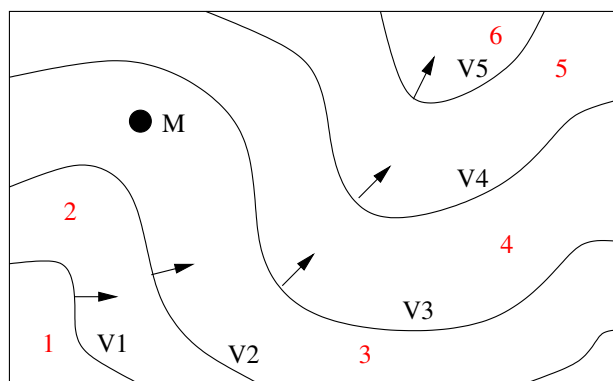


FIG. 3.5 – Détermination de la formation dans le cas d'un seul champ de potentiel.

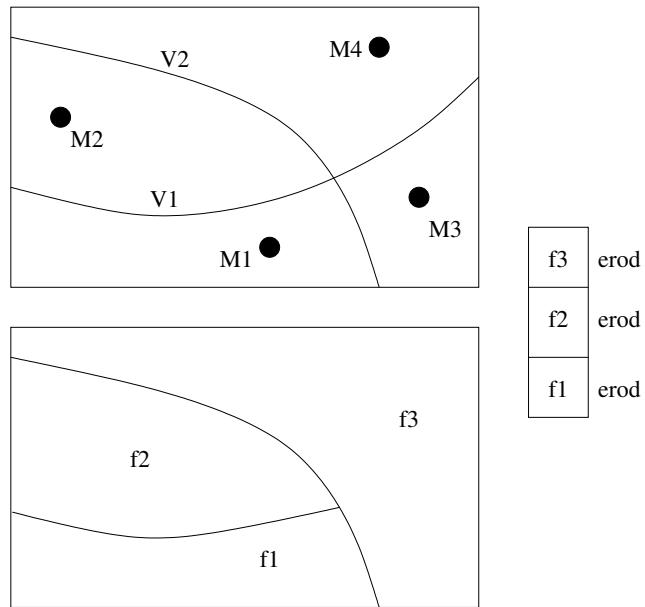


FIG. 3.6 – Traitement des relations et détermination de la formation pour 2 interfaces sécantes avec une relation de type “erod”.

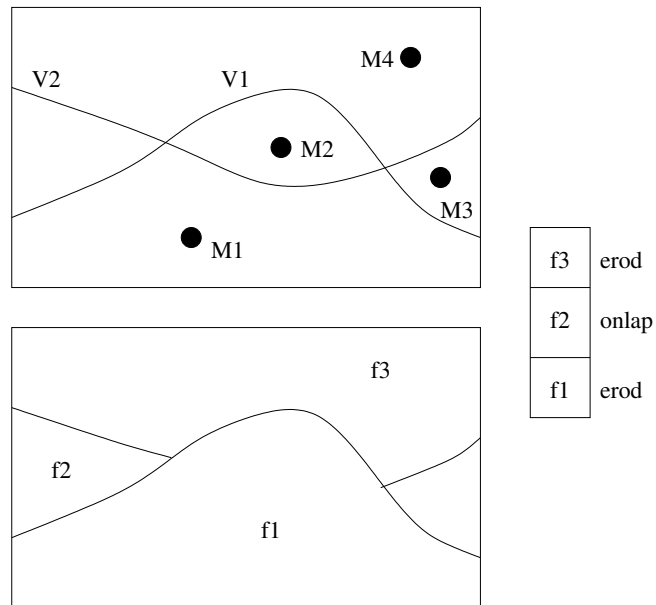


FIG. 3.7 – Traitement des relations et détermination de la formation pour 2 interfaces sécantes avec une relation de type “onlap”.

si $P_2(M) < v_2$ alors M est dans f2, sinon il est dans f3.

3.6 Non-stationnarité et discontinuités par fonctions de dérive

3.6.1 Prise en compte de la non-stationnarité

La fonction aléatoire “potentiel” peut être non stationnaire, c’est-à-dire que son espérance varie spatialement. On dit qu’on est en présence de dérive. On considère une dérive du type $m(\mathbf{x}) = \sum_l e_l f^l(x)$ et des fonctions de dérive f_l qui sont des monômes de degré d avec $0 \leq d \leq 2$.

Comme la verticale joue généralement un rôle particulier en géologie, le degré de la dérive polynomiale est souvent plus grand verticalement qu’horizontalement et la covariance est anisotrope. Dans le cas assez fréquent d’un modèle multicouche où les interfaces sont sub-horizontales et subparallèles, il est logique de considérer une dérive verticale linéaire du type : $m(\mathbf{x}) = e_0 + e_1 z$ avec les fonctions de dérive de base $f^0(\mathbf{x}) = 1$ et $f^1(\mathbf{x}) = z$. Un corps géologique avec une forme ellipsoïdale (ex : intrusion granitique) correspondra plutôt à une dérive quadratique³. Enfin, il faut noter que la fonction de dérive $f^0(\mathbf{x}) = 1$ peut être abandonnée dans tous les cas, car les incréments de potentiel tout comme les dérivées partielles filtrent e_0 .

Pour une présentation plus détaillée du cokrigage universel on se reportera au point 3.1.3 et aussi à l’annexe A pour l’incorporation dans le système matriciel.

3.6.2 Prise en compte des failles

De plus, il est possible également de tenir compte des discontinuités, c’est-à-dire des failles. Dans le cadre de la méthode des potentiels, une faille est une discontinuité non pas d’altitude, mais du potentiel. Les discontinuités sont connues en un certain nombre de points et sont considérées comme des interfaces à part entière. Elles sont interpolées au préalable avec pour chacune un champ de potentiel distinct. Il faut ensuite, lorsqu’on modélise les toits et murs des formations, prendre en compte l’effet de la discontinuité sur la valeur du potentiel. Pour cela, on utilise des fonctions de dérive dans le système de cokrigage, on parle alors de fonctions de faille.

Nous allons voir à présent comment les failles sont décrites dans l’*Editeur*, cependant il existe d’autres façons de les définir. La faille dite “infinie” traverse tout le champ partageant l’espace en deux, un bloc surélevé et/ou un bloc abaissé. On utilise pour cela une classique fonction créneau avec 1 d’un côté de la discontinuité et 0 de l’autre côté.

Mais, on peut aussi avoir une faille “finie”, c’est-à-dire qui n’affecte pas toute la zone d’étude. La description de ce type de discontinuité est en 3D et non en 2D. Ce type d’accident est modélisé par une fonction discontinue avec comme valeur 0 d’un côté de la discontinuité et une décroissance de 1 à 0 de l’autre côté en fonction des distances à la faille et au bord de celle-ci. La fonction de faille matérialise de cette manière la zone d’influence de

³Dans $\mathbb{R}^3$, il y a 10 fonctions de base dont le degré est inférieur ou égal à 2.

la discontinuité (Figures 3.8, 3.9 et 3.10). Le type de faille (son mode de fonctionnement) influe aussi sur le choix de la fonction à utiliser. Par exemple, si une translation rigide est supposée, on choisira plutôt une fonction décroissante seulement en fonction du bord de la faille.

Concrètement, on définit une faille par plusieurs paramètres :

- son bord (ou trace) formé par des points ;
- son rejet le long de la ligne de bord qui traduit la discontinuité ;
- un vecteur plongement qui combiné avec le bord forme la surface de faille (ou miroir) ;
- enfin, un vecteur d'atténuation qui donne la direction d'influence de la faille. Cette influence s'atténue progressivement (définition d'une fonction d'atténuation) suivant la direction de ce vecteur. Surface et vecteur d'atténuation définissent le volume d'action de la faille.

3.7 L'environnement *Open CAS.CADE*

Pour finir ce chapitre consacré à la méthode des potentiels et à l'*Editeur Géologique*, l'environnement *Open CAS.CADE* de l'*Editeur* est succinctement présenté. *Open CAS.CADE*, version “*open source*” de *CAS.CADE*, est une plate-forme de développement pour la création de logiciels de modélisation 3D et une bibliothèque de classes C++. De nombreux modules nécessaires à l'élaboration d'une application de modélisation 3D sont disponibles : gestion des objets géométrique (courbes, surfaces, volumes) et de la topologie, algorithmes de modélisation, visualisation, échange de données, classes de fondation etc. Cette bibliothèque est stable et robuste, mais son fonctionnement peut s'avérer lourd et complexe. Signalons également que le langage de spécification CDL gère automatiquement le squelette des classes ce qui est gain de temps et d'efficacité considérable. Enfin, un environnement de développement particulier WOK permet d'optimiser l'utilisation d'*Open CAS.CADE*. Lors du travail de thèse, certains modules, développés pour répondre à certaines questions posées, ont ainsi été directement intégrés dans une version “recherche” de l'*Editeur*.

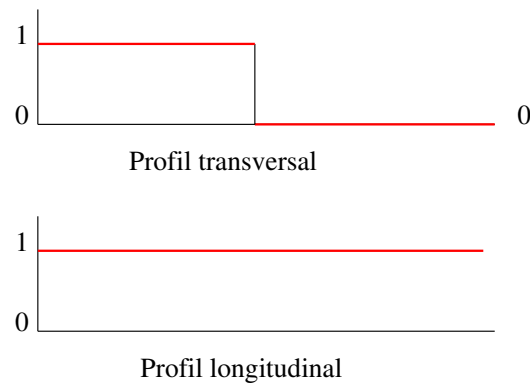


FIG. 3.8 – Exemple de modèle d'influence pour les failles infinies. Coupe transversale en haut, coupe longitudinale en bas.

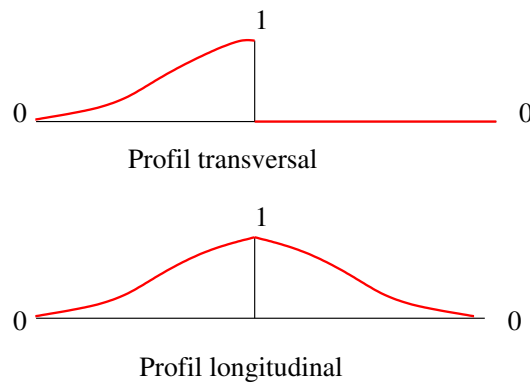


FIG. 3.9 – Exemple de modèle d'influence pour les failles finies. Coupe transversale en haut, coupe longitudinale en bas.

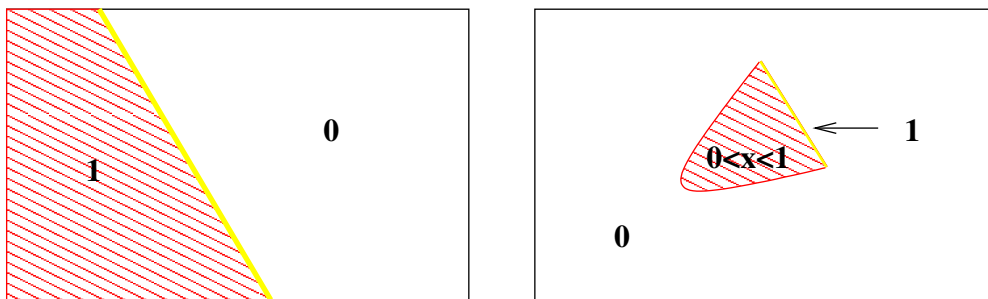


FIG. 3.10 – Zones d'influence des failles vues de haut, faille infinie à gauche, faille finie à droite.

Chapitre 4

Le modèle de covariance

L'interpolation du potentiel par cokrigage exige la connaissance de la covariance du potentiel¹. La démarche habituelle consiste à ajuster un modèle de covariance ou de variogramme expérimental de la variable étudiée. Elle n'est pas applicable ici, car on ne dispose pas de mesures directes du potentiel : les seules données de potentiel disponibles sont des accroissements entre des points situés sur une même interface, c'est-à-dire des accroissements qui se trouvent être nuls par construction et sont donc impropres à un calcul de variogramme.

Deux solutions s'offrent alors à nous :

1. Choisir un modèle *a priori*, c'est la façon dont a été développé l'*Editeur* ; le cokrigage est alors utilisé de manière heuristique, sans prétendre à une quelconque optimalité, même si le choix des paramètres peut être rationalisé.
2. Exploiter le lien entre le potentiel et son gradient pour inférer la covariance du potentiel à partir des variogrammes expérimentaux des données de gradient ; c'est un objectif important de ce travail, qui permet ensuite de qualifier la précision du modèle géologique construit.

4.1 Utilisation des dérivées

4.1.1 Le cas de la méthode des potentiels

L'interpolation du potentiel est un cokrigage, car elle exploite deux variables : des données de potentiel et des données de gradient. Or, notre cadre de travail est un peu particulier par rapport au cokrigage classique, car les variables utilisées sont algébriquement liées : les composantes du gradient sont les dérivées partielles du potentiel. Ce cas de dépendance intervient par exemple pour la profondeur et le pendage mesuré en forage

¹Si on se place dans le cadre de l'hypothèse de stationnarité d'ordre 2, on parlera aussi bien de covariance que de variogramme. Le terme covariance sera plus spécifiquement utilisé pour la théorie, et variogramme pour les applications. En effet, on calcule des variogrammes expérimentaux qui sont ensuite modélisés, mais c'est une matrice de covariance qui est employée dans l'algorithme de cokrigage implémenté dans l'*Editeur*.

(Renard et Ruffo, 1993) ou pour le géopotentiel et les composantes du vent en météorologie (Chauvet *et al.*, 1976). Cette relation fonctionnelle se répercute sur les covariances.

Théoriquement, la covariance croisée entre une variable aléatoire et une de ses dérivées partielles est impaire, donc nulle à l'origine. Le modèle linéaire de corégionalisation, souvent utilisé en multivariable, ne peut donc pas s'appliquer. Le modèle multivariable dérive en fait intégralement de la covariance de $Z(x)$ qui elle peut être ajustée classiquement par un modèle gigogne.

Avant de voir plus concrètement le modèle de covariance utilisé dans la méthode des potentiels, nous allons définir la notion de dérivée dans le cadre des fonctions aléatoires.

4.1.2 Différentiabilité d'une fonction aléatoire

Définissons la différentiabilité d'une fonction aléatoire Z au point x de $\mathbb{R}$ au sens de la moyenne quadratique :

$$\exists Z'(x); \lim_{h \rightarrow 0} E \left[\frac{Z(x+h) - Z(x)}{h} - Z'(x) \right]^2 = 0 \quad (4.1)$$

La variable aléatoire $Z'(x)$ est la dérivée en moyenne quadratique de Z . Pour une fonction aléatoire stationnaire, la différentiabilité en tout point impose à la covariance d'avoir une dérivée seconde en 0.

4.1.3 Dérivée directionnelle d'une fonction aléatoire

On définit de même la dérivée directionnelle d'une fonction aléatoire Z au point x de $\mathbb{R}^n$ au sens de la moyenne quadratique, u étant un vecteur unitaire de $\mathbb{R}^n$:

$$\exists \frac{\partial Z}{\partial u}(x); \lim_{k \rightarrow 0} E \left[\frac{Z(x+ku) - Z(x)}{k} - \frac{\partial Z}{\partial u}(x) \right]^2 = 0 \quad (4.2)$$

Pour une fonction aléatoire stationnaire, la condition d'existence de cet objet mathématique est que la covariance de $Z(x)$ soit deux fois différentiable à l'origine, ce qui interdit d'utiliser les modèles linéaires à l'origine.

4.1.4 Les modèles admissibles pour la covariance du potentiel

Nous considérons ici des interfaces géologiques qui résultent de processus de dépôt ou d'érosion, ou parfois de la mise en place d'un corps granitique. A l'échelle cartographique considérée dans les applications, ces interfaces sont des surfaces régulières. C'est l'hypothèse qui est faite implicitement quand on considère les données structurales comme des données de gradient. Mathématiquement, on suppose que le potentiel est une fonction aléatoire continue et dérivable (le cas de discontinuités localisées - failles - est traité à part). Notons que ces hypothèses ne seraient pas valides pour des interfaces fractales (front de diffusions, minéral délimité par une coupure sur la teneur...); dans de tels cas on ne disposerait d'ailleurs pas de données structurales.

Compte tenu des définitions de la dérivabilité d'une fonction aléatoire en moyenne quadratique, on ne considérera comme admissible pour représenter la covariance du potentiel que les modèles stationnaires ou intrinsèques à comportement parabolique à l'origine. En effet, la fonction de covariance du potentiel doit être un modèle au moins deux fois différentiable afin que Z le soit au moins une fois. Dans ce cas, il faut que $C'_Z(0) = 0$.

Parmi les modèles classiques de covariance en h^2 et h^3 à l'origine on a :

- la covariance stationnaire cubique ;
- le pentamodèle ;
- la covariance généralisée en h^3 .

On dispose aussi de modèles plus réguliers comme :

- le modèle gaussien
- la covariance généralisée en $h^4 \text{Log}(h)$.

Ces derniers modèles sont sans doute d'une régularité excessive pour représenter des interfaces géologiques.

Dans le cas d'une fonction de covariance stationnaire, notons : $C_Z(h) = \sigma_Z^2 C(h)$ où $C(h)$ est par exemple la covariance cubique de palier unité.

4.2 Le modèle heuristique

4.2.1 Le modèle heuristique initial

Lors des premiers développements de l'*Éditeur*, un modèle dit heuristique initial a été employé. Comme le modèle de covariance n'a fait l'objet d'aucune modélisation préalable à partir d'un variogramme expérimental, tous les modèles de covariance admissibles ainsi que les paramètres (portée, palier, effet de pépité) sont laissés au libre choix de l'utilisateur. Cela entraîne une procédure d'essai/erreur fastidieuse, des incohérences dans les paramètres etc.

Lors des développements suivants de l'*Éditeur*, un modèle "conventionnel", modèle choisi parmi d'autres, a été utilisé pour simplifier la démarche. Il consiste à rationaliser les choix du modèle heuristique initial.

4.2.2 Rationalisation du choix

Il s'agit de choisir un modèle et des paramètres pour les valeurs par défaut, tout en se laissant la liberté de pouvoir les modifier. Tout d'abord, regardons ce que donne les résultats de l'interpolation pour les différents types de covariance. Les covariances cubique et généralisée en h^3 donnent de meilleurs résultats que des modèles plus réguliers (Hallouin, 2001). En effet, il apparaît que le modèle gaussien trop régulier (il est indéfiniment dérivable et le reste donc à l'origine) est inadapté à notre problème et entraîne des instabilités numériques. Les autres modèles (pentamodèle et covariance généralisée en $h^4 \text{Log}(h)$) ont montré sur des coupes des structures circulaires peu crédibles. De plus, la covariance stationnaire cubique de très grande portée peut être comparée à la covariance généralisée cubique. Par conséquent, on ne retient que le modèle stationnaire cubique.

Pour le choix du palier, nous allons exploiter la relation entre la variance du potentiel et

la variance du gradient. Un vecteur unitaire d'orientation aléatoire a une espérance nulle et une variance unité et ses composantes ont, dans un repère orthonormé, une espérance nulle et une variance égale à $1/3$. Ainsi pour un champ de potentiel sans dérive, on choisit σ_Z^2 tel que $\sigma_{\frac{\partial Z}{\partial x}}^2 = \frac{1}{3}$, x pouvant être une direction quelconque. Or $\sigma_{\frac{\partial Z}{\partial x}}^2 = -C_Z''(0) = \frac{\alpha \sigma_Z^2}{a^2}$, α étant un paramètre dépendant du modèle et a la portée². Par exemple, pour une covariance cubique, $\alpha = 14$. À partir du moment où la portée est fixée, on connaît donc la variance du potentiel qui vaut $\frac{a^2}{3\alpha}$. La portée est alors le seul paramètre du modèle à déterminer (avec les effets de pépité).

Si le champ de potentiel présente une dérive, il est possible de prendre une valeur plus faible que celle que l'on vient de déterminer, car une partie de la variabilité du gradient est prise en compte par le gradient de cette dérive. D'où : $\sigma_{\frac{\partial Z}{\partial x}}^2 \leq \frac{1}{3}$.

Les éventuelles erreurs de mesure sur les gradients et les données de potentiel sont prises en compte par les effets de pépité.

Jusqu'à présent les paramètres étaient laissés au choix de l'utilisateur. Compte tenu des remarques ci-dessus, le programme propose par défaut pour le modèle "conventionnel" :

- modèle cubique isotrope ;
- portée arbitrairement fixée à la taille du champ ;
- variance du gradient unité ;
- effet de pépité sur le gradient de 0.01 ;
- effet de pépité sur le potentiel de 10^{-7} ;
- dérive linéaire.

Le paramètre par défaut sans doute le plus arbitraire est la portée, qui devait faire l'objet d'une validation par l'utilisateur. (En fait, il se contentait souvent de cette valeur). Ce modèle, même s'il n'est pas choisi au hasard, n'est pas déterminé de façon optimale, en particulier en exploitant l'information contenue dans les données. L'objectif de cette étude est donc de déterminer ces paramètres en ajustant les variogrammes des données de gradient.

4.3 La détermination du modèle de covariance à partir des données d'orientation

4.3.1 Lien entre covariance du potentiel et covariance du gradient

Soit $h \begin{pmatrix} h_x \\ h_y \\ h_z \end{pmatrix}$ un vecteur et r son module : $r = \sqrt{h_x^2 + h_y^2 + h_z^2}$.

La covariance de la composante u du gradient de potentiel C_{G^u} est alors liée à la covariance du champ de potentiel par la relation suivante :

$$C_{G^u}(\mathbf{h}) = -\frac{\partial^2 C_Z(\mathbf{h})}{\partial h_u^2} \quad (4.3)$$

²La relation entre covariance du potentiel et covariance du gradient est explicitée au point 4.3.1.

où h_u est la composante de $\mathbf{h}$ dans la direction u (Les détails de calcul sont en Annexe B.1).

Plaçons-nous pour simplifier dans le cas où la covariance du potentiel est stationnaire et isotrope : $C_Z(h) = C(r)$, on a donc :

$$C_{G^u}(\mathbf{h}) = - \left(\frac{C''(r)}{r^2} h_u^2 + C'(r) \left[\frac{1}{r} - \frac{h_u^2}{r^3} \right] \right) \quad (4.4)$$

Des formules plus générales existent pour des covariances anisotropes.

La covariance (directe) de la composante du gradient dans la direction u varie selon l'orientation de $\mathbf{h}$ par rapport à u entre :

$$\begin{aligned} C_{\parallel}(\mathbf{h}) &= -C''(r) & \text{si } h \parallel u \\ C_{\perp}(\mathbf{h}) &= -\frac{C'(r)}{r} & \text{si } h \perp u \end{aligned} \quad (4.5)$$

4.3.2 Choix du schéma cubique

Nous avons vu précédemment que le modèle de covariance choisi pour les données de potentiel est un schéma cubique. C'est un modèle de transition, c'est-à-dire qu'il possède un palier, autrement dit on se place dans une hypothèse de stationnarité d'ordre deux. Il a de plus un comportement parabolique à l'origine et caractérise des fonctions aléatoires dérivables en moyenne quadratique. Il est aussi parfaitement adapté pour représenter des contours géologiques. Son expression est la suivante et il est représenté à la Figure 4.1.

$$C(r) = \begin{cases} C_0 \left(1 - 7\left(\frac{r}{a}\right)^2 + \frac{35}{4}\left(\frac{r}{a}\right)^3 - \frac{7}{2}\left(\frac{r}{a}\right)^5 + \frac{3}{4}\left(\frac{r}{a}\right)^7 \right) & \text{pour } 0 \leq r \leq a \\ 0 & \text{pour } r \geq a \end{cases} \quad (4.6)$$

où C_0 est le palier et a la portée.

4.3.3 Calcul des composantes dans le cas du modèle cubique

Nous allons donc calculer les variogrammes des trois composantes du gradient. Soit γ_{G_x} le variogramme de la composante suivant x du gradient et on définit de même γ_{G_y} et γ_{G_z} . Même si on fait l'hypothèse de l'isotropie de la variable "potentiel", alors les variogrammes des composantes du gradient sont anisotropes. On va donc regarder le variogramme d'une composante dans la direction perpendiculaire et dans la direction parallèle à cette composante.

Ainsi pour la composante x , on calcule le variogramme simple directionnel suivant sa direction perpendiculaire soit y . On obtient :

$$\gamma_{G_x}(h_y) = \begin{cases} \frac{14C_0}{a^2} \left[\frac{15}{8} \frac{h_y}{a} - \frac{5}{4} \left(\frac{h_y}{a} \right)^3 + \frac{3}{8} \left(\frac{h_y}{a} \right)^5 \right] & \text{pour } h_y \leq a \\ \frac{14C_0}{a^2} & \text{pour } h_y \geq a \end{cases} \quad (4.7)$$

On reconnaît là le schéma sphérique de $\mathbb{R}^5$, qui est linéaire à l'origine.

De même, on regarde le comportement du variogramme suivant la direction parallèle à la

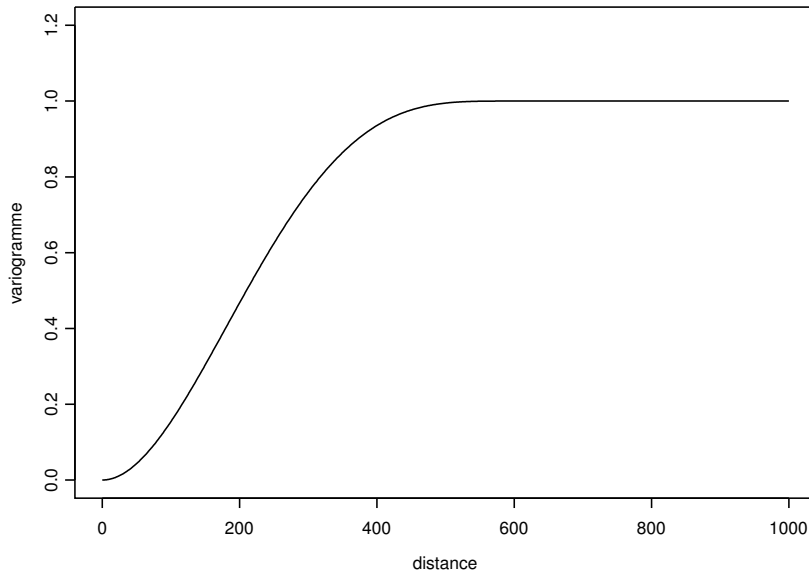


FIG. 4.1 – Modèle cubique de portée $a = 600m$ et de palier $C_0 = 1$.

composante :

$$\gamma_{G_x}(h_x) = \begin{cases} \frac{14C_0}{a^2} \left[\frac{15}{4} \frac{h_x}{a} - 5 \left(\frac{h_x}{a} \right)^3 + \frac{9}{4} \left(\frac{h_x}{a} \right)^5 \right] & \text{pour } h_x \leq a \\ \frac{14C_0}{a^2} & \text{pour } h_x \geq a \end{cases} \quad (4.8)$$

Ce modèle est un schéma à effet de trou. La Figure 4.2 montre ces modèles pour différentes valeurs de portées. Ces deux cas extrêmes (angle entre la direction de la composante et la direction de calcul du variogramme de 0 ou 90 degrés) forment les enveloppes des variogrammes pour des angles variant de 0 à 90 degrés. Les détails des calculs sont exposés dans l'Annexe B.2.

On obtient de même $\gamma_{G_y}(h_x)$ et $\gamma_{G_y}(h_y)$ en remplaçant h_x par h_y et vice-versa. On verra au paragraphe 4.4.3 les modalités pratiques du calcul de ces variogrammes.

Grâce aux paramètres déterminés sur les variogrammes des composantes du gradient, le modèle de covariance du potentiel est intégralement connu. Il faut bien noter aussi que l'on est sensé travailler dans un cadre multivariable. Or, à l'instar des variogrammes des données d'interface qui correspondent aux incréments de potentiel nuls, il n'est pas non plus possible de calculer de variogrammes croisés... On se contentera donc de ce modèle, même s'il est incomplet. Le progrès est d'ores et déjà significatif par rapport à la situation antérieure.

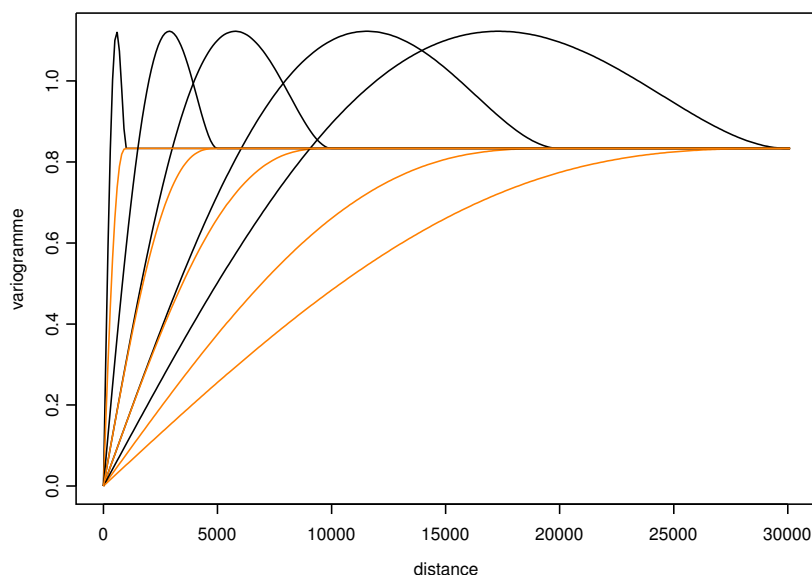


FIG. 4.2 – Modèles de variogrammes de portées 1000m, 5000m, 10000m, 20000m, et 30000m pour les composantes du gradient. En rouge, le schéma sphérique de $\mathbb{R}^5$ correspondant à une composante du gradient dans la direction perpendiculaire et en noir, le schéma à effet de trou correspondant à une composante du gradient dans la direction parallèle.

4.4 Des données d'orientation aux données de gradient

4.4.1 La nature des données d'orientation

Le gradient caractérise le caractère non-uniforme du champ de potentiel autrement dit ses variations spatiales. Les données de gradient utilisées dans cette méthode d'interpolation de surfaces implicites sont des vecteurs normaux aux surfaces isopotentielles. Ces données proviennent de mesures angulaires d'orientation des plans de litage prises sur le terrain à l'aide d'une boussole. Les trois grands types de litages à savoir la stratification (découpage en bancs des roches sédimentaires), la schistosité (découpage de la roche en feuillets de même nature minéralogique dans les roches métamorphiques) et la foliation (découpage de la roche en feuillets de nature minéralogique différente dans les roches métamorphiques à une profondeur plus importante que la schistosité) sont concernés. On fait l'hypothèse que la surface à modéliser est parallèle au litage. Ce cas de figure est comme on l'a vu précédemment très souvent rencontré en géologie.

Il existe de nombreux travaux concernant l'estimation de vecteurs d'orientation (Young, 1987), (Lee et Angelier, 1994), (van den Boogaart et Schaeben, 2002). Par contre comme on l'a vu dans notre cas de figure, les données de gradient servent de variables auxiliaires pour l'interpolation du potentiel auquel elles sont algébriquement liées.

4.4.2 Le passage aux données de gradient

En un point, on fournit le pendage, c'est-à-dire l'angle entre l'horizontale et la ligne de plus grande pente, ainsi que l'azimut, autrement dit la direction de cette ligne par rapport au nord magnétique (Figure 4.3). L'information fournie par la polarité (+1 ou -1) complète la donnée d'orientation en indiquant le sens de la surface, donc le sens de croissance du potentiel.

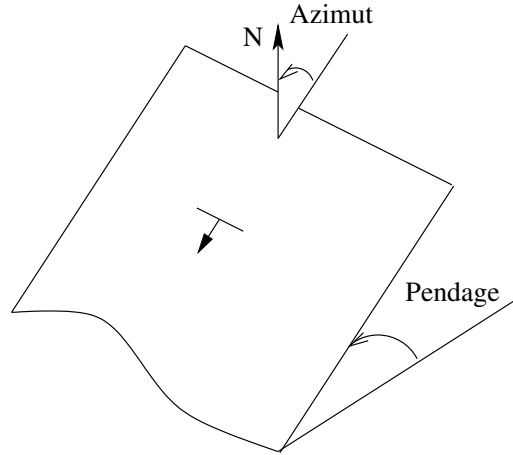


FIG. 4.3 – Données d'orientation : azimut et pendage.

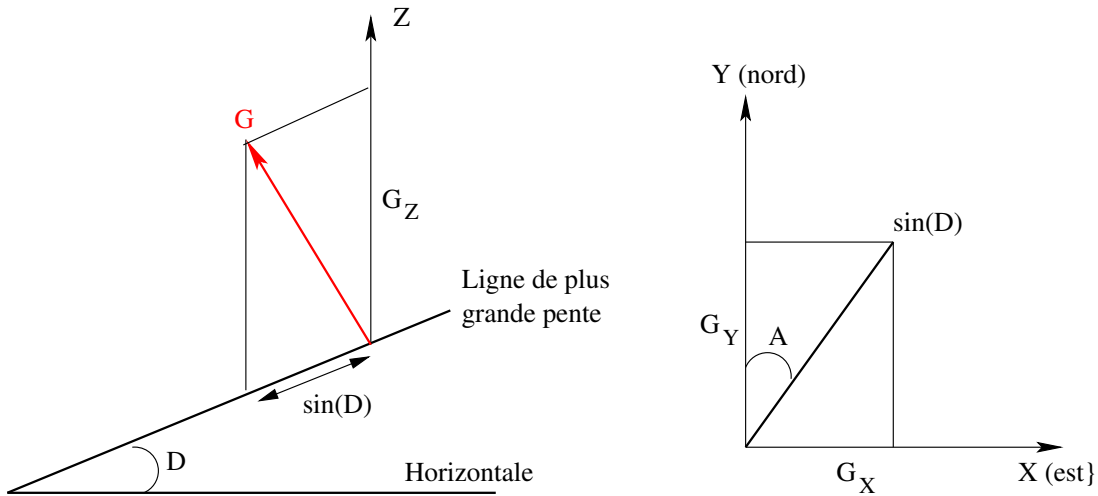


FIG. 4.4 – Projections d'un vecteur normal sur les trois axes de coordonnées.

Les statistiques élémentaires ne pouvant s'appliquer directement aux angles, on passe aux vecteurs unitaires, comme on va le voir ci-dessous. En effet, si on calcule par exemple la moyenne de 2 angles, l'un orienté à 1° et l'autre à 359° , on trouve 180° , alors qu'on s'attendrait plutôt à 0° . On transforme donc ces données d'orientation en composantes

4.5. QUELQUES ÉLÉMENTS SUR LA MODÉLISATION : ANISOTROPIE ET AJUSTEMENT AUTOMATIQUE

d'un vecteur unitaire, en projetant ce vecteur suivant les axes nord-sud (G_y) et est-ouest (G_x), ainsi que suivant la verticale (G_z) (cf. Figure 4.4) avec D la mesure du pendage et A celle de l'azimut :

$$\begin{cases} G_x &= \sin(D) \times \sin(A) \times \text{polarite} \\ G_y &= \sin(D) \times \cos(A) \times \text{polarite} \\ G_z &= \cos(D) \times \text{polarite} \end{cases} \quad (4.9)$$

G_x , G_y et G_z sont les coordonnées du vecteur sur la sphère unité. Ces projections sont celles qui sont habituelles puisque l'on travaille avec des coordonnées géographiques, mais elles sont arbitraires. On peut très bien imaginer des projections sur d'autres axes orthonormés où peut-être la variabilité du gradient serait plus aisée à examiner. La variabilité d'une composante peut être supérieure dans certaines directions, comme on le verra pour les données du Limousin (cf. 5.2.2).

4.4.3 Le calcul pratique des variogrammes des composantes de gradient

Dans trois des exemples d'application (cf. 5.1, 5.2 et 5.3), les mesures de foliation sont cartographiques et la topographie est considérée comme plane. Donc, comme l'extension horizontale est très grande devant l'extension verticale, les variogrammes sont calculés et modélisés dans les directions horizontales. Si le relief est plus marqué, le calcul s'effectue par tranches d'épaisseur (cf. 5.4). Même dans ce cas, la hauteur de la tranche est souvent négligeable devant les dimensions du champ.

Le variogramme expérimental de G_z est calculé de façon omnidirectionnelle dans le plan horizontal (X, Y) donc perpendiculairement à la verticale et les variogrammes de G_x et G_y sont des variogrammes directionnels calculés dans 2 directions : parallèle et perpendiculaire à la composante considérée.

Le calcul du variogramme expérimental fait intervenir différents paramètres. Outre les directions de calcul, il faut définir un pas de calcul et des tolérances angulaire et en distance. Pour la tolérance angulaire, on peut partager l'espace grâce à deux droites nord-est/sud-ouest et nord-ouest/sud-est. On dispose ainsi de 4 secteurs de 90° chacun : nord, est, sud et ouest, définis de manière à ce que toutes les classes soient contiguës dans le plan de référence. Pour la tolérance en distance, on divise l'espace en cercles concentriques (de rayons des multiples de $1000m$ par exemple). La Figure 4.5 montre schématiquement deux paires de données G_x prises en compte dans le calcul du variogramme dans la direction perpendiculaire, puisque elles appartiennent au même secteur de tolérance angulaire englobant le secteur nord de l'espace. Par contre, elles dépendent de deux classes de distance distinctes.

4.5 Quelques éléments sur la modélisation : anisotropie et ajustement automatique

4.5.1 L'anisotropie zonale

Les données étudiées aux sections 5.1, 5.2 et 5.3 étant cartographiques et les couches sub-horizontales, la variabilité du gradient vertical est plus faible que celle des gradients

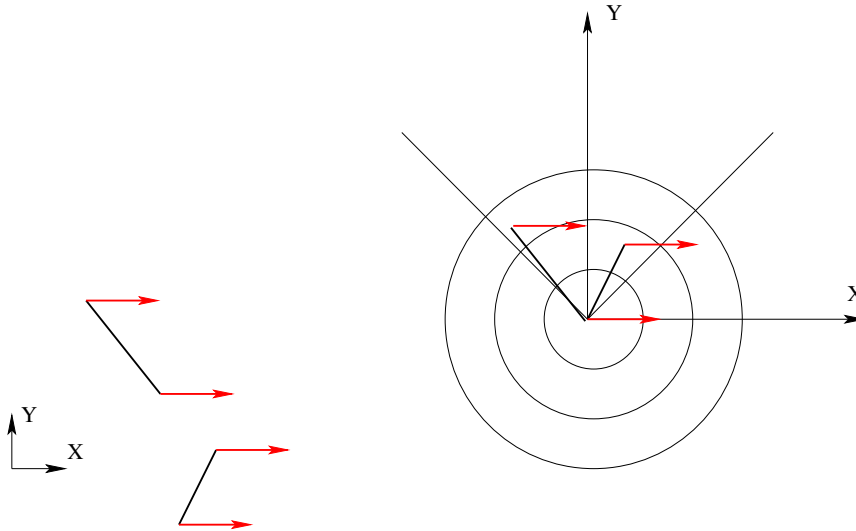


FIG. 4.5 – Exemple de calcul pour le variogramme de G_x suivant la direction perpendiculaire à l'axe (Ox).

horizontaux, comme on peut le constater sur les variogrammes expérimentaux. En effet, suivant la verticale, la moyenne du gradient vertical n'est pas nulle. Les variogrammes des composantes du gradient ont alors des paliers différents. Leur modélisation peut *a priori* se faire en introduisant une anisotropie géométrique ou une anisotropie zonale (ou encore une combinaison des deux qui compliquerait cependant le problème). L'anisotropie géométrique apparaît ici trop contraignante, car la multiplication d'une portée par un certain coefficient entraîne une division du palier par le carré de ce coefficient. Cette différence de palier, va pouvoir être modélisée par contre par une anisotropie zonale, qui est de plus géologiquement acceptable.

On est en présence d'un modèle d'anisotropie zonale si le variogramme ne dépend pas de toutes les composantes du vecteur $\mathbf{h}$. Afin d'obtenir des paliers différents pour les trois composantes, on va donc utiliser des modèles gigognes. Les axes principaux d'anisotropie coïncident avec les axes de projection des composantes du gradient.

Supposons, par exemple³, que le palier du variogramme de G_z est inférieur à celui de G_x . Pratiquement, on modélise d'abord le variogramme dont le palier est le plus faible à savoir celui de G_z , avec une structure isotrope γ_3 de portée a_3 et de palier $\frac{14C_3}{a_3^2}$. On a donc $\gamma_{G_z}(\mathbf{h}) = \gamma_3(\mathbf{h})$.

Ensuite, le variogramme de G_x est ajusté en ajoutant à la structure précédente une structure γ_2 de portée a_2 , de palier $\frac{14C_2}{a_2^2}$ avec une anisotropie géométrique de grande portée verticale. Ainsi γ_2 est quasiment nul dans cette direction et le palier du variogramme de G_x est $14 \left(\frac{C_3}{a_3^2} + \frac{C_2}{a_2^2} \right)$ et on a : $\gamma_{G_x}(\mathbf{h}) = \gamma_3(\mathbf{h}) + \gamma_2(\mathbf{h})$. On peut enfin ajouter une troisième structure γ_1 si le palier de G_y est supérieur au palier de G_x . Il est à noter que si une

³Nous allons voir par la suite au chapitre 5 que ce cas de figure intervient dans toutes les études variographiques effectuées.

structure du modèle gigogne est à faible portée, son poids relatif prend de l'importance du fait du palier en $14 \frac{C}{a^2}$.

Le modèle définitif du potentiel est donc un modèle cubique gigogne :

$$\Gamma(\mathbf{h}) = \Gamma_3 \left(\sqrt{h_x^2 + h_y^2 + h_z^2} \right) + \Gamma_2 \left(\sqrt{h_x^2 + h_y^2} \right) + \Gamma_1(h_y) \quad (4.10)$$

N.B. : la notation Γ se réfère au modèle cubique du potentiel et γ à la structure pentasphérique dérivée du précédent et utilisée pour l'ajustement des composantes du gradient du potentiel.

Lors des cas d'étude du chapitre 5, les tableaux contenant les paramètres du modèle fournissent uniquement les paramètres inférés sur les données de gradient. Ce sont ces valeurs qui apparaissent dans la fenêtre de l'*Editeur* concernant le modèle. Pour le potentiel, les portées restent identiques et les variances se déduisent des paliers et des portées ainsi déterminés.

Le seul paramètre lié au potentiel et que l'on ne peut obtenir par l'étude des variogrammes de gradient est l'effet de pépite sur les données d'incrémentes. Il est fixé à 10^{-6} par expérience, on discutera de cette valeur au chapitre 6.

4.5.2 Un programme d'ajustement automatique

Dans une optique d'utilisation du logiciel par un public non nécessairement initié à la géostatistique, il est nécessaire d'automatiser les étapes de création de la matrice de covariance du système. A partir d'un fichier de données angulaires, on doit pouvoir mettre à jour les paramètres du modèle. Il faut donc intégrer dans l'*Editeur* un programme de calcul des variogrammes expérimentaux, ainsi qu'un algorithme d'ajustement automatique. Ce dernier est basé sur une méthode de régression non linéaire : l'algorithme de Levenberg-Marquardt (Marquardt, 1963). On se place dans un cadre où le modèle dépend non linéairement d'un jeu de M paramètres $\mathbf{a}$.

Soit y_i le variogramme expérimental pour la distance x_i $1 \leq i \leq N$ (les "mesures") et $y(x_i, \mathbf{a})$ le modèle de variogramme considéré⁴. Il s'agit de minimiser, dans l'espace vectoriel de dimension M des paramètres ajustables, la distance métrique entre les données et le modèle. Soit :

$$\chi^2(\mathbf{a}) = \sum_{i=1}^N \left[\frac{y_i - y(x_i, \mathbf{a})}{\sigma_i} \right]^2 \quad (4.11)$$

Les poids σ_i sont des incertitudes connues en chaque point de mesure y_i .

Cressie (1985) choisit par exemple :

$$\sigma_i = \frac{\sqrt{2y(x_i, \mathbf{a})}}{\sqrt{n(x_i)}} \quad (4.12)$$

où $(n(x_i))$ est le nombre de paires au pas i . Ainsi, plus le nombre de paires pour un pas est élevé et plus ce pas a du poids.

⁴Le type de variogramme n'est pas un paramètre, puisque l'on choisit toujours un modèle cubique.

Pour déterminer le vecteur de paramètres qui minimise cette quantité, écrivons les composantes du gradient du χ^2 :

$$\frac{\partial \chi^2}{\partial a_k} = -2 \sum_{i=1}^N \frac{[y_i - y(x_i, \mathbf{a})]}{\sigma_i^2} \frac{\partial y(x_i, \mathbf{a})}{\partial a_k} \quad k = 1, \dots, M$$

ainsi que les éléments de la matrice hessienne :

$$\frac{\partial^2 \chi^2}{\partial a_k \partial a_l} = 2 \sum_{i=1}^N \frac{1}{\sigma_i^2} \left[\frac{\partial y(x_i, \mathbf{a})}{\partial a_k} \frac{\partial y(x_i, \mathbf{a})}{\partial a_l} - [y_i - y(x_i, \mathbf{a})] \frac{\partial^2 y(x_i, \mathbf{a})}{\partial a_l \partial a_k} \right]$$

Posons :

$$\beta_k \equiv -\frac{1}{2} \frac{\partial \chi^2}{\partial a_k} \quad \text{et} \quad \alpha_{kl} \equiv \frac{1}{2} \frac{\partial^2 \chi^2}{\partial a_k \partial a_l}$$

ou en termes de matrices : $\vec{\beta} = -\frac{1}{2} \vec{\nabla} \chi^2$ et $[\alpha] = \frac{1}{2} \mathbf{H}$.

$[\alpha]$ est ce qu'on appelle la matrice de courbure.

La méthode envisagée est dite à gradient adaptatif. Si on se trouve suffisamment près du minimum, on exploite une approximation quadratique du χ^2 pour le préciser. Dans le cas contraire on se contente de se rapprocher du minimum selon la ligne de plus grande pente. L'Annexe C revient rapidement sur la minimisation d'une forme quadratique ou quelconque.

Notons $\delta \mathbf{a} = \mathbf{a}_{\min} - \mathbf{a}_0$, l'incrément à réaliser pour passer du jeu de paramètres initial à celui permettant la minimisation.

Si χ^2 est bien approximé par une forme quadratique (Equation C.3), on résout le système linéaire :

$$[\alpha] \delta \mathbf{a} = \vec{\beta} \quad (4.13)$$

En revanche, si l'approximation ne peut pas être considérée comme quadratique (Equation C.4), la méthode est celle des plus fortes pentes qui donne simplement :

$$\delta \mathbf{a} = C \vec{\beta} \quad (4.14)$$

où C est fixé arbitrairement. On utilise donc cette dernière méthode loin du minimum et on la remplace progressivement par le schéma d'inversion de la matrice hessienne au fur et à mesure du rapprochement vers le minimum. Toute l'astuce de la méthode de Levenberg-Marquardt repose sur l'utilisation optimale de ces deux approches.

La convergence est presque toujours assurée, mais parfois il peut s'agir d'un minimum local. Un balayage des valeurs d'effets de pépité, de portées et de paliers est effectué afin d'obtenir les résultats les plus significatifs.

Dans le chapitre suivant sont présentés différents jeux de données afin d'illustrer la caractérisation de la matrice de covariance à partir des données d'orientation.

Chapitre 5

Variographie des données d'orientation - Exemples des foliations métamorphiques

Plusieurs jeux de données d'orientation vont être présentés. Ils proviennent de différentes régions du Massif central : Velay, Limousin, Cévennes. La structure du Massif central français résulte d'un empilement de nappes. Cette tectonique chevauchante a ainsi entraîné le développement de la foliation métamorphique. La variographie des données d'orientation est aussi illustrée par un exemple suisse. Tous ces jeux contiennent suffisamment de données pour que le calcul des variogrammes expérimentaux soit robuste. Ces derniers sont alors ajustés manuellement et/ou automatiquement avec les structures définies précédemment. Toutes les sections reprendront le même plan classique d'une étude variographique à savoir une présentation des données avec une étude exploratoire, le calcul des variogrammes expérimentaux, l'ajustement de ceux-ci et le rappel des paramètres du modèle de covariance du potentiel.

5.1 Le Velay

5.1.1 Présentation du jeu de données

Localisation géographique et cadre géologique.

Le champ situé dans la région du Velay à l'est du Massif central est approximativement un carré de 105 km sur 105 km et compte $N = 5848$ données de foliation (cf. Figure 5.1). Elles échantillonnent une surface qui est le toit d'un ensemble de roches métamorphiques. En un mot, signalons aussi que le dôme du Velay est un des plus grands dôme de granite - migmatite de la chaîne varisque dont la formation remonte autour de 300 Ma (Ledru *et al.*, 2001). La foliation métamorphique correspond au plan de cristallisation préférentielle des minéraux métamorphiques comme la biotite (Figure 5.2, photographies tirées de l'article de Ledru *et al.* (2001)), la muscovite ou le grenat.

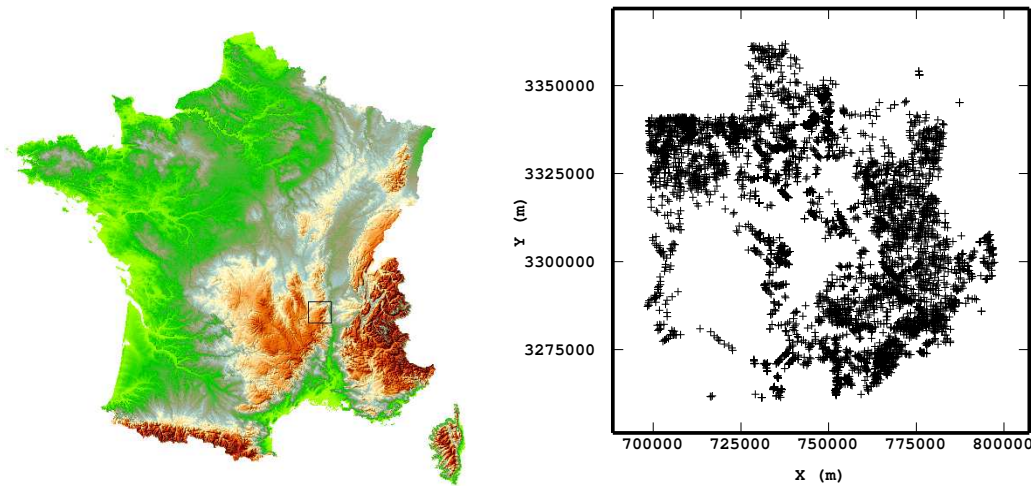


FIG. 5.1 – Plan de position des données de foliation de la région Velay.

L'échantillonnage est très irrégulier avec des zones non couvertes et d'autres très densément représentées. En effet, à certains endroits aucune mesure sur le terrain n'est possible, car la formation géologique n'est pas accessible ou alors inexistante. Les grandes plages sans données au SW correspondent à des terrains sédimentaires (bassins post-stéphanien). A d'autres endroits par contre, où les affleurements sont accessibles et bien visibles, les données sont abondantes. Ces mesures effectuées à des distances très proches sont parfois très différentes entraînant comme on va le voir une variabilité élevée.

Étude exploratoire des données d'orientation

La Figure 5.3(a) est un diagramme en rose des vents. Cette représentation graphique permet de mieux apprécier la répartition des azimuts. Le cercle des angles de 0 à 360° a été divisé en classes de 10°. Le nombre de valeurs de pendage dans chaque classe est compté. Ainsi l'extension radiale est proportionnelle à la fréquence dans chaque classe d'intervalle et peut être exprimée en pourcentage. On constate que toutes les directions sont bien représentées. Le pétale le plus large ne compte que 220 valeurs soit moins de 4% de l'ensemble des valeurs.

On dispose aussi de la Figure 5.3(b) qui présente l'histogramme du pendage. Toutes les gammes de pendage sont présentes et aucune valeur n'est prépondérante, seules les valeurs les plus faibles sont un peu moins nombreuses.

Étude exploratoire des données de gradient

Comme expliqué au paragraphe 4.4.2, on calcule ensuite les composantes du gradient par projection. Ce sont ces données qui vont être utilisées dans l'analyse structurale. Les statistiques élémentaires sont résumées dans le Tableau 5.1. De plus, les Figures 5.4 complètent ces informations. On remarque immédiatement que la composante verticale est très différente des composantes horizontales. En gros, les composantes G_x et G_y s'étalent



FIG. 5.2 – Photographies de granites. La foliation est marquée par l'orientation préférentielle de la biotite (S) et par l'étirement des enclaves micacées (m). Privas, marge méridionale du dôme du Velay. D'après Ledru *et al.*, 2001.

entre -1 et 1 avec une moyenne proche de 0 et une variance proche de 0.3, alors, qu'étant donnée une polarité vers le haut pour toutes les données, G_z est toujours positive avec une moyenne proche de 0.6 et une variance 3.5 fois plus faible. Les composantes G_x et G_y suivent une distribution proche d'une loi uniforme, car tous les azimuts sont représentés de façon égale. Par contre pour G_z qui ne dépend que du pendage, on note l'influence des mesures arrondies qui sont souvent des dizaines d'où l'aspect en dent de scie de l'histogramme.

Calculons le vecteur moyen qui permet de définir une orientation moyenne des vecteurs (Priest, 1993). Définissons $S_x = \sum G_x$, $S_y = \sum G_y$, $S_z = \sum G_z$ et soit $R^2 = S_x^2 + S_y^2 + S_z^2$. On définit ensuite le vecteur moyen par ses composantes $(\frac{S_x}{R}, \frac{S_y}{R}, \frac{S_z}{R})$. Pour ce jeu de données, le vecteur moyen a pour coordonnées $(-0.0752, -0.0516, 0.9958)$.

La concentration des vecteurs autour du pôle moyen peut être approchée par $\frac{N-1}{N-R}$. Plus cette valeur est grande et plus les vecteurs sont concentrés autour du pôle moyen. On trouve ici une valeur de 2.46, ce qui est plutôt faible.

On peut aussi calculer un autre paramètre : le parallélisme qui est défini par $\frac{100(2R-N)}{N}$.

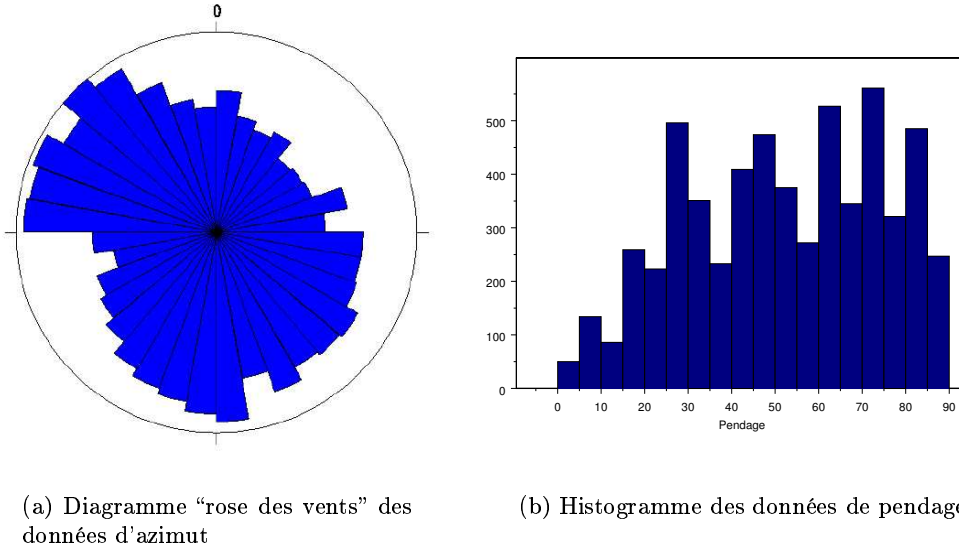


FIG. 5.3 – Histogrammes des données de foliation du Velay.

	Minimum	Maximum	Moyenne	Variance
G_x	-0.9974	0.9980	-0.0446	0.2836
G_y	-0.9962	0.9962	-0.0307	0.2855
G_z	0.0175	0.9998	0.5915	0.0781

TAB. 5.1 – Statistiques élémentaires des données de gradient pour le Velay.

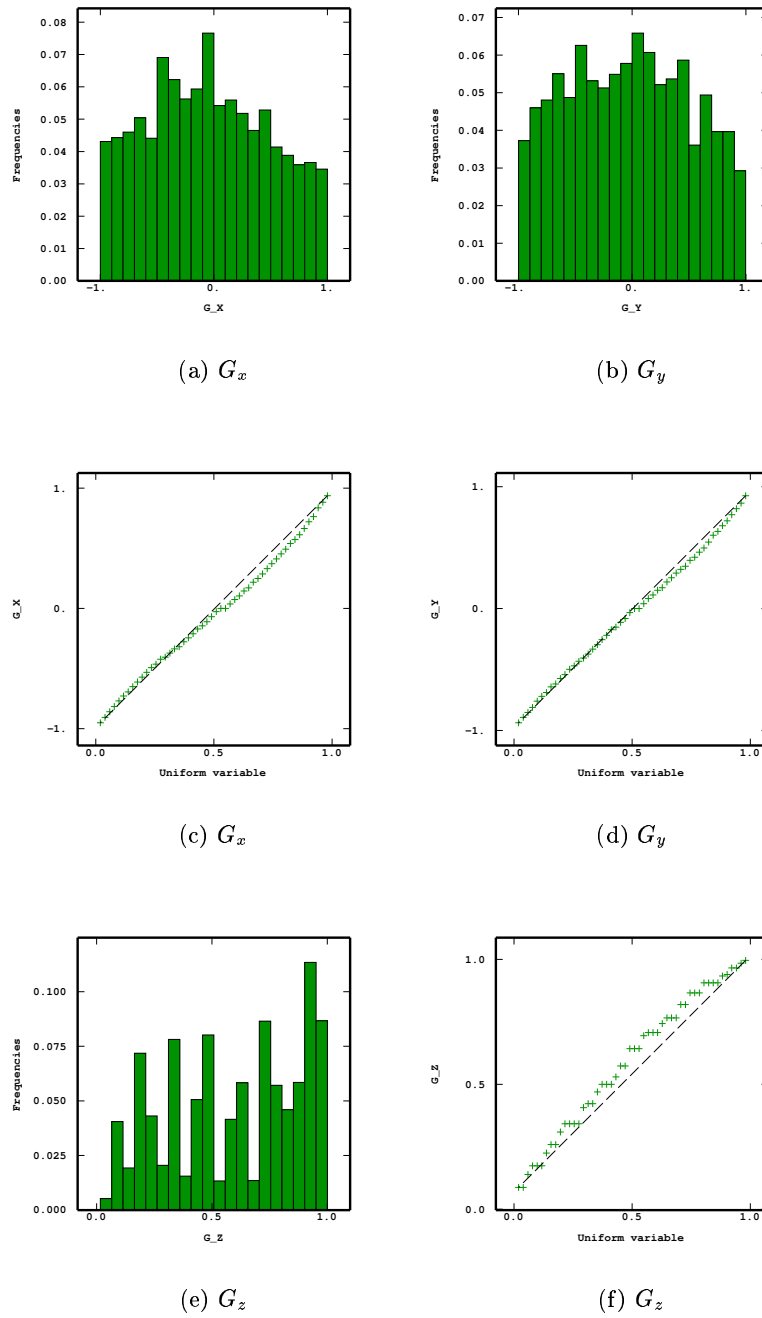


FIG. 5.4 – Histogrammes et diagrammes quantile contre quantile (distribution expérimentale vs distribution uniforme) des composantes du gradient de la région Velay.

Par exemple, des vecteurs tous colinéaires ont un parallélisme de 100%. On trouve ici une valeur de 16%, ce qui est faible. Toutes ces valeurs pourront être comparées à celles obtenues, par exemple avec le jeu de données du Limousin (cf. section 5.2) ou des Cévennes (5.3).

5.1.2 Variogrammes expérimentaux

Sur la Figure 5.5, on peut observer les variogrammes expérimentaux calculés avec les données d'orientation du Velay. Le pas utilisé est de $1000m$ et le nombre de pas est de 35. On définit, pour les variogrammes des composantes horizontales du gradient, les signes $\perp$ et $\parallel$, pour respectivement la direction perpendiculaire et la direction parallèle à la direction de dérivation. Par exemple, $\gamma_{x\perp}$ est le variogramme de la composante x du gradient calculé suivant la direction y (perpendiculairement à la direction x). On constate immédiatement

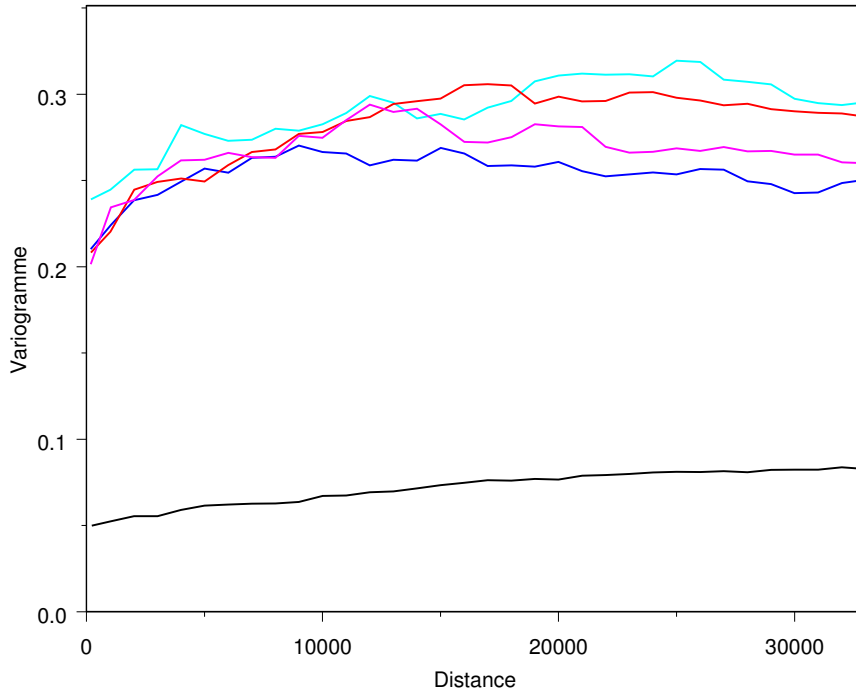


FIG. 5.5 – Variogrammes des données de gradient pour la région Velay. Le variogramme γ_z de G_z est en noir, les variogrammes de G_x sont dans des teintes bleues (bleu pour $\gamma_{x\perp}$ et cyan pour $\gamma_{x\parallel}$), les variogrammes de G_y sont dans des teintes rouges (rouge pour $\gamma_{y\perp}$ et magenta pour $\gamma_{y\parallel}$).

une différence importante entre le variogramme de G_z et les variogrammes des composantes horizontales. En effet, le palier du variogramme de la composante verticale qui est autour de 0.085 est trois à quatre fois plus faible que pour G_x ou G_y . On remarque aussi un

important effet de pépité sur toutes les composantes, indice d'une forte variabilité à courte distance dans les mesures. Cet effet de pépité représente de 60 à 75% de la variance totale. Il faut savoir que le nombre de paires pour le premier pas est de l'ordre de 10000. Les pas suivants ont un nombre de paires de l'ordre de plusieurs dizaines de milliers, voire plus de 100000. Ce point est donc significatif. Regardons ce qu'il se passe à petite échelle : on ne remarque pas de microstructure, le nombre de paires pour des distances très petites est d'ailleurs important. Il faut bien voir que 1998 échantillons sont à moins d'un mètre d'un autre. Si une microstructure existe sa portée est très courte (Figures 5.6). Enfin, l'effet de trou présent dans le modèle pour les directions parallèles aux directions de dérivation n'est pas visible sur les variogrammes expérimentaux.

5.1.3 L'ajustement des variogrammes

L'ajustement s'effectue en commençant toujours par le variogramme dont le palier est le plus faible (à savoir γ_z). De plus, il concerne le seul variogramme dans la direction perpendiculaire pour ceux calculés dans le plan horizontal ($\gamma_{x\perp}$ et $\gamma_{y\perp}$) tout en essayant d'ajuster au mieux la direction parallèle ($\gamma_{x\parallel}$ et $\gamma_{y\parallel}$). Les Tableaux 5.2 et 5.3 rassemblent

	Portée (en m)	Palier $\frac{14C_i}{a_i^2}$
γ_3	37000	0.032
γ_2	6500	0.038

TAB. 5.2 – Portées et paliers des structures de base du modèle de variogramme pour les données du Velay.

	Effet pépité	Palier structure	Palier total
γ_{G_z}	0.051	0.032	0.083
γ_{G_x} ou γ_{G_y}	0.21	0.07	0.28

TAB. 5.3 – Effets de pépité et paliers des variogrammes des composantes du gradient pour les données du Velay.

les paramètres du modèle pour chaque composante du gradient. Rappelons que le modèle est composé de structures de base emboîtées (cf. paragraphe 4.5.1 sur l'anisotropie zonale) : $\gamma_{G_z} = \gamma_3$, $\gamma_{G_x} = \gamma_{G_y} = \gamma_3 + \gamma_2$ s'il n'y a que deux structures, $\gamma_{G_y} = \gamma_3 + \gamma_2 + \gamma_1$ s'il y en a trois. Le palier de chaque structure de base du modèle de variogramme de gradient $\frac{14C_i}{a_i^2}$ est relié au palier C_i de chaque structure de base du modèle de variogramme du potentiel qui est aussi un modèle gigogne, a_i étant la portée de chaque structure. Les Figures 5.7 et 5.8 montrent les variogrammes ajustés avec les paramètres de ces tableaux. Tout d'abord, deux structures suffisent, il n'y a pas lieu de distinguer les modèles dans le plan horizontal. La composante verticale se modélise très facilement et on peut considérer que pour les composantes horizontales l'ajustement est satisfaisant. On remarque aussi

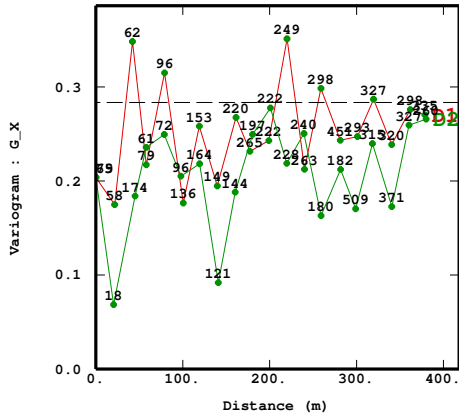
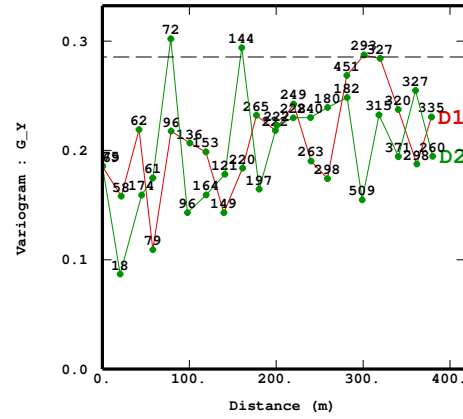
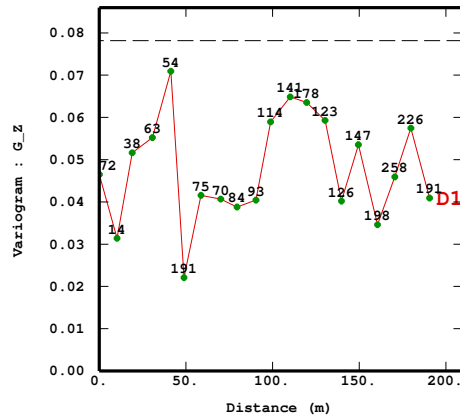
(a) G_x (b) G_y (c) G_z

FIG. 5.6 – Velay : Variogrammes des composantes du gradient à courte échelle : détection d'une éventuelle microstructure.

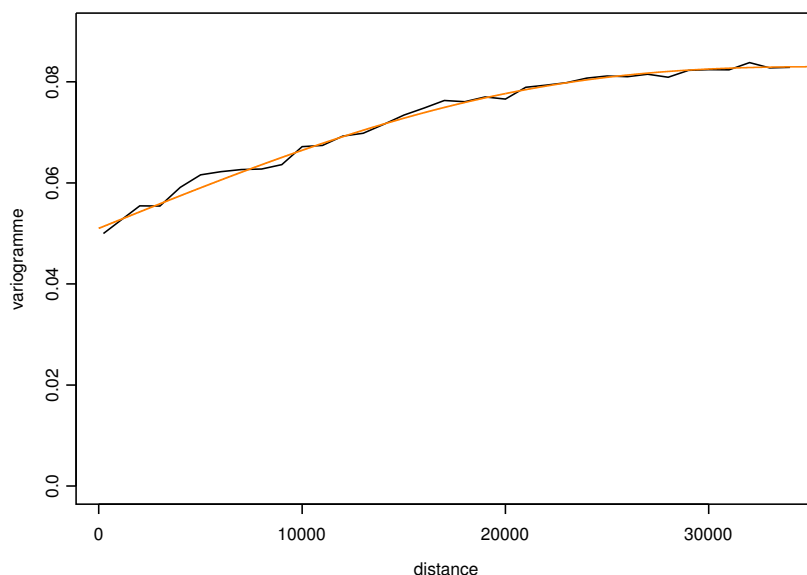


FIG. 5.7 – Velay : Ajustement du variogramme de la composante verticale du gradient.

que la portée dans le plan horizontal est bien plus faible que suivant la verticale, ce qui explique le pic observé sur les variogrammes de $G_{x||}$ et $G_{y||}$.

Comparaison modèle ajusté/modèle heuristique.

Nous allons comparer dans un premier temps les résultats de l'ajustement des variogrammes expérimentaux avec ceux obtenus avec les paramètres du modèle heuristique (cf. section 4.2). On remarque tout d'abord que les portées sont bien inférieures à la taille du champ qui est de 170000m, portée prise par défaut précédemment. Quant à l'effet de pépité, il est bien supérieur à la valeur originelle de 0.01, en particulier pour les composantes horizontales. L'effet de pépité important s'explique essentiellement par des erreurs de mesure et une variabilité locale très forte (cf. paragraphe 5.1.2). Enfin, la variance totale du gradient est inférieure à 1, en grande partie à cause du comportement de G_z .

Comparaison ajustement manuel/automatique.

Comparons maintenant l'ajustement classique manuel et la procédure entièrement automatique. Les résultats sont résumés dans les Tableaux 5.4 et 5.5 et ont été obtenus avec la méthode de Levenberg-Marquardt (cf. paragraphe 4.5.2). Cette chaîne de traitement qui à partir des données d'orientation fournit les paramètres du modèle de covariance, est destinée à être implémentée dans l'*Editeur*. Les illustrations de ces résultats correspondent aux Figures 5.9. On retrouve la facilité d'ajustement du variogramme du gradient vertical

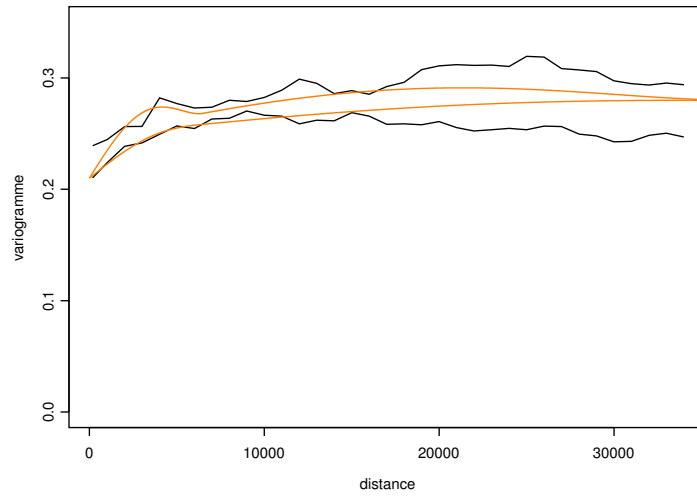
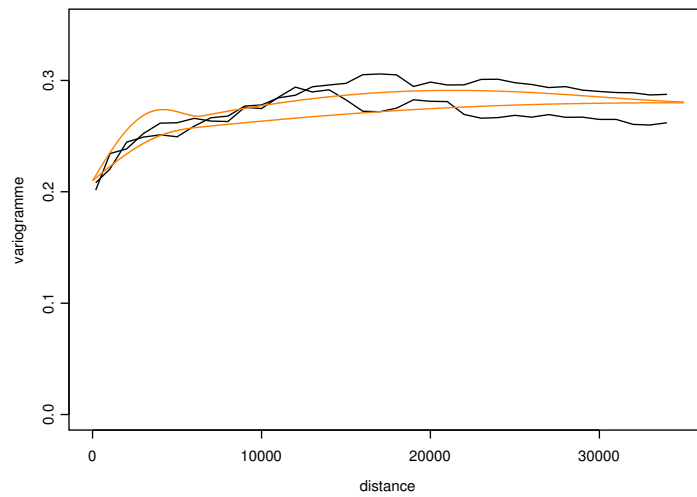
(a) G_x (b) G_y

FIG. 5.8 – Velay : Ajustement des variogrammes des composantes horizontales du gradient.

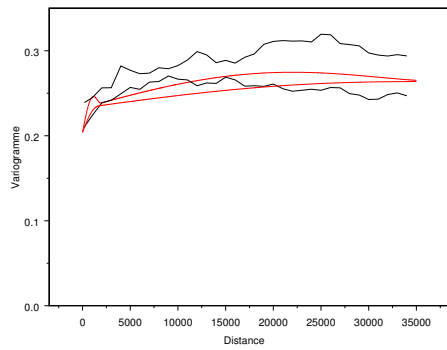
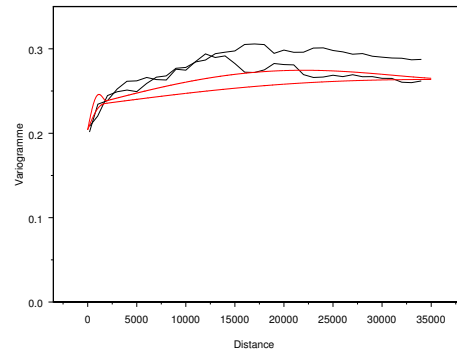
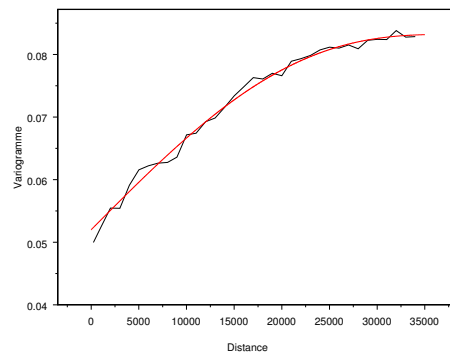
(a) G_x (b) G_y (c) G_z

FIG. 5.9 – Velay : Ajustement automatique des variogrammes des composantes du gradient.

	Portée (en m)	Palier $\frac{14C_i}{a_i^2}$	χ^2
γ_3	38147	0.0312	451
γ_2	1919	0.0286	4224

TAB. 5.4 – Portées et paliers des structures de base du modèle de variogramme pour les données du Velay obtenus par le programme d'ajustement automatique (avec valeur du χ^2).

	Effet pépité	Palier structure	Palier total
G_z	0.052	0.0312	0.0832
G_x ou G_y	0.204	0.0598	0.2638

TAB. 5.5 – Effets de pépité et paliers des variogrammes des composantes du gradient pour les données du Velay obtenus par le programme d'ajustement automatique.

dans la valeur du χ^2 qui est faible par rapport à la valeur trouvée pour la composante horizontale. On constate, en effet que les paramètres du modèle pour le gradient vertical sont très similaires à ceux obtenus par ajustement manuel. Par contre, les composantes horizontales du gradient s'ajustent plus difficilement, en partie à cause des paramètres de G_z qui sont fixés avant détermination de ceux de G_x et G_y . Il est, en effet, plus difficile d'ajuster une courbe par un modèle déjà contraint. Portée et palier sont en outre plus faibles que lors de la procédure manuelle.

5.1.4 Modèle de la covariance du potentiel

Le modèle de la covariance du potentiel est composé de deux structures cubiques emboîtées. La première, isotrope, est de portée $37000m$ et de palier 31.10^5 . La seconde qui ne dépend que des composantes horizontales de la distance a une valeur de $6500m$ pour la portée et un palier de 10^5 .

5.2 Le Limousin

5.2.1 Présentation du jeu de données

Localisation géographique et cadre géologique.

Le champ situé dans la région du Limousin, à l'ouest du Massif central, est approximativement un carré de 70 km sur 70 km et compte $N = 1485$ données de foliation (cf. Figure 5.10). Elles échantillonnent comme pour le Velay le toit d'un ensemble de roches métamorphiques qui correspond à une unité lithotectonique de la zone interne du Massif central. La formation est appelée l'Unité inférieure des Gneiss que l'on abrégera par "uig". Pour information, on peut signaler que cette zone interne est découpée en 4 unités (dont uig) et

est basé sur les lithologies, âges et caractéristiques structurales et métamorphiques (Roig et Faure, 2000) et (Roig *et al.*, 1998).

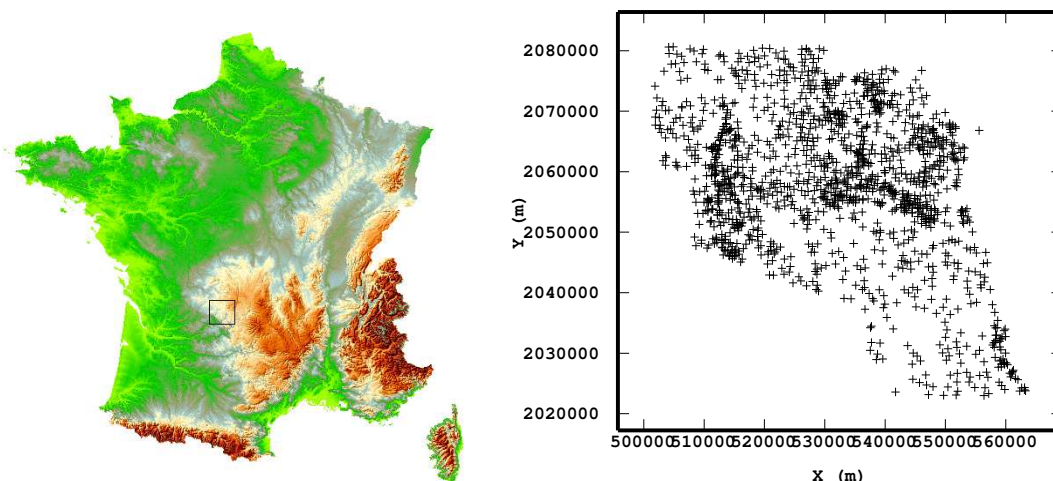


FIG. 5.10 – Plan de position des données de foliation de la région Limousin.

Étude exploratoire des données d'orientation

Observons les deux diagrammes de la Figure 5.11. Sur la rose des vents on voit que tous les angles (par secteur de 10°) sont représentés, mais certains angles le sont mieux que d'autres, comme ce pic que l'on note autour de 200° qui contient 126 valeurs soit plus de 8% du total.

Sur l'histogramme du pendage, on observe qu'il n'y a pratiquement pas de pendages très élevés (aucune valeur supérieure à 72°). Il y a de plus un pic à 60° , qui représente plus de 23% des valeurs.

Étude exploratoire des données de gradient

On calcule alors comme précédemment les composantes du gradient, les statistiques élémentaires étant résumées dans le Tableau 5.6 et la Figure 5.12. On remarque à nouveau que

	Minimum	Maximum	Moyenne	Variance
G_x	-0.8913	0.8659	0.0108	0.1876
G_y	-0.8659	0.8932	-0.1104	0.2076
G_z	0.3168	1	0.7491	0.0314

TAB. 5.6 – Statistiques élémentaires des données de gradient pour le Limousin.

la composante verticale est très différente des composantes horizontales, mais la variabilité est moins grande que dans le cas précédent. En gros, les intervalles où les composantes

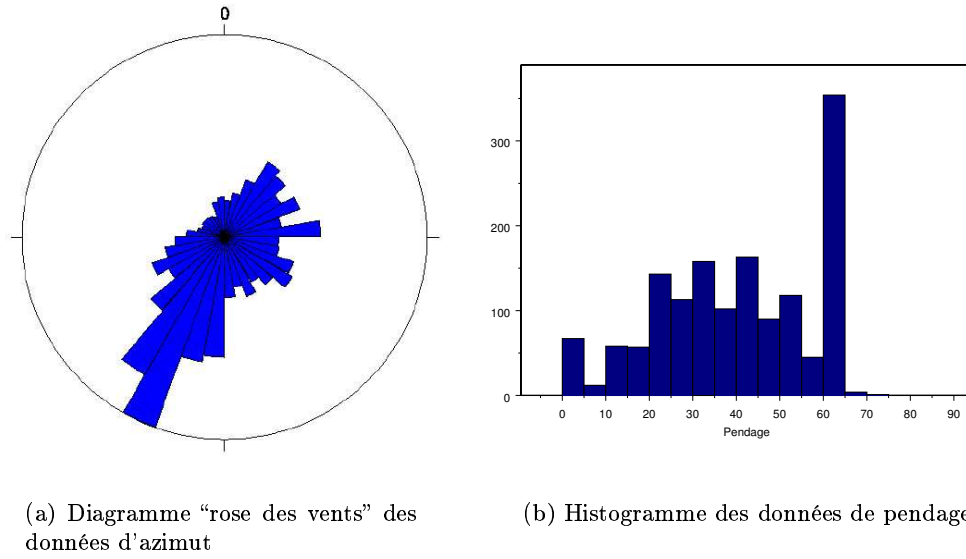


FIG. 5.11 – Histogrammes des données de foliation du Limousin.

G_x et G_y prennent leurs valeurs sont plus petits et la variance est aussi plus faible. A l'instar du Velay, la composante G_z est toujours positive. La moyenne est proche de 0.75 et l'écart avec les variances des composantes horizontales est encore plus grand avec cette fois un facteur 6. Enfin, les trois composantes ont une distribution qui s'éloigne plus de la loi uniforme.

Le vecteur moyen a pour coordonnées $(0.0142, -0.1457, 0.9892)$.

La concentration est autour de 4.12, les vecteurs sont donc plus concentrés autour du pôle moyen. Les valeurs se ressemblent donc plus.

De même, le paramètre de parallélisme a une valeur plus élevée que dans le Velay avec 51%.

Intéressons nous maintenant aux variogrammes expérimentaux des composantes du gradient.

5.2.2 Variogrammes expérimentaux

La Figure 5.13 présente les variogrammes expérimentaux du Limousin. Le pas utilisé est de $1000m$ et le nombre de pas est de 30. Le nombre de paires pour le premier pas est de l'ordre de 300, alors qu'il est de plusieurs milliers pour les suivants.

Si on compare ces variogrammes à ceux du Velay, on constate qu'ils sont plus structurés, la part de l'effet de pépite dans la variabilité totale est en effet plus faible. Cette part chute jusqu'à ne plus représenter que 30 à 50% de la variance totale. Cet effet de pépite moins important s'explique par des erreurs de mesure beaucoup plus faibles, une grande part de la variabilité à courte portée étant supportée par une microstructure dont l'ordre

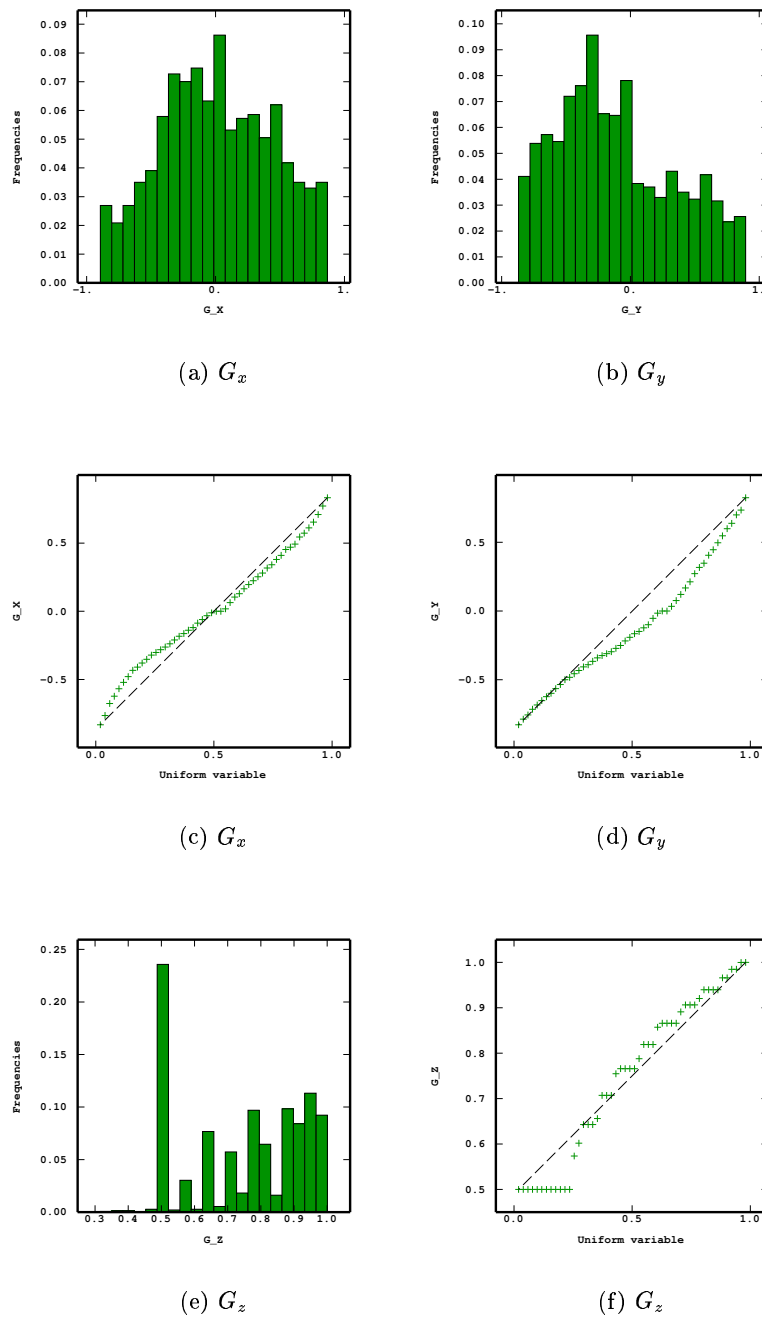


FIG. 5.12 – Histogrammes et diagrammes quantile contre quantile (distribution expérimentale vs distribution uniforme) des composantes du gradient de la région Limousin.

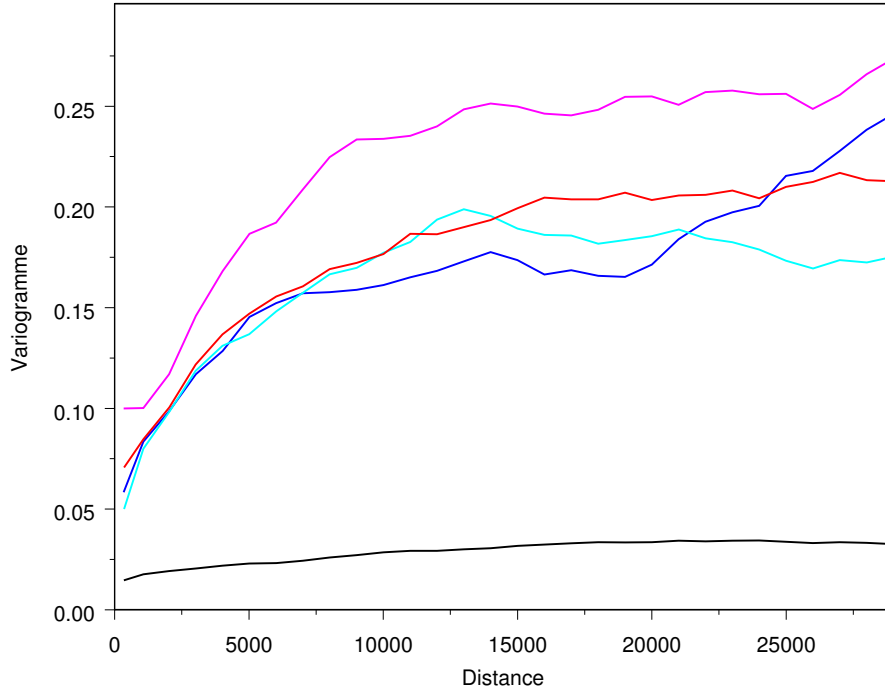
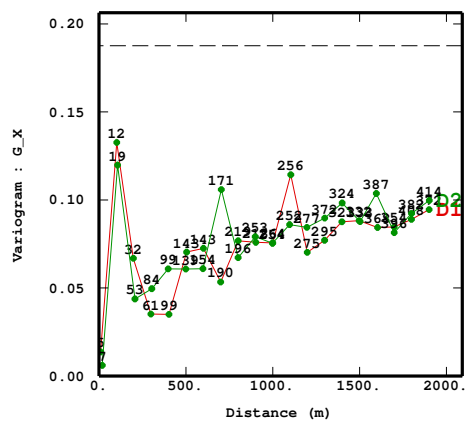


FIG. 5.13 – Variogrammes des données de gradient pour la région Limousin. Le variogramme γ_z de G_z est en noir, les variogrammes de G_x sont dans des teintes bleues (bleu pour $\gamma_{x\perp}$ et cyan pour $\gamma_{x\parallel}$), les variogrammes de G_y sont dans des teintes rouges (rouge pour $\gamma_{y\perp}$ et magenta pour $\gamma_{y\parallel}$).

de grandeur de la portée s'exprime en centaine de mètres, donc *grosso modo* 100 fois plus faible que la portée de la structure principale (cf. Figure 5.14). Il faut noter que seuls 4 échantillons sont à moins d'un mètre d'un autre.

On remarque aussi que les variogrammes de G_x et G_y n'ont pas le même comportement, notamment avec une différence de palier, ce qui traduit une anisotropie de la covariance du potentiel dans le plan horizontal. Enfin, les portées sont bien inférieures pour chaque composante à la taille du champ.

Remarque : Etant donné qu'une troisième structure suivant une direction particulière (nord-sud) est rajoutée, on peut se poser la question du choix de cette direction. En effet, la projection du vecteur d'orientation dans le plan horizontal sur les axes nord-sud et est-ouest est arbitraire. On peut regarder sur les Figures 5.15 et 5.16 ce qu'il se passe pour différents angles de projection pour les données du Limousin. Si 0° correspond à la projection sur l'axe est-ouest (X) et nord-sud (Y), alors 45° correspond, par exemple à la projection sur l'axe NE-SW (X) et NW-SE (Y). De grosses différences de palier apparaissent entre les



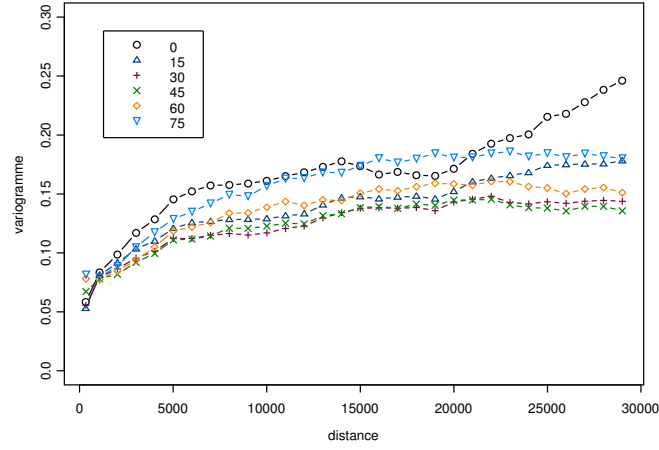
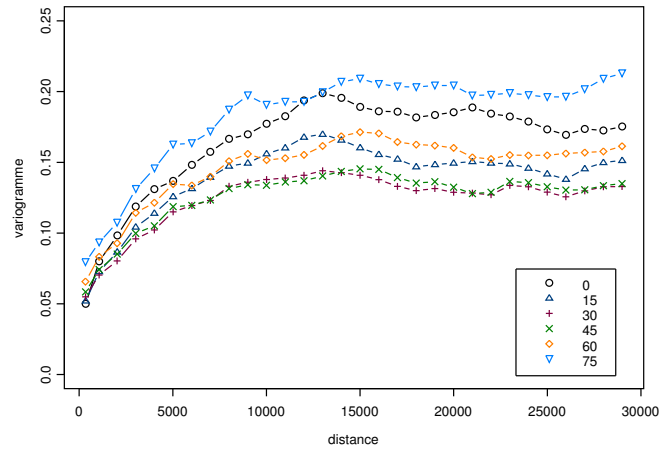
(a) $G_{x\perp}$ (b) $G_{x\parallel}$

FIG. 5.15 – Limousin. Variogrammes des données de la composante x du gradient pour différentes projections.

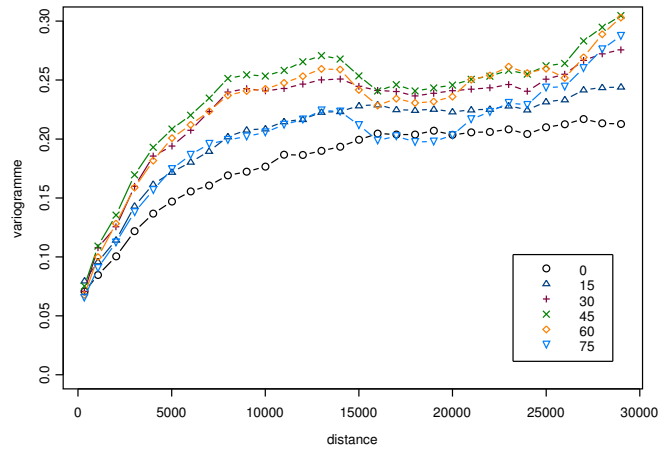
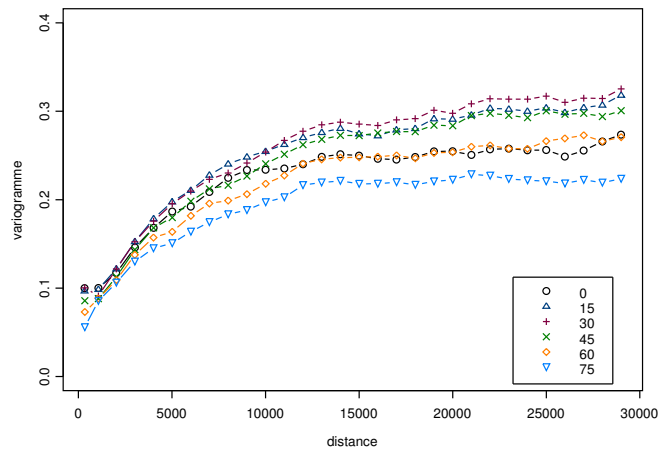
(a) $G_{y\perp}$ (b) $G_{y\parallel}$

FIG. 5.16 – Limousin. Variogrammes des données de la composante y du gradient pour différentes projections.

variogrammes en fonction de l'angle de projection. La plus grande variabilité entre les variogrammes suivant la direction perpendiculaire et la direction parallèle est atteinte pour une direction proche de 30° . Ces résultats sont destinés à l'*Editeur Géologique*, or ce dernier ne prend pas en compte les axes différents de nord-sud et est-ouest. On utilisera donc dans la suite de cette étude les résultats des données projetées sur les axes géographiques. Passons maintenant à la modélisation des variogrammes expérimentaux.

5.2.3 L'ajustement des variogrammes

Premier ajustement.

Les Tableaux 5.7 et 5.8 rassemblent les paramètres du modèle pour chaque composante du

	Portée (en m)	Palier $\frac{14C_i}{a_i^2}$
γ_3	25000	0.0175
γ_2	9500	0.0855
γ_1	55000	0.0570

TAB. 5.7 – Portées et paliers des structures de base du modèle de variogramme pour les données du Limousin.

	Effet pépite	Palier structure	Palier total
γ_{G_z}	0.0165	0.0175	0.034
γ_{G_x}	0.065	0.103	0.168
γ_{G_y}	0.065	0.160	0.225

TAB. 5.8 – Effets de pépite et paliers des variogrammes des composantes du gradient pour les données du Limousin.

gradient. Les Figures 5.17 et 5.18 montrent les variogrammes ajustés avec les paramètres de ces tableaux.

Une troisième structure est nécessaire ici, car le palier du variogramme de G_y est supérieur à celui de G_x . L'effet de pépite quoique important, est presque trois fois plus faible que pour les données du Velay.

La variance totale est bien plus faible que pour les données du Velay, on s'éloigne un peu du cadre théorique du vecteur aléatoire de variance unité. Les données sont donc plus structurées avec un effet de pépite et une variance plus faibles. Il y a moins de variabilité locale des orientations, aussi bien verticalement (surface globalement plus "horizontale") qu'horizontalement avec l'anisotropie observée entre G_x et G_y .

La portée faible nécessaire pour ajuster le variogramme de $G_{x\perp}$ crée un pic sur celui de $G_{x\parallel}$. La composante verticale se modélise toujours aussi facilement, par contre l'ajustement est plus malaisé pour les composantes horizontales. En effet, après avoir modélisé le variogramme de G_z , on constate que, si on ajuste $G_{x\perp}$, $G_{x\parallel}$ ne l'est pas vraiment. Il est

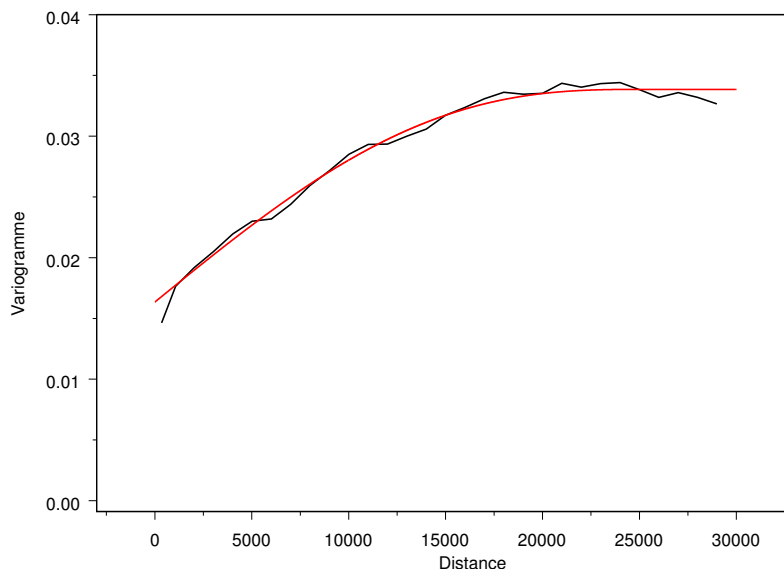


FIG. 5.17 – Limousin : Ajustement du variogramme de la composante verticale du gradient.

ensuite difficile d'obtenir de bons résultats pour $G_{y||}$.

Second ajustement et comparaison modèle ajusté/modèle heuristique.

Un compromis a été adopté pour G_x , comme on peut le voir sur la Figure 5.19(a),

	Portée (en m)	Palier $\frac{14C_i}{a_i^2}$
γ_2	17000	0.0825
γ_1	55000	0.0500

TAB. 5.9 – Portées et paliers des structures de base du nouveau modèle pour les variogrammes de G_x et G_y pour les données du Limousin.

	Effet pépite	Palier structure	Palier total
γ_{G_x}	0.075	0.100	0.175
γ_{G_y}	0.075	0.150	0.225

TAB. 5.10 – Effets de pépite et paliers du nouveau modèle pour les variogrammes de G_x et G_y pour les données du Limousin.

permettant ensuite un bon ajustement pour G_y (Figure 5.19(b)). Les Tableaux 5.9 et 5.10

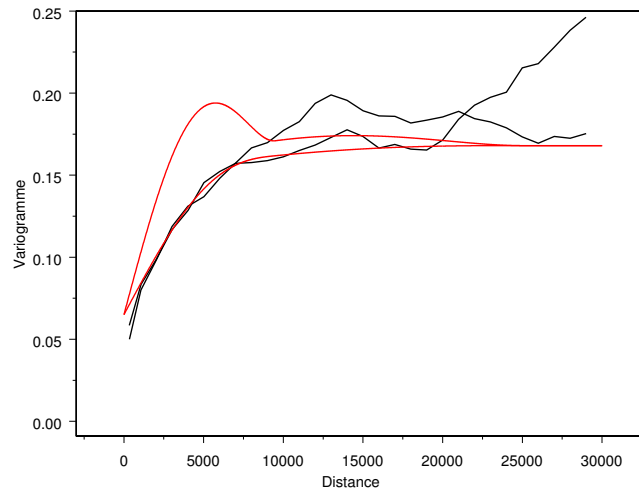
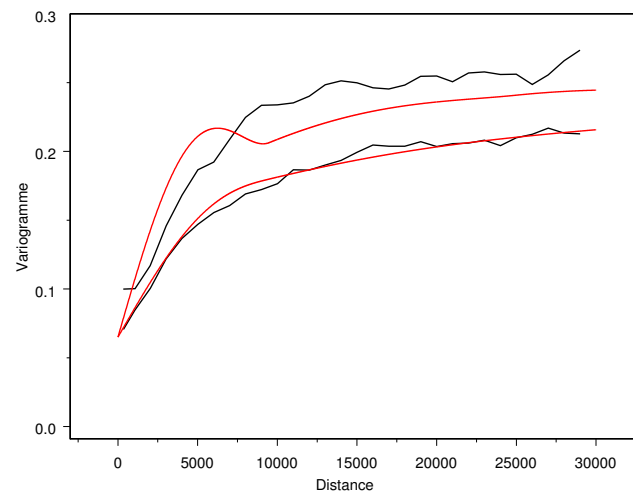
(a) G_x (b) G_y

FIG. 5.18 – Limousin : Premier ajustement des variogrammes horizontaux des composantes du gradient.

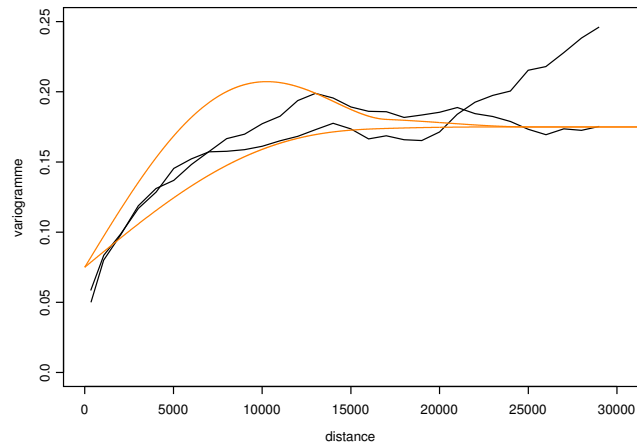
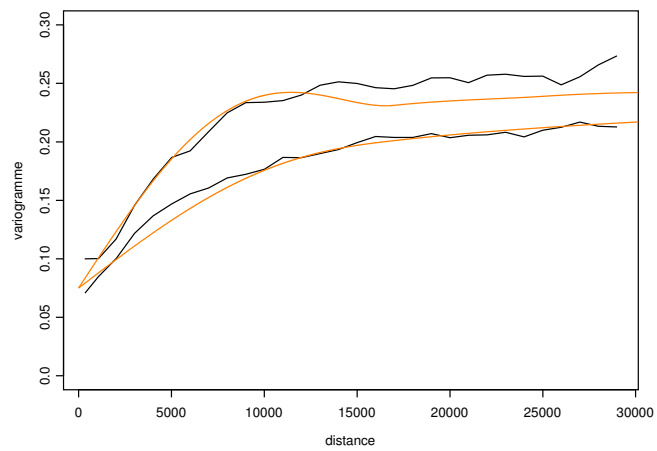
(a) G_x (b) G_y

FIG. 5.19 – Limousin : Second ajustement des variogrammes des composantes horizontales du gradient.

rassemblent les paramètres de ce nouveau modèle pour les deux composantes horizontales du gradient.

Ce sont ces paramètres qui seront retenus pour la suite du travail, en particulier pour illustrer le chapitre 6. Si on compare les paramètres obtenus pour cet ajustement et ceux du modèle “conventionnel”, on constate que, comme pour le Velay, la portée est bien inférieure à la taille du champ qui est de 98000m. On peut faire le même type de remarques concernant l’effet de pépite et la valeur du palier.

Comparaison ajustement manuel/automatique

Les Tableaux 5.11 et 5.12 et la Figure 5.20 présentent les résultats obtenus par le programme d’ajustement automatique. Seules deux structures ont été déterminées. On trouve une excellente concordance entre les ajustements automatique et manuel de G_z , par contre des différences existent pour les autres composantes. L’ajustement automatique fournit pour G_x un résultat proche du premier ajustement avec une portée plus courte et un palier total également plus petit. Néanmoins, l’ajustement de G_y n’est pas très satisfaisant, le programme ne parvenant pas à obtenir une troisième structure pour modéliser l’anisotropie entre les deux composantes horizontales.

	Portée (en m)	Palier $\frac{14G_i}{a_i^2}$	χ^2
γ_3	26707	0.0179	47
γ_2	8560	0.0733	377

TAB. 5.11 – Portées et paliers des structures de base du modèle de variogramme pour les données du Limousin obtenus par le programme d’ajustement automatique (avec valeur du χ^2).

	Effet pépite	Palier structure	Palier total
G_z	0.0166	0.0179	0.0345
G_x ou G_y	0.0707	0.0912	0.1619

TAB. 5.12 – Effets de pépite et paliers des variogrammes des composantes du gradient pour les données du Limousin obtenus par le programme d’ajustement automatique.

Ajustement des données projetées sur d’autres axes¹

A titre d’information (cf. paragraphe 5.2.2 et Figures 5.15 et 5.16), les Tableaux 5.13 et 5.14 fournissent les paramètres pour les données projetées sur les axes ayant subi une rotation de 30°. C’est avec cette rotation que l’écart entre la variance totale de G_x et de G_y est le plus fort. La différence entre les paliers de G_x et G_y atteint 0.13 alors que pour les données projetées sur les axes géographiques elle n’est que de 0.05.

¹Je rappelle que ce modèle n’a pas été retenu pour la suite de l’étude, car l’*Editeur* ne permet pas encore la prise en charge de données suivant d’autres projections.

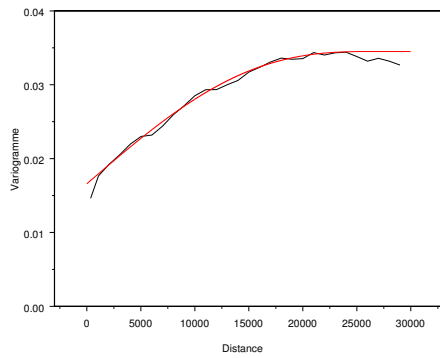
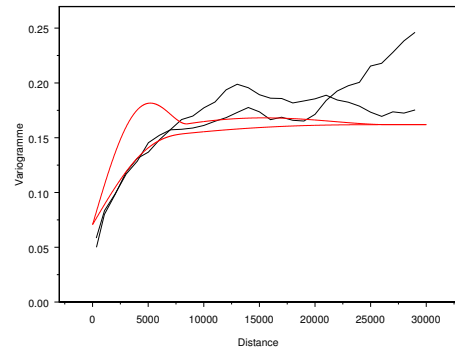
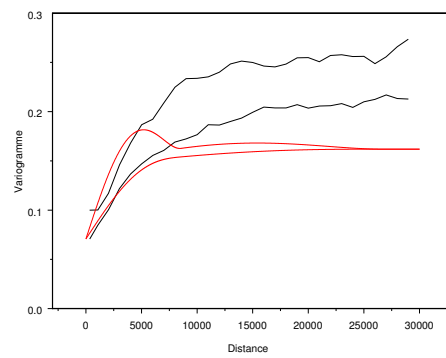
(a) G_z (b) G_x (c) G_y

FIG. 5.20 – Limousin : Ajustement automatique des variogrammes des composantes du gradient.

	Portée (en m)	Palier $\frac{14C_i}{a_i^2}$
γ_3	25000	0.0175
γ_2	18000	0.0375
γ_1	21000	0.13

TAB. 5.13 – Portées et paliers des structures de base du modèle de variogramme pour les données du Limousin projetées sur des axes après rotation de 30° .

	Effet pépite	Palier structure	Palier total
γ_{G_z}	0.0165	0.0175	0.034
γ_{G_x}	0.075	0.055	0.13
γ_{G_y}	0.075	0.185	0.26

TAB. 5.14 – Effets de pépite et paliers des variogrammes des composantes du gradient pour les données du Limousin projetées sur des axes après rotation de 30° .

5.2.4 Modèle de la covariance du potentiel

Le modèle de covariance du potentiel retenu est un schéma cubique gigogne avec trois composantes. La première, isotrope, est de portée $25000m$ et a pour palier la valeur de 8.10^5 . La deuxième dans le plan horizontal est de portée $17000m$ et de palier 17.10^5 . Enfin, la troisième structure suivant la direction nord-sud, a une portée de $55000m$ et un palier de 108.10^5 .

5.3 Les Cévennes

5.3.1 Présentation du jeu de données

Localisation géographique et cadre géologique.

Le champ situé dans les Cévennes au sud-est du Massif central est approximativement un rectangle de 40 km sur 50 km et compte $N = 275$ données de foliations métamorphiques (Figure 5.21). Les données sont encore une fois cartographiques et, comme pour le Velay, de larges plages ne sont pas couvertes.

Étude exploratoire des données d'orientation

Sur la rose des vents des azimuts (Figure 5.22(a)) on voit que tous les angles (par secteur de 10°) sont représentés, avec une majorité dans la direction nord-sud. Le pétale le plus grand contient 15 valeurs soit 5% du total.

Sur l'histogramme du pendage (Figure 5.22(b)), on observe qu'il n'y a pas de pendages très élevés (aucune valeur supérieure à 70°), la majorité des pendages étant inférieure à 45° . Le pic est autour de 30° , ce qui représente autour de 20% des valeurs.

Étude exploratoire des données de gradient

Les statistiques élémentaires sont résumées dans le Tableau 5.15. La composante verticale

	Minimum	Maximum	Moyenne	Variance
G_x	-0.8529	0.8138	-0.0118	0.1205
G_y	-0.8660	0.9063	-0.0504	0.1380
G_z	0.3420	1	0.8498	0.0166

TAB. 5.15 – Statistiques élémentaires des données de gradient pour les Cévennes.

est toujours très différente des composantes horizontales et les variances des trois composantes sont les plus faibles des trois jeux de données du Massif central. La composante G_z est là aussi toujours positive. Sa moyenne est proche de 0.85 et l'écart avec les variances des composantes horizontales est encore plus grand avec cette fois un facteur de presque 8. Enfin, les deux composantes horizontales ont une distribution qui se rapproche plus de la loi normale.

Le vecteur moyen a pour coordonnées $(-0.0139, -0.0591, 0.9981)$.

La concentration est autour de 6.71, les vecteurs sont donc encore plus concentrés autour du pôle moyen.

De même, le paramètre de parallélisme a une valeur très élevée avec 70%.

On remarque donc que statistiquement les données des Cévennes sont moins dispersées que les données du Velay ou même du Limousin, il y a moins de variabilité dans les mesures. Intéressons nous maintenant aux statistiques spatiales avec les variogrammes expérimentaux des composantes du gradient.

5.3.2 Variogrammes expérimentaux

La Figure 5.24, montre les variogrammes expérimentaux pour les données des Cévennes. Le pas utilisé est de 1000m et le nombre de pas est de 25. Le variogramme de G_z a encore une fois un palier bien plus faible que les variogrammes des composantes horizontales. On remarque que les courbes sont moins "lisses", car les nombres de paires impliquées dans le calcul sont beaucoup moins élevés. L'élément le plus remarquable est, comme on le pressentait lors de l'étude exploratoire, la structuration plus importante des variogrammes par rapport aux deux jeux de données précédents. L'effet de pépité ne représente plus que 25% de la variabilité totale de chaque composante. De plus, il n'y a pas vraiment d'anisotropie marquée entre G_x et G_y .

5.3.3 L'ajustement des variogrammes

Les Tableaux 5.16 et 5.17 rassemblent les paramètres du modèle pour chaque composante du gradient. Les Figures 5.25 et 5.26 montrent les variogrammes ajustés avec les paramètres de ces tableaux. D'une manière générale, les variogrammes ont été plus difficiles à ajuster. D'autant plus, que les variogrammes dans les directions perpendiculaire et

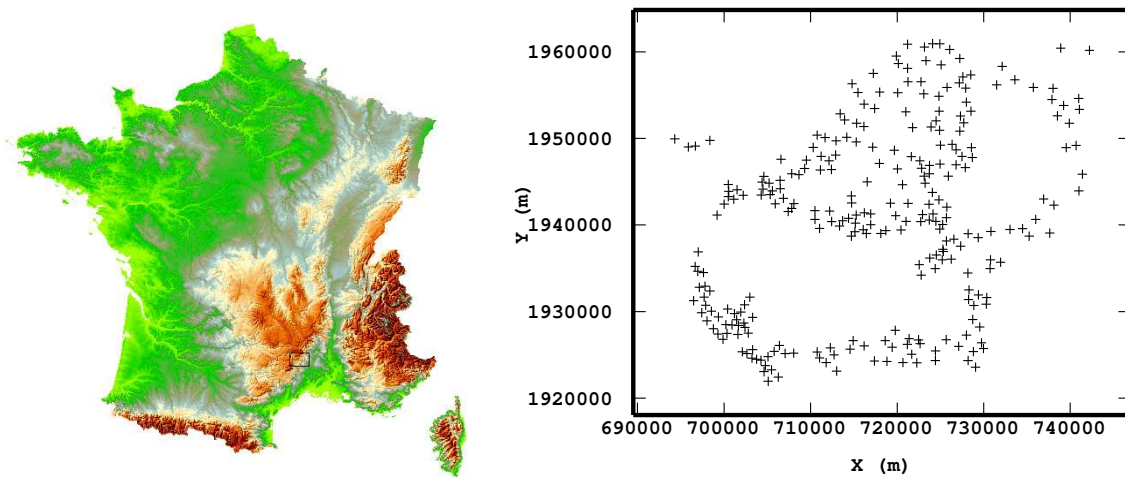
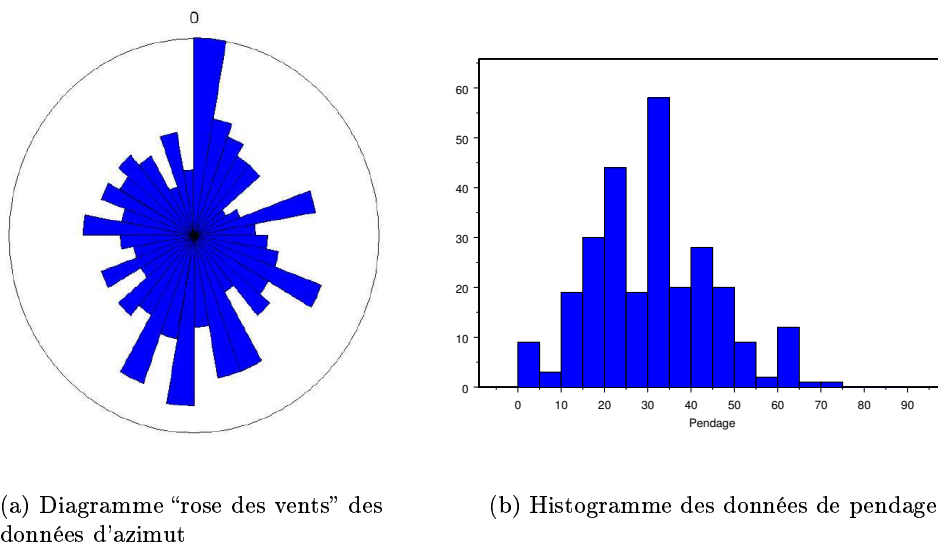


FIG. 5.21 – Plan de position des données de foliation de la région des Cévennes.



(a) Diagramme “rose des vents” des données d'azimut

(b) Histogramme des données de pendage

FIG. 5.22 – Histogrammes des données de foliation des Cévennes.

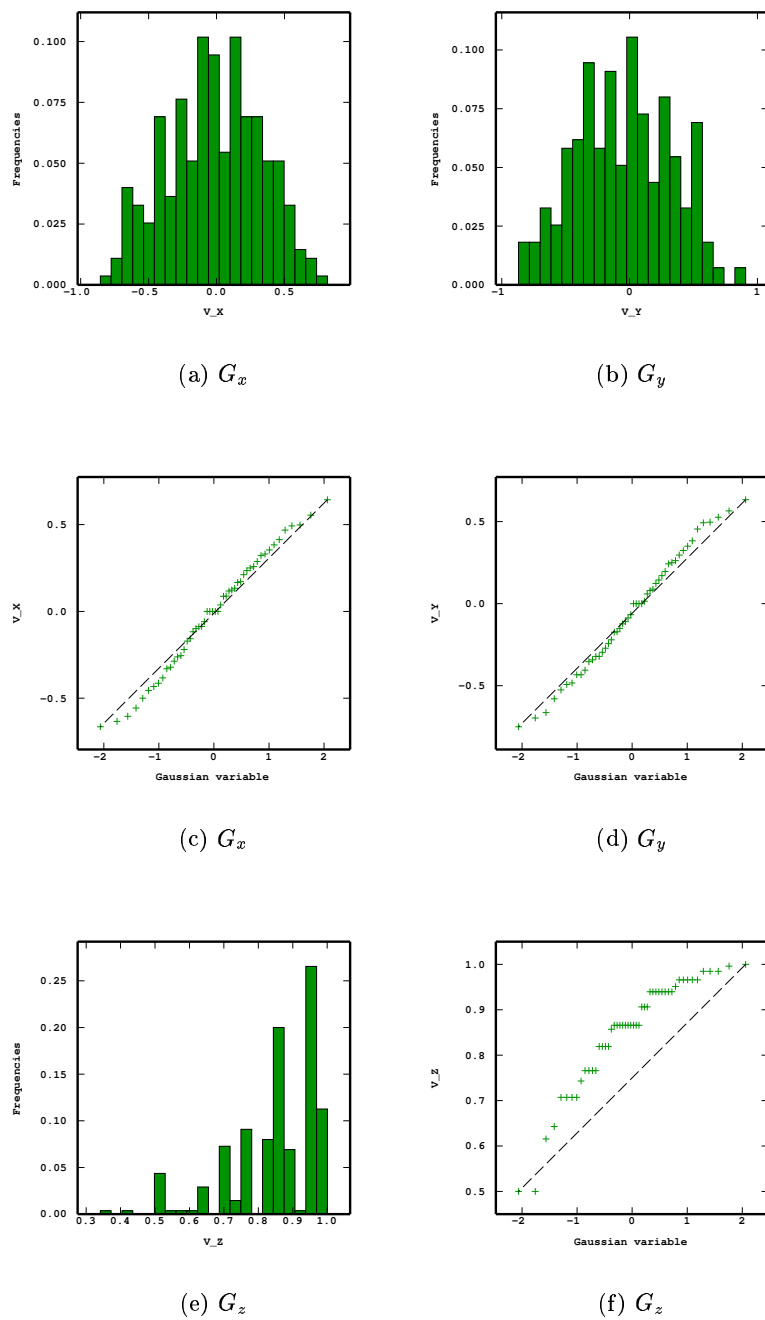


FIG. 5.23 – Histogrammes et diagrammes quantile contre quantile (distribution expérimentale vs distribution gaussienne) des composantes du gradient de la région Cévennes.

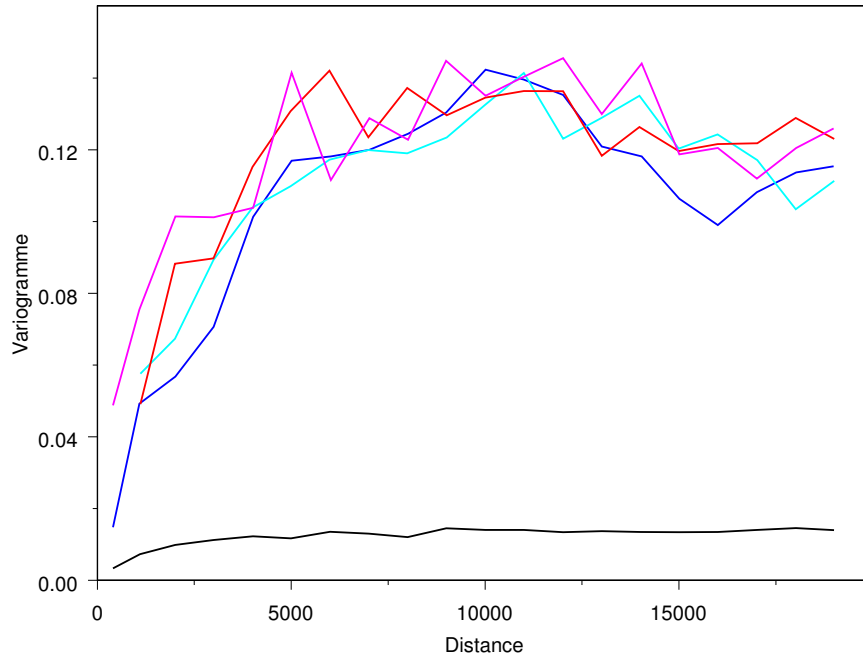


FIG. 5.24 – Variogrammes des données de gradient de la région Cévennes. Le variogramme γ_z de G_z est en noir, les variogrammes de G_x sont dans des teintes bleues (bleu pour $\gamma_{x\perp}$ et cyan pour $\gamma_{x\parallel}$), les variogrammes de G_y sont dans des teintes rouges (rouge pour $\gamma_{y\perp}$ et magenta pour $\gamma_{y\parallel}$).

parallèle sont assez semblables.

5.3.4 Modèle de la covariance du potentiel

Le modèle de la covariance du potentiel est une structure cubique composée de deux structures emboîtées. La première, isotrope, est de portée $5500m$ et de palier 2.10^4 . La seconde, dans le plan horizontal est de portée $13000m$ et de palier 96.10^4 .

	Portée (en m)	Palier $\frac{14C_i}{a_i^2}$
γ_3	5500	0.0102
γ_2	13000	0.0798

TAB. 5.16 – Portées et paliers des structures de base du modèle de variogramme pour les données des Cévennes.

	Effet pépite	Palier structure	Palier total
γ_{G_z}	0.003	0.0102	0.0132
γ_{G_x} ou γ_{G_y}	0.03	0.09	0.12

TAB. 5.17 – Effets de pépite et paliers des variogrammes des composantes du gradient pour les données des Cévennes.

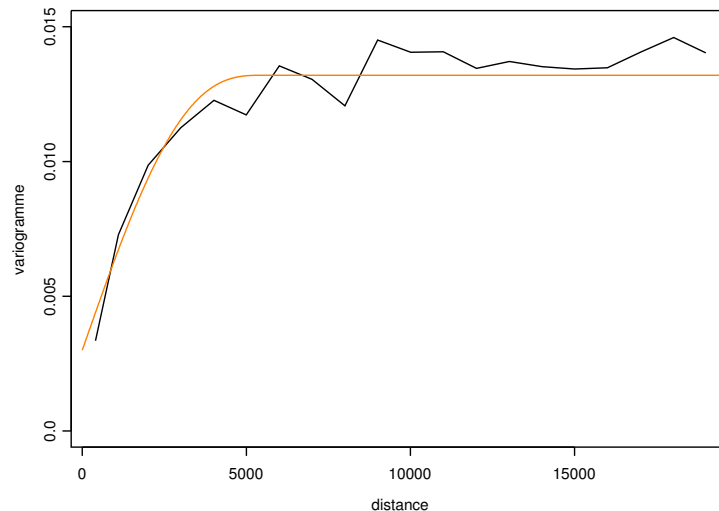


FIG. 5.25 – Cévennes : Ajustement du variogramme de la composante verticale du gradient.

5.4 Les Alpes Lepontines

5.4.1 Présentation du jeu de données

Localisation géographique et cadre géologique.

Ce jeu de données n'est pas situé en France, mais dans les Alpes centrales suisses dans le canton du Tessin (Figure 5.27). Différentes phases de déformation ont affecté les Alpes Léopontines et il en résulte un empilement de différentes nappes. Les données de foliation métamorphiques de plusieurs unités sont rassemblées au sein d'une unité tectonique majeure (Maxelon, 2004).

Le champ est approximativement un rectangle de 30 km sur 55 km et compte $N = 2148$ données de foliations. Les données sont aussi cartographiques, mais du fait du relief elles s'étagent sur un peu plus de 2 km. De plus, l'échantillonnage fait apparaître des zones plus ou moins bien couvertes par les données.

Étude exploratoire des données d'orientation

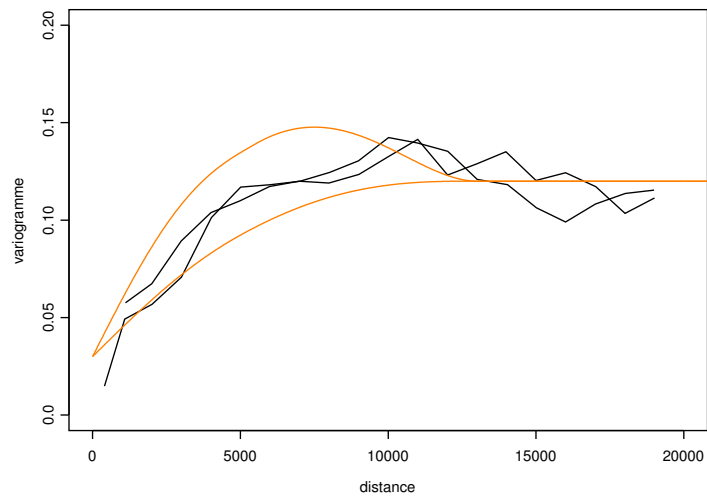
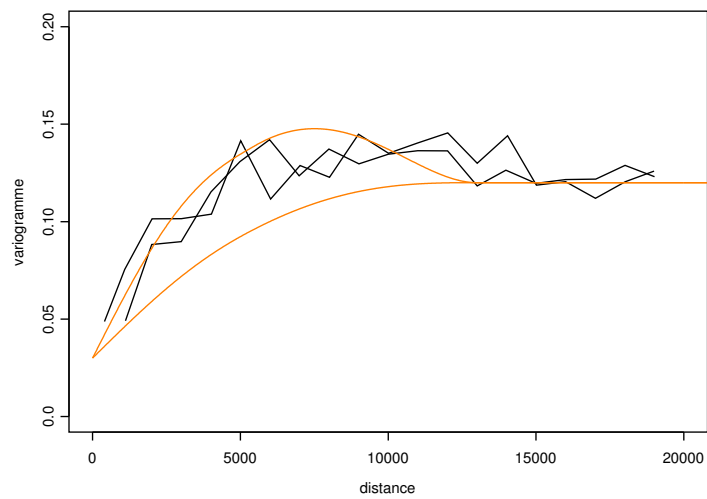
(a) G_x (b) G_y

FIG. 5.26 – Cévennes : Ajustement des variogrammes des composantes horizontales du gradient.

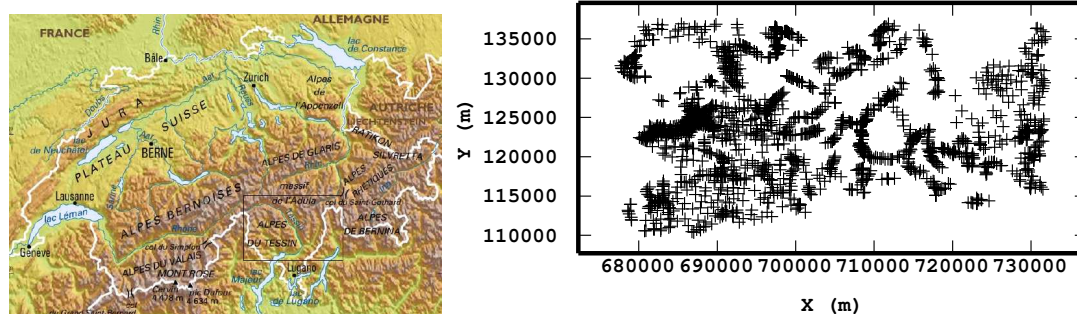
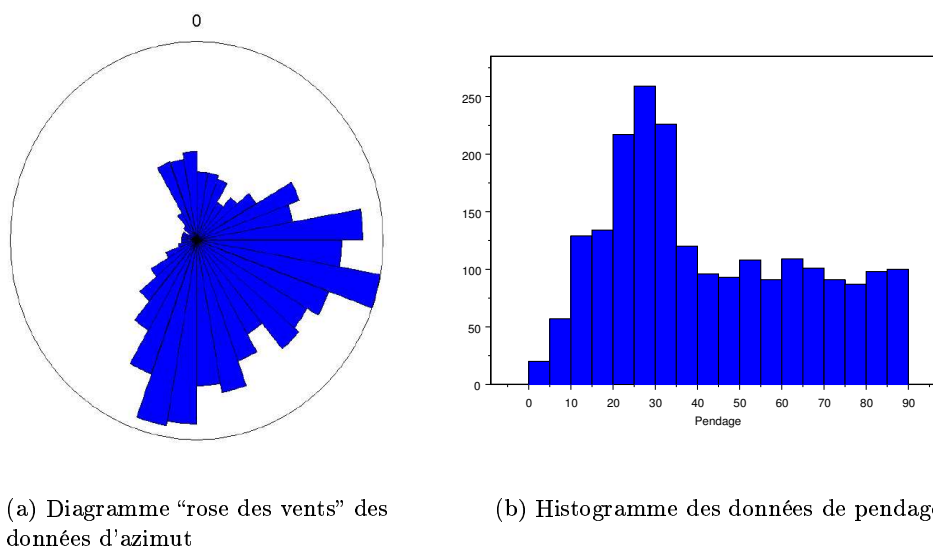


FIG. 5.27 – Plan de position des données de foliation de la région des Alpes centrales suisses.

La rose des vents des azimuts (Figure 5.28(a)) montre une distribution dissymétrique de ceux-ci. Toutes les directions comprises dans la moitié ouest sont quasiment absentes. La grande partie des données est située dans un grand quart sud - est.

Sur l'histogramme du pendage (Figure 5.28(b)), on observe que toutes les gammes de pendage sont représentées, et de manière égale entre 35° et 90° . On constate aussi qu'il y a un peu moins de pendages inférieurs à 10° et un maximum entre 20° et 35° .



(a) Diagramme “rose des vents” des données d’azimut

(b) Histogramme des données de pendage

FIG. 5.28 – Histogrammes des données de foliation des Alpes centrales suisses.

Étude exploratoire des données de gradient

Les statistiques élémentaires sont résumées dans le Tableau 5.18 et sont complétées par la

Figure 5.29. Les moyennes des composantes G_x et G_y ne sont plus aussi proches de 0 que

	Minimum	Maximum	Moyenne	Variance
G_x	-0.9779	0.9811	0.1400	0.1385
G_y	-1	1	-0.1265	0.2770
G_z	0	1	0.6820	0.0837

TAB. 5.18 – Statistiques élémentaires des données de gradient pour les Alpes Lépointines.

dans les autres jeux de données. La composante G_z est par contre toujours positive, mais sa variance est élevée. On retrouve plus de valeurs élevées correspondant aux pentages de la gamme 20°-35°. L'écart avec les variances des composantes horizontales est donc plus faible. La distribution de G_x est proche d'une distribution gaussienne, par contre G_y a plus de valeur proches de ± 1 . Le vecteur moyen a pour coordonnées (0.1979, -0.1788, 0.9638), ce qui est un peu différent de ce qu'on a observé dans les autres jeux de données.

La concentration est autour de 3.42, les vecteurs sont donc assez peu concentrés autour du pôle moyen.

Quant au paramètre de parallélisme, il a une valeur assez moyenne avec 42%.

Regardons à présent les variogrammes expérimentaux des composantes du gradient.

5.4.2 Variogrammes expérimentaux

La Figure 5.30 montrent les variogrammes expérimentaux du jeu de données des Alpes centrales suisses. Le pas utilisé est de 1000m et le nombre de pas est de 20. Le variogramme de G_z a toujours un palier plus faible que les variogrammes des composantes horizontales, mais l'écart est bien moins important qu'auparavant. L'anisotropie est marquée, avec un palier plus élevé pour G_y et l'emploi de trois structures pour l'ajustement sera nécessaire. Les portées semblent grandes et les effets de pépite plus faibles que dans les autres jeux, "Cévennes" excepté. On remarque aussi l'effet de trou important sur $G_{y||}$.

5.4.3 L'ajustement des variogrammes

	Portée (en m)	Palier $\frac{14C_i}{a_i^2}$
γ_3	23000	0.096
γ_2	7000	0.049
γ_1	9000	0.105

TAB. 5.19 – Portées et paliers des structures de base du modèle de variogramme pour les données des Alpes Lépointines.

Les Tableaux 5.19 et 5.20 rassemblent les paramètres du modèle pour chaque composante du gradient. Les Figures 5.31 et 5.32 montrent les variogrammes ajustés avec les paramètres de ces tableaux.

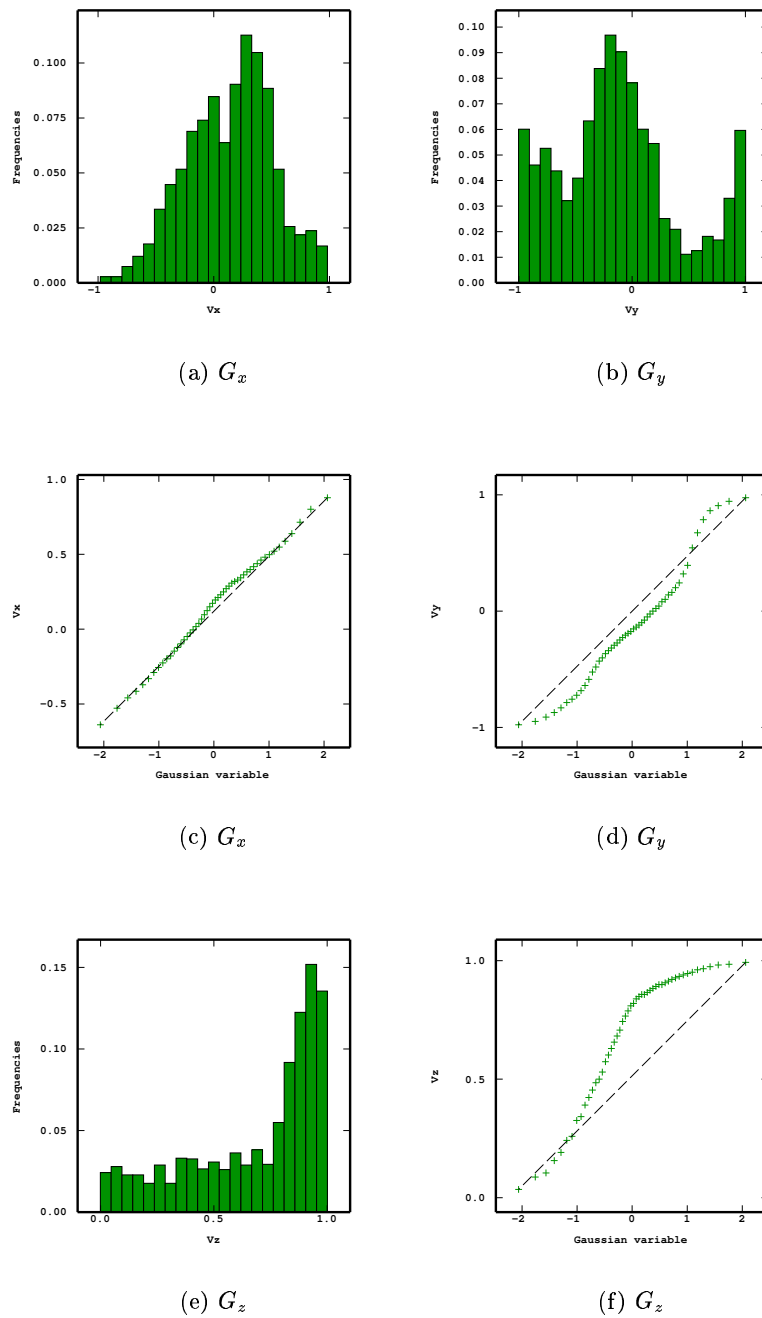


FIG. 5.29 – Histogrammes et diagrammes quantile contre quantile (distribution expérimentale vs distribution gaussienne) des composantes du gradient de la région des Alpes centrales suisses.

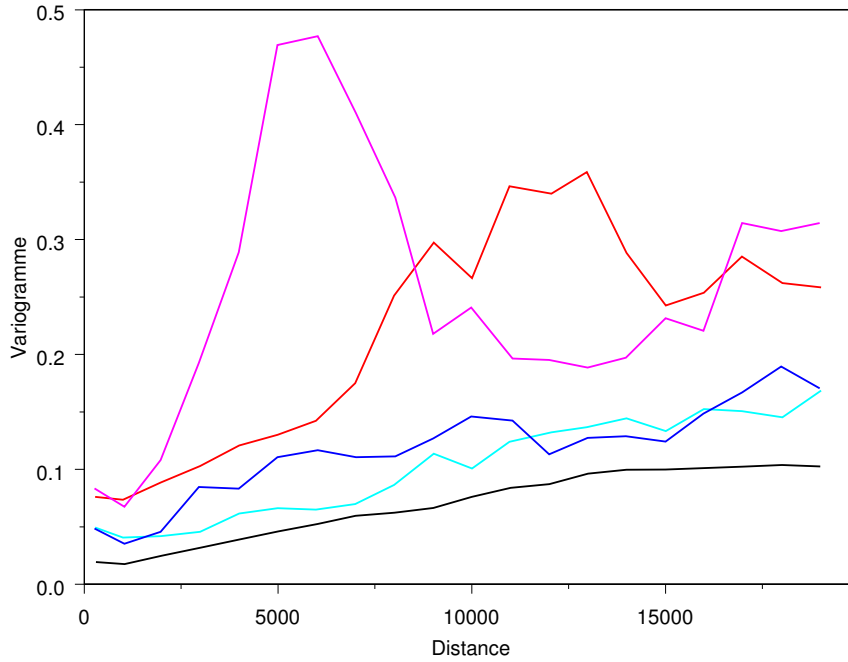


FIG. 5.30 – Variogrammes des données de gradient de la région des Alpes centrales suisses. Le variogramme γ_z de G_z est en noir, les variogrammes de G_x sont dans des teintes bleues (bleu pour $\gamma_{x\perp}$ et cyan pour $\gamma_{x\parallel}$), les variogrammes de G_y sont dans des teintes rouges (rouge pour $\gamma_{y\perp}$ et magenta pour $\gamma_{y\parallel}$).

Comme toujours la composante verticale du gradient s'ajuste bien. Les composantes horizontales s'ajustent plus difficilement, d'une part du fait du passage du variogramme de $G_{x\perp}$ au-dessus de celui de $G_{x\parallel}$ et d'autre part à cause de l'effet de trou important sur le variogramme expérimental de $G_{y\parallel}$. On a donc décidé d'ajuster au mieux le variogramme $G_{x\perp}$ sans tenir compte de $G_{x\parallel}$. On constate aussi que l'effet de trou du modèle pour $G_{y\parallel}$ est moins marqué à cause de la grande portée de G_z .

	Effet pépité	Palier structure	Palier total
γ_{G_z}	0.009	0.096	0.105
γ_{G_x}	0.02	0.145	0.165
γ_{G_y}	0.02	0.250	0.270

TAB. 5.20 – Effets de pépité et paliers des variogrammes des composantes du gradient pour les données des Alpes Lépointines.

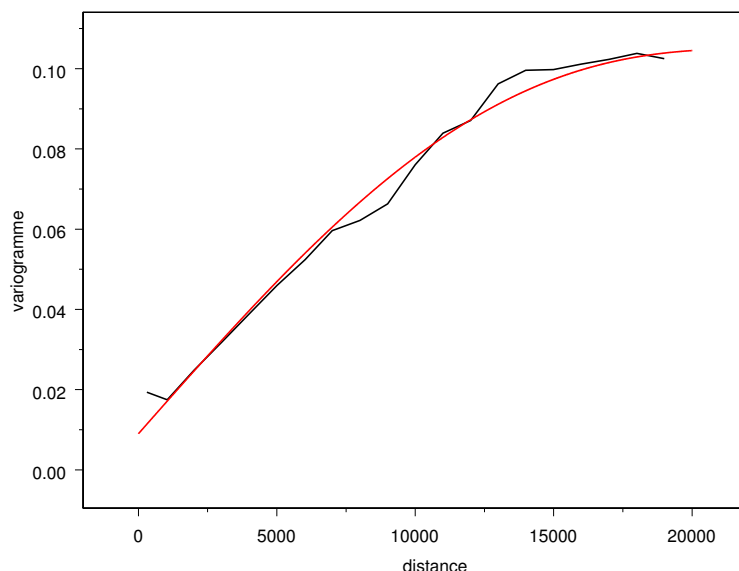


FIG. 5.31 – Alpes centrales suisses : Ajustement du variogramme de la composante verticale du gradient.

5.4.4 Modèle de la covariance du potentiel

Le modèle de covariance du potentiel retenu est un schéma cubique gigogne avec trois composantes. La première, isotrope, est de portée $23000m$ et a pour palier la valeur de 36.10^5 . La seconde qui ne dépend que des composantes horizontales de la distance a une valeur de $7000m$ pour portée et un palier de 17.10^4 . Enfin, la troisième structure suivant la direction nord-sud, a une portée de $9000m$ et un palier de 6.10^5 .

Selon Maxelon, qui a réalisé un modèle géologique de cette région avec le modèle de covariance conventionnel et celui inféré, il n'y a pas de changements majeurs qui remettent en cause l'interprétation géologique régionale. Par contre, dans le détail, le nouveau modèle fournit une interface géologique avec une géométrie plus proche de l'idée que s'en fait le géologue, en particulier les structures plissées semblent mieux restituées.

5.5 Synthèse de la variographie des quatre jeux de données

Ces quatre jeux de données ont permis d'observer les variogrammes des données de gradient et de déterminer les modèles de covariance du potentiel.

Les principales remarques sont les suivantes :

Nous avons étudié dans ce chapitre trois jeux de données de foliations métamorphiques de régions distinctes du Massif central et d'une région des Alpes centrales suisses qui présentent des caractéristiques différentes du point de vue statistique :

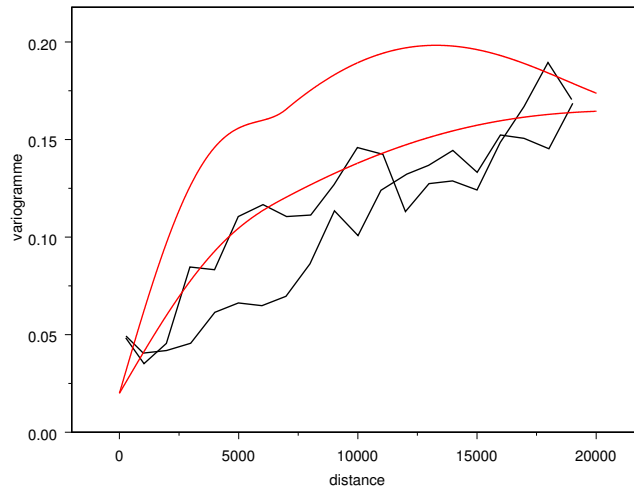
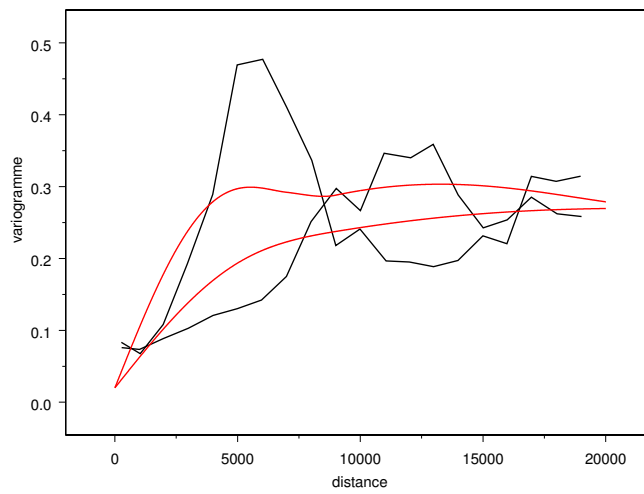
(a) G_x (b) G_y

FIG. 5.32 – Alpes centrales suisses : Ajustement des variogrammes des composantes horizontales du gradient.

- Velay - Toutes les directions d'azimut sont présentes et tous les pendages sont représentés ;
- Limousin - Des directions d'azimut sont privilégiées et les forts pendages sont absents ;
- Cévennes - Toutes les directions d'azimut sont présentes et les forts pendages sont absents ;
- Alpes centrales suisses - Des directions d'azimut sont privilégiées et tous les pendages sont représentés ;

Du point de vue variographique, on constate que tous les variogrammes expérimentaux des données d'orientation ont de nombreux points communs :

- en règle générale portée, palier et effet de pépite sont bien discernables (hormis peut-être les composantes horizontales dans le jeu suisse) ;
- l'effet de trou est peu ou pas visible sur les variogrammes calculés suivant la direction parallèle à la direction de dérivation ;
- le palier du variogramme de la composante verticale du gradient est toujours inférieur à ceux des composantes horizontales ;
- une grande partie de la variabilité est soutenue par l'effet de pépite ;
- la portée des différentes structures est de l'ordre de la dizaine de milliers de mètres (en gros de 5000 à 50000m) ;
- la variance totale du gradient est toujours bien inférieure à 1.

Une différence importante entre les différents jeux de données est la structuration plus ou moins importante suivant les régions (par exemple les données du Velay le sont très peu avec jusqu'à 75% de la variabilité supportée par l'effet de pépite contre 25% pour les Cévennes voire moins de 10% en Suisse.) Les effets de pépite varient en effet beaucoup suivant les champs étudiés de 0.003 à 0.051 pour G_z et de 0.02 à 0.21 pour les composantes horizontales.

Les variogrammes calculés suivant d'autres directions que celles des axes géographiques sur le jeu du Limousin ont montré que l'anisotropie peut être importante entre les deux composantes horizontales. Si on faisait une étude identique sur les données du Velay par exemple, on constaterait des écarts bien moins grands, signe d'une anisotropie peu marquée. Par rapport au modèle heuristique, cette approche introduit de l'anisotropie zonale avec deux ou trois structures suivant les jeux. Portée et palier sont bien inférieurs aux valeurs par défaut et l'effet de pépite est souvent plus grand.

A partir du moment où le modèle de covariance est inféré, il est possible de construire la surface et de déterminer les incertitudes sur sa position en profondeur, c'est l'objet du chapitre suivant.

Chapitre 6

La détermination des incertitudes sur la position des interfaces

Lorsque la covariance était choisie *a priori* par l'utilisateur, le cokrigeage était un interpolateur parmi d'autres. L'inférence de la covariance du potentiel à partir des données d'orientation a permis de rendre cet interpolateur "optimal". Il est de plus possible à présent de déterminer les incertitudes à partir des variances de cokrigeage, puisque celles-ci ont désormais un sens. Mais revenons en premier lieu, sur la comparaison des interpolations obtenues avec les deux modèles : conventionnel et ajusté, pour un cas concret.

6.1 Impact du modèle de covariance sur l'interpolation dans le cas du Limousin

6.1.1 Présentation

Nous allons reprendre le cas d'étude du Limousin présenté à la section 5.2. La Figure 6.1 montre l'ensemble des données de la zone Limousin présentes sur la surface topographique pour le toit de l'interface "uig" : les données d'orientation décrites précédemment augmentées de données d'interface. L'information issue de la carte géologique est dense, une grande partie des contours se dessinant très bien. Par contre aucune donnée de forage n'est utilisée.

Sur la Figure 6.1 sont également représentées les traces de trois coupes est-ouest utilisées ultérieurement. La coupe 1 à l'extrême nord de la carte n'a pas de données d'orientation dans son voisinage et la partie orientale est vide de toute donnée. La coupe 2 au sud traverse une grande zone sans donnée à l'ouest, et à l'est on trouve à la fois des données d'interface et des données d'orientation. Enfin, la coupe A centrale est environnée de données sur tout son parcours.

L'interpolation est effectuée avec un rayon de simplification nul, autrement dit toutes les données interviennent sans fusion de certaines d'entre elles. Les paramètres du modèle ajusté sont donnés au paragraphe 5.2.4. Quant au modèle conventionnel, il consiste en une seule structure isotrope cubique de portée $98000m$ et de palier 229.10^6 , l'effet de pépité sur

les gradients étant fixés à 0.01. L'effet de pépite sur les incréments de potentiel est dans tous les cas de 10^{-6} .

6.1.2 Comparaison des interpolations

Les tracés obtenus en carte (altitude constante correspondant au plan topographique) de l'isosurface modélisant le toit de "uig" à partir des deux modèles, conventionnel et ajusté, sont très similaires en apparence (Fig. 6.2). Les champs de potentiel interpolés sont pourtant différents ; regardons ce qu'il se passe au niveau des isopotentielles. Les Figures 6.3 présentent justement dix isopotentielles pour les deux cas de figure, le toit de "uig", qui est une isopotentielle particulière, possède la valeur de potentiel la plus élevée. Les valeurs absolues de potentiel étant différentes, les écarts entre deux isopotentielles pour les deux modèles ne sont pas les mêmes. Deux trajectoires ont un écart de 0.021 pour le modèle conventionnel et de 0.005 pour le modèle ajusté. Si on fait abstraction de cela, on constate que les différences ont essentiellement lieu en extrapolation.

Si on se concentre uniquement sur l'interface considérée, on note aussi quelques différences au nord et à l'est où aucune donnée n'est présente. En revanche, la courbe fermée au centre a été trouvée dans les deux modèles alors qu'il n'y a pas une seule donnée d'interface. Les données d'orientation suffisent seules ici à reconstruire cet élément de la surface. Evidemment, la différence entre les modèles sera d'autant plus grande qu'il y aura peu de données d'interface pour contraindre l'interpolation. Si on compare maintenant les tracés du toit de l'interface au niveau de la coupe 1, on remarque que le décalage entre les modèles est plus grand qu'il n'y paraît. La différence en profondeur entre les deux modèles peut en effet atteindre 450 m. On considère le modèle ajusté comme plus raisonnable puisque sa covariance a été obtenue à partir des données disponibles, ce qui permet de calculer des écart-types de cokrigage sensés.

6.2 Détermination des incertitudes sur la position des interfaces

6.2.1 Une approche empirique

Peu de travaux existent sur l'évaluation des incertitudes en modélisation géologique 3D. Il faut néanmoins citer les travaux de l'Ecole Polytechnique Fédérale de Lausanne (Tacher et Parriaux, 1997) et (Pomian-Srzednicki, 2001). Ne s'intéressant pas à l'interpolation, ils récupèrent un modèle 3D (*EarthVision*) considéré comme étant le plus probable, nommé *best guess*. Ils considèrent alors les écarts par rapport à cette position de l'interface en déterminant une variance de krigeage. Ainsi tout point de la surface est décrit par une fonction aléatoire composée d'un terme déterministe (position prédite par le modèle) et d'un terme aléatoire (fluctuations autour de la position). La fonction aléatoire étudiée est la coordonnée z exprimée dans un repère local à la surface. Le modèle de covariance est un modèle gaussien dont les paramètres sont déterminés empiriquement (amplitude et période des structures plissées). Il faut signaler aussi que des informations géologiques comme l'épaisseur des formations et la forme des plis sont aussi pris en compte. Les probabilités

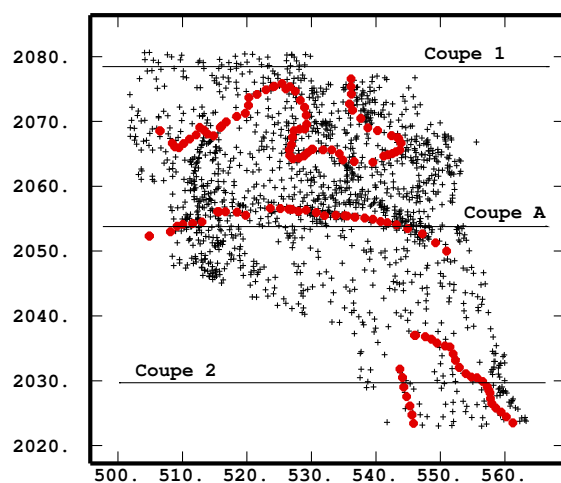


FIG. 6.1 – Plan de position des données de la région Limousin. Croix noires : données de foliation. Disques rouges : données d'interface.

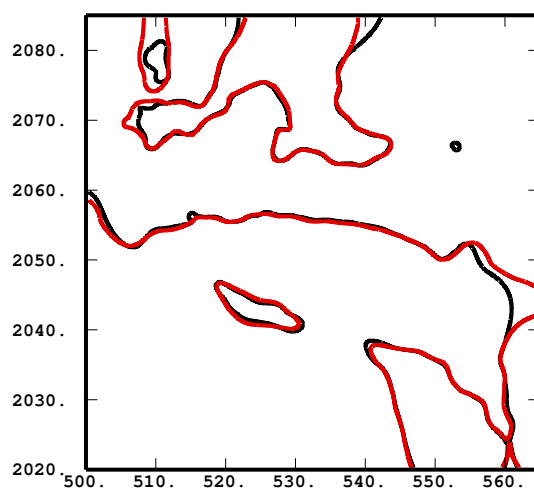
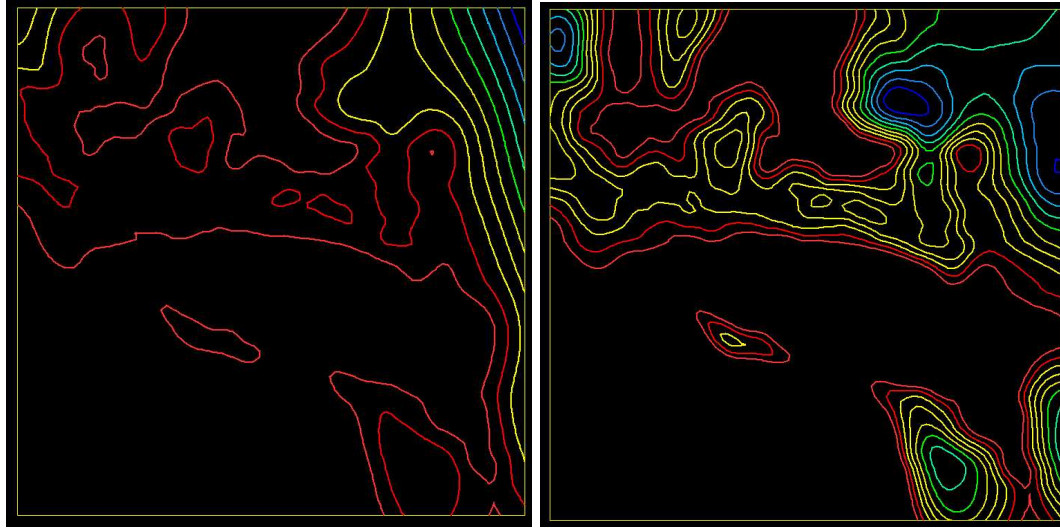


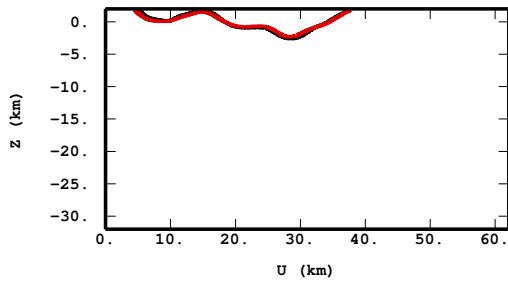
FIG. 6.2 – Comparaison de l'interpolation du toit de "uig" (Limousin) pour deux modèles différents. En noir, modèle "conventionnel". En rouge, modèle ajusté.



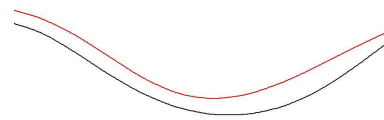
(a) Modèle conventionnel.

(b) Modèle ajusté.

FIG. 6.3 – Comparaison des isopotentielles issues du champ de potentiel servant à modéliser le toit de “uig” (Limousin) pour deux modèles différents, conventionnel et ajusté.



(a) Coupe 1.



(b) Zoom sur la zone la plus profonde de la Coupe 1 (vers 2500m de profondeur.) On constate un écart de l'ordre de 450m.

FIG. 6.4 – Comparaison de l'interpolation du toit de “uig” (Limousin) pour deux modèles différents dans une coupe. En noir, modèle “conventionnel”. En rouge, modèle ajusté.

d'occurrence des différents types de roche sont calculées permettant la détermination d'une information volumique sur l'incertitude. Cette dernière démarche est semblable à celle utilisée dans la suite de notre travail pour cartographier le potentiel réduit (6.2.5). Pour plusieurs interfaces les probabilités sont combinées en prenant en compte les règles de dépôt et d'érosion.

6.2.2 Retour à la méthode du potentiel : variance d'estimation

A partir du moment où la covariance du potentiel découle des données et a un sens physique, il est possible de se pencher sur le problème crucial de la détermination des incertitudes du modèle. Il est important lorsqu'on utilise un modèle de connaître les erreurs susceptibles d'entacher le résultat. Les incertitudes existent à différents niveaux. Ainsi elles peuvent concerner les données (par exemple sur la localisation de celles-ci), les mesures (qualité) ou les méthodes de calculs (par exemple l'interpolation). Dans notre cas de figure, la question se pose concernant les incertitudes sur la position (en profondeur) de l'interface calculée. La variance de cokrigage fournit une mesure de l'erreur associée à l'estimateur de cokrigage, mais n'est pas disponible dans l'optique de cokrigage dual implémentée dans l'*Editeur*. Un module a été ajouté dans le logiciel afin de choisir de déterminer ou non les incertitudes. Si elles doivent être déterminées, on résout le système de cokrigage classique, ce qui augmente le temps de calcul de manière significative. La matrice de cokrigage est identique à celle du système dual, c'est le second membre qui change. L'inversion de la matrice de covariance, faite une fois pour toute, peut également servir lors du traitement des données d'inégalité. Les variances de cokrigage en un point $\mathbf{x}_0$ sont alors calculées classiquement à l'aide de l'expression :

$$\sigma_{CK}^2 = \sigma_{00} - \sum_{\alpha} \lambda_{\alpha} \sigma_{\alpha 0} - \sum_l \mu_l f_0^l \quad (6.1)$$

Il faut cependant attirer l'attention sur le fait qu'il s'agit de la variance d'estimation d'un incrément de potentiel. Par exemple, le terme constant σ_{00} de l'expression précédente s'écrit :

$$\sigma_{00} = 2[C(0) - C(\mathbf{x}_0 - \mathbf{x}_{\text{ref}})] \quad (6.2)$$

où $C(0)$ est la variance *a priori* et $C(\mathbf{x}_0 - \mathbf{x}_{\text{ref}})$ est la covariance entre le point courant et le point de référence de l'interface.

Il s'agit maintenant de transposer cette incertitude sur le potentiel en incertitude sur la géométrie des formations ou des interfaces.

6.2.3 Probabilité d'appartenance à une formation

Avant de définir la probabilité d'appartenance d'un point à une interface, définissons cette notion au niveau de la formation géologique. Soit $P(\mathbf{x})$ la valeur du potentiel en un point $\mathbf{x}$ quelconque de l'espace et $\sigma_{CK}(\mathbf{x})$ l'écart-type d'estimation en ce même point. Supposons qu'une formation géologique F soit définie par l'ensemble des points $\mathbf{x}$ tels que :

$$P_1 < P(\mathbf{x}) \leq P_2 \quad (6.3)$$

où P_1 et P_2 sont deux valeurs de potentiel particulières correspondant au mur et au toit de F .

Si on fait l'hypothèse que le champ de potentiel est une fonction aléatoire gaussienne, la probabilité qu'un point $\mathbf{x}$ appartienne à la formation F est :

$$\begin{aligned} Pr(P_1 < P(\mathbf{x}) \leq P_2) \\ = G\left(\frac{P_2 - P^*(\mathbf{x})}{\sigma_{CK}(\mathbf{x})}\right) - G\left(\frac{P_1 - P^*(\mathbf{x})}{\sigma_{CK}(\mathbf{x})}\right) \end{aligned} \quad (6.4)$$

où G est la fonction de répartition de la loi normale centrée réduite.

6.2.4 Probabilité d'appartenance à une interface - Définition du potentiel réduit

Raisonnons de manière similaire en nous intéressant maintenant au toit de la formation F (par exemple l'interface passant par le point $\mathbf{x}_0$) : $P(\mathbf{x}_0) = P_2$. On s'occupe donc de l'ensemble des points $\mathbf{x}$ tels que : $P(\mathbf{x}) - P(\mathbf{x}_0) = 0$. Définissons le potentiel réduit $\Phi(\mathbf{x})$ par :

$$\Phi(\mathbf{x}) = \frac{P^*(\mathbf{x}) - P^*(\mathbf{x}_0)}{\sigma_{CK}(\mathbf{x})} \quad (6.5)$$

Cette variable représente l'écart réduit au potentiel en un point donné. Comme $P(\mathbf{x}) - P(\mathbf{x}_0) = P^*(\mathbf{x}) - P^*(\mathbf{x}_0) + ECK$ avec ECK l'erreur de cokrigeage, on voit que $\mathbf{x}$ appartient à l'interface si et seulement si $P^*(\mathbf{x}) - P^*(\mathbf{x}_0) = -ECK$ ou si $\Phi(\mathbf{x}) = -\frac{ECK}{\sigma_{CK}(\mathbf{x})}$.

Comment interpréter le potentiel réduit en terme de probabilité ?

Considérons l'hypothèse $H_0 = \{\mathbf{x} \in I_{\mathbf{x}_0}\} : x$ est sur l'interface passant en x_0 . Alors sous H_0 , et si le champ de potentiel est gaussien, Φ est une variable gaussienne centrée réduite. Or le fait d'observer $|\Phi|$ grand est contraire à H_0 . On doit donc rejeter cette hypothèse si $Pr\{|Y| > |\Phi|\} < 0.05 = G(-|\Phi|) + 1 - G(|\Phi|) = 2G(-|\Phi|)$ et $2G(-|\Phi|)$ est donc la p-valeur associée.

Les courbes $\Phi = \pm 2$, par exemple, localisent l'intervalle de confiance à 95% autour de la position de l'interface si on se place dans l'hypothèse gaussienne¹. Quoique l'intervalle de confiance soit ici purement conventionnel, on peut cependant relativiser cette assertion, car si on prend un intervalle de largeur $\pm 3\sigma$, on couvre une probabilité supérieure ou égale à 95% pour toutes les distributions continues et unimodales (Chilès et Delfiner, 1999).

6.2.5 Cartographie du potentiel réduit dans le cas du Limousin

Le potentiel réduit permet de mieux se représenter les choses en transposant les incertitudes sur le potentiel en incertitudes sur la position des interfaces. La cartographie de cette nouvelle variable consiste donc à en tracer des isovaleurs. Les figures suivantes (Figures 6.5, 6.6, 6.7, 6.8, 6.9) montrent des cartes ou coupes obtenues avec les données du Limousin.

Le tracé de l'interface est en noir ($\Phi(\mathbf{x}) = 0$), les tons bleus correspondent aux points

¹On fait tout de même cette hypothèse bien qu'une variance d'estimation ne soit pas un intervalle de confiance sur l'estimation, ce dernier nécessitant la connaissance de la loi de probabilité de l'erreur

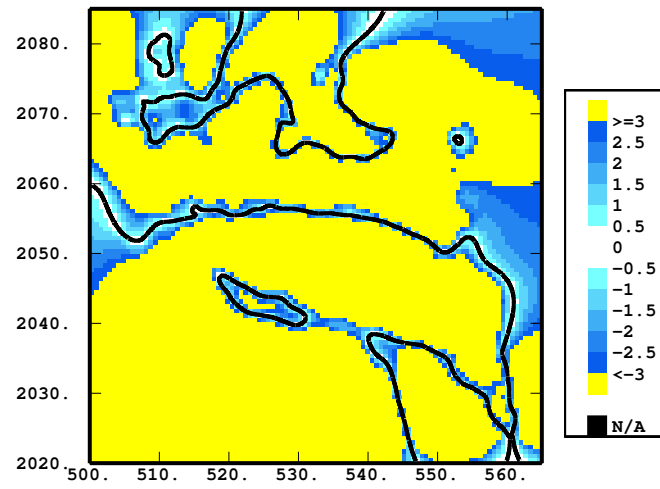


FIG. 6.5 – Limousin - modèle “par défaut” : carte du potentiel réduit.

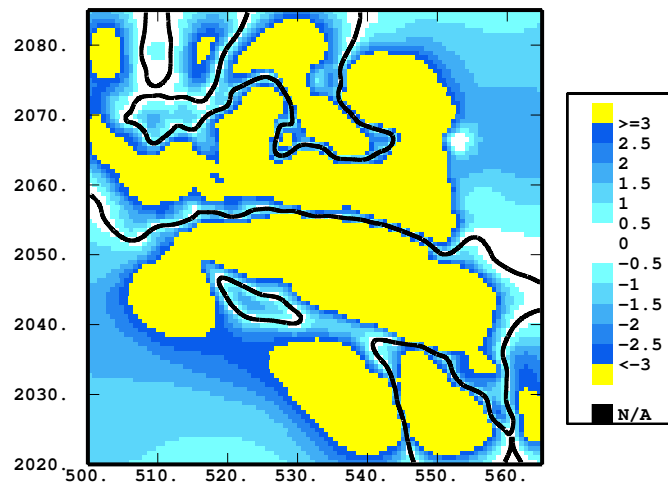


FIG. 6.6 – Limousin - modèle ajusté : carte du potentiel réduit.

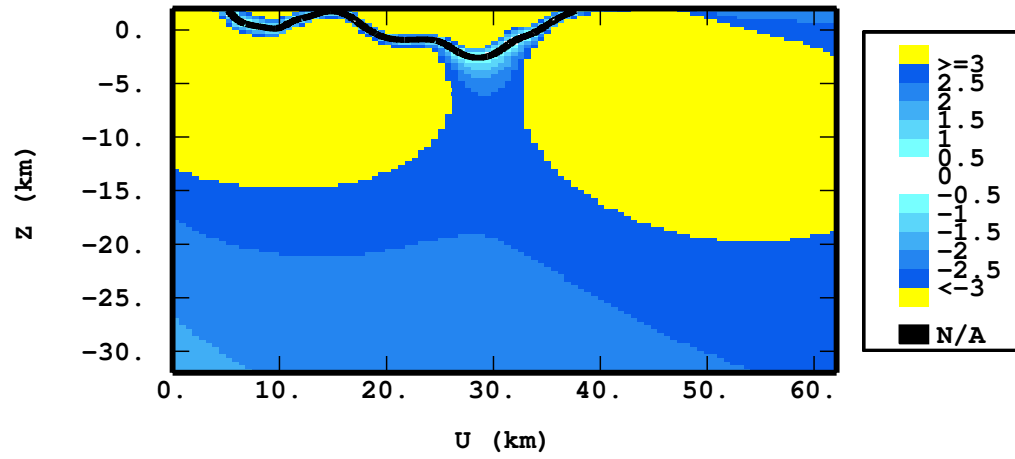


FIG. 6.7 – Limousin, Coupe 1 - modèle “par défaut” : carte du potentiel réduit.

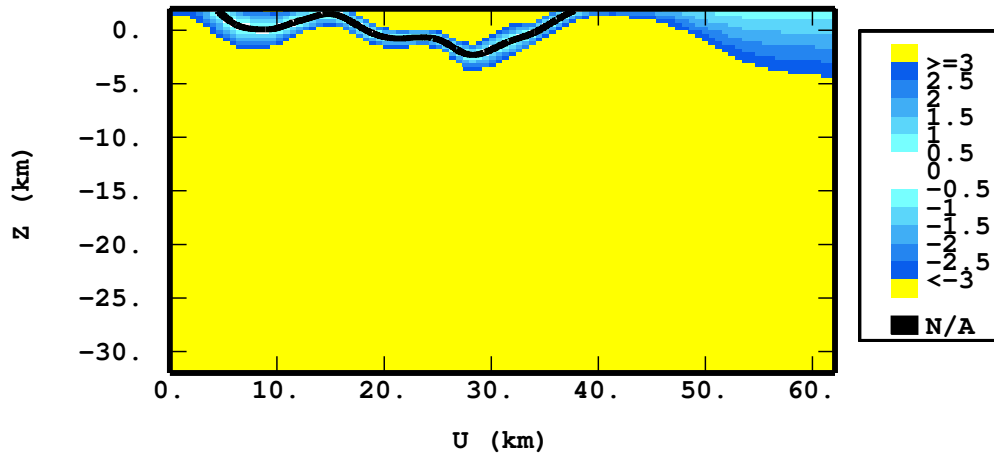


FIG. 6.8 – Limousin, Coupe 1 - modèle ajusté : carte du potentiel réduit.

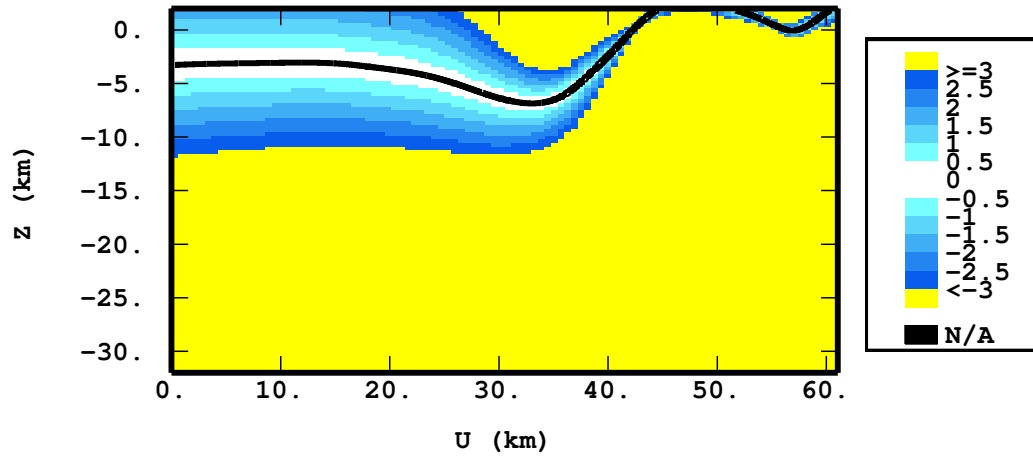


FIG. 6.9 – Limousin, Coupe 2 - modèle ajusté : carte du potentiel réduit.

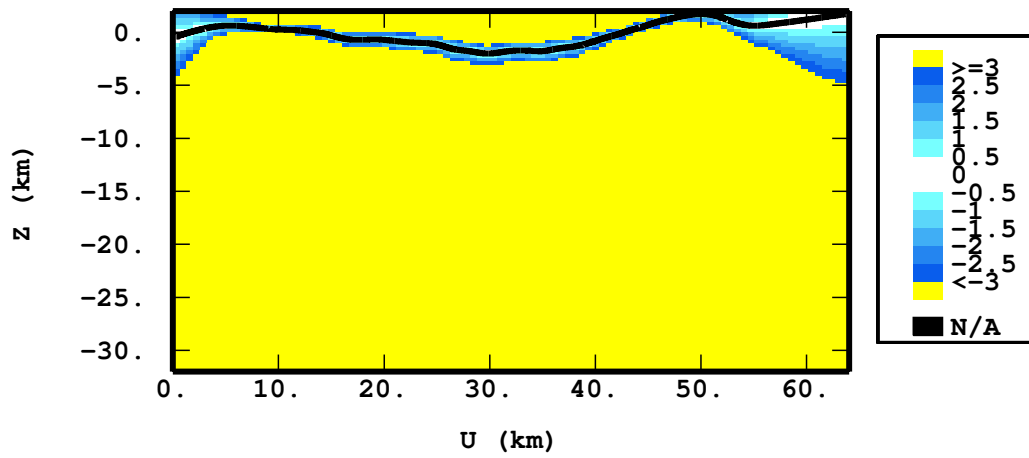


FIG. 6.10 – Limousin, Coupe A - modèle ajusté : carte du potentiel réduit.

où le potentiel réduit est compris entre ± 3 . L'aire définie par $|\Phi(\mathbf{x})| < 3$ correspond à 99% des réalisations possibles de l'interface dans une hypothèse gaussienne, le tracé a donc très peu de chance de passer dans une plage jaune qui représente une zone "interdite".

On observe bien sur les coupes, les zones où l'incertitude est grande (plages de bleu importantes) de celles bien contraintes par les données (tracé de l'interface quasiment au contact de l'aire interdite en jaune). Ces images peuvent être très utiles pour localiser par exemple les endroits où l'incertitude est forte afin d'y effectuer un sondage supplémentaire.

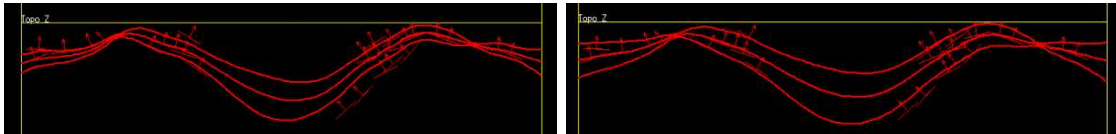
On constate de plus la présence à la Figure 6.7 d'une zone de confiance plus faible en profondeur, ce qui se traduit par un "trou", car la courbe boucle. Ce trou est par ailleurs inexistant à la Figure 6.8. Une autre illustration de cette cartographie des incertitudes est fournie par la Figure 6.9 où une grande partie de la coupe à l'ouest est en extrapolation.

On remarque aussi que les incertitudes sont plus importantes pour le modèle ajusté. On a de plus réduit le domaine dans lequel les données de gradient prennent leurs valeurs puisque la variance du gradient est plus faible. Le poids des erreurs de mesure est donc plus important.

Ces différences dans les résultats d'incertitudes entre modèle ajusté et modèle conventionnel s'expliquent aussi par les effets de pépite sur les composantes du gradient plus forts dans le cas du modèle ajusté. Cela se voit particulièrement bien à la Figure 6.11 où un petit test de sensibilité montre l'influence de ce paramètre du modèle sur les incertitudes. A gauche, les courbes sont interpolées avec la valeur de l'effet de pépite par défaut. A droite, cette valeur est 10 fois plus grande.

Nota Bene : dans les deux cas l'effet de pépite imposé est isotrope.

Globalement, les incertitudes sont donc plus faibles pour le modèle par défaut, mais ce modèle n'a pas de sens physique. On retient donc comme résultat final, les cartes obtenues avec le modèle ajusté.



(a) Effet de pépite sur le gradient = 0.01

(b) Effet de pépite sur le gradient = 0.1

FIG. 6.11 – Influence de l'effet de pépite du gradient sur les incertitudes.

6.3 Ecart-types en profondeur et effet de pépité sur les données de potentiel

6.3.1 Détermination des écart-types en profondeur

Comme on ne dispose pas d'information quantitative exprimée en mètres, puisque l'on a travaillé jusqu'à présent avec les potentiels, un écart-type empirique en profondeur σ a été calculé à partir de l'écart-type de potentiel σ_{CK} . Nous nous sommes alors placés dans une coupe où le tracé de l'interface a été obtenu. Il suffit ensuite d'estimer G^z , la valeur de la composante suivant z du gradient, aux nœuds de la grille passant par les points interpolés de l'interface. Cette approche est généralisable à des valeurs de gradient perpendiculairement au contour. L'incertitude en profondeur ne peut donc être fournie que pour des surfaces subhorizontales. On a alors la relation, en première approximation :

$$\sigma = \frac{\sigma_{CK}}{G^z} \quad (6.6)$$

La Figure 6.12 montre les plages où les écart-types prennent leur valeur dans la Coupe A (Fig.6.10).

On dispose donc désormais de la donnée en mètres de l'écart-type pour chaque point de

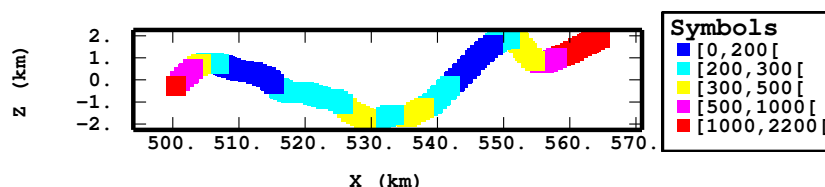


FIG. 6.12 – Limousin, Coupe A - modèle ajusté : carte de l'écart-type en profondeur. (Le rapport entre longueur et profondeur n'est pas respecté.)

l'interface. On peut prolonger cette idée en le calculant en tous points de la grille 3D, afin d'obtenir un bloc tridimensionnel d'écart-types.

6.3.2 L'effet de pépité sur les données de potentiel

L'effet de pépité sur les données de potentiel traduit les erreurs de mesure sur les points positionnés sur les interfaces, ce qui entraîne des incréments légèrement différents de 0. Regardons dans le Tableau 6.1 les statistiques élémentaires obtenues sur la Coupe 1 pour les écart-types en profondeur avec différents effets de pépité².

La Figure 6.14 est la courbe de la moyenne des écart-types en fonction de l'effet de pépité tracée entre autres à partir des valeurs du Tableau 6.1. Rappelons que toutes les données sont cartographiques, il n'y a sur ces exemples aucun point en profondeur. Dans toutes les

²L'interpolation a été obtenue avec un rayon de simplification de 3000m.

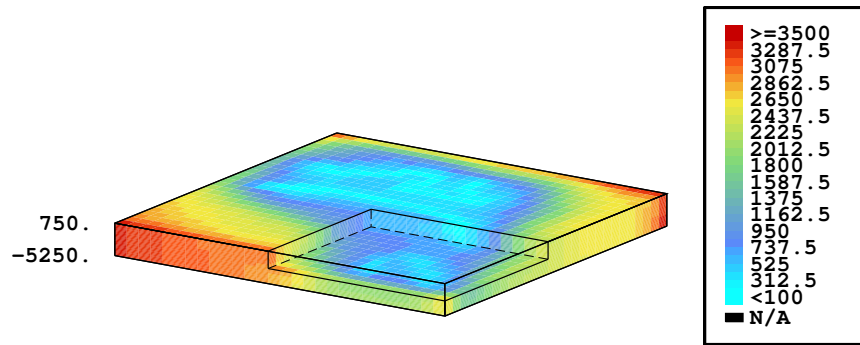


FIG. 6.13 – Limousin, modèle ajusté : bloc 3D de l'écart-type en profondeur.

Effet pépite	Minimum	Maximum	Moyenne	Ecart-type
10^{-4}	1277	1924	1488	183
10^{-5}	606	1504	922	268
10^{-6}	435	1445	815	301
10^{-7}	412	1438	801	306
10^{-8}	410	1437	800	307

TAB. 6.1 – Limousin, Coupe 1. Statistiques élémentaires des écart-types en profondeur du toit de “uig” en fonction de l'effet de pépite sur le potentiel.

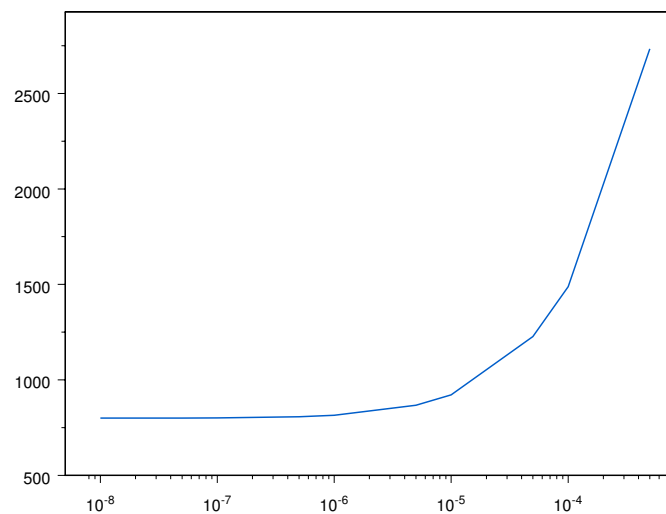


FIG. 6.14 – Limousin. Courbe de la moyenne des écart-types sur l'interface de la Coupe 1 en fonction de l'effet de pépite (échelle semi-logarithmique).

études précédentes, c'est la valeur de 10^{-6} qui a été utilisée.

On observe que pour 10^{-7} et 10^{-8} , les valeurs obtenues sont très proches, l'effet de pépite sur les données de potentiel étant trop petit pour avoir une réelle influence.

De même, en un point de donnée la variance d'estimation sera non nulle et d'autant plus forte que l'effet de pépite sera grand. Plaçons un seul point sur la coupe C1 au plus près de l'interpolation obtenue sans ce point et recalculons la surface. Sur cet exemple, l'écart-type est multiplié par 10 si l'effet de pépite est multiplié par 100, passant de 27m pour 10^{-7} à 275m pour 10^{-5} . Pour une valeur de 10^{-6} , l'écart-type est de 87m. Ces chiffres sont importants, mais il faut voir qu'aucune contrainte n'existe en profondeur, l'interface étant parfois à plusieurs kilomètres de la surface. Ces résultats ont été obtenus avec uniquement les données de foliation qui ont permis d'inférer le modèle de covariance ainsi que quelques points provenant de la carte géologique. Dans un cas d'étude plus poussé, les données de forages, voire géophysiques auraient également été utilisées.

Nous allons voir dans la suite de ce travail qu'il est possible aussi de déterminer les incertitudes affectant les paramètres du modèle de covariance que nous avons déterminés une fois pour toute au chapitre 5.

Chapitre 7

Les incertitudes sur les paramètres du modèle par l'approche bayésienne

Dans une approche géostatistique classique, on considère les variables régionalisées d'intérêt comme des réalisations de fonctions aléatoires. Si on fait l'hypothèse de stationarité d'ordre deux, la covariance expérimentale obtenue à l'aide des données est ajustée. Les paramètres du modèle sont donc choisis "au mieux". Dans l'analyse bayésienne, on considère les paramètres du modèle de covariance également comme aléatoires. La randomisation de ces paramètres permet ainsi d'utiliser des techniques de simulation et donc de déterminer les incertitudes relatives à leur estimation.

7.1 Rappels sur l'estimation par approche bayésienne dans le cas de champs aléatoires gaussiens

7.1.1 L'approche bayésienne en géostatistique : une rencontre relativement récente

Il existe depuis presque une vingtaine d'années une abondante littérature sur la géostatistique bayésienne. L'incertitude sur les paramètres du modèle de covariance est très souvent abordée, en particulier dans (Kitanidis, 1986) et (Handcock et Stein, 1993). De même, deux thèses récentes soutenues au Centre de Géostatistique ont largement utilisé les techniques bayésiennes. L'une pour étudier l'incertitude attachée à l'estimation de la géométrie d'une scène à partir d'un couple stéréoscopique d'images (Sénégas, 2002), l'autre pour incorporer l'incertitude de modélisation dans l'estimation des réserves d'un projet minier (Goria, 2004). Nous avons suivi dans ce travail les procédures mises en place par cette dernière. L'approche bayésienne permet, en outre, de fournir une méthode statistique de mise à jour de l'estimation lorsque de nouvelles données sont disponibles. Enfin, elle autorise l'incorporation au processus d'estimation des connaissances géologiques préalables. Le krigeage bayésien est une technique qui mixe la théorie du krigeage et l'analyse bayésienne ; il est présenté entre autres dans (Omre, 1987) et (Omre et Halvorsen, 1989).

7.1.2 La formule de Bayes

Pour une revue plus complète des méthodes bayésiennes, on se reportera utilement à (Droesbeke *et al.*, 2002) et à (Berger, 1985), car on se bornera ici à présenter quelques notions de base sur le formalisme bayésien.

Toute l'inférence repose sur la densité *a posteriori* des paramètres considérés conditionnellement aux données et aux informations *a priori* disponibles.

Soit $Z(x)$ le champ aléatoire d'intérêt (champ de potentiel) et $\mathbf{Z} = {}^t(Z(x_1), \dots, Z(x_n))$ un jeu de n observations d'une seule réalisation du champ aléatoire. Soit $\boldsymbol{\theta}$, le vecteur des paramètres du modèle. Ils sont vus comme une variable aléatoire vectorielle de densité *a priori* $\pi(\boldsymbol{\theta})$. Cette densité *a priori* tient compte des connaissances à notre disposition sur $\boldsymbol{\theta}$ avant la prise en compte des observations.

La densité *a posteriori* reflète quant à elle l'information sur les paramètres mise à jour après l'expérimentation. Elle s'écrit :

$$\pi(\boldsymbol{\theta}|\mathbf{Z}) = \frac{\pi(\boldsymbol{\theta})f(\mathbf{Z}|\boldsymbol{\theta})}{\pi(\mathbf{Z})} \quad (7.1)$$

Comme le dénominateur ne dépend pas de $\boldsymbol{\theta}$, il représente la constante d'intégration (ou de normalisation) de la formule suivante :

$$\pi(\boldsymbol{\theta}|\mathbf{Z}) \propto \pi(\boldsymbol{\theta})f(\mathbf{Z}|\boldsymbol{\theta}) \quad (7.2)$$

Il s'agit de la formule de Bayes qui permet de réactualiser les probabilités. On voit qu'elle combine connaissance *a priori* et information contenue dans les données par l'intermédiaire de la fonction de vraisemblance $f(\mathbf{Z}|\boldsymbol{\theta})$ notée aussi $L(\boldsymbol{\theta}|\mathbf{Z})$.

Si on s'intéresse à l'estimation du potentiel, on cherche à calculer :

$$f(\mathbf{Z}_0|\mathbf{Z}) = \int f(\mathbf{Z}_0|\mathbf{Z}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{Z})d\boldsymbol{\theta} \quad (7.3)$$

Cette expression prend en compte l'incertitude sur les paramètres en moyennant la densité de probabilité conditionnelle sur l'espace des paramètres. C'est une moyenne pondérée par la densité *a posteriori* des paramètres du modèle. En général, on a recours à des méthodes de type Monte-Carlo par chaînes de Markov (MCMC) pour effectuer ces calculs.

7.1.3 Application au cas d'un champ aléatoire gaussien

Nous faisons l'hypothèse que $Z(\cdot)$ est un champ aléatoire gaussien avec pour moyenne $E\{Z(\mathbf{x})\} = {}^t\boldsymbol{\beta}f(\mathbf{x})$ où $\boldsymbol{\beta}$ est le vecteur des p coefficients inconnus de la dérive et f le vecteur des fonctions de base. Sa covariance est $\alpha^{-1}K_\nu(\mathbf{x}, \mathbf{x}')$, α étant la précision (inverse du palier) et ν la portée et les paramètres de forme. Notons $\boldsymbol{\theta} = {}^t(\boldsymbol{\beta}, \alpha, \nu)$ le vecteur des paramètres du modèle.

La fonction de vraisemblance de $\boldsymbol{\theta}$ connaissant les observations $\mathbf{Z}$ est donnée par :

$$L(\boldsymbol{\theta}|\mathbf{Z}) = (2\pi)^{-\frac{n}{2}} \alpha^{\frac{n}{2}} |K_\nu|^{-\frac{1}{2}} \exp\left(-\frac{\alpha}{2} {}^t(\mathbf{Z} - F\boldsymbol{\beta})K_\nu^{-1}(\mathbf{Z} - F\boldsymbol{\beta})\right) \quad (7.4)$$

où $|K_\nu|$ représente le déterminant¹ de la matrice de corrélation K_ν et F la matrice contenant les fonctions de base de la dérive calculées aux points observés. Le modèle de la matrice de corrélation peut être choisi dans une famille plus large que les modèles habituellement utilisés en géostatistique, ceci afin de couvrir une large gamme de comportements à l'origine (par exemple on prendra la classe des fonctions de corrélation de Matérn² (Handcock et Stein, 1993)). Nous avons néanmoins décidé de ne retenir que le schéma cubique qui a modélisé jusqu'à présent la structure spatiale du potentiel parce que le comportement à l'origine est ici imposé. De même, nous avons choisi de déterminer les incertitudes pour les trois paramètres. Ribeiro Jr et Diggle (1999) calculent dans le cadre gaussien les densités *a posteriori* pour différents cas de figures envisagés (β inconnu, α inconnu, β et α inconnus, β , α et ν inconnus). Nous ne développerons les calculs à la section 7.1.5 que dans le cas où tous les paramètres sont inconnus. Notons enfin que la densité *a posteriori* (ou conjuguée) peut s'écrire comme le produit des densités conditionnelles des paramètres :

$$\pi(\beta, \alpha, \nu | \mathbf{Z}) = \pi(\beta | \mathbf{Z}, \alpha, \nu) \pi(\alpha | \mathbf{Z}, \nu) \pi(\nu | \mathbf{Z}) \quad (7.5)$$

Mais commençons dans un premier temps par définir les lois *a priori*.

7.1.4 Les lois *a priori*

L'utilisation des densités de probabilité décrivant l'état des connaissances sur les paramètres avant la prise en compte des données observées est une des caractéristiques majeures de l'analyse bayésienne. Le choix de ces densités *a priori* est le problème le plus délicat de cette approche. En effet, il est très souvent difficile de déterminer exactement cette densité du fait du peu d'information précise disponible.

On suppose que les paramètres du modèle de covariance sont *a priori* indépendants. Il existe deux cas extrêmes. Si l'information disponible est pauvre, une densité non-informative peut être choisie. Par contre, lorsque le paramètre est bien connu, on prend une densité centrée sur sa valeur et très peu étalée. Seules les deux types de densité *a priori* les plus utilisées, à savoir conjuguées et non informatives, sont présentées.

Densités *a priori* conjuguées

Il s'agit de choisir une famille de densités *a priori* telle que celles-ci aient la même forme fonctionnelle que la fonction de vraisemblance. Ce choix obéit plus à des considérations techniques, le calcul de la densité *a posteriori* étant grandement simplifié, qu'à une réelle adéquation avec la connaissance *a priori* disponible. Ce type de loi *a priori*, en l'occurrence la famille Normale-Gamma, a été choisi pour β et α . Cette famille suppose une dépendance entre β et α , et une relation entre les paramètres du modèle de la forme :

$$(\beta | \alpha, \nu) \sim \mathcal{N}(\mathbf{b}, \alpha^{-1} S) \quad \text{et} \quad (\alpha | \nu) \sim \Gamma(a_1, \frac{1}{a_2})$$

¹Il est souvent nécessaire de passer par le logarithme lors du calcul du déterminant vue la taille de celui-ci.

²Les schémas gaussien et exponentiel appartiennent à cette classe.

et on a aussi :

$$(\boldsymbol{\beta}, \alpha | \nu) \sim \mathcal{N}(\mathbf{b}, \alpha^{-1} S) \Gamma(a_1, \frac{1}{a_2}) \quad (7.6)$$

Ainsi, $\mathbf{b}$ et $\alpha^{-1} S$ sont respectivement la moyenne et la variance *a priori* de $(\boldsymbol{\beta} | \alpha, \nu)$. De même, a_1 et $\frac{1}{a_2}$ sont respectivement l'indice et le paramètre de $(\alpha | \nu)$. La loi Gamma est bien adaptée ici car seules des valeurs positives de palier sont acceptées. De plus, la forme de la densité est flexible et permet de s'adapter à un grand nombre de situations.

Densités *a priori* non informatives

Elles sont utilisées pour décrire l'état d'ignorance de l'information *a priori* à notre disposition avant l'obtention des données. Ce type de densité *a priori* ne "favorisent" pas certaines valeurs de paramètres par rapport à d'autres. La plus simple et la plus utilisée est la densité uniforme où toutes les valeurs des paramètres sont équiprobables sur le domaine de définition. Une densité uniforme sera employée pour la densité *a priori* de la portée en se restreignant à un intervalle raisonnable.

Le problème de la densité uniforme est qu'elle n'est pas invariante par transformation. C'est pourquoi d'autres types de densité ont été introduites comme la densité *a priori* de Jeffreys : $\pi(\theta) \propto \frac{1}{\theta}$. Il faut prendre garde au fait que pour notre problème de paramètres du modèle de covariance, l'utilisation d'une densité telle que $\pi(\boldsymbol{\beta}, \alpha, \nu) \propto \frac{1}{\alpha}$ conduit à des distributions *a posteriori* impropres³ pour $(\boldsymbol{\beta}, \alpha, \nu)$. Certains auteurs (Berger *et al.* (2000), etc.) introduisent le terme de *reference prior* pour désigner des densités *a priori* non informatives qui conduisent à des densités *a posteriori* qui ne sont pas impropres. Une distribution *a posteriori* propre peut néanmoins être obtenue même si la densité *a priori* est impropre.

Enfin, Mosegaard et Tarantola (2002) ne recommandent pas d'utiliser le terme "non informatif" pour désigner les lois *a priori*, car elles sont toutes aussi informatives que d'autres. Dans le cadre de paramètres positifs et mutuellement réciproques (période/fréquence, résistivité/conductivité...), rebaptisés paramètres de Jeffreys, ils emploient plutôt le terme de densités de probabilité homogènes.

7.1.5 Lois *a posteriori* des paramètres de la covariance

Calculons à présent les densités conditionnelles pour les paramètres du modèle (cf. équation 7.5), à savoir $\pi(\boldsymbol{\beta} | \mathbf{Z}, \alpha, \nu)$, $\pi(\alpha | \mathbf{Z}, \nu)$ et $\pi(\nu | \mathbf{Z})$. Le calcul analytique est possible pour des densités *a priori* conjuguées et non-informatives.

Rappelons tout d'abord l'expression de la densité *a posteriori* :

$$\begin{aligned} \pi(\boldsymbol{\beta}, \alpha, \nu | \mathbf{Z}) &\propto \pi(\boldsymbol{\beta}, \alpha, \nu) f(\mathbf{Z} | \boldsymbol{\beta}, \alpha, \nu) \\ &= \pi(\boldsymbol{\beta}, \alpha, \nu) \alpha^{\frac{n}{2}} |K_\nu|^{-\frac{1}{2}} \exp \left(-\frac{\alpha}{2} (\mathbf{Z} - F\boldsymbol{\beta}) K_\nu^{-1} (\mathbf{Z} - F\boldsymbol{\beta}) \right) \end{aligned}$$

³L'intégrale de la densité de probabilité n'est pas égale à 1.

Introduisons ensuite la notation suivante :

$${}^t(\mathbf{Z} - F\boldsymbol{\beta})K_\nu^{-1}(\mathbf{Z} - F\boldsymbol{\beta}) = \|\mathbf{Z} - F\boldsymbol{\beta}\|_{K_\nu^{-1}}^2$$

a) Cherchons dans un premier temps la loi de $\pi(\boldsymbol{\beta}|\mathbf{Z}, \alpha, \nu)$

On a : $\pi(\boldsymbol{\beta}|\alpha, \nu) \propto \exp\left(-\frac{\alpha}{2}{}^t(\boldsymbol{\beta} - \mathbf{b})S^{-1}(\boldsymbol{\beta} - \mathbf{b})\right)$

Avec la notation introduite précédemment on peut écrire :

$$\pi(\boldsymbol{\beta}|\alpha, \nu) \propto \exp\left(-\frac{\alpha}{2}\|\boldsymbol{\beta} - \mathbf{b}\|_{S^{-1}}^2\right)$$

La formule de Bayes donne :

$$\begin{aligned} \pi(\boldsymbol{\beta}|\mathbf{Z}, \alpha, \nu) &\propto \pi(\boldsymbol{\beta}|\alpha, \nu)f(\mathbf{Z}|\boldsymbol{\beta}, \alpha, \nu) \\ &\propto \text{const} \exp\left(-\frac{\alpha}{2}\left(\|\boldsymbol{\beta} - \mathbf{b}\|_{S^{-1}}^2 + \|\mathbf{Z} - F\boldsymbol{\beta}\|_{K_\nu^{-1}}^2\right)\right) \end{aligned} \quad (7.7)$$

On peut montrer que :

$$\|\mathbf{Z} - F\boldsymbol{\beta}\|_{K_\nu^{-1}}^2 + \|\boldsymbol{\beta} - \mathbf{b}\|_{S^{-1}}^2 = S(\nu) + \|\boldsymbol{\beta} - \mathbf{b}_c\|_{S_c^{-1}}^2 \quad (7.8)$$

avec

$$\mathbf{b}_c = S_c({}^tFK_\nu^{-1}\mathbf{Z} + S^{-1}\mathbf{b}) \quad , \quad S_c = (S^{-1} + {}^tFK_\nu^{-1}F)^{-1} \quad (7.9)$$

et

$$S(\nu) = {}^tbS^{-1}b + {}^t\mathbf{Z}K_\nu^{-1}\mathbf{Z} - {}^tb_cS_c^{-1}b_c$$

On a donc :

$$(\boldsymbol{\beta}|\mathbf{Z}, \alpha, \nu) \sim \mathcal{N}_p(\mathbf{b}_c, \alpha^{-1}S_c) \quad (7.10)$$

b) Poursuivons avec la loi de $\pi(\alpha|\mathbf{Z}, \nu)$

On a :

$$\begin{aligned} \pi(\boldsymbol{\beta}, \alpha|\mathbf{Z}, \nu) &\propto \pi(\boldsymbol{\beta}, \alpha|\nu)f(\mathbf{Z}|\boldsymbol{\beta}, \alpha, \nu) \\ &\propto \alpha^{\frac{n}{2}}\alpha^{a_1 + \frac{n}{2}} \exp\left(-\frac{\alpha}{2}\left(\|\boldsymbol{\beta} - \mathbf{b}\|_{S^{-1}}^2 + \|\mathbf{Z} - F\boldsymbol{\beta}\|_{K_\nu^{-1}}^2\right)\right) \exp(-a_2\alpha) \end{aligned}$$

En utilisant 7.8 et en intégrant sur $\boldsymbol{\beta}$, on obtient que :

$$(\alpha|\mathbf{Z}, \nu) \sim \Gamma(a_{11}, \frac{1}{a_{22}}) \quad (7.11)$$

avec comme paramètres :

$$a_{11} = a_1 + \frac{n}{2} \quad \text{et} \quad a_{22} = a_2 + \frac{S(\nu)}{2} \quad (7.12)$$

On voit que le choix de a_1 a peu d'importance lorsque la taille n de l'échantillon est grande.

c) Terminons avec la loi de $\pi(\nu|\mathbf{Z})$

Encore une fois l'application du théorème de Bayes entraîne :

$$\pi(\nu|\mathbf{Z}) = \frac{\pi(\boldsymbol{\beta}, \alpha, \nu|\mathbf{Z})}{\pi(\boldsymbol{\beta}, \alpha|\mathbf{Z}, \nu)}$$

Ce qui donne pour la densité *a posteriori* de ν :

$$\pi(\nu|\mathbf{Z}) \propto \pi(\nu) |K_\nu|^{-\frac{1}{2}} |S_c|^{\frac{1}{2}} \left(a_2 + \frac{S(\nu)}{2} \right)^{-(a_1 + \frac{n}{2})} \quad (7.13)$$

7.2 Application au cas du Limousin

Maintenant que nous disposons des expressions des densités *a posteriori*, nous allons calculer des statistiques pour les différents paramètres du modèle de covariance. Le gradient vertical et une des composantes horizontales du gradient pour le jeu de données du Limousin vont illustrer ces calculs. Il est possible ensuite d'incorporer ces résultats pour l'interpolation du champ de potentiel dans le cadre du krigeage bayésien ou de simulations conditionnelles.

7.2.1 Les paramètres et l'algorithme utilisés

On considère une dérive constante, le coefficient $\boldsymbol{\beta} = \beta$ de cette dérive suivant une loi normale centrée réduite. La précision α suit une loi gamma d'indice 1 et de paramètre 10^{-5} . Le nombre de données étant élevé (1485), la valeur de l'indice intervient peu dans les calculs. L'allure de cette distribution ressemble à une loi uniforme. On introduit aussi comme paramètre du modèle l'effet de pépite ζ . La matrice de covariance s'écrit à présent : $\alpha^{-1} K_\nu(\mathbf{x}, \mathbf{x}') + \zeta I$, I étant la matrice identité. On introduit l'effet de pépite relatif $\tau = \alpha\zeta$ dans les calculs.

L'algorithme utilisé ici est donné par (Ribeiro Jr et Diggle, 1999) :

Étape 1 : On discrétise la densité de $(\nu, \tau|\mathbf{Z})$, autrement dit on choisit des valeurs pour (ν, τ) dans des intervalles raisonnables vues les valeurs obtenues lors des ajustements et une densité uniforme discrète pour (ν, τ) sur ces intervalles. Les valeurs de ν sont comprises entre $8000m$ et $45000m$ avec un pas de $1000m$ et celles de τ entre 0.1 et 1.1 au pas de 0.1 dans le cas du gradient vertical. Des tests ont montré que la probabilité d'avoir $\tau = 0$ avec ce jeu de données est nulle.

Étape 2 : On calcule les probabilités *a posteriori* sur les intervalles considérés en utilisant 7.13 (on remplace ν par (ν, τ)). Les résultats définissent une approximation de la densité *a posteriori* discrète $\tilde{\pi}(\nu, \tau|\mathbf{Z})$.

Étape 3 : On simule un couple de valeurs (ν, τ) à partir de $\tilde{\pi}(\nu, \tau|\mathbf{Z})$.

Étape 4 : Le couple de valeurs simulées est introduit dans $\pi(\boldsymbol{\beta}, \alpha|\mathbf{Z}, \nu, \tau)$ en employant la relation 7.6 adaptée pour (ν, τ) .

Étape 5 : Les étapes (3) et (4) sont itérées autant de fois que d'échantillons (quadruplets $(\boldsymbol{\beta}, \alpha, \nu, \tau)$) désirés (ici 1000) à partir des densités *a posteriori* des paramètres.

7.2.2 Composante verticale du gradient

Les résultats des simulations des paramètres sont fournis dans le Tableau 7.1 et à la Figure 7.1. Cette Figure présente les histogrammes et une estimation de la densité *a posteriori* des différents paramètres estimés (α , β , ν , τ) ainsi que du palier ($\frac{1}{\alpha}$) et de l'effet de pépité ($\frac{\tau}{\alpha}$).

On remarque que la moyenne expérimentale des données (0.75) est assez bien retrouvée

	β	α	ν	τ	palier	pepité
minimum	0.563	18.54	11000	0.2	0.0115	0.0108
maximum	0.887	89.69	45000	1.5	0.0539	0.0188
moyenne	0.688	50.46	18979	0.779	0.0209	0.0153
médiane	0.692	50.07	18000	0.8	0.0200	0.015
mode	0.696	49.70	15000	0.75	0.0175	0.0155
variance	0.001	127.88	28821380	0.041	3e-5	1e-6

TAB. 7.1 – Limousin : Statistiques *a posteriori* du modèle de variogramme pour les données du gradient vertical.

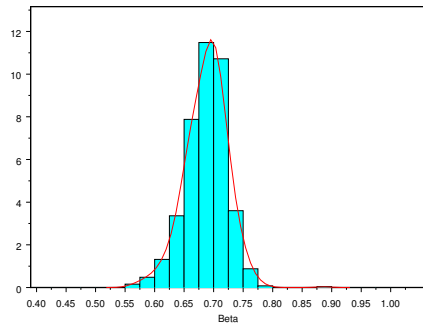
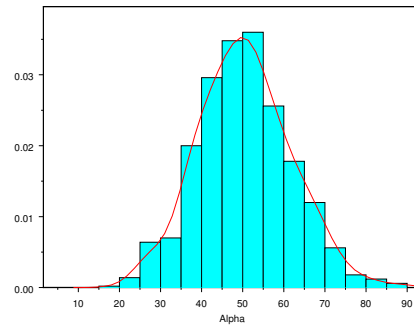
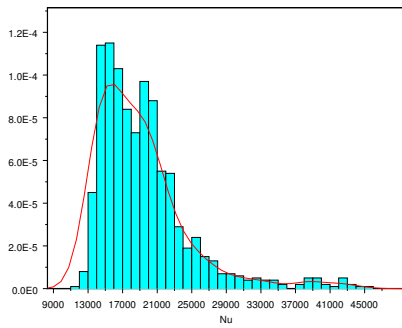
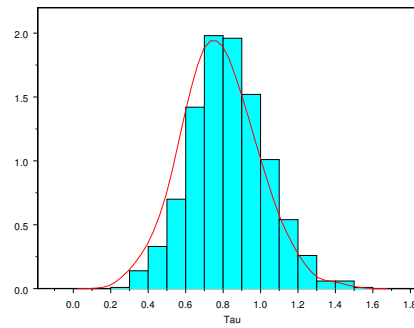
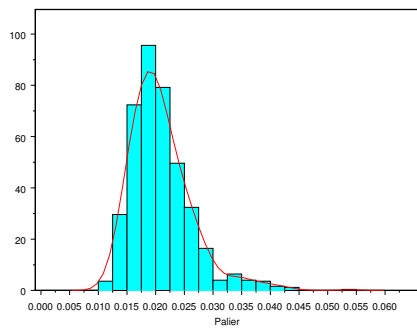
quoique légèrement sous-estimée (0.7). Le palier estimé par l'approche bayésienne (mode de la densité *a posteriori*⁴) est exactement le même que celui obtenu par l'ajustement manuel, à savoir 0.0175. L'effet de pépité est lui aussi très bien estimé avec une valeur de 0.0155 contre 0.0165 pour la valeur trouvée précédemment. Par contre la valeur de la portée (ν) est sous-estimée (si on admet que la portée empirique est exacte) avec des valeurs comprises entre 15000m et 19000m suivant les estimateurs choisis contre une valeur de 25000m déterminée par l'approche classique.

La plupart des densités ont une allure gaussienne, les densités du palier et surtout de la portée présentent une queue pour les valeurs fortes. Les plages où les paramètres prennent leurs valeurs sont relativement importantes. Les variances donnent une idée de l'incertitude relative à l'estimation des paramètres de la covariance.

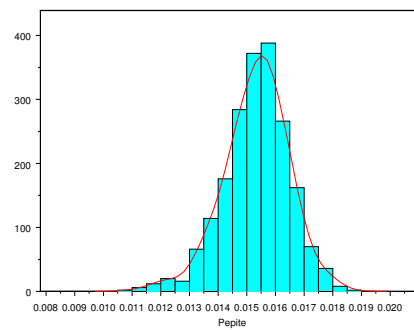
La Figure 7.2 montre le variogramme expérimental et le modèle obtenu avec les paramètres simulés, l'estimateur retenu étant le mode. Avec d'autres estimateurs (moyenne ou médiane), le résultat serait légèrement différent à cause des dissymétries dans les densités de la portée, du palier et de l'effet de pépité. Le modèle ainsi déterminé est très proche du modèle choisi en 5.2, on remarque cependant la plus faible portée du modèle obtenu par l'approche bayésienne.

En conclusion, le modèle du variogramme du gradient vertical des données du Limousin obtenu par l'approche bayésienne est très proche de celui choisi manuellement (ou calculé par ajustement automatique). Cette manière d'aborder l'inférence du modèle de covariance est beaucoup plus lourde en matière de calculs (simulations), mais a l'avantage de fournir des intervalles de valeurs acceptables ainsi que des incertitudes sur la détermination de ces paramètres.

⁴Maximum général.

(a) β (b) α (c) ν (d) τ 

(e) palier



(f) pépîte

FIG. 7.1 – Limousin : Histogrammes de la distribution *a posteriori* des paramètres pour la composante verticale.

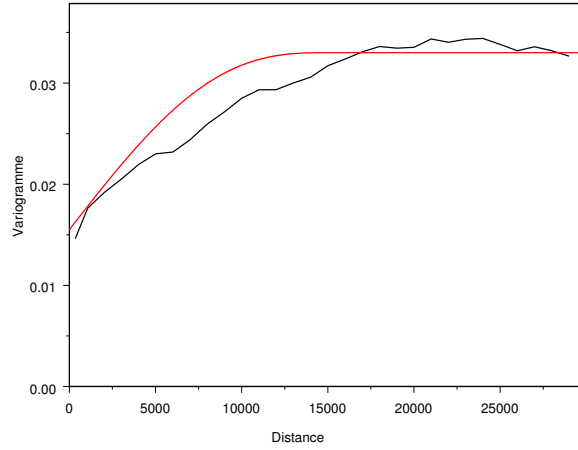


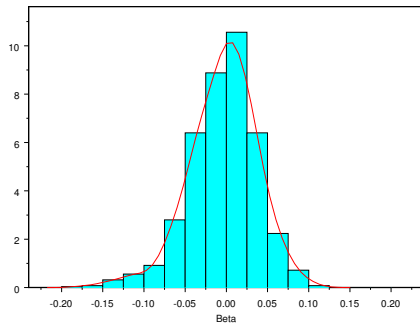
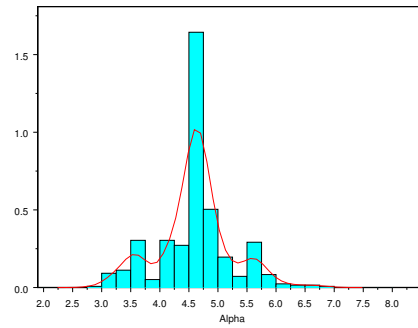
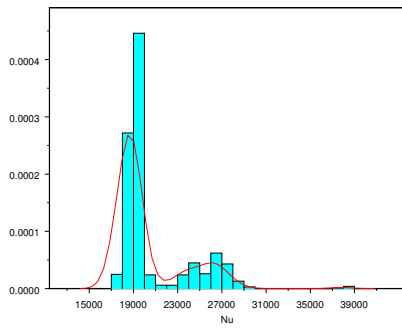
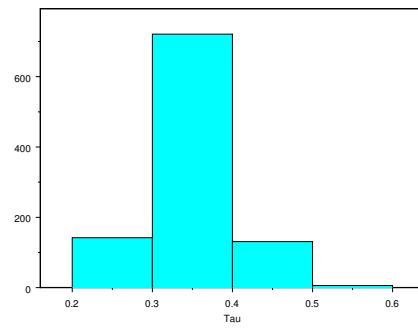
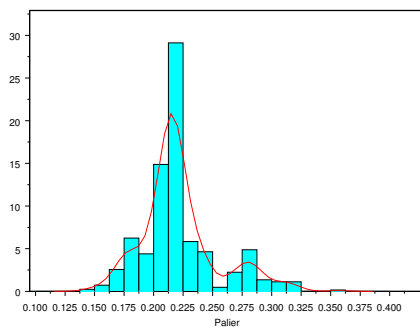
FIG. 7.2 – Limousin : Variogramme expérimental (en noir) et variogramme modélisé (en rouge) en utilisant les modes *a posteriori* des paramètres de la composante verticale du gradient.

Illustrons maintenant cette approche par un autre exemple : la composante horizontale G_x du gradient pour ce même jeu de données du Limousin.

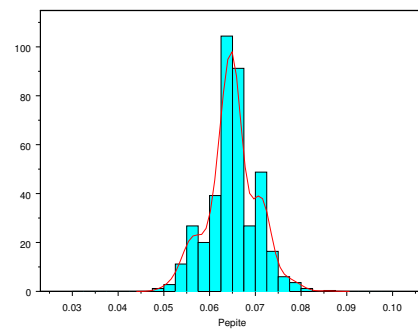
7.2.3 Composante horizontale du gradient

Nous allons effectuer maintenant les simulations sur les composantes horizontales du gradient et illustrer cela avec la composante G_x . Les calculs sont les mêmes que précédemment, les bornes de l'intervalle pour la portée sont désormais 8000m et 38000m. La Figure 7.3 et le Tableau 7.2 montrent les résultats des simulations des paramètres du modèle de variogramme pour la composante G_x . Ce modèle ne tient pas compte des résultats obtenus pour le gradient vertical, c'est-à-dire que le palier qui sera trouvé sera un palier total. De même, la portée déduite précédemment pour le gradient vertical n'a pas été introduite dans la covariance recherchée pour G_x .

La moyenne est à nouveau bien estimée (proche de 0) quoique très légèrement inférieure à la moyenne expérimentale des données. Le palier total (toutefois sans l'effet de pépité) estimé par l'approche bayésienne (0.212) est comme on le prévoyait supérieur à celui du gradient vertical, par contre il est assez différent de celui obtenu par l'ajustement manuel, à savoir 0.175. On constate aussi de telles surestimations de la variance dans (Diggle *et al.*, 2003) et (Ribeiro Jr et Diggle, 1999). On notera surtout que la variance de ce paramètre est beaucoup plus forte que pour G_z . L'effet de pépité est bien estimé avec une valeur de 0.065 contre 0.075 pour la valeur trouvée précédemment. Contrairement à ce qui avait été constaté pour G_z , la valeur de la portée (ν) est surestimée avec des valeurs autour de 19000m - 20000m suivant les estimateurs choisis contre une valeur de 17000m déterminée par l'approche classique. L'écart entre les valeurs de portée obtenues par approche bayésienne et traditionnelle est cependant plus faible que pour G_z . On note aussi que la portée

(a) β (b) α (c) ν (d) τ 

(e) palier



(f) pépîte

FIG. 7.3 – Limousin : Histogrammes de la distribution *a posteriori* des paramètres pour la composante horizontale G_x .

	β	α	ν	τ	palier	pepite
minimum	-0.178	2.803	17000	0.2	0.144	0.049
maximum	0.111	6.922	38000	0.5	0.357	0.086
moyenne	-0.004	4.601	20228	0.3	0.222	0.065
médiane	-2e-5	4.618	19000	0.3	0.217	0.065
mode	0	4.5	19000	0.3	0.212	0.062
variance	0.002	0.390	10332348	0.003	0.001	3e-5

TAB. 7.2 – Limousin : Statistiques *a posteriori* du modèle de variogramme pour les données de la composante horizontale G_x du gradient.

est le seul paramètre qui a une variance plus faible pour G_x que pour G_z . Seule la densité du paramètre β a une allure bien gaussienne, les autres densités présentent des maxima relatifs. La densité de ν possède une queue importante pour les valeurs fortes. La Figure 7.4 montre le variogramme expérimental et le modèle obtenu avec les paramètres

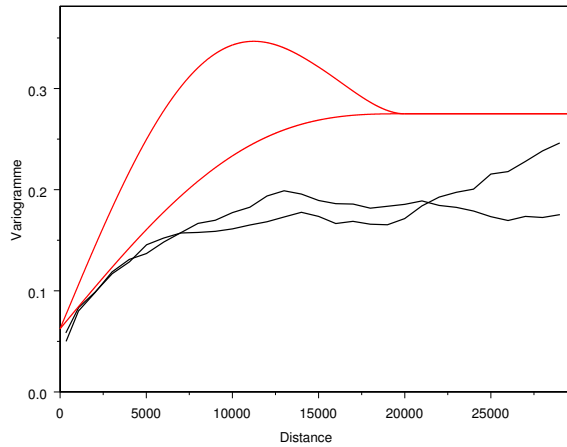


FIG. 7.4 – Limousin : Variogramme expérimental (en noir) et variogramme modélisé (en rouge) en utilisant les modes *a posteriori* des paramètres de la composante horizontale G_x .

simulés. On remarque donc que comme pour les ajustements manuel et automatique où cette covariance s'ajustait plus difficilement que celle du gradient vertical, les incertitudes déterminées par l'approche bayésienne sont plus grandes. On constate que le palier est plutôt mal retrouvé par rapport aux autres paramètres.

N.B. : Comme le modèle porte sur le potentiel et que les composantes du gradient s'en déduisent, un seul jeu de paramètres pourrait être considéré et l'estimation devrait alors intégrer toutes les composantes du gradient de potentiel. On ajusterait donc le variogramme du potentiel sur ceux de G_x , G_y et G_z (ainsi que les 3 croisés) et on pourrait prendre en compte aussi le fait que le module du gradient est inconnu. D'une manière générale,

l'approche bayésienne appliquée à notre problème de gradient mériterait une étude plus approfondie.

7.2.4 Conclusions

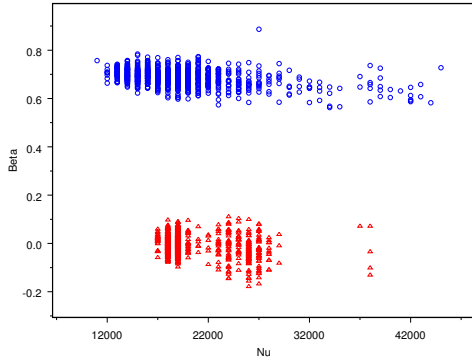
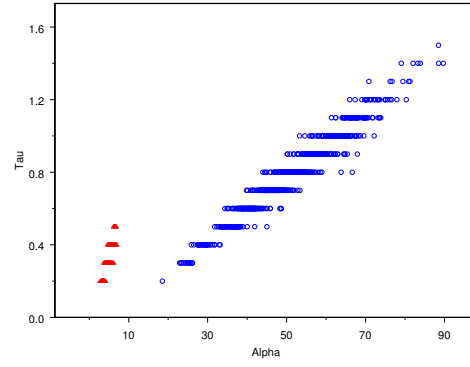
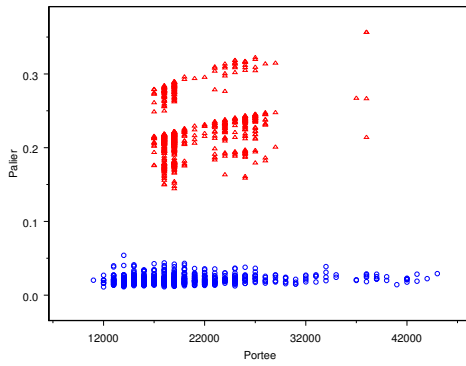
Les paramètres dont l'estimation par l'approche bayésienne est proche de celle obtenue plus classiquement sont la moyenne et l'effet de pépité. D'une manière générale, la portée est le paramètre qui présente le plus de différences avec l'approche traditionnelle. Elle est soit sous-estimée, soit surestimée, mais dans l'ensemble la valeur choisie lors de l'ajustement manuel est dans l'intervalle des valeurs obtenues par la technique bayésienne. Quant au palier, les deux estimations de sa valeur pour le gradient vertical sont similaires, mais il n'en est pas de même pour la composante G_x où l'écart est significatif.

La latitude dans le choix des paramètres du modèle est importante et peut entraîner localement des différences notables dans l'interpolation. Ainsi, en prenant comme modèle de covariance du potentiel un modèle gigogne composé de deux schémas cubiques avec comme paramètres les modes *a posteriori* déterminés dans ce chapitre, on observe dans la Coupe 1 (paragraphe 6.1.1) que l'interface "uig" est plus profonde de 250m qu'avec la covariance ajustée à vue.

La Figure 7.5 montre des nuages de corrélation entre des paramètres du modèle pour G_z et G_x . On retrouve bien les résultats attendus, à savoir qu'il semble y avoir indépendance entre β et les paramètres de covariance (Fig. 7.5(a)), qu'il existe une relation linéaire entre τ et α (Fig. 7.5(b)) et une dépendance entre palier et portée (Fig. 7.5(c)). Néanmoins, cette dernière n'est pas aussi forte que ce que l'on pourrait attendre. En effet, la portée est sensée augmenter avec la variance comme on le constate classiquement sur un variogramme. Cette mauvaise corrélation est due au paramètre τ et à sa discrétisation qui créent des maxima relatifs. Par contre, si on travaille à τ fixé, on constate que la dépendance apparaît plus nettement, comme par exemple avec $\tau=0.3$ pour G_x (Fig. 7.5(d)). En particulier, on voit se dessiner une limite supérieure au-delà de laquelle il n'existe plus de valeur de palier à portée fixée.

En utilisant une approche bayésienne et sans avoir une idée précise de la distribution *a priori* des paramètres du modèle de covariance, il est possible d'inférer ces paramètres et d'avoir une information sur leur incertitude.

L'étape suivante serait d'exploiter ces résultats pour l'estimation du potentiel avec le krigage bayésien ou à l'aide de simulations conditionnelles. Cela ne fait pas l'objet de cette thèse, mais pourrait l'être lors d'études ultérieures.

(a) $\beta = f(\nu)$ (b) $\tau = f(\alpha)$ 

(c) portée=f(palier)

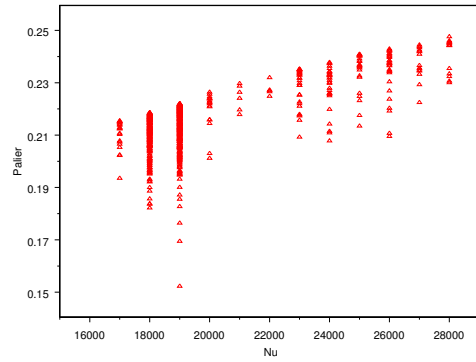
(d) portée=f(palier) à τ fixé

FIG. 7.5 – Limousin : Nuages de corrélation entre différents paramètres en utilisant les échantillons de la densité *a posteriori*. (Cercle bleu : G_z - Triangle rouge : G_x .)

Chapitre 8

Illustration du problème des vecteurs normés par l'exemple du Limousin

8.1 Rappel de l'hypothèse

8.1.1 Utilisation des gradients

La méthode des potentiels utilise les données structurales relevées sur le terrain et amène à considérer les vecteurs normaux aux surfaces isopotentielles comme des gradients du potentiel (cf. 4.4). Ils sont projetés sur les axes géographiques et leur module est fixé à 1 par construction. Or, dans la réalité le module de ce vecteur gradient n'est pas forcément constant. Il faut donc s'assurer, au niveau de la variographie, que l'amalgame entre vecteurs normés et non-normés, c'est-à-dire le passage entre des vecteurs unitaires en espérance à des vecteurs tous unitaires, n'est pas trop invalidant..

8.1.2 Le module du gradient

Le module du gradient nous renseigne localement sur les zones où les isopotentielles tendent à se resserrer ou au contraire à s'élargir. Autrement dit il apporte une information sur l'intensité relative de la déformation liée à la foliation. Ainsi, plus une roche est déformée, plus les isopotentielles sont resserrées. Pour des terrains sub-horizontaux, on pourrait interpréter l'information apportée par le module du gradient comme étant reliée à la sédimentation. Des isopotentielles lâches seraient signes d'un apport de matériaux important ou de compaction faible. Cette interprétation des isopotentielles repose sur l'analogie entre isopotentielles et isochrones. Il est donc possible de raisonner ainsi, mais il faut pouvoir le mesurer sur le terrain ou alors l'estimer. On utilise les simulations pour étudier l'impact de l'utilisation des vecteurs unitaires au lieu des vrais gradients sur l'analyse variographique. On pourrait également le faire pour le cokrigage¹.

¹Sur une simulation, comparer le cokrigage des données de potentiel avec les gradients réels d'une part, avec les gradients simulés d'autre part.

8.2 Simulations des vecteurs normés et non-normés : l'exemple du Limousin

8.2.1 Simulations non conditionnelles du potentiel

L'opération préalable à la simulation est la création d'une grille sensiblement de la taille du champ du Limousin, à savoir $66km$ sur $61km$. Un pas de 100 m a été choisi afin d'obtenir des résultats suffisamment précis. Les données du Limousin sont cartographiques, mais il va falloir aussi calculer une composante verticale du gradient. On décide donc de superposer une grille au-dessus et une autre en-dessous séparées chacune de $10m$ de la grille considérée. On dispose donc d'une grille tridimensionnelle de $1\,207\,800$ nœuds. Des simulations non-conditionnelles (cf. Annexe D) par bandes tournantes sont entreprises avec le modèle défini pour le Limousin en 5.2.4. Les paramètres du modèle sont rappelés au Tableau 8.1 (gauche). Le potentiel est à présent connu en tout point de la grille. Les Figures 8.1 et 8.2 montrent les cartes du potentiel² et les histogrammes obtenus pour deux simulations.

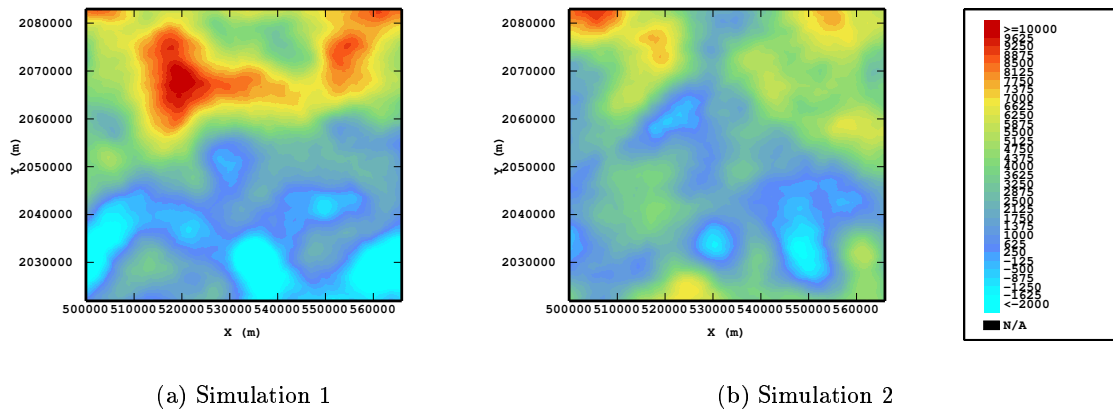


FIG. 8.1 – Cartes du potentiel pour deux simulations.

Afin d'obtenir une configuration spatiale proche de celle du cas réel, les localisations géographiques des données d'orientation sont migrées aux points de la grille les plus proches. Certaines données étant trop proches entre elles, elles ne seront pas prises en compte. Il en résulte un total de 1477 positions où les gradients vont être calculés.

²Elles correspondent à l'altitude de localisation des données.

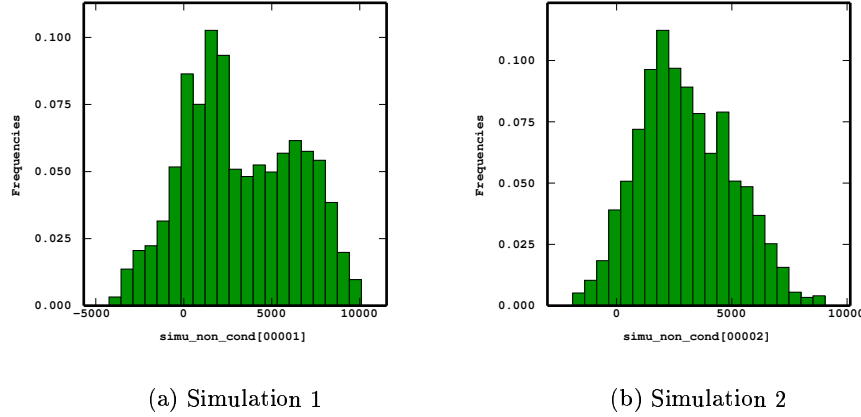


FIG. 8.2 – Histogrammes du potentiel pour deux simulations.

8.2.2 Calcul des gradients

Les composantes continues des gradients sont calculées aux points retenus par différences finies³. Il faudrait, cependant s'assurer que lors de la différentiation numérique le comportement HF de la variable est bien simulé. Les statistiques élémentaires des gradients obtenus pour deux simulations sont résumées aux Tableaux 8.2 et 8.3 (mention "continu"). Pour comparaison, les statistiques des données réelles sont rappelées au Tableau 8.1 (droite). Afin de ressembler au mieux aux données réelles, on va ajouter une dérive et un effet de pépité aux données de gradient non normé. En effet, les données simulées sont d'espérance nulle et de variance égale au palier du gradient.

La dérive du potentiel est supposée linéaire. Sur le gradient elle se traduit par une valeur moyenne d'espérance non nulle. Cette valeur est prise égale à la moyenne expérimentale des données réelles. Pour l'effet de pépité nous utilisons une erreur de loi uniforme⁴ (espérance nulle et variance égale à l'effet de pépité du gradient). Les statistiques des composantes du gradient "complété" sont aussi présentées aux Tableaux 8.2 et 8.3.

Enfin les données sont normées à 1 et les statistiques à nouveau compilées dans ces deux mêmes tableaux. Les données normées sont sensées reproduire le comportement des véritables données et vont être comparées aux données non-normées complétées. On remarque que la variance des composantes continues du gradient n'est pas égale à la variance théorique, car on travaille sur un champ fini. L'ajout de la dérive permet de corriger le biais sur la moyenne, en particulier pour le gradient vertical et l'effet de pépité de corriger la variance.

³Nous appelons continues les composantes du gradient non-normé obtenues directement après la simulation, complétées les composantes non-normées après ajout de la dérive et de l'effet de pépité et normées celles obtenues après normation.

⁴Une erreur de loi normale fait tout aussi bien l'affaire.

	Portée	Palier		Minimum	Maximum	Moyenne	Variance
γ_3	25000	8.10^5	G_x	-0.8913	0.8659	0.0108	0.1876
γ_2	17000	17.10^5	G_y	-0.8659	0.8932	-0.1104	0.2076
γ_1	55000	108.10^5	G_z	0.3168	1	0.7491	0.0314

TAB. 8.1 – Rappel des paramètres du modèle de covariance (gauche) et des statistiques élémentaires des données de gradient (droite) pour le Limousin.

	Minimum	Maximum	Moyenne	Variance
G_x continu	-0.9226	0.8128	0.0012	0.0964
G_y continu	-0.9266	1.1382	0.1693	0.1676
G_z continu	-0.2831	0.4464	0.0062	0.0125
G_x complété	-1.2313	1.1634	0.0151	0.1622
G_y complété	-1.3725	1.3657	0.0533	0.2446
G_z complété	0.3092	1.3242	0.7608	0.0289
G_x normé	-0.8843	0.8020	0.0110	0.1464
G_y normé	-0.9439	0.8934	0.0548	0.2056
G_z normé	0.3075	0.9996	0.7902	0.0208

TAB. 8.2 – Statistiques élémentaires des données de gradient simulées (simulation 1).

	Minimum	Maximum	Moyenne	Variance
G_x continu	-1.1044	0.9120	0.0247	0.1036
G_y continu	-1.0276	1.0266	0.0591	0.1099
G_z continu	-0.4144	0.4623	-0.0336	0.0199
G_x complété	-1.4281	1.3215	0.0386	0.1834
G_y complété	-1.3199	1.2150	-0.0569	0.1821
G_z complété	0.1958	1.3188	0.7210	0.0368
G_x normé	-0.8951	0.9102	0.0406	0.1814
G_y normé	-0.9424	0.9028	-0.0475	0.1805
G_z normé	0.2949	0.9996	0.7826	0.0219

TAB. 8.3 – Statistiques élémentaires des données de gradient simulées (simulation 2).

8.2.3 Les variogrammes

Les variogrammes expérimentaux des composantes du gradient complété et normé sont calculés comme lors de l'étude variographique des données réelles du Limousin. Les résultats sont exposés aux Figures 8.3 et 8.4. Notons tout d'abord les fluctuations statistiques des simulations par rapport au modèle utilisé. Observons également l'effet de trou (bien visible pour la simulation 1) sur les variogrammes des composantes horizontales calculés dans la direction parallèle à la direction de dérivation, alors qu'il n'était pas très visible sur les variogrammes expérimentaux dans le cas réel. Ensuite, on constate que la normation des gradients a peu d'effets sur les variogrammes des composantes horizontales. Par contre, la portée et surtout le palier du variogramme du gradient vertical sont plus faibles. Le contraste est particulièrement bien visible sur la simulation 2.

8.2.4 Conséquences

Il faut bien garder à l'esprit que cet exemple illustre le cas d'étude du Limousin et ne peut donc prétendre à l'universalité.

On remarque toutefois ici que les composantes des vecteurs normés - et surtout de la composante verticale - ont une variance plus faible que les composantes des vecteurs non normés. En assimilant les vecteurs normés au gradient et en utilisant ces données pour définir le variogramme du potentiel, on sous-estime le palier de ce dernier. On peut montrer que si le variogramme est sous-estimé d'un facteur global constant, du fait que les données de base sont des accroissements nuls, cela n'a pas d'impact sur la forme des isopotentielles.

Dans le cas plus général, on peut imaginer d'autres solutions :

- mesurer le module du gradient (l'intensité de la déformation) et travailler sur de véritables données de gradient. Mais cela est rarement réalisable ;
- modéliser la covariance du potentiel à partir des données d'orientation en calculant les variogrammes théoriques de G_x , G_y , G_z comme étant relatifs à des gradients normés (soit analytiquement si c'est possible, soit par simulation) et réaliser un cokrigage ;
- une solution *ad hoc* est d'augmenter le palier de la structure γ_3 pour les covariances des données de potentiel, mais pas pour les données du gradient.

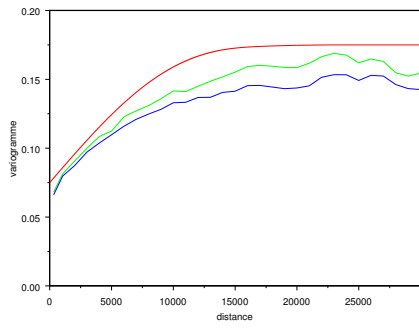
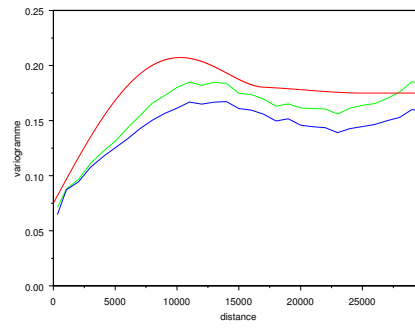
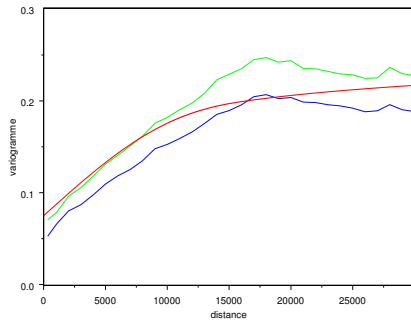
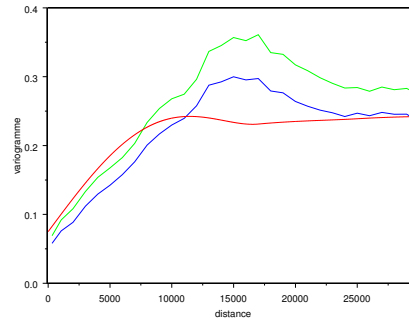
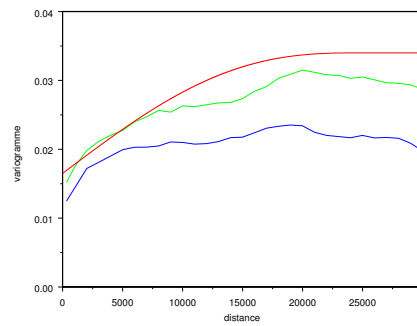
(a) $G_x \perp$ (b) $G_x \parallel$ (c) $G_y \perp$ (d) $G_y \parallel$ (e) G_z

FIG. 8.3 – Variogrammes de la simulation 1. Le modèle de variogramme est en rouge, les variogrammes des gradients simulés non normés (complétés) sont en vert et normés en bleu.

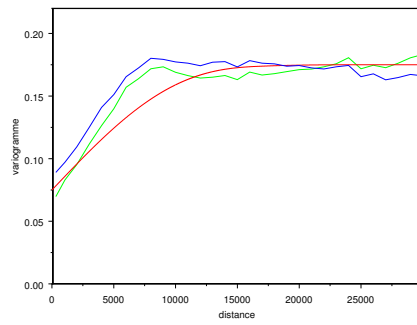
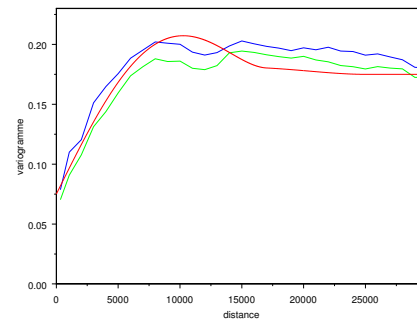
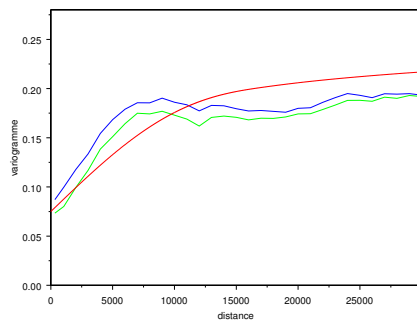
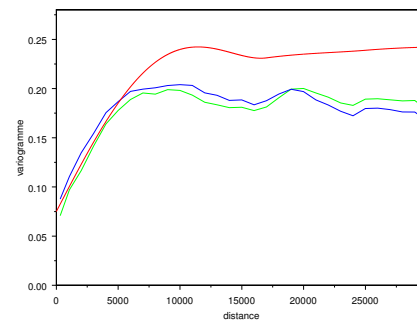
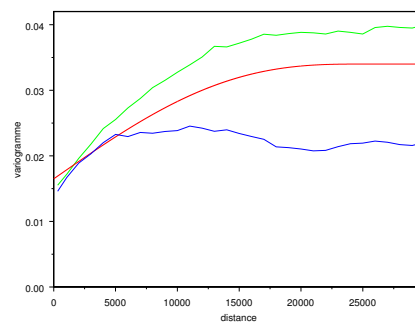
(a) $G_x \perp$ (b) $G_x \parallel$ (c) $G_y \perp$ (d) $G_y \parallel$ (e) G_z

FIG. 8.4 – Variogrammes de la simulation 2. Le modèle de variogramme est en rouge, les variogrammes des gradients simulés non normés (complétés) sont en vert et normés en bleu.

Chapitre 9

Le traitement des sondages incomplets - Aspects théoriques

Nous allons à présent traiter un autre type d'incertitude rencontré en modélisation géologique, à savoir les incertitudes sur les données. Plus précisément, ce chapitre traite des données d'inégalité présentes très fréquemment en géologie par l'intermédiaire des passes incomplètes dans les sondages.

9.1 Les données d'inégalité

9.1.1 Les types de données

Un critère légitime de classification de l'information est la notion de qualité (Alabert, 1987a). Il peut se traduire par la question : "quelle confiance accorder à l'information ?" Les bonnes données sont donc celles que l'on considère comme suffisamment précises et exactes pour l'objectif à atteindre. Lorsque le nombre de données est limité, l'introduction d'information moins précise contribue à l'amélioration du résultat final.

Journel (1986) distingue les données dures (*hard* en anglais) dont l'incertitude attachée à la mesure de la valeur est considérée comme négligeable et les données souples (*soft*). Ces dernières que l'on peut qualifier de contraintes doivent être accompagnées d'une information sur leur incertitude ou sur leur imprécision. Il est possible de distinguer 3 types de données souples, quoique les limites entre chaque groupe puissent apparaître floues ou arbitraires.

- La donnée imprécise est une valeur unique (l'imprécision pouvant être due aux mesures, estimations...). Un exemple est une estimation indirecte de la porosité par des techniques d'inversion sismique. L'incertitude est fournie dans ce cas par la donnée d'un indice de confiance ou d'une distribution *a posteriori*.
- La donnée souple peut être aussi disponible avant toute mesure par l'intermédiaire d'une loi de probabilité. Par exemple, il est possible de connaître, au préalable à toute mesure, la répartition d'une zone en plusieurs faciès grâce à des études similaires et donc par l'intermédiaire de distribution *a priori*.

– Enfin, la donnée d'inégalité n'est connue que par un intervalle.

En prospection pétrolière, on distingue communément les données dures fournies par les forages (rares données précises connues ponctuellement) des données souples fournies par la sismique 3D (données échantillonnées de façon dense, mais moins précises). D'une manière générale, on a souvent tendance à appeler souples toutes données qualitatives (en gros l'information résultant d'une interprétation géologique) par opposition aux données quantitatives, c'est-à-dire dont on puisse donner une valeur (en faisant abstraction de possibles erreurs de mesure).

Pour notre problème, les données dures sont fournies par l'intersection entre un sondage et une interface géologique et les données souples par des relations d'inégalités entre des points du forage et des points de la surface à interpoler.

9.1.2 Les inégalités

Il existe deux types d'inégalité : les contraintes locales et les contraintes globales. Les contraintes globales sont inhérentes à la définition même de ces variables (épaisseur d'une formation géologique positive ou porosité comprise entre 0 et 1). Par contre, les contraintes locales dépendent de la situation, les bornes de l'intervalle sont variables. La mesure de l'imprécision de l'information contenue dans ce type de donnée est fournie par la taille de l'intervalle. Nous nous plaçons avec notre problème de passes incomplètes dans les sondages dans ce dernier cas de figure.

Regardons maintenant en quoi consistent ces données d'inégalités dans l'approche multicouche et dans la méthode des potentiels.

La Figure 9.2 est un schéma permettant d'illustrer les données d'inégalité dans les deux approches.

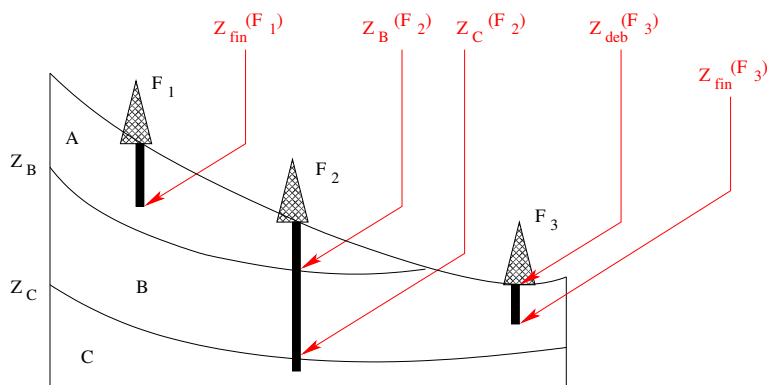


FIG. 9.1 – Contraintes d'inégalité dans les forages

Rappelons que l'approche multicouche considère les surfaces géologiques comme des surfaces d'élévation $z = f(x, y)$.

Soit F_i ($i = 1, 2, 3$) trois forages verticaux implantés aux points de coordonnées 2D $\mathbf{x}_i$, et A, B, C trois formations géologiques. Soit $z_{deb}(F_i)$ et $z_{fin}(F_i)$ les cotes de début et de fin de sondage. On ne s'intéressera pas au toit de la formation A qui est entièrement érodée,

mais au toit de B et C . Soit z_B et z_C , respectivement les altitudes des toits de B et C ; ce sont des variables dépendant de $\mathbf{x}_i$.

On cherche à exprimer les relations existant entre ces valeurs et des altitudes liées aux forages. Le forage F_2 traverse l'interface séparant les deux formations A et B , la relation entre z_B et l'altitude $z_B(F_2)$ à laquelle ce forage rencontre l'interface est une relation d'égalité : $z_B(x_2) = z_B(F_2)$. Il s'agit donc là d'une donnée dure. On peut écrire de même : $z_C(x_2) = z_C(F_2)$.

Or, le forage F_1 ne traverse pas l'interface entre A et B , car il s'arrête trop tôt. La relation entre z_B et $z_{fin}(F_1)$ se traduit par une inégalité : $z_B(x_1) < z_{fin}(F_1)$. De même, pour le sondage F_3 qui n'atteint pas l'interface C , on peut écrire : $z_C(x_3) < z_{fin}(F_3)$.

Enfin, pour le toit de B une autre information peut être exploitée. En effet, le forage F_3 démarre directement dans la formation B qui a été érodée (ou alors le forage est tubé dans les premiers mètres, la formation traversée n'étant alors pas reconnue) d'où la relation avant érosion $z_B > z_{deb}(F_3)$.

Dans la méthode des potentiels, on ne va plus raisonner à partir des altitudes, mais des incréments de potentiel. De cette manière, si on reprend la Figure 9.2, les variables z_B et z_C correspondent respectivement aux valeurs du potentiel sur les interfaces (surfaces isopotentielles) B et C ; ce sont cette fois-ci des valeurs uniques. Un point dans un sondage est désormais défini par ses coordonnées dans l'espace 3D et par une valeur de potentiel inconnue. Il est alors possible d'exprimer une relation entre un point du forage et un point d'une interface par l'intermédiaire d'une différence de potentiel. Si le sondage rencontre l'interface, la différence de potentiel entre le point du forage et tout point de l'interface est nulle : $z_B - z_B(F_2) = 0$.

Pour les données souples, les points du sondage ne rencontrent pas l'interface et on en déduit les relations d'inégalités suivantes : $z_B - z_{fin}(F_1) < 0$, $z_C - z_{fin}(F_3) < 0$, $z_B - z_{deb}(F_3) > 0$. Le sens des inégalités dépend du sens arbitraire donné au potentiel qui est décroissant avec la profondeur.

On peut définir de telles relations avec tous les points du sondage, mais dans la pratique, on se limite aux points d'intersection avec les formations ainsi qu'aux extrémités du forage.

9.1.3 Bornes

Les bornes des intervalles sont porteuses de l'incertitude relative à ces données. L'intervalle d'une donnée dure se réduit ainsi à un singleton, les deux bornes de l'intervalle étant égales. Généralement lorsqu'on dispose de données souples, seule une des deux bornes de l'intervalle est connue. Ainsi, pour un sondage s'arrêtant avant l'objectif, seule la borne supérieure est présente, c'est la cote de fin de sondage. Pour des questions pratiques et afin d'éviter de prendre une valeur infinie, il a été décidé de fixer la borne inconnue. On prendra alors pour borne inférieure de l'intervalle, la valeur minimale de l'ensemble des bornes inférieures des données certaines moins un certain pourcentage de cette valeur. Il est possible bien sûr d'affecter une valeur différente définie par l'utilisateur en fonction de sa connaissance de la zone étudiée. De même, un forage démarrant sous l'objectif n'a pas de borne supérieure, sa borne inférieure étant la cote de début de sondage. La borne supérieure de l'intervalle de simulation a alors été fixée à la valeur maximale de l'ensemble des

bornes supérieures et des données certaines plus un certain pourcentage de cette valeur. N.B. : Cette étape est néanmoins facultative, puisqu'il est possible de simuler $z|Z > 0$.

9.1.4 L'importance de la validation des contraintes

Il arrive fréquemment de disposer de peu de données exactes (dures) réparties dans l'espace, car elles sont souvent très chères à acquérir. Les données souples, moins précises, sont quant à elles beaucoup plus nombreuses, car d'un coût moins élevé. Un exemple frappant sont les forages pétroliers qui sont en nombre limité sur un champ, alors que la zone est couverte par des milliers de mesures sismiques régulièrement échantillonnées. Si on effectue une estimation à partir des seules données précises, on va donc négliger une plus ou moins grande quantité d'information. La Figure 9.2, inspirée d'un résultat de Bazzana *et al.* (1995), montre l'effet de l'information supplémentaire sur la géométrie d'une interface géologique en coupe. On suppose que le forage représenté ne rencontre pas l'interface que l'on désire modéliser, par contre on sait qu'il se trouve à une profondeur comprise entre les deux étoiles. Le krigeage avec les seules données exactes est le trait noir continu, il ne respecte pas l'intervalle. Il en est de même du krigeage en utilisant la sismique en dérive externe, cette information supplémentaire nous apportant néanmoins la forme anticlinale de l'interface (trait pointillé noir). En rouge (sans dérive externe, trait continu ; avec, trait pointillé), l'utilisation de la contrainte d'inégalité force l'interpolation à respecter l'intervalle.

Ainsi, lorsque l'intervalle n'est pas respecté, non seulement de l'information est mise de côté, mais il apparaît aussi une contradiction avec la connaissance du phénomène considéré.

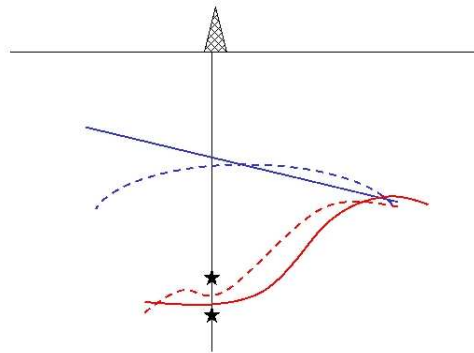


FIG. 9.2 – Coupe d'un forage pétrolier - Trace d'une interface en fonction de l'information utilisée

9.2 Les méthodes utilisant des données d'inégalité

On trouvera dans l'ouvrage de Chilès et Delfiner (1999) une revue des différentes techniques d'estimation ou de simulation maniant des inégalités. L'objectif principal est d'utiliser dans les estimations à la fois des données exactes et des données d'inégalité afin de

préciser le résultat. Ces différentes approches se distinguent par les objets sur lesquels on impose les conditions.

9.2.1 Vérification *a posteriori*

La technique la plus classique consiste à utiliser les données souples *a posteriori*. L'estimation est effectuée sur les données exactes, mais il arrive souvent qu'aux points de données souples le résultat ne vérifie pas les inégalités. Une correction locale est alors entreprise "à la main". Cependant des techniques plus élaborées existent.

9.2.2 Conditions sur les poids

Les conditions sur les poids sont utilisées afin d'éviter le non respect des contraintes globales dû à la présence de poids de krigeage négatifs. Une solution est alors d'imposer uniquement des poids de krigeage positifs soit directement dans le système de krigeage ("krigeage positif" de Barnes et Johnson (1984)) ou en utilisant un modèle de covariance *ad hoc*. Une autre approche est proposée par Diamond (1988) qui krige directement des intervalles.

9.2.3 Méthodes non-linéaires

Des méthodes non-linéaires comme le cokrigeage d'indicatrices prennent en compte naturellement les données d'inégalités (Journel, 1986). L'intervalle de variabilité est discrétisé en un grand nombre de bornes auxquelles on adjoint une distribution de probabilité *a priori*. Malheureusement, ce grand nombre de seuils de coupure rend difficile l'utilisation pratique de cette méthode, car il y a autant de variogrammes simples et croisés à modéliser. Elle est de plus limitée actuellement au cas stationnaire. Il semblerait en réalité que ce type d'approche échoue à obtenir des valeurs dans les intervalles.

9.2.4 Conditions sur les estimations

Une autre technique concernant les contraintes locales consiste à effectuer une minimisation sous contraintes. Dubrule et Kostov (1986) (et Kostov et Dubrule (1986)) proposent d'affecter aux points connus par un encadrement une valeur estimée par minimisation d'une forme quadratique. Cette méthode a pour cadre théorique l'analogie avec les splines en utilisant le formalisme du dual. Si la valeur estimée est dans l'intervalle, elle est conservée (contrainte inactive). Sinon, la borne de l'intervalle la plus proche est affectée au point considéré (contrainte active). Cette méthode simple et rapide n'est cependant pas satisfaisante pour notre problème, à cause de ce calage aux bornes. Ce calage artificiel peut entraîner des artefacts sur une carte isohypse. De plus, on ne dispose pas, avec le krigeage dual, de prédiction des erreurs.

9.2.5 Les inégalités en tant que données

Une autre optique, celle qui est choisie dans notre cas de figure, revient à considérer les inégalités comme des données et à les remplacer par leur espérance conditionnelle. Plus précisément, l'idéal est de remplacer chaque inégalité soit par l'espérance conditionnelle de la valeur compte tenu de l'ensemble des données (dures et souples), soit par une valeur obtenue par tirage au sort dans la loi conditionnelle.

Dans sa thèse, Langlais (1990) développe une méthode incrémentale. Elle estime ou simule la variable au point soumis à une contrainte conditionnellement aux données exactes et aux données d'inégalité déjà générées. Cette méthode rapide, n'est néanmoins pas totalement satisfaisante, car à chaque étape une partie de l'information (celle qui n'a pas encore été transformée) est négligée.

Freulon (1992) propose une méthode itérative basée sur les simulations par chaînes de Markov qui est une solution correcte au problème pour les fonctions aléatoires gaussiennes de moyenne connue. En effet, il est difficile, en présence d'inégalités d'utiliser une méthode directe de simulation. On va donc procéder par approximation en construisant une suite convergente vers la loi désirée, autrement dit une simulation non satisfaisante est modifiée jusqu'à ce qu'une solution satisfaisante soit obtenue. Si les modifications ne dépendent que de l'état courant, on parle alors de chaînes de Markov. L'échantillonneur de Gibbs qui est une méthode de Monte-Carlo par chaînes de Markov (MCMC) est justement une technique itérative qui permet, par approximations successives de simuler une contrainte conditionnellement aux données exactes et aux données d'inégalité.

Dans le domaine du traitement des inégalités on peut encore citer les travaux de Abrahamson et Benth (2001) qui utilisent aussi une technique MCMC à savoir l'algorithme *Data Augmentation* qui permet de simuler paramètres du modèle et données manquantes (en reflétant correctement l'incertitude relative à celles-ci) selon certaines distributions.

Dans la suite de ce travail, nous allons reprendre la méthode de Freulon pour l'adapter au problème des passes incomplètes dans les sondages et ceci dans les deux approches définies précédemment. Nous allons dès à présent définir les différents concepts et méthodes employés.

9.3 La génération des gaussiennes tronquées et les chaînes de Markov

Définissons tout d'abord les données en termes de variables aléatoires.

Soit $X = (X_i)_{1 \leq i \leq n}$ un vecteur aléatoire multigaussien. C'est le vecteur des données : chaque composante est une variable aléatoire, randomisée de la variable régionalisée altitude dans l'approche multicouche, de la différence de potentiel dans la méthode du potentiel.

$(X_1, \dots, X_p) = (x_1, \dots, x_p)$ sont les données exactes et $(X_{p+1}, \dots, X_n) \in B_{p+1} \times \dots \times B_n$ sont les contraintes d'inégalité, les B_i étant des réunions finies d'intervalles. Précisons que dans notre cas de figure, il n'y a qu'un intervalle par contrainte.

L'objectif est de simuler¹ un vecteur multigaussien $X = (X_i)_{1 \leq i \leq n}$ tronqué, c'est-à-dire qu'on impose au vecteur la contrainte $X \in B = \{x_1\} \times \dots \times \{x_p\} \times B_{p+1} \times \dots \times B_n$.

9.3.1 Les gaussiennes tronquées

Un vecteur aléatoire est multigaussien si et seulement si :

- toute combinaison linéaire des composantes du vecteur est une variable aléatoire gaussienne ;
- la densité du vecteur X est :

$$\forall \mathbf{x} \in \mathbb{R}^n, \quad g(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} {}^t(\mathbf{x} - \mathbf{m}) \Sigma^{-1} (\mathbf{x} - \mathbf{m})\right) \quad (9.1)$$

avec $\mathbf{m}$, vecteur réel des moyennes, et Σ , matrice des covariances, qui sont tels que :

$$\begin{aligned} \mathbf{m} &= E[\mathbf{X}] \\ \Sigma &= E[(\mathbf{X} - \mathbf{m})^t (\mathbf{X} - \mathbf{m})]. \end{aligned}$$

La densité conditionnelle du vecteur $\mathbf{X}$ sachant $\mathbf{X} \in B$ est :

$$\mathbf{x} \in \mathbb{R}^n, \quad f(\mathbf{x}) = \frac{g(\mathbf{x}) 1_B(\mathbf{x})}{\int_B g(\mathbf{u}) d\mathbf{u}} \quad (9.2)$$

Un vecteur gaussien tronqué par B est un vecteur aléatoire qui a pour densité f .

9.3.2 La méthode d'acceptation et rejets

Nous allons être amené à un moment donné à simuler une gaussienne scalaire U_i de moyenne nulle et de variance 1 dans un intervalle $[\alpha, \beta]$, donc tronquée. La méthode retenue est une simulation par acceptation et rejets. On désire simuler une variable gaussienne : $X \sim g$. On définit tout d'abord le sous-graphe S_g de g par :

$$S_g = (x, y) \in \mathbb{R} \times \mathbb{R} ; 0 \leq y \leq g(x) \quad (9.3)$$

On simule alors uniformément un couple (X, Y) sur le domaine S_g , ce qui garantit que X soit gaussienne. Pratiquement, on utilise des majorations. Ainsi, on suppose que l'on connaît une densité f et une constante C telles que :

$$\forall x, \quad g(x) \leq C f(x) \quad (9.4)$$

Le principe consiste à générer des points uniformes dans le sous-graphe de Cf et à éliminer les points qui ne sont pas dans le sous-graphe de g . Dans notre cas, la majoration s'effectue par loi uniforme dans la partie centrale de la gaussienne et par $C \cdot |y| \cdot \exp\left(\frac{-y^2}{2}\right)$ sur les côtés. La Figure 9.3 présente le fonctionnement de la méthode pour la partie centrale de la gaussienne, c'est-à-dire lorsque les bornes de l'intervalle sont comprises entre $[-S, S]$ où S est une valeur seuil supérieure à 1, par exemple 2.

¹Des rappels sur les simulations sont exposés en Annexe D.

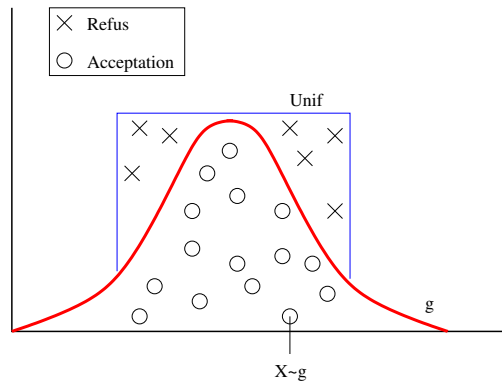


FIG. 9.3 – Majoration de la densité de Gauss par loi uniforme pour la partie centrale de la gaussienne.

Regardons maintenant la manière d'aborder le problème pour une gaussienne tronquée. Soit une variable $U_1 \sim Unif$ (loi uniforme sur $[0, 1]$), il lui correspond une variable $X = \alpha + (\beta - \alpha)U_1$ dans l'intervalle $[\alpha, \beta]$ qui suit une loi uniforme sur $[\alpha, \beta]$ qui constitue la loi f sur cet intervalle. Soit $Y = U.f(X)$ avec $U \sim Unif$. On élimine alors les points (X, Y) qui ne sont pas dans le sous-graphe de g . L'abscisse X des points acceptés suit la loi normale que l'on cherche à simuler. Autrement dit, si on a la majoration $Y \leq \exp(-\frac{X^2}{2})$, la valeur est acceptée, sinon d'autres variables sont tirées jusqu'à ce que la relation soit vérifiée.

9.3.3 Définition et propriétés des chaînes de Markov

L'algorithme utilisé pour la validation des contraintes fait partie des méthodes MCMC qui sont des méthodes de simulations stochastiques très puissantes (Gilks *et al.* (1996), Guyon (1999)). Comme pour toute technique de simulation, il s'agit de générer des échantillons de densité de probabilité π . Or si celle-ci est trop complexe pour une simulation directe, l'idée de base est de construire une chaîne de Markov $X^{(\nu)}$, en partant d'un état initial $X^{(0)}$ et d'un noyau de transition qui garantit la convergence en loi vers π . Au bout d'un grand nombre d'itérations, les échantillons simulés sont considérés comme suffisamment représentatifs de la loi utilisée. La convergence est garantie par une propriété d'ergodicité.

Soit B un sous-ensemble de $\mathbb{R}^n$ de mesure non nulle et $\mathcal{B}_B$ la tribu des Boréliens de B . Une chaîne de Markov à temps discret et états continus à valeurs dans B est un processus $(X^{(\nu)})_{\nu \in \mathbb{N}}$ tel que :

$$\forall \mathcal{A} \in \mathcal{B}_B, P \left\{ X^{(\nu)} \in \mathcal{A} | X^{(0)}, \dots, X^{(\nu-1)} \right\} = P \left\{ X^{(\nu)} \in \mathcal{A} | X^{(\nu-1)} \right\}. \quad (9.5)$$

Cela signifie que la connaissance de tous les états passés, n'est pas nécessaire pour prédire l'état du phénomène à un moment donné; seul suffit la connaissance de l'état précédent. L'avantage de la mise en œuvre d'une telle méthode est de ne stocker en mémoire que l'itération courante et non toutes celles qui la précèdent.

Si la loi qui permet de passer de $X^{(\nu-1)}$ à $X^{(\nu)}$ ne dépend pas de ν , c'est-à-dire si

$P\{X^{(\nu)} \in \mathcal{A} | X^{(\nu-1)} = x\}$ ne dépend pas de ν , on dit alors que la chaîne est homogène. Elle ne dépend donc plus que de x et de $\mathcal{A}$ et on définit une chaîne homogène par :

- la loi de probabilité de son état initial p_0 ;
- son noyau ou probabilité de transition $P(x, \mathcal{A}) = P\{X^{(\nu)} \in \mathcal{A} | X^{(\nu-1)} = x\}$.

On définit, pour une chaîne homogène, $P^{(\nu)}$ la probabilité de transition d'ordre ν :

$$P^{(\nu)}(x, \mathcal{A}) = P\{X^{(k+\nu)} \in \mathcal{A} | X^{(k)} = x\} \quad (9.6)$$

En posant $P^{(1)}(x, \mathcal{A}) = P(x, \mathcal{A})$, on définit le noyau de ν transitions par la relation de récurrence suivante :

$$P^{(\nu)}(x, \mathcal{A}) = \int P^{(\nu-1)}(u, \mathcal{A}) P(x, du) \quad \text{pour } \nu > 1 \quad (9.7)$$

Définissons tout d'abord ce qu'est l'ergodicité. Ergodique signifie que la suite formée par les lois des itérations successives converge vers une loi $\pi > 0$:

$$\forall(x, \mathcal{A}) \in (B, \mathcal{B}_B) \quad \lim_{\nu \rightarrow \infty} P^{(\nu)}(x, \mathcal{A}) = \pi(\mathcal{A}) \quad (9.8)$$

Rappelons que l'objectif est de générer une chaîne de Markov ergodique de loi invariante π . Cela est assuré par le théorème suivant : si P est irréductible, apériodique et admet π comme loi invariante alors P est ergodique. Définissons les propriétés qui interviennent dans ce théorème :

- Irréductible : $\exists \nu; P^{(\nu)}(x, \mathcal{A}) > 0$
Ce qui signifie que tout point de l'espace d'état peut être atteint par la chaîne de Markov. Cette propriété permet de réduire grandement la dépendance de la chaîne vis à vis des conditions initiales p_0 et $X^{(0)}$.
- Apériodique : on a $P^{(\nu)}(x, \mathcal{A}) > 0$ à partir d'un certain rang.
Cette propriété évite à la chaîne de boucler sur un sous-ensemble des états possibles.
- Invariante : la loi π est invariante par P si $X^{(\nu)} \sim \pi \Rightarrow X^{(\nu+1)} \sim \pi$.

Pour le problème qui nous occupe, ce résultat nous assure donc que, si on se place dans ces conditions, la loi des simulations successives converge vers la loi d'un vecteur gaussien tronqué si la probabilité de transition est ergodique. Il nous reste à trouver ce noyau.

9.3.4 La probabilité de transition de Geman et l'échantillonneur de Gibbs

Le principe de cette probabilité de transition introduite par les frères Geman pour la restauration d'image (Geman et Geman, 1984), est d'oublier la valeur initiale de la composante à modifier. Elle est remplacée par une valeur simulée selon sa loi conditionnelle. On utilise donc des densités conditionnelles pour générer les états successifs.

Soient $\mathbf{x} \in \mathbb{R}^n$ et $\mathbf{y} \in \mathbb{R}^n$. Soit P_i la probabilité de transition du vecteur $\mathbf{x}$ au vecteur $\mathbf{y}$ et ne modifiant que la composante i . On note x_i cette i -ème composante et $\overline{\mathbf{x}}_i$ le vecteur privé de sa i -ème composante : $\overline{\mathbf{x}}_i = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$. On peut donc écrire $\mathbf{x} = (x_i, \overline{\mathbf{x}}_i)$

et $\mathbf{y} = (y_i, \overline{x_i})$. Si f est la densité de la gaussienne tronquée sur B , $f(\cdot|\overline{x_i})$ est la densité conditionnelle de la i -ème composante à $\overline{x_i}$ fixé. On a alors :

$$dP_i(\mathbf{x}, \mathbf{y}) = \delta_{\overline{x_i}}(\overline{y_i}) f(y_i|\overline{x_i}) dy_i \quad \text{où } \delta_{(\cdot)} \text{ représente la mesure de Dirac} \quad (9.9)$$

On appelle échantillonneur de Gibbs l'algorithme de simulation utilisant le noyau de transition de Geman. Regardons ce que devient la probabilité de transition de Geman dans notre cas de figure. Pour un vecteur multigaussien tronqué, on a :

$$f(x_i|\overline{x_i}) \propto g(x_i|\overline{x_i}) 1_{B_i}(x_i) \quad (9.10)$$

En détaillant, on obtient :

$$f(x_i|\overline{x_i}) = \frac{f(x_i, \overline{x_i})}{f(\overline{x_i})} = \frac{g\left(\frac{x_i - x_i^{KS}(\overline{x_i})}{\sigma_{KS}(i)}\right) 1_{B_i}(x_i) \frac{1}{\sigma_{KS}(i)}}{\int_{B_i} g\left(\frac{u_i - x_i^{KS}(\overline{x_i})}{\sigma_{KS}(i)}\right) \frac{du_i}{\sigma_{KS}(i)}} \quad (9.11)$$

d'où :

$$dP_i(\mathbf{x}, \mathbf{y}) = \delta_{\overline{x_i}}(\overline{y_i}) \frac{g\left(\frac{y_i - x_i^{KS}(\overline{x_i})}{\sigma_{KS}(i)}\right) 1_{B_i}(y_i) dy_i}{\int_{B_i} g\left(\frac{u_i - x_i^{KS}(\overline{x_i})}{\sigma_{KS}(i)}\right) du_i} \quad (9.12)$$

On en déduit en termes de variables aléatoires :

$$X_i^{(\nu+1)} = X_i^{KS} + \sigma_{KS}(i) U_i \quad (9.13)$$

X_i^{KS} est le krigeage simple de X_i par $(X_1^{(\nu+1)}, \dots, X_{i-1}^{(\nu+1)}, X_{i+1}^{(\nu)}, \dots, X_n^{(\nu)})$ et U_i désigne une gaussienne scalaire tronquée sur $B_i^* = \{u_i \in \mathbb{R}, X_i^{KS} + \sigma_{KS}(i)u_i \in B_i\}$. Cela signifie que la $i^{ème}$ composante du vecteur à simuler se calcule à partir de l'estimateur de krigeage des valeurs des autres composantes plus le produit de l'écart-type de krigeage par un résidu gaussien. Quant à l'estimateur de krigeage de la $i^{ème}$ composante à l'itération $(\nu + 1)$, il s'obtient en utilisant les valeurs des $(i - 1)^{èmes}$ premières composantes déjà mises à jour lors de cette itération et des valeurs des autres composantes déterminées à l'itération précédente.

On détermine ainsi la procédure à suivre pour un cycle (itération sur ν) :

- choix d'un site i (balayage systématique ou aléatoire des données d'inégalité)
- relaxation du site i : $y_i \sim f(y_i|\overline{x_i})$
- itération sur l'ensemble des sites.

Ainsi, la composante X_i du vecteur multigaussien $\mathbf{X}$ suit la loi normale de moyenne le krigeage (X_i^{KS}) et de variance la variance de krigeage ($\sigma_{KS}^2(i)$). De plus, comme on veut que $X_i^{inf} \leq X_i \leq X_i^{sup}$, on doit donc avoir : $\alpha = \frac{X_i^{inf} - X_i^{KS}}{\sigma_{KS}(i)} \leq U_i \leq \frac{X_i^{sup} - X_i^{KS}}{\sigma_{KS}(i)} = \beta$ pour les bornes de l'intervalle de simulation de la gaussienne scalaire.

9.4 L'algorithme de validation des contraintes

9.4.1 Résumé des différentes étapes de l'algorithme

Cet algorithme va consister à transformer les données d'inégalité en données d'égalité par une approche de validation des contraintes. Après lecture des données et mise à jour des bornes des intervalles (cf. 9.1.3), la validation des contraintes peut commencer par la phase d'initialisation. Celle-ci comprend le calcul de la matrice de covariance, sa triangulation (LU), son inversion et l'initialisation proprement dite de l'échantillonneur de Gibbs. Cet état initial doit simplement être cohérent avec toutes les données (égalités et inégalités), l'échantillonneur de Gibbs se chargeant alors d'introduire la structure spatiale. La méthode nécessite donc la connaissance des poids de krigeage et de la variance en tous points en voisinage unique. Le calcul des poids de krigeage utilise une méthode directe de résolution de systèmes linéaires (méthode des sous-matrices) dite technique des partitionnements (Gastinel, 1966). En effet, pour l'application de l'échantillonneur de Gibbs dans le cas gaussien, poids de krigeage et variance de krigeage dépendent uniquement de l'inverse de la matrice de covariance. Ainsi on a :

$$y_i = \sum_{j \neq i} \lambda_j^i y_j \text{ avec } \lambda_j^i = \frac{-b^{ij}}{b^{ii}} \text{ pour } i \neq j$$

où b^{ij} est l'élément générique de l'inverse de la matrice de covariance. De même, on a : $\sigma^2(i) = \frac{1}{b^{ii}}$. Cette méthode est bien évidemment limitée par la taille de la matrice donc du nombre de données utilisées.

L'initialisation se fait en tirant aléatoirement une valeur dans l'intervalle. Ensuite, chaque contrainte est passée en revue et simulée selon l'équation 9.13 lors d'un cycle. Ce processus est répété un certain nombre de fois (N simulations sont obtenues en conservant une valeur simulée tous les p cycles, donc N*p itérations au total) afin de garantir la convergence vers la loi. Or la moyenne de toutes les simulations conditionnelles convergent vers l'espérance conditionnelle. On effectue ainsi la moyenne de ces N simulations de p cycles pour chaque contrainte afin d'obtenir une estimation de la valeur "validée" de chaque contrainte (approche de type Monte-Carlo). En effet, d'après la loi forte des grands nombres, on a convergence presque sûre :

$$\lim_{\nu \rightarrow +\infty} \frac{1}{\nu} \sum_{i=1}^{\nu} \tilde{X}^{(i)} = \mathbf{E}[X|X \in B] \quad (9.14)$$

avec $\tilde{X}^{(i)}$ la i-ème simulation de X

Il faut bien voir que dans les méthodes MCMC on distingue deux processus. Les premières itérations, qui ont pour but une convergence rapide vers l'équilibre, ne doivent pas être prises en compte. Après cette phase de chauffe (ou *burn-in*), la chaîne de Markov est supposée avoir atteint son régime stationnaire. On ne garde alors que les états de la deuxième étape de simulation qui sont des échantillons de la loi simulée afin de garantir une bonne estimation par Monte-Carlo. Au bout de ce processus on a donc transformé chaque donnée d'intervalle en une donnée unique. Le problème de la vitesse de convergence est abordé

de manière expérimentale au chapitre suivant (cf. 10.2.1). Pour être totalement tranquille vis-à-vis de l'indépendance entre les différentes simulations, il est conseillé de lancer les simulations en réinitialisant à chaque fois.

9.4.2 La prise en compte des inégalités dans la méthode des potentiels

Jusqu'à présent seules les données d'interface dures à savoir les différences de potentiel nulles étaient utilisées dans l'*Editeur Géologique*. Le respect du modèle vis à vis des sondages incomplets était vérifié *a posteriori*. Le calcul du modèle géologique était réitéré après avoir donné une valeur arbitraire aux inégalités les plus violées et ceci jusqu'au respect de tous les sondages (essais/erreurs). La convergence de cette méthode est lente et non assurée, et l'affectation des valeurs n'est de plus qu'approximative. La méthode de simulation apporte donc une certaine souplesse, tout en déterminant les nouvelles valeurs en accord avec les autres données. Les inégalités sont repérées dans les données de sondage par un algorithme exploitant la pile stratigraphique et les relations de type érosion/dépôt. Il permet ainsi à tout point dans un sondage d'être désormais défini, outre par ses coordonnées dans l'espace, par un entier (0,1 ou -1) suivant que la différence de potentiel entre ce point et la référence de l'interface est nulle, positive ou négative.

Les valeurs de potentiel vont être simulées en ces points, même s'ils n'appartiennent pas à une interface. Les schémas de la Figure 9.4 décrivent la différence d'approche entre la méthode classique et la méthode des potentiels. On cherche à interpoler le toit de B et on dispose de données d'inégalité. L'étape de validation des contraintes est engagée au préalable. Dans le point de vue multicouche, la valeur de l'altitude de l'intersection entre le sondage F et l'interface $Z(B/F)$ est simulée dans un intervalle dont le point de fin de sondage est ici la borne supérieure. Dans l'approche de l'*Editeur*, on simule le potentiel $Z_{fin}(F)$ au point de fin de sondage dont on sait que la différence de potentiel avec tout point de l'interface est ici positive.

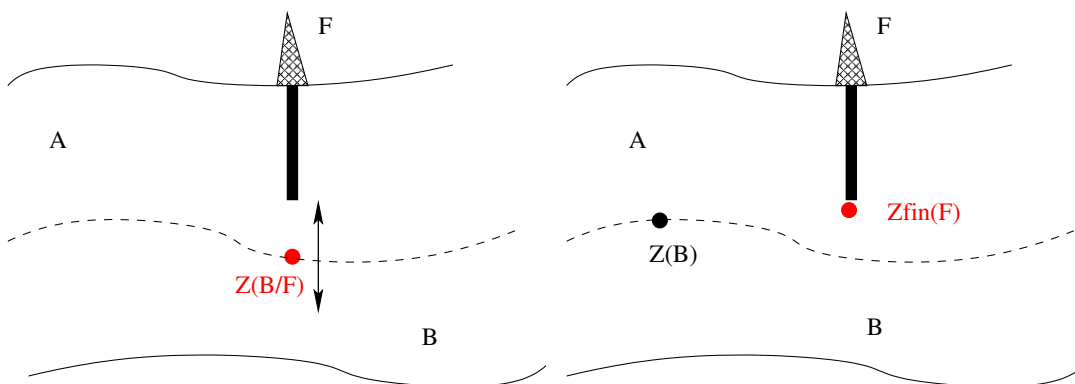


FIG. 9.4 – Comparaison de la validation des contraintes dans l'approche multicouche à gauche, et dans la méthode des potentiels à droite.

Le problème se complique lorsqu'il s'agit de prendre en compte plusieurs interfaces en présence de lacunes. Il faut dans ce cas exploiter les relations érosion/dépôt. Les données

souples ne sont alors pas seulement les extrémités du forage, mais les relations avec les intersections à l'intérieur de celui-ci peuvent aussi s'exprimer par des inégalités.

L'expérience montre que les différences de potentiel ne prennent jamais de valeurs très grandes (quelques unités ou dixièmes d'unités), car les coordonnées des points sont ramenées entre 0 et 1. Les bornes des intervalles de simulation sont alors fixées à la plus petite valeur proche de 0 et à ± 10 pour la limite. Les valeurs simulées sont ensuite récupérées pour augmenter le jeu de données et effectuer le calcul du potentiel et le tracé de ses iso-valeurs.

Le second membre du système de cokrigeage (cf. Partie A.6 de l'Annexe A) contient donc outre les valeurs des gradients (vecteur G), les incréments de potentiel (vecteur I) nuls pour les données dures et non-nuls pour les données souples, ces dernières ayant été obtenues après l'étape de validation des contraintes.

Le chapitre suivant présente des applications de cette technique. N.B. : Cette remarque concerne les deux approches et les incertitudes. En effet, il ne faut pas utiliser les données obtenues après validation comme des données dures dans le cadre de la détermination des incertitudes, car il n'est plus assuré que la variance soit minimale. La variance calculée lors de l'étape "Monte-Carlo" fournit une information sur leur incertitude, les vraies données dures ayant une incertitude considérée comme négligeable.

Chapitre 10

Le traitement des sondages incomplets - Applications

Deux types d'applications vont être présentés dans ce chapitre. Le premier concerne l'approche géostatistique classique (on travaille donc sur la cote et non le potentiel) où la validation des contraintes est testée sur un cas réel : la ville de Lyon. Le deuxième est un ensemble de cas synthétiques destinés à illustrer son implémentation dans la méthode des potentiels.

10.1 Le cas de la ville de Lyon : Etudes préliminaires

10.1.1 Présentation des données

La méthode présentée au chapitre précédent a été testée sur un jeu de 652 données réelles localisées dans un secteur de la ville de Lyon (Figure 10.1). La taille du champ est approximativement de $9km$ sur $8km$. Il se trouve en bordure du socle du Massif central : des terrains sédimentaires, pour la plupart des formations récentes (Quaternaire), recouvrent le socle primaire. En particulier on trouve une pseudo-terrasse alluviale (dite villafranchienne) réentaillée par la Saône, qui coule dans la partie sud-est de la zone d'étude. La Figure 10.2 est une photographie aérienne de la partie sud où la concentration de données est importante. On localise particulièrement bien le bâti, les voies de circulation, la rivière et la végétation.

La variable étudiée est l'altitude de l'interface entre le socle et les séries sédimentaires. Elle provient de l'étude de sondages de différentes sources (Banque du Sous-Sol, Bureau de Recherches Géologiques et Minières, Communauté Urbaine de Lyon). Les séries sédimentaires correspondent à tous les terrains déposés sur le socle primaire. Les données d'égalité (sondages ayant atteint le socle) sont au nombre de 271 contre 381 données d'inégalité majoritairement des sondages arrêtés avant d'avoir atteint le toit du socle. L'échantillonnage est irrégulier, on constate la présence de nombreux groupes de sondages (*clusters*) et de points isolés. Certaines données s'alignent le long de directions correspondant à des voies de communication (autoroutes). Certaines parties du champ apparaissent dépourvues de

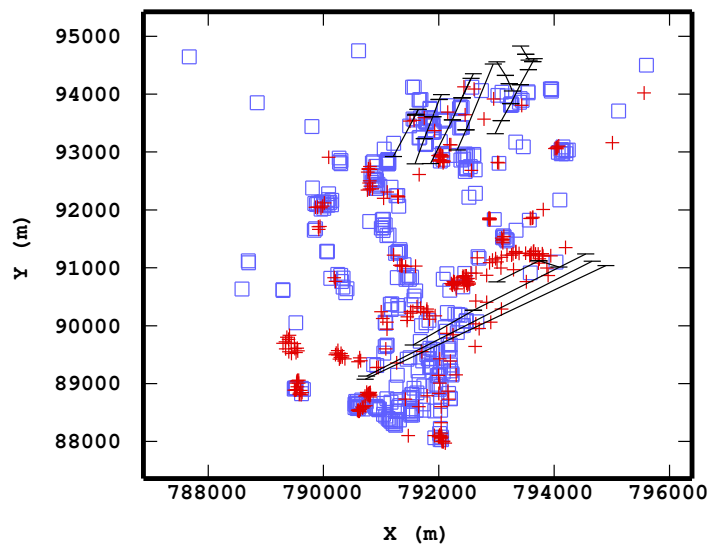


FIG. 10.1 – Carte de localisation des données - Légende : + rouge : données exactes, $\square$ bleu : données d'inégalité.

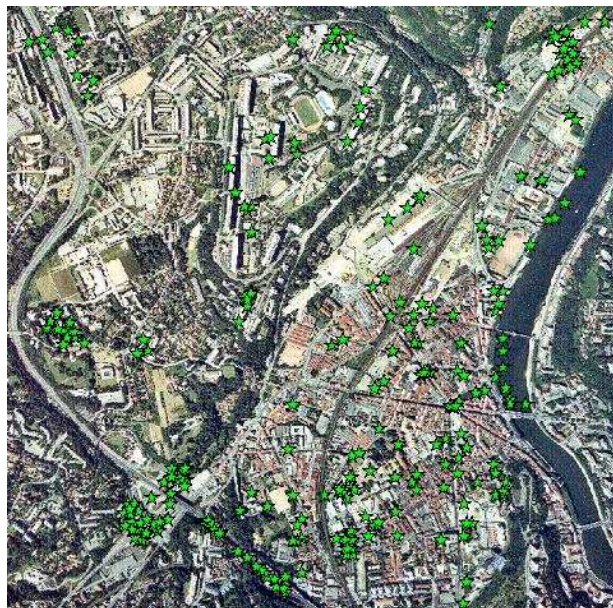


FIG. 10.2 – Zoom de la carte de localisation des données avec photo aérienne en fond de carte.

toute information, en particulier la partie centrale. De plus, la répartition entre données dures et souples est irrégulière. On observe des zones composées uniquement de données d'égalité, d'autres uniquement de données d'inégalité, d'autres enfin avec les deux types de données. On constate, en particulier, une grande concentration de données d'égalité juste au nord des grandes failles du sud. La longueur moyenne des forages est de l'ordre de 15m. On dispose également des traces de 11 failles supposées verticales. Les failles les plus longues sont au sud et sont orientées à peu près ENE / WSW (70° N). Ce sont celles qui ont le rejet le plus important. Lors de l'extension oligocène, elles ont créé de très fins bassins qui se sont remplis de molasses. Certains blocs ont été surélevés lors d'une phase de compression postérieure. Les autres failles au nord sont orientées NNE / SSW et NNW / SSE et correspondent à de la distension.

10.1.2 Analyse exploratoire

L'objectif de l'analyse exploratoire est d'étudier la distribution et l'homogénéité des données dans l'espace. On met ainsi en évidence les observations les plus " atypiques ".

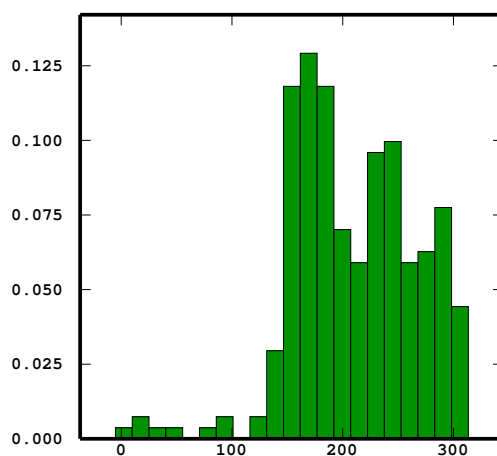


FIG. 10.3 – Histogramme des données dures.

Minimum	Maximum	Moyenne	Ecart-type
-4.88	313.17	210.10	56.47

Quartiles	Q25 = 170.45	Q50 = 207.22	Q75 = 249.59
-----------	--------------	--------------	--------------

TAB. 10.1 – Statistiques élémentaires des données dures (altitude du toit du socle en m).

L'histogramme des 271 données d'égalité est présenté à la Figure 10.3 et le Tableau 10.1 résume les principaux résultats statistiques. La plupart des valeurs (97%) sont comprises

entre 120 et 310m. A peu près 40% des valeurs sont entre 140 et 190m (1er pic) et moins de 20% sont entre 230 et 240m (2ème pic). Les faibles valeurs (0 à 110m), soit 3% du total, se trouvent essentiellement entre les grandes failles du sud.

Le diagramme quantile contre quantile (Figure 10.4(a)) compare la distribution expérimentale des valeurs avec des distributions données (ici gaussienne). Les 6 valeurs les plus extrêmes n'ont pas été prises en compte. La variable étudiée est en ordonnée, la gaussienne en abscisse. Si la variable avait été gaussienne, on aurait obtenu une droite centrée sur les pointillés. On voit qu'ici les données ne sont pas gaussiennes, si du moins on suppose la moyenne constante, comme le montrait également l'allure de l'histogramme. On ne se trouve donc pas dans le cadre gaussien de moyenne connue pour lequel la méthode de validation des contraintes décrite au Chapitre 9 a été développée. On examinera sa robustesse plus loin.

Les nuages de corrélation entre l'altitude Z du toit du socle et les coordonnées géographiques X et Y mettent en évidence le comportement de la variable le long des directions principales de l'espace (Figure 10.5). Ils permettent de tester la validité d'une hypothèse de stationnarité (locale ou globale) ou au contraire d'observer l'existence d'une dérive. Les courbes de régression expérimentale donnant la valeur moyenne de Z pour différentes classes de X ou Y sont tracées sur le nuage et montrent que dans l'ensemble, Z décroît avec X ($\approx 100m$) et croît avec Y ($\approx 100m$ aussi). On peut penser, en première approximation à l'existence d'une dérive linéaire. Ces nuages permettent également de repérer les observations les plus atypiques c'est-à-dire les points qui s'éloignent le plus de la nuée. Ici ce sont les valeurs faibles situées entre les failles.

Enfin, des nuages de corrélation différée (*h-scatter plot*) ont été déterminés (Figure 10.4(b)). Pour un vecteur h donné, ici de 1500m, c'est le nuage des couples $(z(x_a), z(x_{a+h}))$ tels que x_a et x_{a+h} coïncident avec des sites de données. Les couples de données fort dissemblables correspondent aux points du nuage éloignés de la première bissectrice. Les valeurs atypiques du nuage de corrélation différée correspondent aux importants rejets de ces failles.

10.1.3 Analyse structurale

Le variogramme expérimental omnidirectionnel a été calculé à partir des seules données d'égalité, d'une part sans tenir compte des failles (tous les couples de données sont pris en considération ; Figure 10.6(a)), d'autre part en tenant compte des failles (les couples de points séparés par une faille sont omis ; Figure 10.6(b)). Ces variogrammes sont déterminés à partir d'un grand nombre de couples, mais cela est trompeur, à cause de la présence des *clusters*. On a donc une sur-représentation des zones à forte densité de sondages. Cependant, si on excepte les zones proches des failles, la variabilité au sein des groupes de sondages est faible. Pour éviter le *clustering*, on peut envisager de compenser par pondération ou en ne prenant qu'un point par groupe, puis itérer en changeant de point. Dans les deux cas, le variogramme est continu à l'origine, linéaire, ce qui est conforme avec la physique de la variable altitude. Le palier est atteint vers 3000m : à cette distance deux données ne sont plus corrélées. Vers 1500m, on observe un creux qui est certainement en relation avec de fortes valeurs systématiquement entourées de faibles valeurs. De plus, un changement de pente, qui indique qu'à partir de cette distance la variabilité croît plus

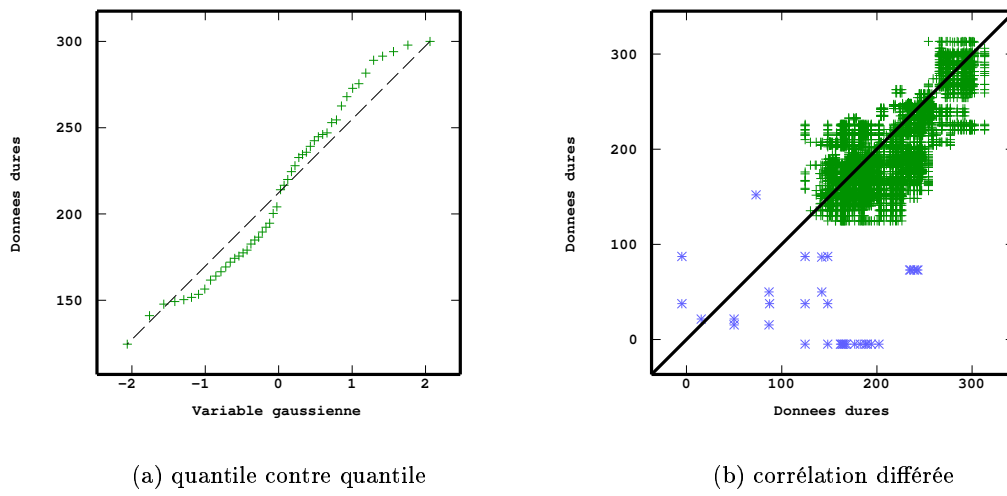


FIG. 10.4 – Diagramme quantile contre quantile (distribution expérimentale vs distribution gaussienne), à gauche et diagramme de corrélation différée ($h = 1500m$), à droite.

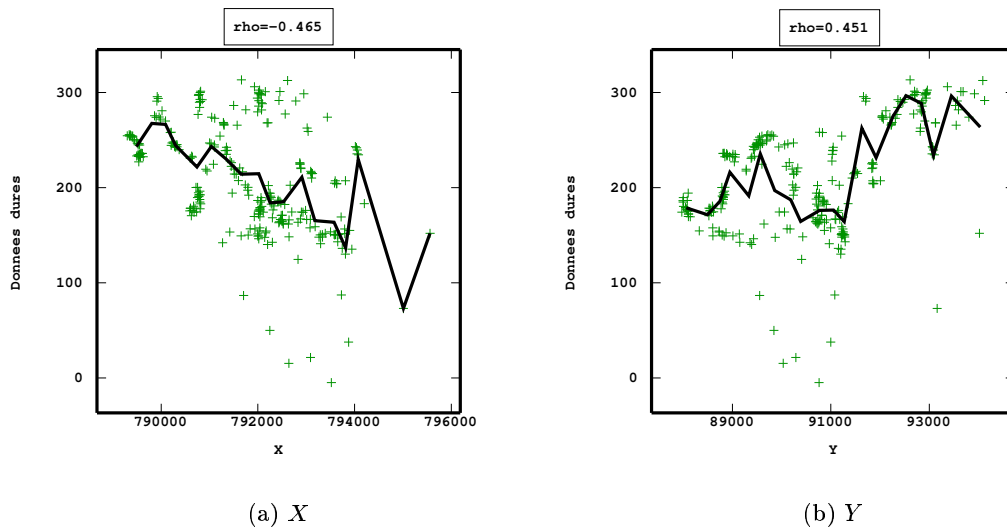
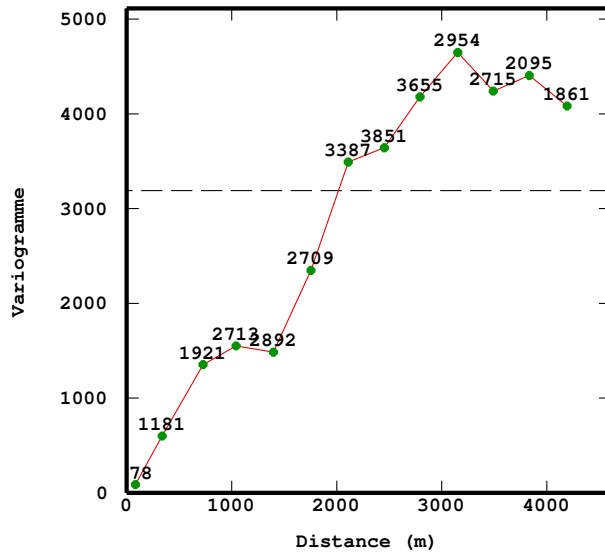
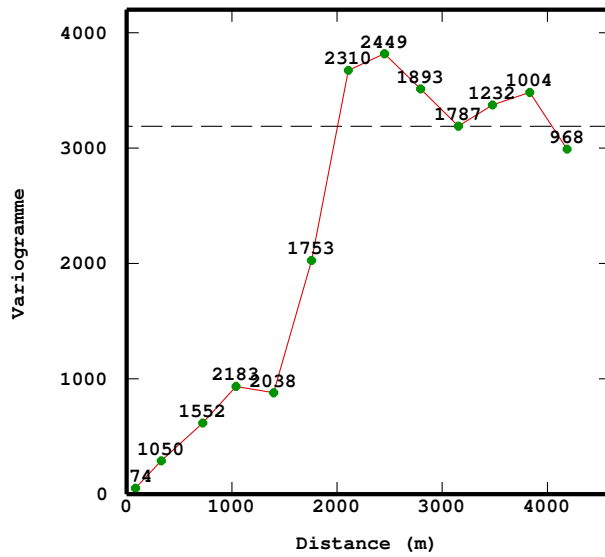


FIG. 10.5 – Nuages de corrélation entre les données d'altitude et les coordonnées géographiques X et Y .



(a) sans failles



(b) avec failles

FIG. 10.6 – Variogramme omnidirectionnel expérimental pour les 271 données dures, en haut sans tenir compte des failles et en bas en tenant compte des failles. C'est celui-ci qui sera utilisé dans la suite de l'étude.

rapidement, est particulièrement flagrant. Cela est dû à l'orientation de la structure et des grandes failles du sud comme on peut le voir sur la carte interpolée (Figure 10.21) et sur les variogrammes directionnels calculés dans les directions principales mises en évidence par la carte variographique (Figure 10.7(a)) : cette carte représente le variogramme expérimental en fonction du vecteur $\mathbf{h}$ (donc de son module et de sa direction). La rupture de

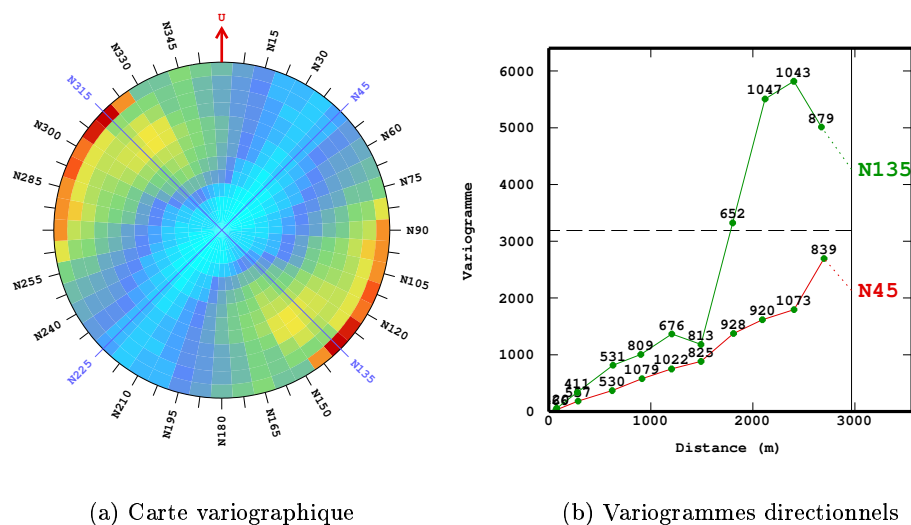


FIG. 10.7 – Carte variographique, à gauche et variogrammes directionnels ($45^\circ N$ et $135^\circ N$) expérimentaux pour les 271 données dures, à droite.

pente s'observe en effet dans la direction N135 où la variabilité est la plus grande. Dans la direction N45, par contre, qui est la direction de continuité, le palier est plus faible et il n'y a pas de changement de pente (Figure 10.7(b)). Dans un premier temps, l'anisotropie n'a pas été prise en compte dans le modèle et seul le variogramme omnidirectionnel a été modélisé.

Très rapidement, vu l'allure du variogramme expérimental, le choix s'est porté vers un modèle sphérique ou cubique. Finalement, un modèle gigogne dont les caractéristiques sont reportées dans le Tableau 10.2 a été retenu. La Figure 10.8 montre que la modélisation

Modèle	Portée	Palier
Sphérique	2200	900
Cubique	3600	2800

TAB. 10.2 – Caractéristiques du modèle retenu

n'a pas été facile et que le modèle n'est pas très bien calé sur le variogramme expérimental, en grande partie à cause de la pente raide après la rupture.

A partir de ce modèle, les altitudes des 381 données d'inégalité ont été simulées.

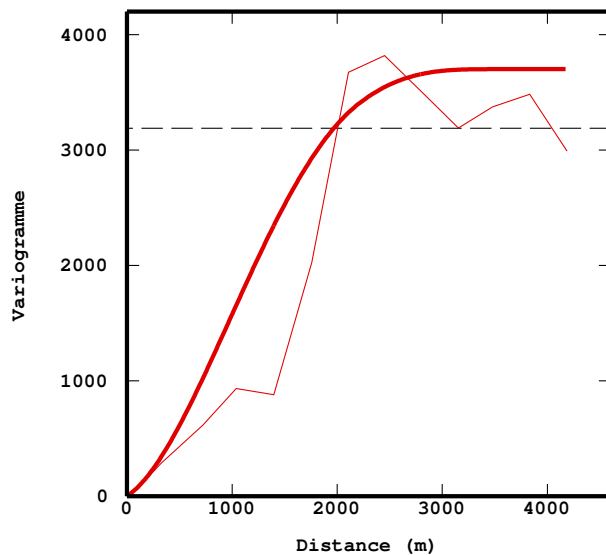


FIG. 10.8 – Variogramme omnidirectionnel modélisé pour les 271 données dures

10.2 Le cas de la ville de Lyon : Résultats de la validation des contraintes

10.2.1 Convergence

La convergence de la méthode vers la loi simulée est garantie pour un nombre d'itérations suffisamment grand. Il n'existe actuellement pas de critère théorique permettant de fixer ce nombre. On sait néanmoins que la vitesse de convergence de l'algorithme dépend des valeurs propres de la matrice d'itération (Galli et Gao, 2001). De nombreuses méthodes existent, deux revues de celles-ci sont présentées dans Cowles et Carlin (1996), et Brooks et Roberts (1998). Un critère empirique de détermination de l'état stationnaire de la chaîne est la convergence de la moyenne et de la variance des itérations. La convergence est plus rapide pour certains sondages que pour d'autres : de quelques itérations à plusieurs centaines pour certains forages isolés. La Figure 10.9 présente les résultats d'une simulation pour le forage A005, situé au sud des grandes failles ($X=791879.93$, $Y=89289.97$) dans une zone densément échantillonnée où les données d'inégalité sont particulièrement nombreuses. On observe sur la Figure 10.10 la moyenne et la variance des simulations. Le processus a été initialisé à $127.38m$ alors que la borne supérieure est à $128.28m$. L'histogramme des valeurs obtenues est une gaussienne tronquée. Une fois le régime stationnaire apparemment atteint, les valeurs possibles de l'altitude du toit du socle varient autour de $120m$, avec une variance de $40m^2$. La valeur krigée sans validation donnerait une valeur de $150m$ avec un écart-type de $12m$, donc en dehors de l'intervalle. De même, la Figure 10.11 présente les résultats pour le forage A224, isolé au nord-ouest ($X=788849.93$, $Y=93850.01$). Les courbes des deux forages qui illustrent ces propos ne sont pas des cas extrêmes, mais

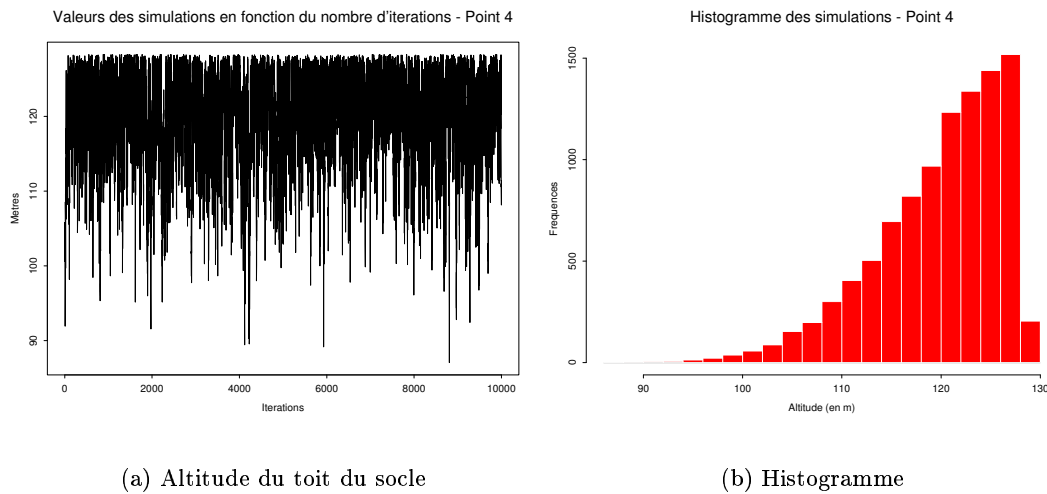


FIG. 10.9 – Forage A005 - Courbe de l’altitude du toit du socle en fonction du numéro d’itération, à gauche, et histogramme des valeurs obtenues pour 10000 itérations, à droite.

sont cependant représentatives. La moyenne des simulations converge plus rapidement pour A005 que pour A224 étant donné leurs voisinages respectifs. La variance pour A005 au bout de 1000 itérations est plutôt faible, alors qu’elle est très élevée pour A224.

Toujours est-il qu’on observe sur ces deux exemples une première phase de convergence dite de “chauffage” (*burn in*) où la loi simulée converge vers la loi stationnaire. Au bout de quelques centaines d’itérations, nous avons l’état stationnaire, on calcule dès lors plusieurs centaines ou milliers de simulations.

Comment s’assurer de la convergence de la chaîne de Markov vers la loi stationnaire ? Le danger est la convergence lente où une unique chaîne de Markov semble avoir convergé alors que d’autres chaînes indépendantes révèlent en fait qu’elle n’est pas atteinte. En effet, des mouvements internes aux chaînes, trop lents pour être détectés, sont la cause de cette convergence apparente. Ainsi deux séquences prises séparément peuvent sembler converger, alors que cela n’est pas le cas (valeurs finales très différentes).

Gelman (1996) propose de contrôler la convergence en détectant le moment où les chaînes de Markov “ont oublié leur point de départ”. Il s’agit donc de comparer des graphes de variance en fonction du nombre d’itérations pour différentes séquences parallèles. La convergence intervient approximativement lorsque la variance entre les différentes chaînes n’est pas plus grande que la variance à l’intérieur de chaque séquence prise individuellement.

Nous utilisons alors n_c chaînes parallèles contenant chacune n_s valeurs simulées. Ainsi pour chaque altitude simulée Z_{ij} (que l’on n’indexe pas sur le sondage considéré pour simplifier, mais sur son appartenance à une séquence (indice i) et sa position dans celle-ci (indice j)), on calcule deux grandeurs :

- la variance “entre séquences” var_E
- la variance “à l’intérieur d’une séquence” var_I .

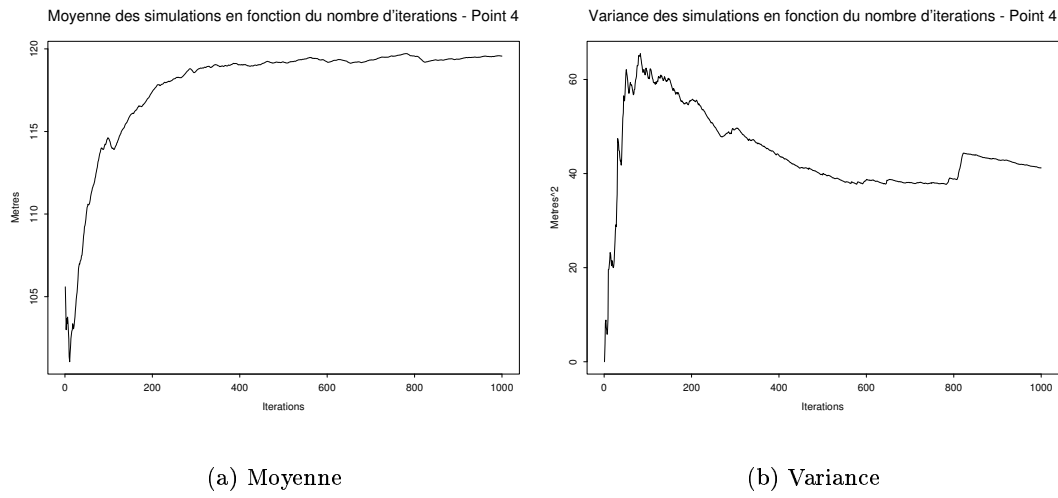


FIG. 10.10 – Forage A005 - Moyenne et variance de l'altitude de l'interface séries sédimentaires/socle en fonction du numéro d'itérations.

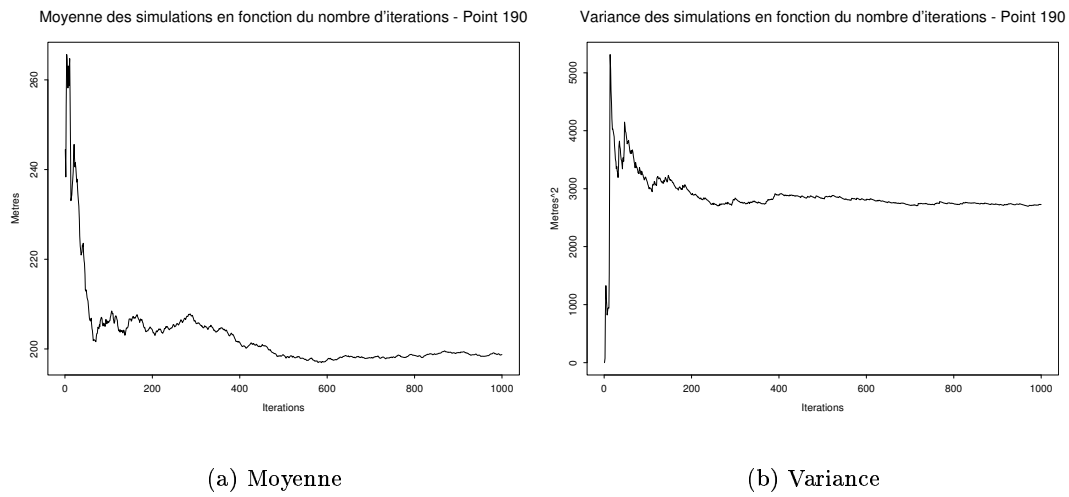


FIG. 10.11 – Forage A224 - Moyenne et variance de l'altitude de l'interface séries sédimentaires/socle en fonction du numéro d'itérations.

On les définit de la manière suivante :

$$\begin{aligned} var_E &= \frac{n_s}{n_c-1} \sum_{i=1}^{n_c} (\bar{Z}_{i.} - \bar{Z}_{..})^2 \text{ où } \bar{Z}_{i.} = \frac{1}{n_s} \sum_{j=1}^{n_s} Z_{ij} \text{ et } \bar{Z}_{..} = \frac{1}{n_c} \sum_{i=1}^{n_c} \bar{Z}_{i.} \\ var_I &= \frac{1}{n_c} \sum_{i=1}^{n_c} \left(\frac{1}{n_s-1} \sum_{j=1}^{n_s} (Z_{ij} - \bar{Z}_{i.})^2 \right) \end{aligned} \quad (10.1)$$

On définit :

$$\widehat{var}(Z) = \frac{n_s-1}{n_s} var_I + \frac{1}{n_s} var_E \quad (10.2)$$

qui est un estimateur de la variance de Z qui surestime cette valeur. De même, en général, var_I est un estimateur de la variance de Z qui sous-estime cette valeur. Donc pour un nombre suffisamment important de simulations n_s , ces deux estimateurs approchent $var(Z)$ par des directions opposées.

Enfin on a :

$$\sqrt{\hat{R}} = \sqrt{\frac{\widehat{var}(Z)}{var_I}} \quad (10.3)$$

Le “facteur de réduction d'échelle potentiel” R , dont $\hat{R}$ est une estimation, décroît vers 1 lorsque la simulation converge. L'augmentation du nombre d'itérations permet alors de diminuer une valeur élevée de R .

La Figure 10.12(a) montre une courbe de $\sqrt{\hat{R}}$ en fonction du nombre d'itérations, et la Figure 10.12(b) présente les courbes de variance qui convergent vers $var(Z)$. Ces résultats ont été obtenus avec $n_c = 10$. D'une manière générale, on a constaté une convergence rapide du processus. On peut également vérifier l'influence de la simulation initiale sur la convergence, car la probabilité de transition de Geman est définie complètement par la loi du vecteur à simuler. Si on initialise la première itération de l'échantillonneur non par une valeur tirée uniformément dans l'intervalle, mais systématiquement par la borne inférieure de chaque intervalle, on constate des différences minimes avec le résultat précédent (de l'ordre de $0.1m$ en moyenne), ce qui montre que la valeur de la première simulation n'influe pas sur le résultat.

Enfin, la convergence de la moyenne des valeurs simulées a aussi été vérifiée. La variance de l'estimateur fait intervenir les corrélations entre différents états de la chaîne. La dernière itération sur un certain nombre (pas d'échantillonnage k) est conservée afin d'assurer un certain mélange dans la chaîne et donc de diminuer la dépendance entre les différents états. Afin de calculer la variance, on emploie ainsi des moyennes par paquets (*batch means*), c'est-à-dire les moyennes des itérations utilisées pour former une réalisation Z_{ik} (k étant la taille du cycle).

Le résultat est l'ensemble des valeurs simulées pour toutes les positions de données souples. Au final pour un forage, on sort 1000 simulations (1 itération sur 15 étant conservée) qui sont ensuite moyennées. Ces chiffres ont été choisis pour un résultat optimal, même si une bonne précision peut être obtenue pour quelques centaines de simulations seulement. De plus, le gain de précision est négligeable si on augmente le nombre de simulations et donc le temps de calcul.

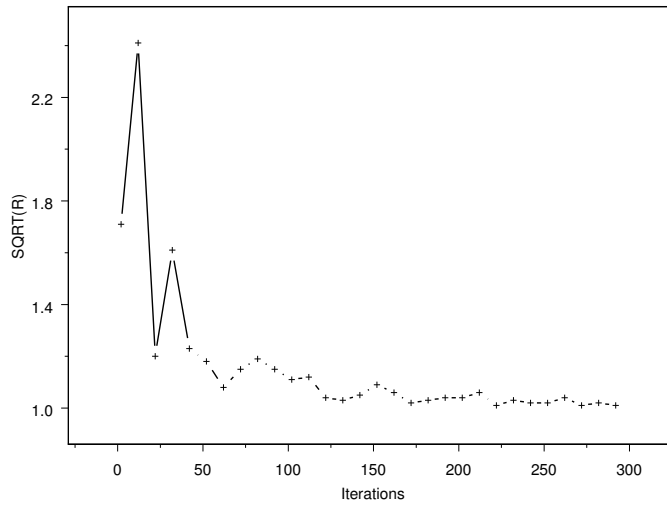
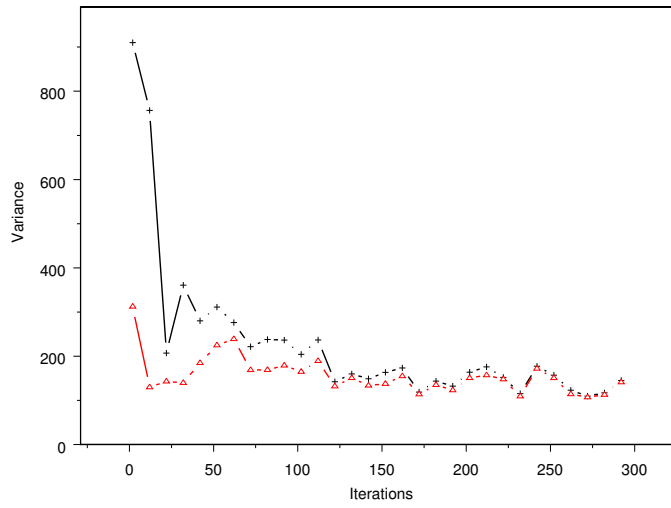
(a) $\sqrt{\hat{R}}$ (b) $\widehat{var}(Z)$ et var_I

FIG. 10.12 – Courbe du paramètre $\sqrt{\hat{R}}$ en fonction du numéro d'itérations et courbes de $\widehat{var}(Z)$ (en noir) et var_I (en rouge) également en fonction du numéro d'itérations.

10.2.2 Krigeage simple

La Figure 10.13 montre la distribution des données d'inégalité seules et de l'ensemble des données (souples et dures) pour une validation faisant intervenir le krigeage simple (KS) comme développé dans le chapitre 9. La moyenne a été fixée à $210m$, ce qui corres-

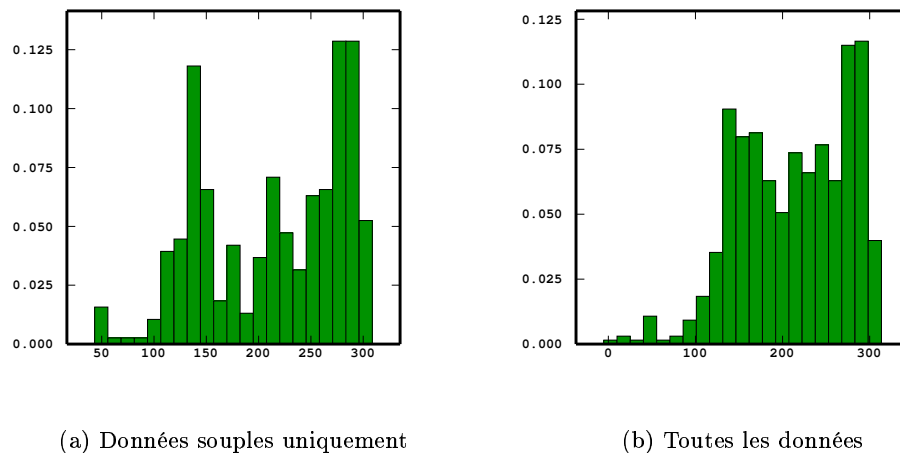


FIG. 10.13 – Histogrammes après validation des contraintes (krigeage simple $m = 210$).

pond à la moyenne des données d'égalité seules. Ces histogrammes présentent une queue pour les faibles valeurs, qui est inexistante pour les fortes. Elle correspond aux valeurs entre les failles (bloc abaissé) qui constituent une population pouvant être à la limite séparée du reste des autres points. Hormis ces valeurs faibles, on constate que les valeurs moyennes (correspondant pour une grande part à la zone centrale moins informée) sont moins fréquentes que les valeurs extrêmes. Ce déficit de valeurs autour de $150 - 250m$ crée deux pics l'un vers $130 - 140m$ et l'autre vers $275 - 290m$. En outre, ce caractère bimodal est plus accentué pour les données souples que pour l'ensemble des données. Si on compare cette distribution des données simulées aux données d'origine (Figure 10.3), on remarque que l'on a généré plus de valeurs élevées (près de 70% des données entre 230 et $300m$ sont souples ; inversement 70% des données entre 140 et $190m$ sont dures). De même, on simule plus de valeurs faibles ; en fait, on simule d'autres valeurs que les données dures mais la moyenne de toutes les données est proche de la moyenne des données d'origine. Il serait intéressant de comparer ces résultats avec ceux de techniques d'estimation de données censurées (par exemple, pour les mesures de fractures en carrière).

La principale difficulté au niveau du KS est le choix de la valeur de l'espérance m . Nous avons regardé alors ce qu'il se passait si la moyenne fournie était très différente de celle utilisée précédemment. La valeur $m = 0$ a été choisie à la place de $m = 210$, on remarque alors qu'une grande majorité des valeurs des altitudes sont quasiment identiques ; seuls quelques points sont mal estimés (Figure 10.14). Ceux-ci correspondent pour la plupart

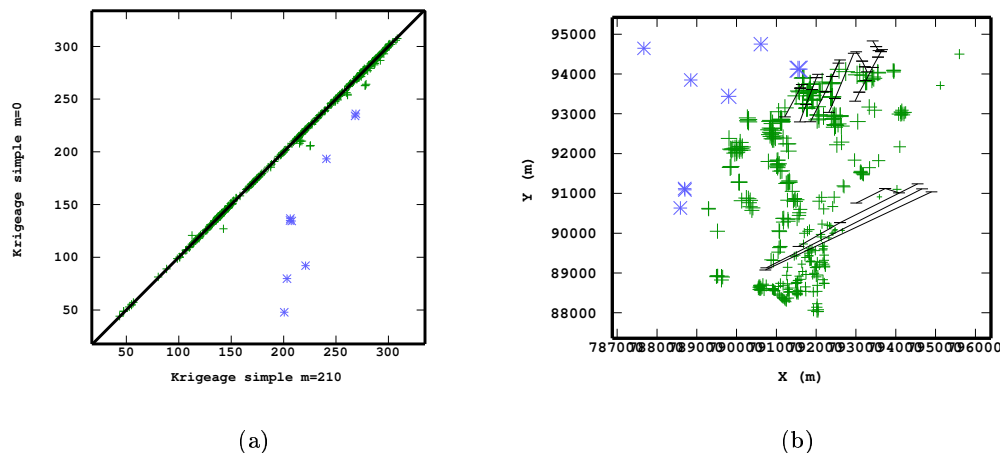


FIG. 10.14 – Nuage de corrélation entre les données totales après validation des contraintes par krigeage simple (KS) pour $m = 210$ et pour $m = 0$, à gauche et carte de localisation, à droite. En bleu les points qui s'éloignent le plus de la première bissectrice.

aux sondages isolés au nord-ouest. Si on avait effectué le krigeage des données dures sans validation des contraintes (cf. section 10.2.4), on aurait trouvé pour ces points la variance d'estimation la plus forte.

Généralement, la somme des poids de krigeage est très proche de 1, ce qui explique un résultat comparable pour $m = 0$ et $m = 210$. La densité d'information est pour une grande part responsable de ce résultat. En effet, lorsqu'on simule une donnée dans un groupe, les distances aux voisins sont très petites et le poids de la moyenne est proche de 0 (Rivoirard, 1984). Comme on est en voisinage unique, les points les plus éloignés ont une influence négligeable dans le KS. On pourrait envisager, dès le prétraitement, de regrouper les données (pour des informations ayant des pondérateurs égaux) au sein des groupes de sondages, ce qui diminuerait la taille de la matrice à inverser. On peut noter que les quelques forages dont les poids de krigeage sont les plus éloignés de 1 font partie des points s'éloignant de la première bissectrice sur la Figure 10.14.

10.2.3 Krigeage ordinaire

Le fait que les pondérateurs fournis par le krigeage simple somment habituellement à une valeur proche de 1 laisse entrevoir une bonne robustesse de l'algorithme si on remplace le KS par le krigeage ordinaire (KO), autrement dit si la moyenne n'est pas considérée connue. En effet, dans ce cas de figure la condition d'universalité impose que la somme des poids soit égale à 1.

On constate (Figure 10.15) qu'il y a très peu de différences entre les deux approches, ce qui montre la robustesse de l'algorithme même si on s'éloigne un peu du cadre théorique

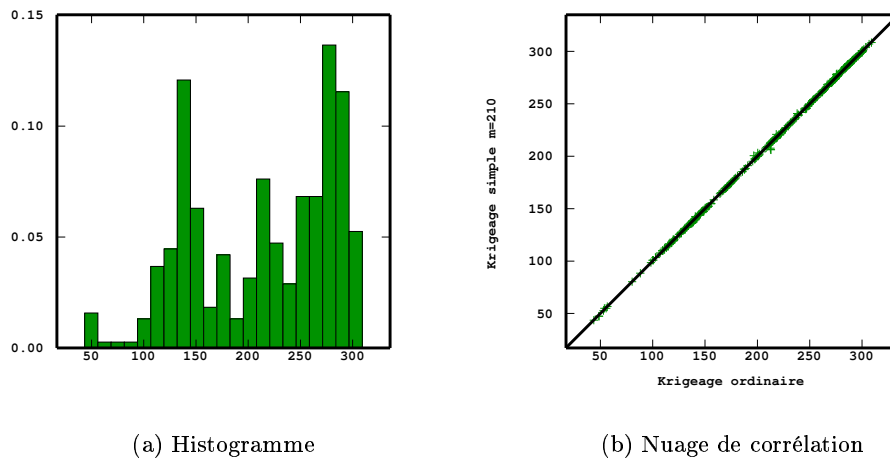


FIG. 10.15 – Histogramme des données souples pour le krigeage ordinaire et nuage de corrélation entre les données souples après validation des contraintes par krigeage simple pour $m = 210$ et par krigeage ordinaire.

strict pour lequel il a été développé.

Le Tableau 10.3 résume les statistiques élémentaires des résultats obtenus pour les 381 données souples simulées.

Approche utilisée	Minimum	Maximum	Moyenne	Ecart-type
Krigeage simple $m=210$	43.59	308.41	213.69	66.49
Krigeage simple $m=0$	44.10	307.43	211.37	67.65
Krigeage ordinaire	43.78	309.00	213.76	66.37

TAB. 10.3 – Statistiques élémentaires des données souples (en m).

La variance d'estimation qui est obtenue avec le résultat de la contrainte validée nous renseigne, comme on l'a vu précédemment, sur la facilité ou non de la convergence. La Figure 10.16 et le Tableau 10.4 montrent les résultats au bout de 1000 simulations, chaque valeur simulée étant retenue après 15 cycles (i.e. 15 balayages de toutes les données) pour le KO.

Minimum	Maximum	Moyenne	Ecart-type
1.24	3345.27	134.06	306.05

TAB. 10.4 – Statistiques élémentaires de la variance d'estimation des données souples (en m^2).

Les variances les plus élevées sont réparties au nord où la concentration des données est

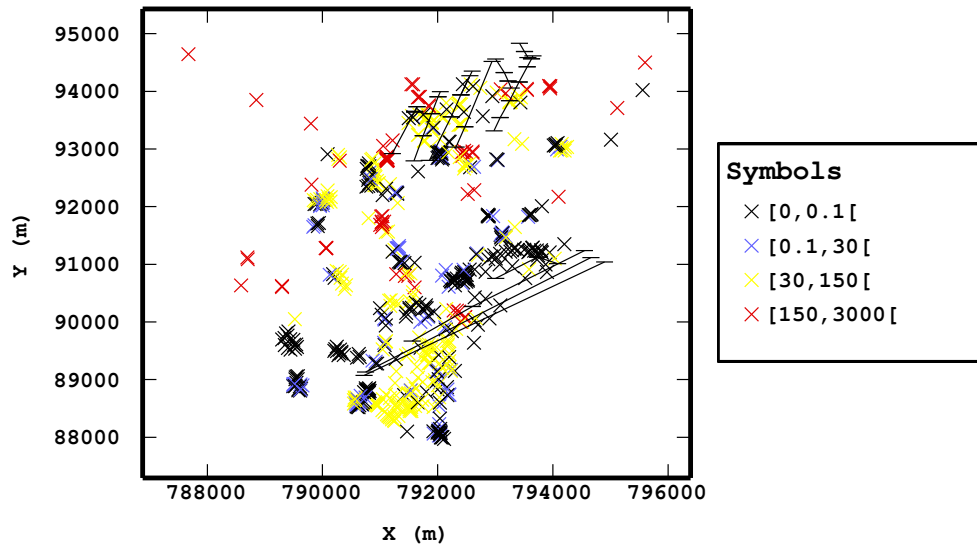


FIG. 10.16 – Carte de localisation de toutes les données et valeur des variances d'estimation. La légende fournit des intervalles pour les valeurs des variances. En noir, les données dures (variance nulle). Les trois autres couleurs concernent les données souples.

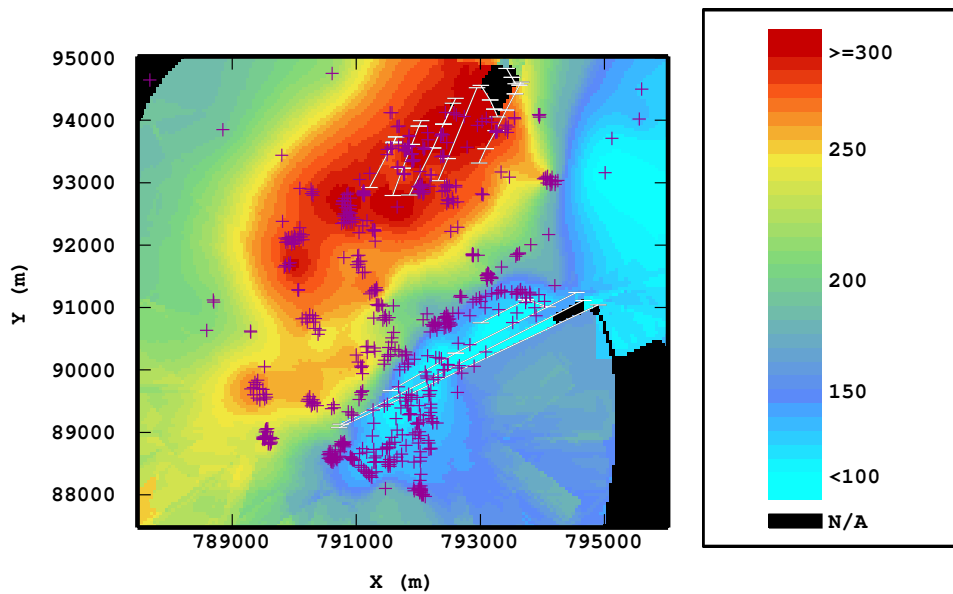


FIG. 10.17 – Carte interpolée à partir des données dures et souples (KO).

moins importante et en particulier au nord-est où ces données sont essentiellement des inégalités. En revanche, les variances les plus faibles concernent des données souples au voisinage immédiat de données dures. De plus, si on regarde le variogramme des données totales, l'augmentation des données, en particulier dans des zones moins informées, contribue à diminuer le changement de pente constaté sur le variogramme expérimental des données dures. A part cela, si on trace le variogramme des données totales, on constate qu'il a même allure que le variogramme des données dures seules.

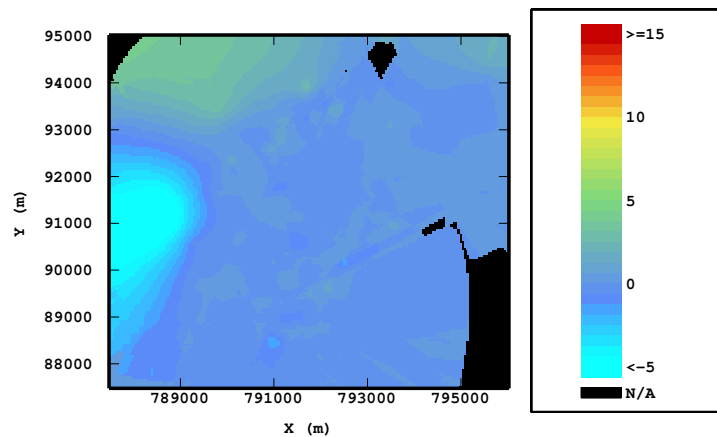


FIG. 10.18 – Carte de différence entre les cartes utilisant toutes les données, l'une avec l'approche de type KS, l'autre le KO.

Une grille (Figure 10.17 pour les données validées avec krigeage ordinaire) a été obtenue par interpolation rapide de l'ensemble des données à l'aide d'un krigeage avec un variogramme linéaire en voisinage glissant (nombre de secteurs angulaires : 10 ; nombre minimum d'échantillons : 8 ; nombre optimum d'échantillons par secteur : 12 ; rayon de l'ellipsoïde de recherche : 3000m). La topographie, ainsi que les relations de type dépôt/érosion n'ont pas été prises en compte pour la création de cette carte. La carte fait apparaître au nord une antiforme faillée et une dépression très faillée au sud. Une grille semblable à celle-ci a été créée pour les données validées par krigeage simple avec $m = 210$. Enfin une grille de différence (Figure 10.18) a été calculée à partir de ces deux grilles. On voit très bien que les différences entre les deux approches sont minimales.

10.2.4 Comparaison avec et sans validation

Nous avons vu comment prendre en compte les inégalités. Mais ce traitement apporte-t-il une amélioration substantielle ? Lorsque nous effectuons la validation des contraintes, tous les intervalles liés aux données d'inégalité sont évidemment respectés. Ils ne le sont plus forcément sans ce traitement des inégalités. Nous avons donc effectué le krigeage en voisinage unique des données souples à partir des données d'égalité seules sans tenir compte de l'information fournie par les inégalités (Figure 10.19 et Tableau 10.5). Les résultats ont été déterminés pour le KO. (Ils sont identiques pour le KS comme on l'a vu plus haut).

Validation	Minimum	Maximum	Moyenne	Ecart-type
sans	15.28	312.39	226.98	60.21
avec	43.78	309.00	213.76	66.37

TAB. 10.5 – Statistiques élémentaires des données souples sans et avec validation des contraintes (en m).

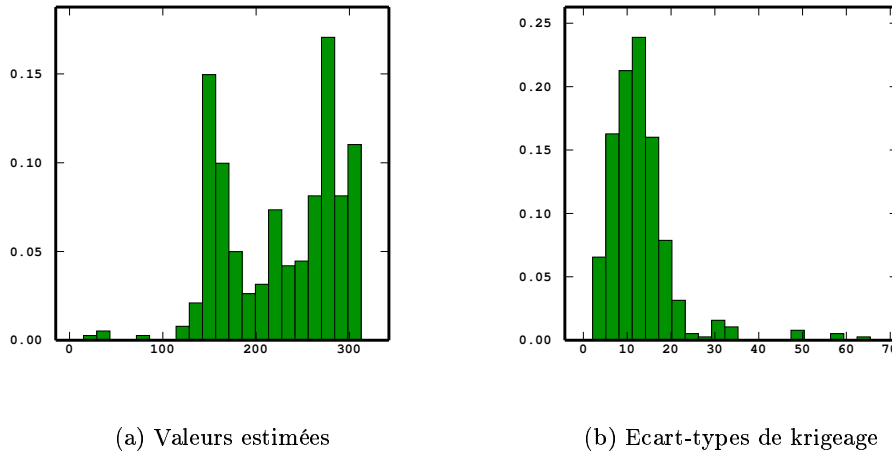


FIG. 10.19 – Histogrammes des données souples sans validation des contraintes.

Une première remarque concerne le respect des contraintes : presque $\frac{1}{3}$ des estimations (120 sur 381) ne respectent pas les intervalles. Les données non satisfaites sont surtout présentes au sud. Ensuite l'allure de l'histogramme des estimations ressemble à celui des simulations. On retrouve le caractère bimodal avec les deux pics vers 150 et 280m.

Pour mieux comparer ces résultats observons la Figure 10.20. On remarque que le nuage de corrélation entre les valeurs obtenues avec et sans validation est systématiquement décalé, de même que l'histogramme des erreurs standardisées. L'erreur standardisée est le rapport de la différence entre la valeur obtenue avec validation et la valeur sans validation par l'écart-type de krigeage sans validation. Le biais se traduit par un décalage par rapport à 0. Les valeurs sans validation sont donc surestimées (13m en moyenne). Cela est dû au fait que la quasi-totalité des contraintes sont des fins de sondages. Le sens du biais s'explique par le fait que la borne supérieure de l'intervalle impose une valeur plus faible. De plus, l'histogramme est dissymétrique, ce qui indique qu'il y a plus de valeurs estimées supérieures aux simulées que le contraire (médiane différente). De même que pour les valeurs non satisfaites, les plus grosses différences entre valeurs obtenues avec et sans validation sont concentrées au sud de l'étude, en particulier autour des failles (blocs légèrement basculés, flexures ?) et aussi au nord-ouest où l'estimation entraîne une extrapolation

des valeurs d'altitude.

Dans certaines zones sur une courte distance, la variabilité est plutôt forte. Regardons ce qui se passe par exemple dans un carré de $150m \times 250m$ de côté au nord-est de l'étude (vers $X=794000$, $Y=93000$). On dispose de 5 données exactes (moyenne $\approx 238m$) au nord-ouest de ce carré, les 10 données souples étant regroupées au sud-ouest. Sans validation la moyenne est de $222m$, avec validation la moyenne tombe à $205m$. Ce sont les bornes des intervalles qui imposent des valeurs plus faibles : par exemple en un des points, on a $z < 218m$. Cette zone correspond à la pente d'un talus d'érosion qui révèle peut-être l'existence d'une faille non reconnue en surface.

Pour visualiser l'apport des données souples, calculons comme précédemment une grille, mais en ne considérant cette fois-ci que les données dures (Figure 10.21). La Figure 10.22 montre la carte de différences entre le cas sans (Figure 10.21) et avec validation des contraintes (Figure 10.17). On constate que l'on a une meilleure image de l'ensellement avec

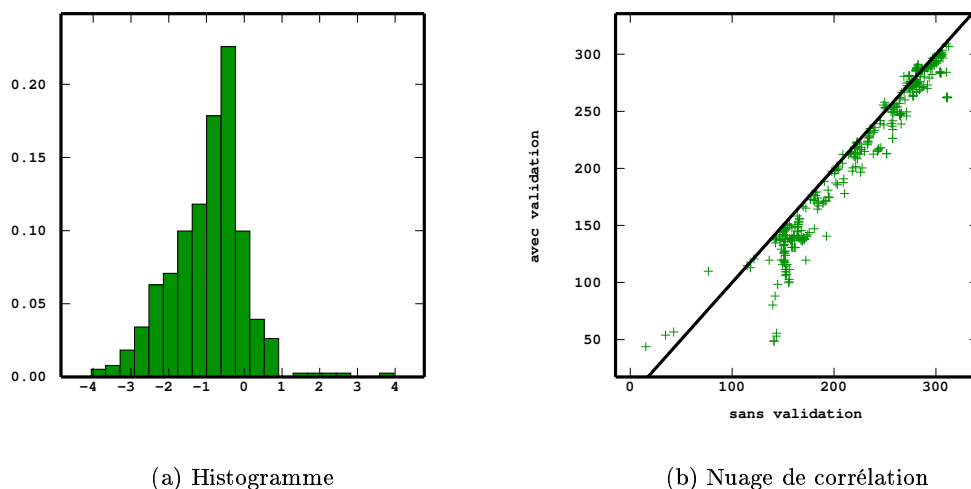


FIG. 10.20 – Histogramme des erreurs standardisées, à gauche et nuage de corrélation entre les valeurs obtenues avec et sans validation, à droite.

la validation, l'utilisation des données supplémentaires au nord-ouest permettant d'éviter l'extrapolation. Les failles au sud apparaissent très nettement. Les surfaces en rouge correspondent aux zones où les valeurs obtenues sans les données souples sont très surestimées par rapport aux valeurs obtenues avec toutes les données disponibles.

10.2.5 Fonctions de dérive et de failles

Nous avons vu au paragraphe 10.1.2 que nos données sont affectées par une dérive linéaire (modérée) et par des failles. Aussi est-il intéressant d'examiner la robustesse de l'algorithme de validation des contraintes lorsqu'on l'applique avec un krigeage universel

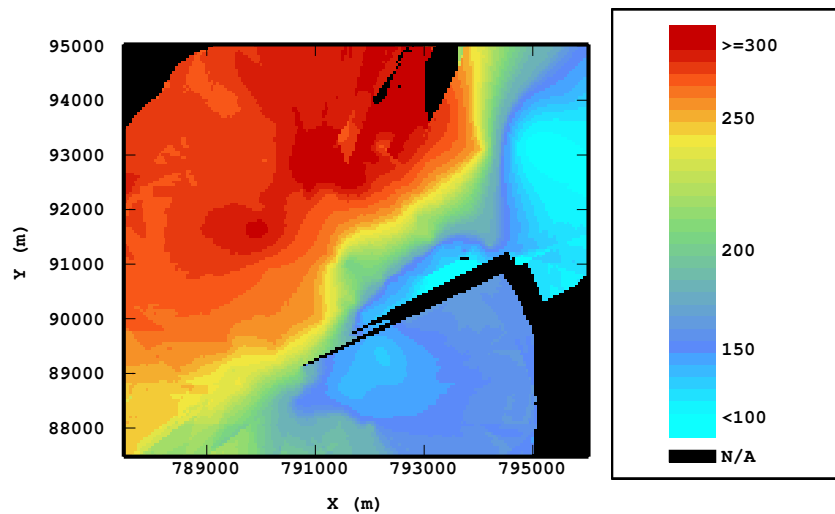


FIG. 10.21 – Carte interpolée à partir des seules données dures.

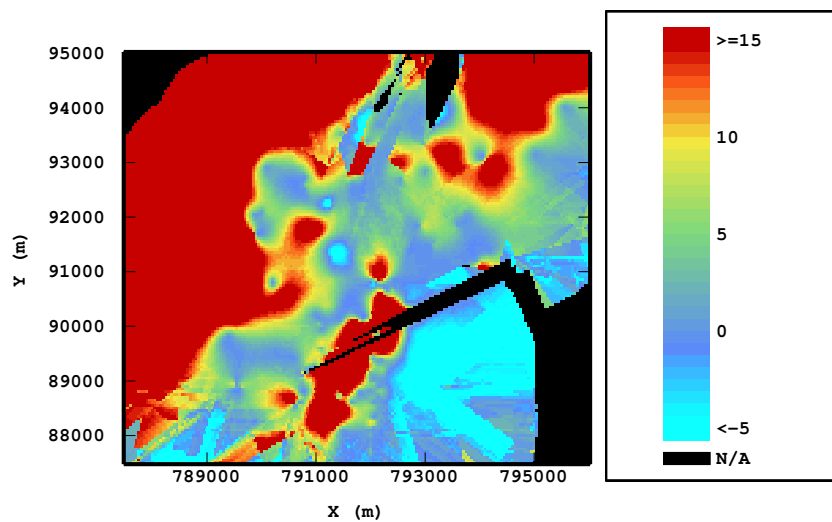


FIG. 10.22 – Carte des écarts entre les altitudes estimées sans validation (Figure 10.21) et avec validation (Figure 10.17).

(KU) en introduisant des fonctions de dérive linéaire ou des fonctions de failles. Commentons par un krigeage universel avec dérive linéaire. Les résultats obtenus sont peu différents

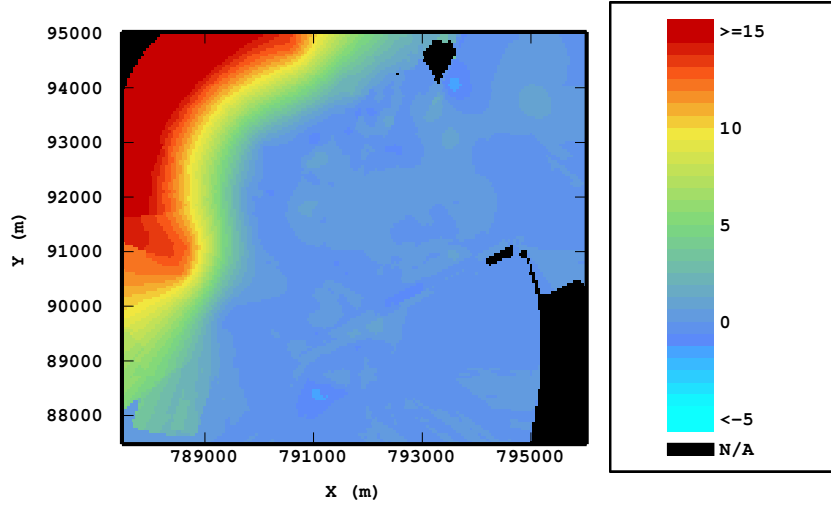


FIG. 10.23 – Carte de différence entre les cartes utilisant toutes les données, l'une avec l'approche de type KU et dérive linéaire, l'autre avec le KO.

de la carte obtenue en recourant au KO, comme le montre la Figure 10.23 qui présente l'écart entre ces deux cartes. Les plus grosses différences sont constatées dans la partie nord-ouest où l'estimation est plus difficile à réaliser vu le peu de données localisées à cet endroit. Cela montre la robustesse de l'algorithme, mais aussi l'apport limité de son extension lorsque comme ici la dérive est peu marquée. Par contre la prise en compte des failles par l'ajout de fonctions de rejet s'est révélée plus fructueuse. Les valeurs entre les failles au sud sont nettement inférieures aux autres valeurs, la simulation doit donc tenir compte de ces informations. En particulier, pour l'espace entre les deux failles les plus au sud (F1 sur la Figure 10.24) ainsi que la partie est de celui qui forme un coude un peu plus au nord (F3). Ces deux zones ont les altitudes les plus basses du champ. Les failles sont supposées verticales. On considère la fonction rejet comme l'intersection d'un plan et d'une fonction de Heaviside H_L définie uniquement le long d'un segment L (Maréchal, 1984) :

$$f(x, y) = a_0 H_L(x, y) + a_1 x H_L(x, y) + a_2 y H_L(x, y)$$

Les coefficients a_0 , a_1 et a_2 étant inconnus, on introduit les trois fonctions de base H_L , xH_L et yH_L . Il est possible aussi d'introduire des fonctions avec amortissement comme dans le cas de l'*Editeur* à la Figure 3.9. On a donc considéré quatre segments, donc trois espaces interfailles F_1 , F_2 et F_3 qui contiennent 21 données. En plus, des fonctions de base 1, x , y globales, on utilise trois fonctions supplémentaires pour chaque domaine F_i du type H_F , xH_F , yH_F où H_F est l'indicatrice du domaine F . Ce qui fait 12 fonctions de base en tout.

Regardons le nuage de corrélation à la Figure 10.25 entre les cas sans et avec fonctions de rejet. Les valeurs les plus différentes sont les valeurs les plus faibles au sud entre ou

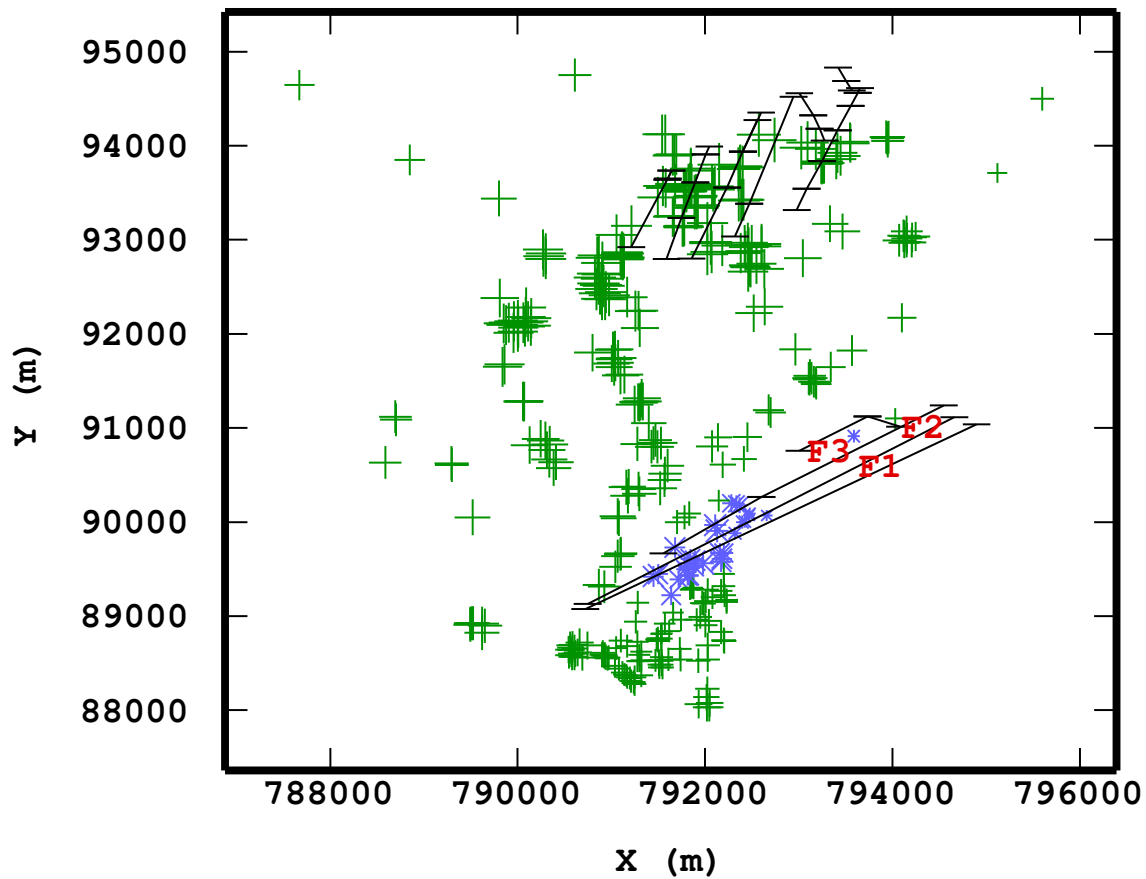


FIG. 10.24 – Carte de localisation des zones de failles. En bleu, les points qui s'éloignent le plus de la première bissectrice à la Figure 10.25.

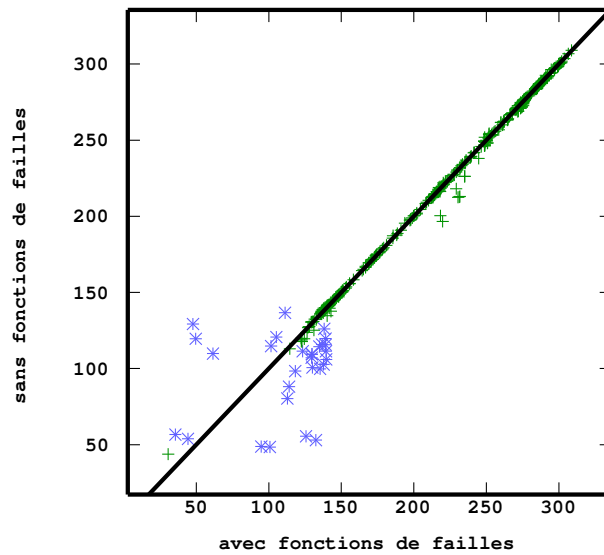


FIG. 10.25 – Nuage de corrélation entre une validation des données d'inégalité sans et avec fonctions de failles. En bleu, les points qui s'éloignent le plus de la première bissectrice.

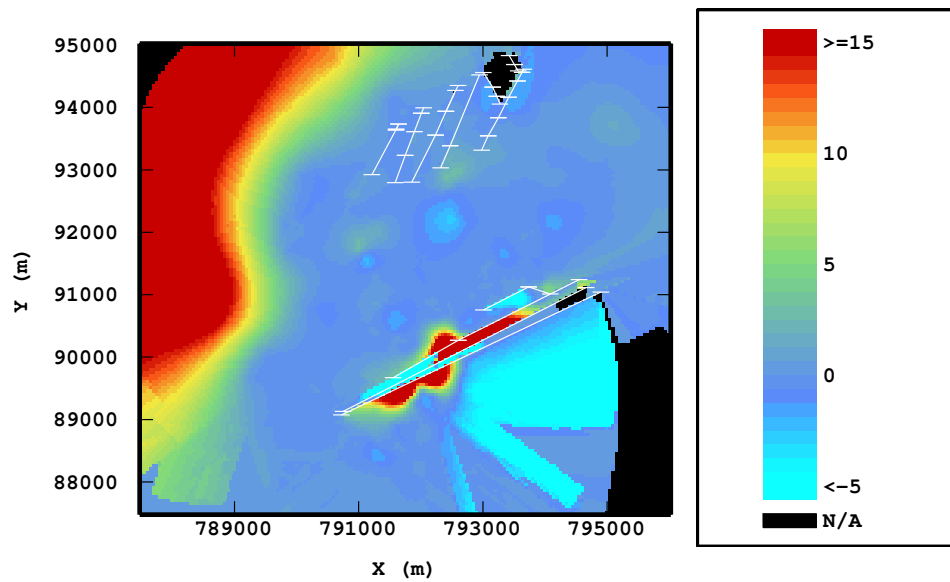
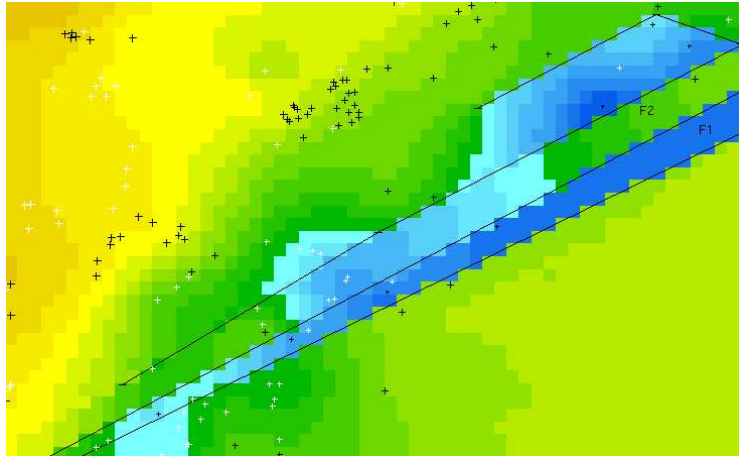


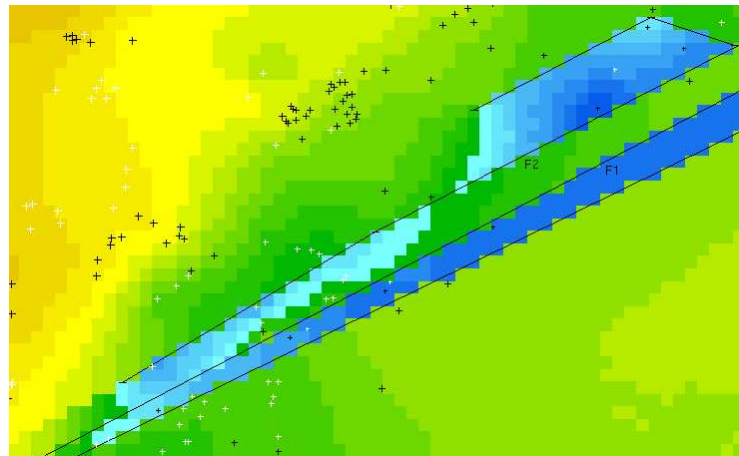
FIG. 10.26 – Carte de différence entre les cartes utilisant toutes les données, l'une sans les fonctions de failles, l'autre avec.

à proximité des failles (Figure 10.24). L'image que l'on a de la zone faillée s'en trouve globalement modifiée, comme on peut le voir sur la carte de différence entre l'utilisation ou non de ces fonctions de failles (Figure 10.26).

Intéressons-nous maintenant à des zones inter-failles en particulier. Par exemple, prenons l'espace F2 dans la région faillée au sud (correspondant à la zone entre les deux failles les plus au nord à la Figure 10.27). La moyenne des 7 données simulées dans cette zone est



(a) sans fonctions de failles



(b) avec fonctions de failles

FIG. 10.27 – Carte interpolée utilisant toutes les données (zoom sur les failles au sud), (a) sans les fonctions de failles, (b) avec les fonctions de failles. Les croix noires correspondent aux données dures et les blanches aux données souples.

de $84m$ sans les fonctions de failles et la variance est forte. En utilisant les fonctions de failles, la moyenne monte à $122m$ et la variance est beaucoup plus faible. Au contraire, si on regarde ce qu'il se passe dans l'espace F1 (l'espace inter-faille au-dessous), la moyenne des 5 données d'inégalité situées dans cet espace est de $89m$ sans les fonctions de failles. Lorsqu'on utilise les fonctions de failles, la moyenne chute à $67m$. En résumé, sans les fonctions de failles, la moyenne des données souples dans les deux espaces inter-failles est de l'ordre de $85 - 90m$ dans chaque bloc. On ne distingue donc pas le bloc effondré du bloc surélevé. Par contre, grâce aux fonctions de failles, les valeurs simulées se calent sur les valeurs dures situées dans le compartiment. Enfin, on peut noter que la prise en compte de l'anisotropie, de la dérive, des failles entraîne à chaque fois une diminution de la moyenne de la variance des simulations.

10.2.6 Validation croisée

Une approche de type validation croisée a été entreprise afin de tester concrètement les résultats de la simulation et de mesurer d'une certaine manière la confiance accordée à cette méthode. Le principe consiste à créer à partir des seules données exactes un nouveau jeu de données dures et souples. Il s'agit ensuite de comparer les résultats simulés aux valeurs connues.

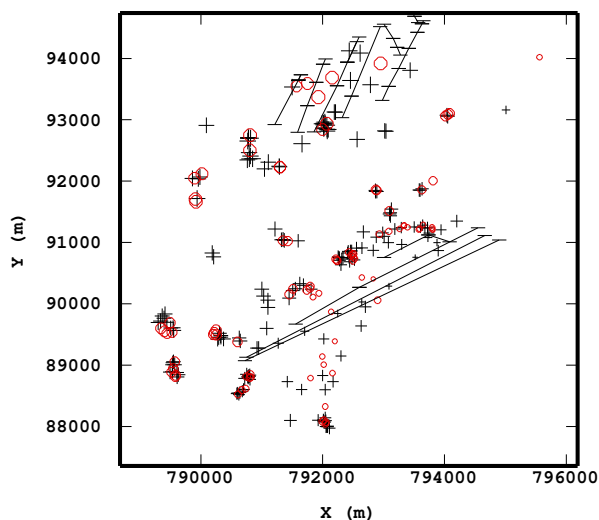
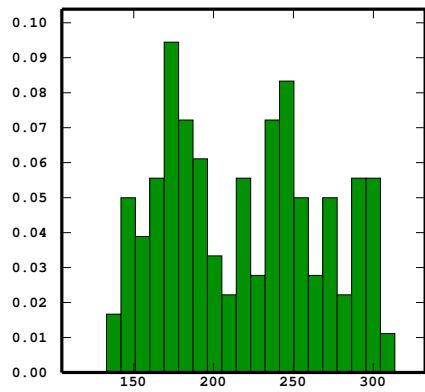


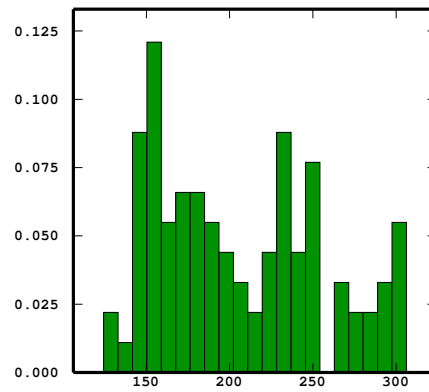
FIG. 10.28 – Validation croisée : carte de localisation des données - Croix noires : données d'égalité ; Cercles rouges : données d'inégalité.

Sur les 271 sondages atteignant l'objectif, 180 sont restés des données dures et 91 ont été transformés en données d'intervalle (Figure 10.28). On qualifiera de "souples réelles" les valeurs avant transformation en données d'intervalle et de "souples simulées" les valeurs reconstituées par la validation des contraintes via un krigeage ordinaire.

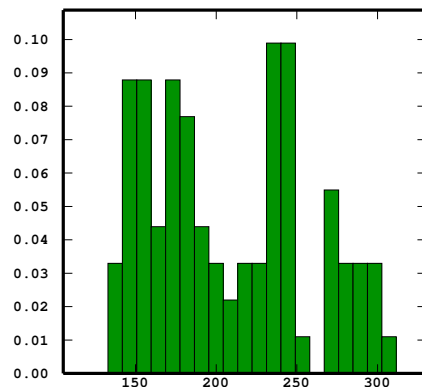
Le Tableau 10.6 présente les principales statistiques et la Figure 10.29(a) montre l'his-



(a) Données dures

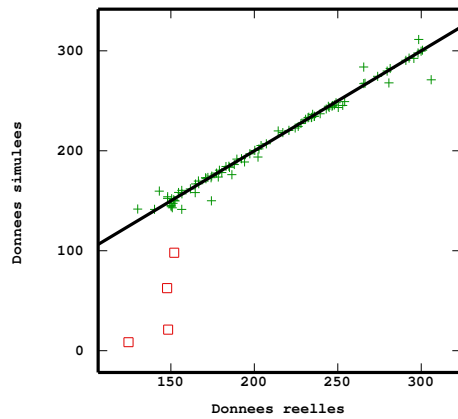


(b) Données souples réelles

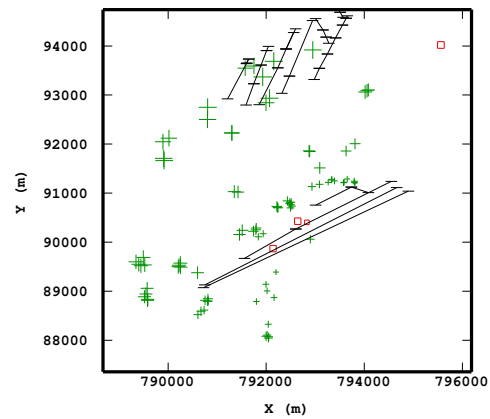


(c) Données souples simulées

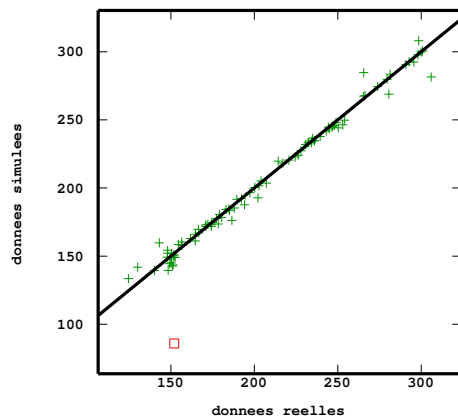
FIG. 10.29 – Validation croisée : histogrammes des données dures et souples.



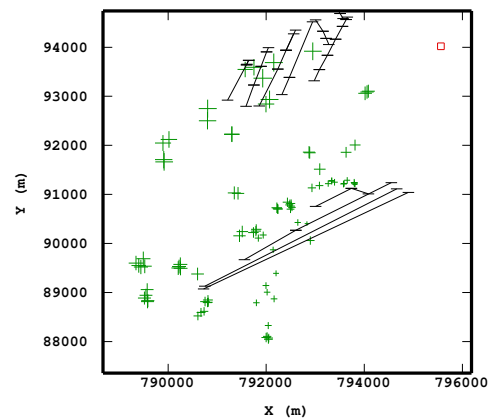
(a) Sans fonctions de failles - Nuage de corrélation



(b) Sans fonctions de failles - Plan de position



(c) Avec fonctions de failles - Nuage de corrélation



(d) Avec fonctions de failles - Plan de position

FIG. 10.30 – Validation croisée : cartes de localisation et nuages de corrélation entre les données simulées et réelles sans fonctions de failles (en haut) et après ajout de fonctions de failles, en bas.

Données	Minimum	Maximum	Moyenne	Ecart-type
Hard	-4.88	313.17	212.01	59.65
Soft réelles	124.58	306.00	206.32	49.35
Soft simulées (KO)	8.56	311.37	201.15	57.98

TAB. 10.6 – Statistiques élémentaires pour la validation croisée (en m)

togramme des 180 données dures restantes. De même, les Figures 10.29(b) et 10.29(c) comparent les histogrammes des 91 données souples réelles et simulées. Les valeurs inférieures à 120 m ne sont pas visibles (au nombre de 4 pour les données souples simulées). Si on regarde les données de départ (dures et souples réelles), on constate que les valeurs des données souples sont en moyenne plus faibles que celles des données dures, mais aussi moins dispersées. En particulier, il n'y a pas, contrairement aux données exactes, de valeurs inférieures à 120 m. Après simulation, la sous-estimation des altitudes est de 5 m en moyenne (on simule des valeurs faibles aux alentours des données dures proches des failles du sud). On remarque également que la distribution des données simulées est très proche de la distribution des données réelles, ce qui est très encourageant.

Regardons à présent les nuages de corrélation entre données souples réelles et simulées. Comme on le voit sur les Figures 10.30(a) et 10.30(b) (sans utilisation de fonctions de failles), seuls 4 points de valeurs faibles s'éloignent de la première bissectrice (carrés rouges). L'un est un point isolé, les autres sont situés dans des blocs entre les failles du sud. Or ces derniers ne sont pas situés dans les blocs les plus bas (cf. Section 10.2.5), mais dans le deuxième espace interfaille à partir du sud (espace F2 sur la Figure 10.27) et dans la partie ouest du bloc formant un coude (F3). Ceci explique pourquoi les valeurs réelles sont plus élevées. L'introduction des fonctions de failles ramène les valeurs les plus atypiques concernées par les failles vers la première bissectrice. Seule la valeur du point isolé n'a pas été corrigée (Figures 10.30(c) et 10.30(d)).

10.3 Méthode des potentiels et données d'inégalité : étude sur des données synthétiques

Nous allons appliquer l'algorithme de validation des contraintes adapté à la méthode des potentiels et observer son comportement sur des exemples assez simples.

Considérons un champ de taille $1000m \times 1000m \times 1000m$. Nous travaillerons essentiellement sur des coupes verticales.

10.3.1 Un forage, une inégalité

On utilise tout d'abord une pile stratigraphique contenant uniquement deux couches géologiques distinctes : B surmontant A. L'objectif est de dessiner la trace de l'horizon dans la coupe. On se donne 3 données d'interface (données dures, autrement dit la différence de

potentiel entre deux points est nulle) plus ou moins alignées dans la section, et une donnée de foliation. On place un forage de $300m$ de profondeur, situé à $300m$ du bord de la section, afin de “perturber” l’estimation. Le sondage est constitué d’une seule passe afin d’être dans le cas d’une donnée souple. Si la passe est composée uniquement de B (en bleu sur la Figure 10.31), la différence de potentiel entre le point situé à la fin du forage et un point de l’interface est positive, en considérant que l’ordre de succession des couches n’est pas inversé par rapport à la pile. Inversement, si la passe est composée seulement de A (en rouge sur la Figure 10.31), la différence de potentiel entre le point situé au début du forage et un point de l’interface avant érosion est négative. On simule donc la différence de potentiel

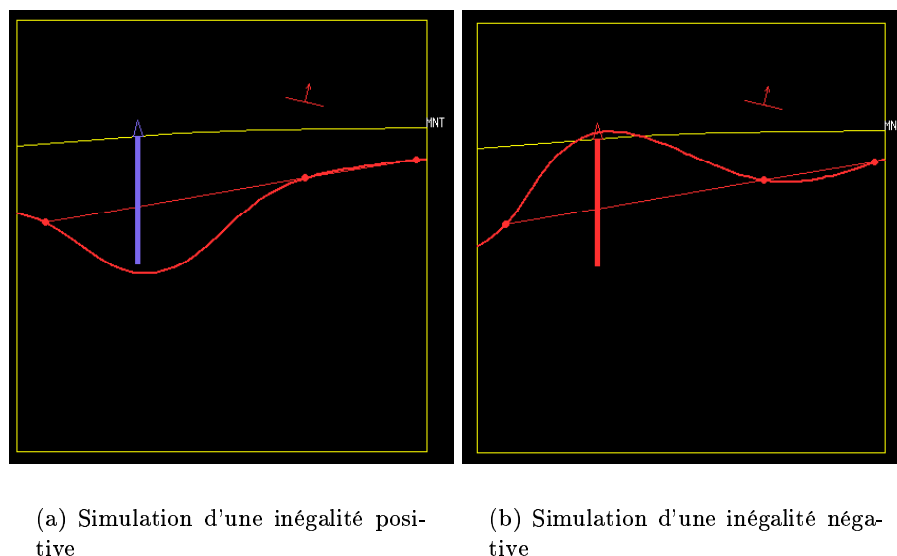


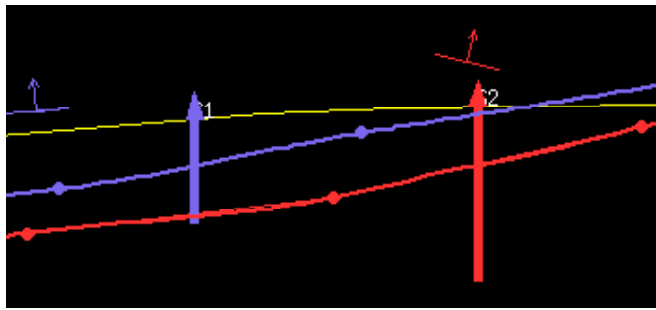
FIG. 10.31 – Tracés de l’interface entre deux couches géologiques.

entre le point situé au début ou à la fin du sondage et un point sur l’interface. La valeur ainsi calculée constitue une donnée supplémentaire du champ de potentiel, même si elle n’appartient pas à l’interface considérée (différence de potentiel non nulle avec un point situé sur l’interface). On interpole alors le potentiel et on trace la surface isopotentielle correspondant à l’interface.

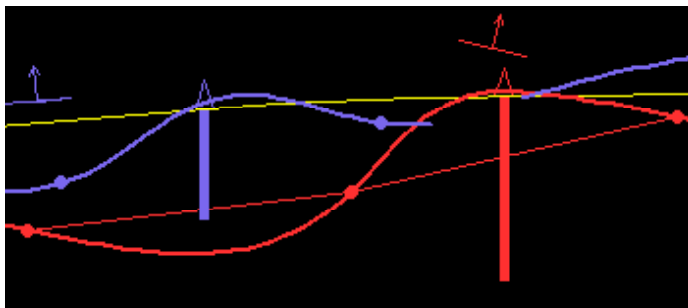
Le variogramme utilisé est un schéma cubique de portée 200 m, de palier unité et avec des effets de pépite sur les gradients et sur les incréments de respectivement 0.01 et 10^{-7} . Une dérive linéaire est également ajoutée. On peut voir sur la Figure 10.31 que les inégalités sont parfaitement respectées, obligeant le tracé à éviter la traversée du sondage.

10.3.2 Deux sondages, deux inégalités

La Figure 10.32 est un cas un peu plus complexe avec une interface rouge (R) ancienne et érodante et le toit plus récent bleu (B) d’une formation qui s’y est déposée. Deux



(a) Sans validation des contraintes



(b) Avec validation des contraintes

FIG. 10.32 – Détermination de 2 interfaces géologiques.

forages, traversant chacun un seul faciès, fournissent des données d'inégalité. Avant prise en compte des données d'inégalité, les données dures sont respectées, mais pas les données souples des deux forages. Elles le sont après simulation de différences de potentiel positive (R) et négative (B) pour le premier sondage et négative (R) pour le second.

10.3.3 Le cas de lacunes

Un dernier exemple présente le cas de figure des lacunes qui peuvent être présentes dans les sondages.

La pile stratigraphique est la suivante du haut vers le bas :

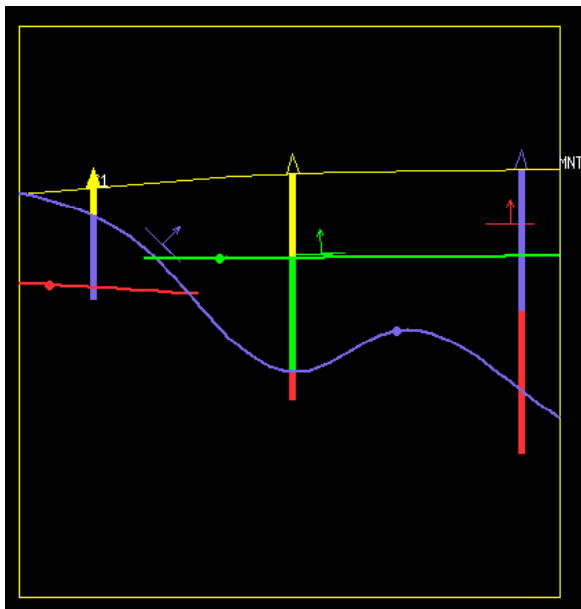
- couche jaune J
- couche verte V en dépôt
- couche bleue B érodante
- couche rouge R en dépôt

L'objectif est de tracer les toits de V, B et R, connaissant une donnée d'orientation et un point sur chaque interface. Trois sondages existent et apportent des informations sûres (4

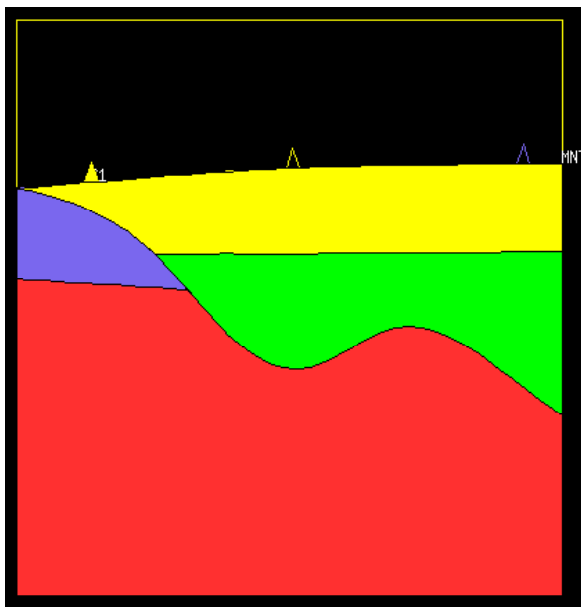
points supplémentaires) et des données d'inégalités (en particulier une contrainte négative pour R sur le premier sondage et une contrainte positive pour B sur le troisième). Des lacunes imposent également des contraintes, car il n'y a pas de dépôt de B dans le deuxième sondage, ni de V dans les premier et troisième sondages.

L'absence de prise en compte des inégalités (Figure 10.33) montre que l'interface R est trop haut et traverse B au sondage 1. De même, l'interface R est trop bas et traverse R dans le sondage 3. Ce qui provoque par conséquent un non respect du hiatus de V au sondage 3. La Figure 10.34 est le résultat du tracé avec prise en compte des contraintes d'inégalité. Le remplissage des couches montre bien l'apport de l'information supplémentaire sur le modèle géologique. Les dépôts de V et J sont maintenant circonscrits à la partie centrale de la coupe et B existe désormais à droite.

Les résultats obtenus sur données synthétiques sont donc encourageants pour un passage à des applications réelles.

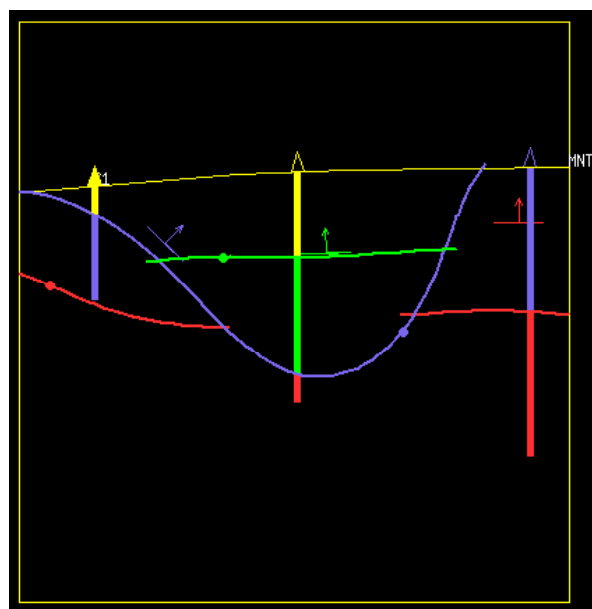


(a) Tracé des interfaces

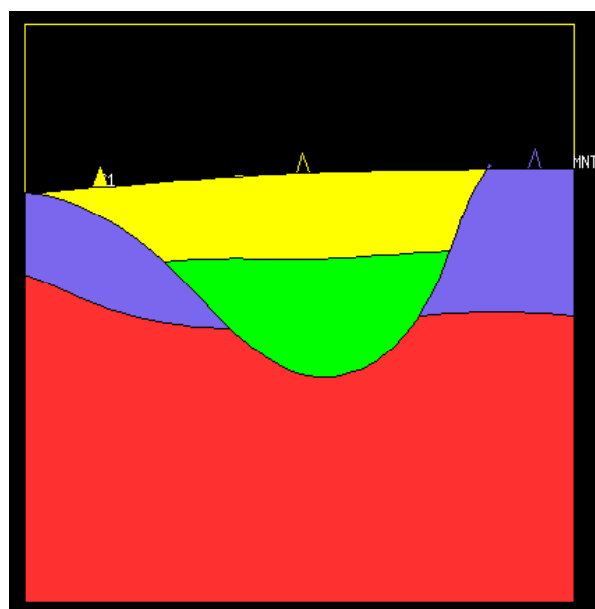


(b) Remplissage des couches

FIG. 10.33 – Détermination des interfaces entre 4 couches géologiques sans validation des contraintes.



(a) Tracé des interfaces



(b) Remplissage des couches

FIG. 10.34 – Détermination des interfaces entre 4 couches géologiques avec validation des contraintes.

Chapitre 11

Conclusions et perspectives

11.1 Un outil de modélisation géologique 3D : l'*Editeur Géologique*

Le travail présenté dans ce mémoire se place dans le cadre de la modélisation tridimensionnelle des corps géologiques, en combinant surfaces implicites et méthodes géostatistiques. Depuis les premiers travaux consacrés à la méthode des potentiels il y a de près de dix ans, un outil développé au BRGM, l'*Editeur Géologique*, a vu le jour et est opérationnel. Il permet de construire des modèles géologiques réalistes, même dans des situations complexes (Maxelon (2004), Guillen *et al.* (2004)). Pour ce faire, il intègre différents types de données : carte géologique, forages, données structurales, failles, en s'assurant de la cohérence géométrique. Les interfaces géologiques sont construites, pour résumer, par un cokrigage à partir de points de passage et de données d'orientation.

L'*Editeur* continue d'évoluer, la recherche se poursuit et cette thèse en est une illustration.

11.2 Les apports principaux de la thèse

Le progrès majeur apporté par cette thèse est la caractérisation de la structure spatiale du champ de potentiel grâce aux données d'orientation et les possibilités qui en découlent :

- (i) la quantification de l'incertitude qui affecte le modèle géologique construit ;
- (ii) la prise en compte des données d'inégalité.

Dans une approche multicouche, la détermination des incertitudes affectant les modèles construits possède une solution connue, puisqu'elle relève des méthodes géostatistiques multivariées classiques 2D.

Par contre, pour la méthode des champs de potentiel, il a fallu au préalable caractériser le modèle de covariance qui était jusqu'à présent choisi de manière heuristique. Ce modèle est maintenant décrit à l'aide des seules données exploitables pour le calcul des variogrammes expérimentaux : les données structurales. Comme celles-ci sont assimilées à des gradients du potentiel et comme la covariance du potentiel est liée à la covariance de son gradient par une dérivée seconde, les données d'orientation permettent en effet de

remonter à la structure du potentiel. Le modèle de covariance du potentiel choisi est une structure cubique gigogne reprenant les structures dérivées des composantes verticales et horizontales du gradient.

Les quatre jeux de données réelles étudiés afin de tester la méthode ont ceci en commun que la composante du gradient suivant z a une variance beaucoup plus faible que les autres composantes et que toutes ont un variogramme bien structuré malgré un effet de pépite relativement important. Le modèle obtenu est généralement très différent du modèle “conventionnel” utilisé auparavant, ce qui a un impact non négligeable sur la qualité de l’interpolation.

Une procédure automatique d’ajustement des variogrammes a été mise en place afin de déterminer les paramètres du modèle, l’optique de ce travail étant toujours d’intégrer à plus ou moins long terme les applications au sein du logiciel. Ces dernières se doivent donc d’être utilisables par les géologues avec un minimum de connaissances géostatistiques.

Les variances de cokrigeage ayant maintenant un sens objectif, il a été possible de cartographier le potentiel réduit, variable qui mesure la probabilité d’appartenance d’un point à une interface. Ces cartes qui permettent de localiser les zones d’incertitudes élevées vont être très utiles pour les utilisateurs des modèles fournis. Une variance d’estimation plus classique, dont la dimension est une distance, peut aussi être obtenue en combinant variance de cokrigeage du potentiel et estimation du gradient vertical.

L’ajustement d’un variogramme comporte aussi une part d’incertitude qu’il est possible de quantifier à l’aide d’une approche bayésienne. Nous nous sommes attachés à caractériser l’incertitude relative aux paramètres du modèle de covariance : portée, palier, effet de pépite, dérive. La distribution *a posteriori* calculée lors de cette étude peut dans un deuxième temps être incorporée au processus de cokrigeage.

Les données structurales sont des mesures d’orientation, mais ne sont pas assorties d’une mesure d’intensité. La méthode des potentiels les considère toutes comme des vecteurs unitaires, ce qui est une approximation. Des simulations sur un exemple donné ont montré l’impact, sur les variogrammes, de l’usage des gradients normés à la place des gradients non-normés. Seul le palier du variogramme du gradient vertical est affecté de manière significative.

Enfin, la connaissance de la covariance du potentiel a permis de résoudre un problème récurrent posé aux géologues : le non-respect des fins de sondage lors de l’interpolation. Le traitement des données incomplètes a ainsi été opéré au sein de l’*Editeur* en mettant en œuvre une procédure de simulation de vecteurs gaussiens tronqués. La première approche envisagée a été de déterminer l’altitude d’une interface séparant deux formations géologiques en 2D, l’altitude étant la variable d’intérêt. La procédure s’étend facilement au cas multicouche. Les résultats obtenus sur un cas réel ont montré l’intérêt de cette démarche afin d’offrir une vision plus précise du sous-sol en utilisant de l’information non-exploitée.

Un des résultats importants de cette étude est la robustesse de l'algorithme, lorsque l'on remplace le krigeage simple par un krigeage ordinaire, universel ou avec l'emploi de fonctions de failles. L'étape suivante a consisté à appliquer cette procédure de validation des contraintes à des données de potentiel dans une optique 3D et pour plusieurs horizons. Les résultats sont réalistes pour des données synthétiques, mais restent à finaliser pour des données réelles.

11.3 Perspectives

Quelques points examinés lors de cette thèse méritent d'être précisés. L'impact de l'utilisation de vecteurs normés à la place des gradients réels, non plus sur les variogrammes, mais sur le cokrigeage, est envisagé. De même, l'étude d'un cas réel concernant le traitement des inégalités dans la méthode des potentiels est prévue.

D'autres fonctionnalités ont été ajoutées par l'équipe "Modélisation 3D" du BRGM. Ainsi, il est désormais possible, entre autres, d'exploiter des données structurales non orientées telles que les axes de pli.

Cependant nombre de nouveaux développements restent à réaliser :

Lors de l'interpolation du champ de potentiel, le cokrigeage s'effectue en voisinage unique, ce qui impose des tailles de matrices très grandes quand le nombre de données est important. Un voisinage glissant permettrait un gain de temps non négligeable, mais il faudra néanmoins être très attentifs au niveau des recouvrements de ces voisinages afin d'éviter tout artefact. On peut imaginer diverses solutions, comme un double cokrigeage et des cokrigeages sur de grands sous-domaines avec recouvrement partiel.

Une alternative complémentaire à l'estimation par cokrigeage est envisageable, comme les simulations conditionnelles. On obtient par ce biais différents modèles, moins lisses, mais tous aussi représentatifs de la variabilité de détail de la réalité sous-jacente inconnue, ce qui permet de tester différents *scenarii*.

Un autre point essentiel est l'amélioration de la gestion des discontinuités qui est en cours d'étude, car les réseaux complexes de failles sont mal pris en compte avec nos modèles simplistes.

De nombreuses informations, également sous forme d'inégalités, ne sont pas exploitées au niveau de la carte géologique. Une approche de type validation des contraintes est dans ce cas également à mettre en œuvre.

L'*Editeur Géologique* comporte en fait deux outils distincts : le modelleur géologique étudié ici, et un outil d'inversion géophysique (données géomagnétiques ou gravimétriques).

Actuellement, le modèle issu du premier outil est pris comme solution initiale du second. Une interaction entre le modeleur et le module d'inversion apporterait sans doute une amélioration significative. D'une manière plus générale, l'intégration d'autres données géophysiques (radar, RMN, sismique par exemple) reste à réaliser.

Annexes

Annexe A

Matrice de covariance utilisée dans la méthode du potentiel

Nous allons expliciter dans le détail l'expression de la matrice de covariance servant à l'interpolation du champ de potentiel. Définissons tout d'abord les paramètres utilisés. On considère p interfaces géologiques et chaque interface I_k $1 \leq k \leq p$ possède n_k points. Comme un point par interface est utilisé comme référence pour les incréments, le nombre total de données d'incrément est : $n = \sum_{k=1}^p (n_k - 1)$. L'indice servant aux données d'incrément est j . On a de plus m données d'orientation qui sont des vecteurs à trois composantes ; l'indice utilisé est i . La variable aléatoire potentiel sera notée Z . La matrice de covariance utilisée pour le cokrigage des données d'interface et de gradient est une matrice-bloc de la forme :

$$\mathcal{C} = \begin{pmatrix} C_G & {}^t C_{GI} & {}^t U_G & {}^t F_G \\ C_{GI} & C_I & {}^t U_I & {}^t F_I \\ U_G & U_I & 0 & 0 \\ F_G & F_I & 0 & 0 \end{pmatrix}$$

avec

C_G la matrice de covariance sur les gradients (dérivées partielles du potentiel) ;

C_I la matrice de covariance sur les interfaces (incrément de potentiel) ;

C_{GI} la matrice de covariance croisée gradients / interfaces ;

U_G et U_I les matrices de conditions d'universalité sur les gradients et les interfaces ;

F_G et F_I les matrices de failles sur les gradients et les interfaces ;

A.1 Les gradients

La matrice de covariance des données de gradient est de taille $(3 * m)^2$ et de la forme :

$$C_G = \begin{pmatrix} C_G^{xx} & C_G^{xy} & C_G^{xz} \\ C_G^{yx} & C_G^{yy} & C_G^{yz} \\ C_G^{zx} & C_G^{zy} & C_G^{zz} \end{pmatrix}$$

Le vecteur reliant les deux points $\mathbf{u}$ et $\mathbf{v}$ est $\mathbf{h} = \mathbf{v} - \mathbf{u}$. Soit r son module et h_x, h_y, h_z ses composantes. Pour une covariance isotrope du potentiel de variance σ_Z^2 et de portée a , on a par exemple :

$$C_G^{xy} = Cov \left[\frac{\partial Z}{\partial x}(\mathbf{u}), \frac{\partial Z}{\partial y}(\mathbf{v}) \right] = \frac{\sigma_Z^2}{a^2} \frac{h_x h_y}{r^2} \left(\frac{1}{r} C'_Z(r) - C''_Z(r) \right)$$

Un effet de pépite est rajouté sur les éléments diagonaux.

A.2 Les incréments de potentiel

On travaille sur les incréments de $Z_\alpha^k - Z_1^k$, où k est l'indice de l'interface, 1 la donnée de référence (arbitrairement la première rencontrée) et α ($2 \leq \alpha \leq n_k$) les autres données. La matrice de covariance des données d'interface est de taille n^2 et de la forme :

$$C_I = \begin{pmatrix} C_{I_1 I_1} & C_{I_1 I_2} & \cdots & C_{I_1 I_p} \\ C_{I_1 I_2} & C_{I_2 I_2} & \cdots & C_{I_2 I_p} \\ \vdots & \vdots & \ddots & \vdots \\ C_{I_1 I_p} & C_{I_2 I_p} & \cdots & C_{I_p I_p} \end{pmatrix}$$

Regardons de plus près la matrice $C_{I_r I_s}$ de taille $(n_r - 1) * (n_s - 1)$:

$$C_{I_r I_s} = \begin{pmatrix} \vdots & & \\ \cdots & f_{I_r I_s}(u, v) & \cdots \\ \vdots & & \end{pmatrix}$$

avec $f_{I_r I_s}(u, v) = Cov[Z_u^r, Z_v^s] - Cov[Z_1^s, Z_u^r] - Cov[Z_v^s, Z_1^r] + Cov[Z_1^r, Z_1^s]$

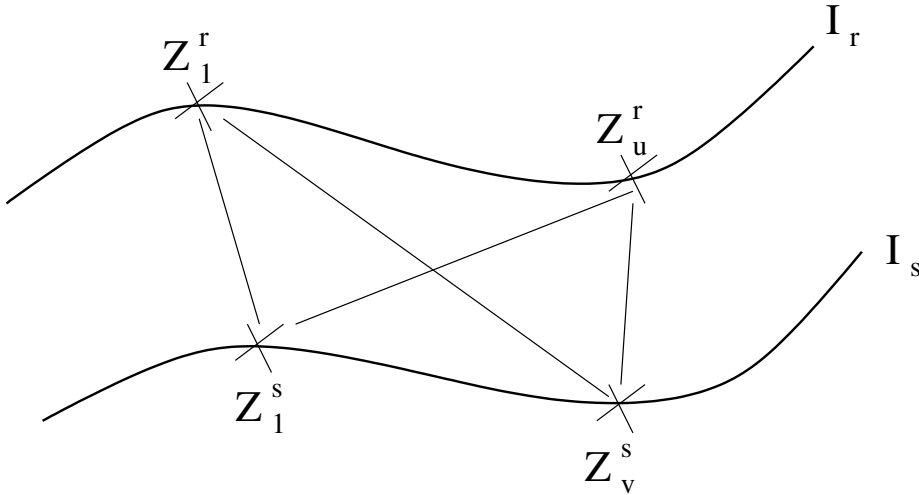


FIG. A.1 – Relations entre points sur des interfaces différents.

A.3 Les interactions gradients / interfaces

La matrice des termes croisés est de taille $3 * m * n$ et de la forme :

$$C_{GI} = \begin{pmatrix} C_{G^x I}^1 & C_{G^y I}^1 & C_{G^z I}^1 \\ \vdots & \vdots & \vdots \\ C_{G^x I}^p & C_{G^y I}^p & C_{G^z I}^p \end{pmatrix}$$

Regardons de plus près la matrice $C_{G^x I}^k$ de taille $m * (n_k - 1)$:

$$C_{G^x I}^k = \begin{pmatrix} \vdots \\ \cdots & g^k(i, j) & \cdots \\ \vdots \end{pmatrix}$$

$$\text{avec } g^k(i, j) = \text{Cov} \left[\frac{\partial Z}{\partial x}(u_i), Z_j^k \right] - \text{Cov} \left[\frac{\partial Z}{\partial x}(u_i), Z_1^k \right] \text{ et } \text{Cov} \left[\frac{\partial Z}{\partial x}(u_i), Z_j^k \right] = \alpha \frac{\sigma_Z^2}{a^2} \frac{h_x}{r} C'_Z(r)$$

A.4 Les conditions d'universalité

Les matrices contenant les fonctions de dérive sont de taille $9 * (3 * m)$ pour les gradients et $9 * n$ pour les incréments sous l'hypothèse d'une dérive polynomiale cubique. Le nombre 9 vient du nombre de fonctions de dérive prises en compte à savoir $x, y, z, x^2, y^2, z^2, xy, xz, yz$. En général, on se contente d'une dérive linéaire. Il est à noter que la condition d'universalité faisant intervenir la dérive constante 1 est automatiquement remplie puisque l'on travaille sur des incréments et elle n'apparaît donc pas dans le système de cokrigeage. La matrice U_G contient les dérivées partielles des fonctions de base aux points de donnée du gradient $u_i = {}^t(x_i, y_i, z_i)$:

$$U_G = \begin{pmatrix} \cdots & 1 & \cdots & \cdots & 0 & \cdots & \cdots & 0 & \cdots \\ \cdots & 0 & \cdots & \cdots & 1 & \cdots & \cdots & 0 & \cdots \\ \cdots & 0 & \cdots & \cdots & 0 & \cdots & \cdots & 1 & \cdots \\ \cdots & 2x_i & \cdots & \cdots & 0 & \cdots & \cdots & 0 & \cdots \\ \cdots & 0 & \cdots & \cdots & 2y_i & \cdots & \cdots & 0 & \cdots \\ \cdots & 0 & \cdots & \cdots & 0 & \cdots & \cdots & 2z_i & \cdots \\ \cdots & y_i & \cdots & \cdots & x_i & \cdots & \cdots & 0 & \cdots \\ \cdots & z_i & \cdots & \cdots & 0 & \cdots & \cdots & x_i & \cdots \\ \cdots & 0 & \cdots & \cdots & z_i & \cdots & \cdots & y_i & \cdots \end{pmatrix}$$

La matrice U_I contient les accroissements des fonctions de base aux points de données sur

les interfaces :

$$U_I = \begin{pmatrix} \cdots & \cdots & x_j^k - x_1^k & \cdots & \cdots \\ \cdots & \cdots & y_j^k - y_1^k & \cdots & \cdots \\ \cdots & \cdots & z_j^k - z_1^k & \cdots & \cdots \\ \cdots & \cdots & x_j^k x_j^k - x_1^k x_1^k & \cdots & \cdots \\ \cdots & \cdots & y_j^k y_j^k - y_1^k y_1^k & \cdots & \cdots \\ \cdots & \cdots & z_j^k z_j^k - z_1^k z_1^k & \cdots & \cdots \\ \cdots & \cdots & x_j^k y_j^k - x_1^k y_1^k & \cdots & \cdots \\ \cdots & \cdots & x_j^k z_j^k - x_1^k z_1^k & \cdots & \cdots \\ \cdots & \cdots & y_j^k z_j^k - y_1^k z_1^k & \cdots & \cdots \end{pmatrix}$$

A.5 Les termes de failles

On dispose de n_f failles, l'indice générique utilisé est l . La matrice contenant les fonctions de failles pour les gradients est de taille $3 * m * n_f$.

$$F_G = \begin{pmatrix} \vdots & \vdots & \vdots \\ \cdots & dh_l x(i) & \cdots & \cdots & dh_l y(i) & \cdots & \cdots & dh_l z(i) & \cdots \\ \vdots & \vdots & \vdots \end{pmatrix}$$

avec $dh_l x(i)$ la valeur de la dérivée par rapport à x de la fonction de dérive liée à la faille l au point i des données de gradient.

La matrice contenant les fonctions de failles pour les interfaces est de taille $n * n_f$. Seule la fonction de dérive de la faille l au $j^{ème}$ point de l'interface k est représentée.

$$F_I = \begin{pmatrix} \vdots \\ \cdots & \cdots & H_j^k(l) - H_1^k(l) & \cdots & \cdots \\ \vdots \end{pmatrix}$$

A.6 Le système de cokrigage

Lorsque nous ne sommes pas intéressés par la variance de cokrigage, pour des raisons d'efficacité, la forme duale du cokrigage est retenue. Le système de cokrigage universel dual utilisé pour estimer un incrément de potentiel a donc la forme :

$$\begin{pmatrix} C_G & {}^t C_{GI} & {}^t U_G & {}^t F_G \\ C_{GI} & C_I & {}^t U_I & {}^t F_I \\ U_G & U_I & 0 & 0 \\ F_G & F_I & 0 & 0 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} G \\ I \\ 0 \\ 0 \end{pmatrix}$$

G est le vecteur de taille $3 * m$ contenant les valeurs des composantes des gradients :

$$G = \begin{pmatrix} \vdots \\ \frac{\partial Z}{\partial x}(u_i) \\ \vdots \\ \vdots \\ \frac{\partial Z}{\partial y}(u_i) \\ \vdots \\ \vdots \\ \frac{\partial Z}{\partial z}(u_i) \\ \vdots \end{pmatrix}$$

et I est le vecteur de taille n des incréments de potentiel. Les composantes de ce vecteur sont en général nulles, sauf lorsque l'on traite des inégalités (cf. chapitre 9).

Dans le cas d'un voisinage unique, ce système est résolu une seule fois afin d'obtenir l'expression explicite de l'interpolateur pour tout le domaine.

Annexe B

Calcul du modèle de variogramme du gradient de potentiel

B.1 Détermination du lien entre covariance du potentiel et covariance du gradient

Considérons deux points de l'espace repérés par leur vecteur de coordonnées $\mathbf{u} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ et $\mathbf{v} \begin{pmatrix} x + h_x \\ y + h_y \\ z + h_z \end{pmatrix}$. Soit $C_P = Cov[P(\mathbf{v}), P(\mathbf{u})]$ la covariance entre ces deux points et $C_x = Cov \left[\frac{\partial P}{\partial x}(v), \frac{\partial P}{\partial x}(u) \right]$ la covariance entre les composantes suivant x du gradient en ces deux mêmes points. $P(\mathbf{u})$ et $P(\mathbf{v})$ sont respectivement la valeur du potentiel P au point $\mathbf{u}$ et au point $\mathbf{v}$. Calculons la covariance entre la composante suivant x du gradient au point $\mathbf{v}$ et la composante suivant x du gradient au point $\mathbf{u}$ suivant le schéma de la Figure B.1.

$$Cov \left[\frac{\partial P}{\partial x}(x + h_x, y + h_y, z + h_z), \frac{\partial P}{\partial x}(x, y, z) \right]$$

$$\begin{aligned}
 &= Cov \left[\lim_{k \rightarrow 0} \frac{P(x + h_x + k, y + h_y, z + h_z) - P(x + h_x, y + h_y, z + h_z)}{k}, \lim_{l \rightarrow 0} \frac{P(x + l, y, z) - P(x, y, z)}{l} \right] \\
 &= \lim_{k \rightarrow 0} \lim_{l \rightarrow 0} \frac{1}{kl} Cov [P(x + h_x + k, y + h_y, z + h_z) - P(x + h_x, y + h_y, z + h_z), P(x + l, y, z) - P(x, y, z)] \\
 &= \lim_{k \rightarrow 0} \lim_{l \rightarrow 0} \frac{1}{kl} [C_P(h_x + k - l, h_y, h_z) - C_P(h_x - l, h_y, h_z) - C_P(h_x + k, h_y, h_z) + C_P(h_x, h_y, h_z)] \\
 &= \lim_{l \rightarrow 0} \frac{1}{l} \left[\frac{\partial C_P}{\partial h_x}(h_x - l, h_y, h_z) - \frac{\partial C_P}{\partial h_x}(h_x, h_y, h_z) \right] \\
 &= -\frac{\partial^2 C_P}{\partial h_x^2}(h_x, h_y, h_z)
 \end{aligned}$$

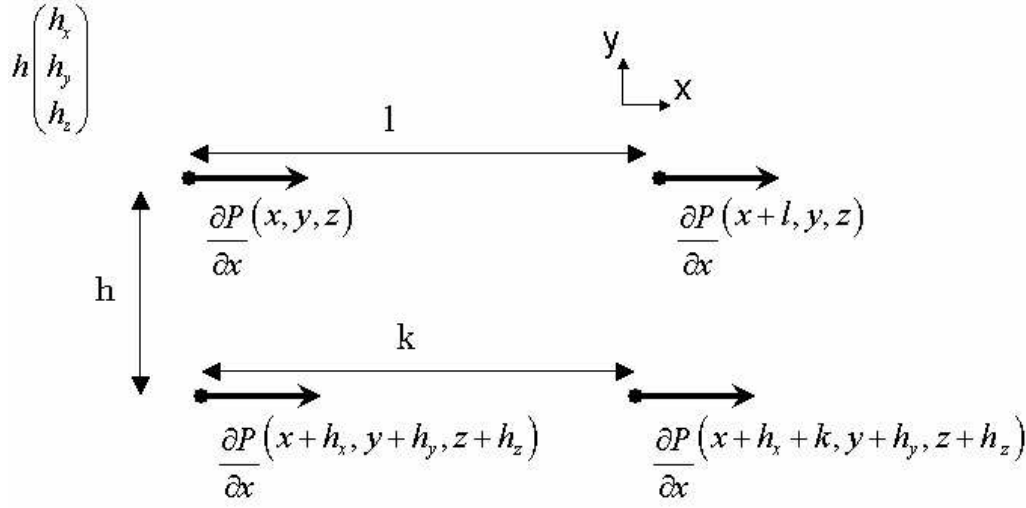


FIG. B.1 – Configuration de calcul.

B.2 Le cas du modèle cubique

On considère que le modèle de covariance de la variable “potentiel” est un modèle cubique isotrope.

$$C(r) = \begin{cases} C_0 \left(1 - 7\left(\frac{r}{a}\right)^2 + \frac{35}{4}\left(\frac{r}{a}\right)^3 - \frac{7}{2}\left(\frac{r}{a}\right)^5 + \frac{3}{4}\left(\frac{r}{a}\right)^7 \right) & \text{pour } 0 \leq r \leq a \\ 0 & \text{pour } r \geq a \end{cases}$$

avec $r = \sqrt{h_x^2 + h_y^2 + h_z^2}$.

D'où

$$\frac{\partial C_p}{\partial h_x} = \frac{\partial C}{\partial r} \frac{\partial r}{\partial h_x} = C_0 \left(-\frac{14h_x}{a^2} + \frac{105}{4} \frac{h_x r}{a^3} - \frac{35}{2} \frac{h_x r^3}{a^5} + \frac{21}{4} \frac{h_x r^5}{a^7} \right)$$

$$\frac{\partial^2 C_p}{\partial h_x^2}(h_x, h_y, h_z) = \frac{C_0}{a^2} \left(-14 + \frac{105}{4} \frac{r}{a} - \frac{35}{2} \frac{r^3}{a^3} + \frac{21}{4} \frac{r^5}{a^5} + \frac{h_x^2}{r^2} \left(\frac{105}{4} \frac{r}{a} - \frac{105}{2} \frac{r^3}{a^3} + \frac{105}{4} \frac{r^5}{a^5} \right) \right)$$

Comme $C_x(h_x, h_y, h_z) = -\frac{\partial^2 C_p}{\partial h_x^2}(h_x, h_y, h_z)$ et $\gamma_x(h_x, h_y, h_z) = C_x(0) - C_x(h_x, h_y, h_z)$

avec $C_x(0) = \frac{14C_0}{a^2}$, on a donc :

$$\gamma_x(h_x, h_y, h_z) = \frac{C_0}{a^2} \left(\frac{105}{4} \frac{r}{a} - \frac{35}{2} \frac{r^3}{a^3} + \frac{21}{4} \frac{r^5}{a^5} + \frac{h_x^2}{r^2} \left(\frac{105}{4} \frac{r}{a} - \frac{105}{2} \frac{r^3}{a^3} + \frac{105}{4} \frac{r^5}{a^5} \right) \right)$$

On en déduit alors les expressions des variogrammes directionnels perpendiculairement et parallèlement à la composante considérée du gradient, numérotées 4.7 et 4.8 au Chapitre 4. Comme le modèle de variogramme du potentiel est borné, les variogrammes du gradient dans toutes les directions atteignent tous le même palier.

Annexe C

Rappels sur la minimisation d'une forme quadratique ou quelconque

Définissons une fonction scalaire χ^2 quelconque, dépendante de plusieurs variables réelles. Le début du développement de Taylor de χ^2 au voisinage d'un point $\mathbf{a}_0$ s'écrit :

$$\begin{aligned}\chi^2(\mathbf{a}) &= \chi^2(\mathbf{a}_0) + \sum_i \frac{\partial \chi^2}{\partial a_i}(\mathbf{a}_0) a_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 \chi^2}{\partial a_i \partial a_j}(\mathbf{a}_0) a_i a_j + \dots \\ &\simeq c - \mathbf{b} \cdot \mathbf{a} + \frac{1}{2} \mathbf{a} \cdot \mathbf{H} \cdot \mathbf{a}\end{aligned}\tag{C.1}$$

avec $c \equiv \chi^2(\mathbf{a}_0)$, $\mathbf{b} \equiv -\vec{\nabla} \chi^2(\mathbf{a}_0)$, et $H_{ij} \equiv \left[\frac{\partial^2 \chi^2}{\partial a_i \partial a_j}(\mathbf{a}_0) \right]$. $\mathbf{H}$ est la matrice hessienne de χ^2 en $\mathbf{a}_0$.

Si χ^2 est une forme quadratique, l'expression C.1 est exacte et on peut alors exprimer son gradient par l'expression suivante :

$$\vec{\nabla} \chi^2 = \mathbf{H} \mathbf{a} - \mathbf{b}\tag{C.2}$$

Celui-ci s'annule, autrement dit χ^2 atteint un extremum, pour une valeur de $\mathbf{a}$ solution du système : $\mathbf{H} \mathbf{a} = \mathbf{b}$. Dans le cas où l'approximation C.1 est bonne sans être exacte, on passe d'un essai courant $\mathbf{a}_0$ au minimum $\mathbf{a}_{\min}$ grâce à l'équation C.2, d'où l'expression du minimum :

$$\mathbf{a}_{\min} = \mathbf{a}_0 - \mathbf{H}^{-1} \vec{\nabla} \chi^2(\mathbf{a}_0)\tag{C.3}$$

Par contre, si l'approximation de la forme s'avère être mauvaise, tout ce qu'on peut faire est d'utiliser la méthode des plus fortes pentes. Elle consiste à se déplacer d'un pas de longueur défini arbitrairement dans la direction opposée au gradient :

$$\mathbf{a}_1 = \mathbf{a}_0 - C \vec{\nabla} \chi^2(\mathbf{a}_0)\tag{C.4}$$

C étant une constante. Ces deux approches sont utilisées de manière optimisée dans la méthode de Levenberg-Marquardt (Marquardt, 1963).

Annexe D

Simulations

Cette annexe présente diverses méthodes qui ont été utilisées lors des simulations présentées dans le cadre de cette thèse, mais non expliquées dans le corps du texte.

D.1 Les simulations conditionnelles

D.1.1 Généralités

Commençons tout d'abord par rappeler ce que l'on entend par simulations conditionnelles. En géostatistique, on se place pratiquement toujours dans un cadre probabiliste qui permet de formaliser connaissances et incertitudes du phénomène régionalisé étudié. On considère alors la variable régionalisée comme la réalisation d'une fonction aléatoire. Cette dernière est caractérisée par sa loi spatiale, c'est-à-dire l'ensemble des lois multivariées entre ses composantes.

Quand on effectue l'estimation d'une variable à partir de données expérimentales, on recherche une valeur aussi proche que possible de la valeur vraie inconnue.

Les simulations, quant à elles, consistent à créer d'autres réalisations de la fonction aléatoire. Une simulation cherche donc plutôt à reproduire les principales caractéristiques de dispersion de cette variable. Si la simulation est conditionnelle, on exige en plus qu'elle respecte les données expérimentales (cf. Figure D.1). Pour une fonction aléatoire multi-

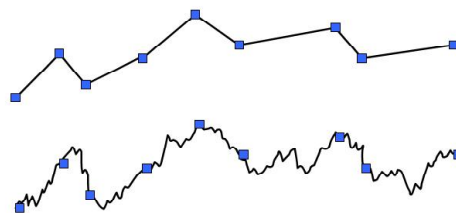


FIG. D.1 – Schéma illustrant l'estimation d'une variable 1D à partir de quelques points de données en haut, simulation conditionnelle en bas.

gaussienne, la loi spatiale est entièrement déterminée par ses deux premiers moments. C'est

dans ce cas idéal que de nombreux algorithmes de simulations non conditionnelles ont été développés comme les bandes tournantes (le plus général), la décomposition matricielle ou la méthode séquentielle gaussienne. On peut aussi ajouter des techniques spécifiques à une dimension telles que la méthode spectrale ou la dilution sur les droites. Pour des données plus complexes, d'autres techniques peuvent être mises en œuvre : simulation séquentielle d'indicatrice, séquentielle isofactorielle, par lancement d'objets (modèle booléen, de dilution...). On se reportera utilement à Chilès et Delfiner (1999) pour la description de ces techniques, ainsi qu'à Lantuéjoul (2001) où de nombreux algorithmes utilisés dans les simulations géostatistiques sont décrits.

D.1.2 Le krigeage conditionnant

Une méthode classique et directe de simulation conditionnelle d'une fonction aléatoire gaussienne est d'effectuer d'une part une simulation non conditionnelle et d'autre part le conditionnement par un double krigeage. Cette méthode produit des simulations qui respectent la covariance et donc la loi spatiale. Nous nous plaçons ici dans le cadre gaussien qui du point de vue pratique est très agréable. En effet, les fonctions aléatoires gaussiennes sont une classe de fonctions aléatoires dont la loi spatiale dépend uniquement des deux premiers moments. Supposons que l'on veuille simuler un vecteur multigaussien centré X dont on connaît les p premières composantes $(X_i)_{1 \leq i \leq p}$. Toute composante dont la valeur n'est pas fixée peut se décomposer sous la forme d'une somme de deux termes indépendants :

$$p < i \leq n, X_i = X_i^{KS} + (X_i - X_i^{KS})$$

avec X_i^{KS} krigeage simple de X_i par les composantes gaussiennes fixées et $X_i - X_i^{KS}$ l'erreur de krigeage.

Comme cette erreur est inconnue mais indépendante de tous les X_i , il suffit alors de simuler non-conditionnellement un vecteur multigaussien centré U_i à $n - p$ composantes corrélées (les corrélations étant les covariances de krigeage), d'où :

$$X_i = \begin{cases} X_i & 1 \leq i \leq p \\ X_i^{KS} + U_i & p < i \leq n \end{cases}$$

On a donc remplacé l'erreur réelle inconnue par l'erreur calculée sur une simulation non conditionnelle.

D.2 Génération d'une variable aléatoire uniforme par la méthode de congruence

Une variable aléatoire uniforme sur $]0,1[$ a une densité de probabilité :

$$f(u) = \begin{cases} 1 & \text{si } 0 < u < 1 \\ 0 & \text{sinon} \end{cases}$$

Pour générer des nombres pseudo-aléatoires à loi uniforme - à une discrétisation près par m valeurs - le principe est d'engendrer une suite pseudo-aléatoire d'entiers x_k compris entre

0 et $m - 1$ et on prend $u_k = \frac{x_k}{m}$. Voyons cela un peu plus en détail.

On se donne deux entiers m (le module) et a (le multiplicateur) avec $a \in [2, m - 1]$. L'objectif est de générer une suite u_k de valeurs entre 0 et $m - 1$, simulant des variables aléatoires uniformes, en utilisant le reste de la division euclidienne de deux entiers. On construit donc notre suite de la façon suivante :

$$u_k = \frac{x_k}{m}, \quad k \in \mathbb{N}$$

avec

$$\begin{cases} x_0 \in [1, m - 1] \\ x_{k+1} \equiv a x_k [m] \end{cases}$$

x_0 (le germe) est choisi arbitrairement et x_{k+1} est le reste de la division de $a x_k$ par m . Or la suite x_k à valeurs dans $[1, m - 1]$ est périodique. On a donc intérêt à ce que la période maximale soit $m - 1$. Pour cela m et $\frac{m-1}{2}$ doivent être des nombres premiers et $a^{\frac{(m-1)}{2}} \equiv -1[m]$. De plus, a ne doit pas diviser m afin d'éviter la valeur nulle pour la suite simulée. Les valeurs choisies sont : $m = 2001179$ et $a = 1000$ (Lantuéjoul, 2001).

D.3 Approximation polynomiale de la fonction de répartition gaussienne

Soit $g(x) = \frac{1}{\sqrt{2\pi}} \exp \frac{-x^2}{2}$ la densité de probabilité de Gauss, d'où se déduit la fonction de répartition gaussienne $G(x) = \int_{-\infty}^x g(t) dt$. Pour tout $x \geq 0$, on peut approcher cette fonction par une décomposition polynomiale suivant l'expression suivante :

$$G(x) = 1 - g(x)(b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5) + \varepsilon(x)$$

avec $t = \frac{1}{1+px}$ et $p = 0.2316419$.

Les coefficients sont donnés par les valeurs $b_1 = 0.31938153$, $b_2 = -0.356563782$, $b_3 = 1.781477937$, $b_4 = -1.821255978$ et $b_5 = 1.330274429$.

Le résidu est tel que : $\varepsilon < 10^{-7}$, on pourra donc le négliger. De plus, si x est strictement négatif, c'est la fonction $1 - G(-x)$ qui est utilisée.

D.4 Simulation de variables aléatoires gaussiennes

Considérons deux variables aléatoires U et V , indépendantes et uniformes sur $]0, 1[$. Les variables X et Y définies par :

$$\begin{cases} X = \sqrt{-2 \log(U)} \cos(2\pi V) \\ Y = \sqrt{-2 \log(U)} \sin(2\pi V) \end{cases}$$

suivent une loi normale centrée réduite ($\mathcal{N}$) et sont indépendantes.

Bibliographie

Références

- Abrahamsen, P. et F. Benth (2001). Kriging with inequality constraints. *Mathematical Geology* 33(6), 719–744.
- Abrahamsen, P. et H. Omre (1994). Random functions and geological subsurfaces. STAT-Report 3/94, Dep. of Math Sci., NTNU, Trondheim.
- Alabert, F. (1987a). Stochastic imaging of spatial distributions using hard and soft data. Thèse Master of Science, Stanford University, Stanford.
- Alabert, F. (1987b). The practice of fast conditional simulations through the LU decomposition of the covariance matrix. *Mathematical Geology* 19(5), 369–386.
- Armstrong, M., A. Galli, G. Le Loch, F. Geffroy, et R. Eschard (2003). *Plurigaussian Simulations in Geosciences*, Berlin : Springer.
- Auerbach, S. et H. Schaeben (1990). Computer-aided geometric design of geologic surfaces and bodies. *Mathematical Geology* 22(8), 957–987.
- Barnes, R. et T. Johnson (1984). Positive kriging. G. Verly et al. (Eds.), *Geostatistics for Natural Resources Characterization*, Volume 1, Dordrecht, p. 231–244. Reidel.
- Bazzana, L., P. Ruffo, et D. Renard (1995). Estimation de cartes de profondeur conditionnées par des inégalités. E. N. S. des Mines de Paris (Ed.), *Compte rendu des journées de Géostatistique 15-16 juin 1995*, Cahiers de Géostatistique Fascicule 5, Fontainebleau, p. 23–43.
- Benito García-Morales, M. (2003). *Non stationnarité dans les modèles de type booléen : application à la simulation d'unités sédimentaires*. Thèse de docteur en géostatistique, Ecole Nationale Supérieure des Mines de Paris, Fontainebleau.
- Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*, New York : Springer.
- Berger, J., V. De Oliveira, et B. Sansó (2000). Objective Bayesian analysis of spatially correlated data. Technical report, Institute of Statistics and Decision Sciences, Duke University, Durham.
- Blanchin, R. et J. Chilès (1993). The Channel Tunnel : Geostatistical prediction of the geological conditions and its validation by the reality. *Mathematical Geology* 25(7), 963–974.
- Boissonnat, J. (1988). Shape reconstruction from planar cross-sections. *Computer Vision, Graphics, and Image Processing* 44(1), 1–29.
- Boissonnat, J. et S. Nullans (1996). Reconstruction of geological structures from heterogeneous and sparse data. Rapport de Recherche 3069, INRIA, Sophia Antipolis.

- Brooks, S. et G. Roberts (1998). Assessing convergence of Markov Chain Monte Carlo algorithms. *Statistics and Computing* 8, 319–335.
- Chauvet, P. (2000). *Aide-mémoire de géostatistique linéaire*, Paris : Presses de l'Ecole des Mines.
- Chauvet, P., J. Pailleux, et J. Chilès (1976). Analyse objective des champs météorologiques par cokrigage. *La Météorologie, Sciences et Techniques*, 6è série(4), 37–54.
- Chilès, J. (1977). *Géostatistique des phénomènes non stationnaires (dans le plan)*. Thèse de docteur-ingénieur, Université de Nancy I, Nancy.
- Chilès, J. (1979). La dérive à la dérive. Note du Centre de Géostatistique N-591, Centre de Géostatistique, Fontainebleau.
- Chilès, J. et P. Delfiner (1999). *Geostatistics : Modeling Spatial Uncertainty*, New York : Wiley. Probability and Statistics.
- Courrioux, G. et P. Calcagno (2001). Modélisation géométrique d'interfaces et volumes géologiques. Notes de Cours.
- Courrioux, G., S. Nullans, A. Guillen, J. Boissonnat, P. Repusseau, X. Renaud, et M. Thibaut (2001). 3D volumetric modelling of Cadomian terranes (Northern Brittany, France) : an automatic method using Voronoï diagrams. *Tectonophysics* (331), 181–196.
- Cowan, E., R. Beatson, H. Ross, W. Fright, T. McLennan, T. Evans, J. Carr, R. Lane, D. Bright, A. Gillman, P. Oshust, et M. Titley (2003). Practical implicit geological modelling. S. Dominy (Ed.), *5th International Mining Geology Conference*, p. 89–99. The Australasian Institute of Mining and Metallurgy.
- Cowles, M. et B. Carlin (1996). Markov chain Monte Carlo convergence diagnostics : a comparative review. *Journal of the American Statistical Association* 91, 883 – 904.
- Cressie, N. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology* 17(5), 563–586.
- Diamond, P. (1988). Interval-valued random functions and the kriging of intervals. *Mathematical Geology* 20(3), 145–165.
- Diggle, P., P. Ribeiro, et O. Christensen (2003). *Spatial statistics and computational methods*, Chap. An introduction to model-based geostatistics, p. 43–86. New York : Springer Verlag.
- Dong, A. (1990). *Estimation géostatistique de phénomènes régis par des équations aux dérivées partielles*. Thèse de docteur en géostatistique, Ecole Nationale Supérieure des Mines de Paris, Fontainebleau.
- Droesbeke, J., J. Fine, et G. Saporta (Eds.) (2002). *Méthodes bayésiennes en statistique* : Editions TECHNIP.
- Dubrule, O. et C. Kostov (1986). An interpolation method taking into account inequality constraints : I. Methodology. *Mathematical Geology* 18(1), 33–51.
- Emery, X. (2001). *Géostatistique linéaire*. Notes de Cours.
- Freulon, X. (1992). *Conditionnement du modèle gaussien par des inégalités ou des randomisées*. Thèse de docteur en géostatistique, Ecole Nationale Supérieure des Mines de Paris, Fontainebleau.

- Freulon, X. (1994). Conditional simulation of a Gaussian random vector with non linear and/or noisy observations. M. Armstrong et P. Dowd (Eds.), *Geostatistical Simulations*, Dordrecht, p. 57–71. Kluwer.
- Freulon, X. et C. de Fouquet (1993). Conditioning a Gaussian model with inequalities. A. Soares (Ed.), *Geostatistics Troia '92*, Volume 1, Dordrecht, p. 201–212. Kluwer.
- Frey, P. et H. Borouchaki (1996). Texel :triangulation des surfaces implicites. Partie I : aspects théoriques. Rapport de recherche 3066, INRIA.
- Galera, C., C. Bennis, I. Moretti, et J. Mallet (2003). Construction of coherent 3D geological blocks. *Computers and Geosciences* (29), 971–984.
- Galli, A. et H. Gao (2001). Rate of convergence of the Gibbs sampler in the Gaussian case. *Mathematical Geology* 33(6), 653–677.
- Galli, A., E. Murillo, et J. Thomann (1984). Dual kriging - its properties and its uses in direct contouring. G. Verly et al. (Eds.), *Geostatistics for Natural Resources Characterization*, Volume 2, Dordrecht, p. 621–634. Reidel.
- Gastinel, N. (1966). *Analyse numérique linéaire*, Paris : Hermann. Enseignement des Sciences.
- Gelman, A. (1996). Inference and monitoring convergence. W. R. Gilks, S. Richardson, et D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, Chap. 8, p. 131–143. Boca Raton : Chapman and Hall/CRC.
- Geman, S. et D. Geman (1984). Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence PAMI-6*(6), 721–741.
- Gilks, W., S. Richardson, et D. Spiegelhalter (1996). *Markov Chain Monte Carlo in practice*, Boca Raton : Chapman and Hall/CRC.
- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*, New-York : Oxford University Press. Applied Geostatistics Series.
- Goria, S. (2004). *Evaluation d'un projet minier : approche bayésienne et options réelles*. Thèse de docteur en géostatistique, Ecole Nationale Supérieure des Mines de Paris, Fontainebleau.
- Green, P. et X. Han (1992). Metropolis methods, Gaussian proposals and antithetic variables. P. Barone, A. Frigessi, et M. Piccioni (Eds.), *Stochastic Models, Statistical Methods and Algorithms in Image Analysis. Lectures Notes in Statist.*, 74, p. 142–164. Berlin : Springer Verlag.
- Guillen, A., G. Courrioux, P. Calcagno, R. Lane, T. Lees, et P. McInerney (2004). Constrained gravity inversion applied to Broken Hill. *Proceedings ASEG 17th Geophysical Conference and Exhibition*.
- Guyon, X. (1999). Méthodes numériques par chaînes de Markov. Ecole d'été de Mathématiques, Merida, Venezuela.
- Halbwachs, Y., G. Courrioux, X. Renaud, et P. Repusseau (1996). Topological and geometric characterization of fault networks using 3-Dimensional Generalized Maps. *Mathematical Geology* 28(5), 625–656.

- Hallouin, S. (2001). Robustification des algorithmes de champs de potentiel. Rapport de stage - Maîtrise MIME, Université d'Orléans.
- Handcock, M. et M. Stein (1993). A Bayesian analysis of kriging. *Technometrics* 35(4), 403–410.
- Journal, A. (1986). Constrained interpolation and qualitative information - The soft kriging approach. *Mathematical Geology* 18(3), 269–286.
- Kernighan, B. et D. Ritchie (1995). *Le langage C* (2è ed.), Paris : Masson.
- Kitanidis, P. (1986). Parameter uncertainty in estimation of spatial functions : Bayesian analysis. *Water Resources Research* 22(4), 499–507.
- Kostov, C. et O. Dubrule (1986). An interpolation method taking into account inequality constraints : II. Practical approach. *Mathematical Geology* 18(1), 53–73.
- Lajaunie, C. et J. Béhaxétéguy (1989). Élaboration d'un programme d'ajustement semi-automatique d'un modèle de corégionalisation. Note du Centre de Géostatistique N21/89/G, Centre de Géostatistique, Fontainebleau.
- Lajaunie, C., G. Courrioux, et L. Manuel (1997). Foliation fields and 3D cartography in geology : principles of a method based on potential interpolation. *Mathematical Geology* 29(4), 571–584.
- Langlais, V. (1990). *Estimation sous contraintes d'inégalités*. Thèse de docteur en géostatistique, Ecole Nationale Supérieure des Mines de Paris, Fontainebleau.
- Lantuéjoul, C. (2001). *Geostatistical Simulation : Models and Algorithms.*, Berlin : Springer.
- Ledez, D. (2003). *Modélisation d'objets naturels par formulation implicite*. Thèse de doctorat de l'institut national polytechnique de lorraine, Ecole Nationale Supérieure de Géologie, Nancy.
- Ledru, P., G. Courrioux, C. Dallain, J. Lardeaux, J. Montel, O. Vanderhaeghe, et G. Vitel (2001). The Velay dome (French Massif Central) : melt generation and granite emplacement during orogenic evolution. *Tectonophysics* 342, 207–237.
- Lee, J. et J. Angelier (1994). Paleostress trajectory maps based on the results of local determinations : the "lissage" program. *Computers and Geosciences* 20(2), 161–191.
- Lopez, S. (2003). *Modélisation de réservoirs chenalisés méandriformes. Approche génétique et stochastique*. Thèse de docteur en géostatistique, Ecole Nationale Supérieure des Mines de Paris, Fontainebleau.
- Lorensen, W. et H. Cline (1987). Marching cubes : A high resolution 3D surface construction algorithm. M. Stone (Ed.), *Computer Graphics (SIGGRAPH'87 Proceedings)*, Volume 21, p. 163–169.
- Mallet, J. (1997). Discrete modeling for natural objects. *Mathematical Geology* 29(2), 199–219.
- Mallet, J. (2002). *Geomodeling*, New York : Oxford University Press. Applied Geostatistics Series.
- Manuel, L. (1995). Interpolation des données directionnelles avec un modèle de potentiel. Mémoire de DEA en Génie Civil et Minier, option Géostatistique S-329, Centre de Géostatistique, Fontainebleau.

- Maréchal, A. (1984). Kriging seismic data in presence of faults. G. Verly et al. (Eds.), *Geostatistics for Natural Resources Characterization*, Volume 1, Dordrecht, p. 271–294. Reidel.
- Marquardt, D. (1963). An algorithm for least-squares estimation of non-linear parameters. *SIAM, Journal on Applied Mathematics* 11(2), 431–441.
- Matheron, G. (1970). La théorie des variables régionalisées et ses applications. Les cahiers du Centre de Morphologie Mathématique Fascicule 5, Centre de Géostatistique, Fontainebleau.
- Matheron, G. (1981). Remarques sur le krigeage et son dual. Note du Centre de Géostatistique N-695, Centre de Géostatistique, Fontainebleau.
- Maxelon, M. (2004). *Developing a three-dimensional structural model of the Lower Lepontine Nappes - Central Alps, Switzerland and Northern Italy*. PhD Thesis, ETH Zürich, Zürich.
- Mosegaard, K. et A. Tarantola (2002). *International Handbook of Earthquake and Engineering Seismology (Part A)*, Chap. Probabilistic approach to inverse problems, p. 237–265. Academic Press.
- Nullans, S. (1998). *Reconstruction géométrique de formes - Application à la géologie*. Thèse de docteur en sciences pour l'ingénieur, Université de Nice, Nice.
- Omre, H. (1987). Bayesian kriging - Merging observations and qualified guesses in kriging. *Mathematical Geology* 19(1), 25–39.
- Omre, H. et K. Halvorsen (1989). The Bayesian bridge between simple and universal kriging. *Mathematical Geology* 21(7), 767–786.
- Pomian-Srzednicki, I. (2001). *Calculation of geological uncertainties associated with 3-D geological models*. PhD Thesis, Ecole Polytechnique Fédérale de Lausanne, Lausanne.
- Press, W., S. Teukolsky, W. Vetterling, et B. Flannery (1992). *Numerical Recipes in C : The Art of Scientific Computing* (2nd ed.), Cambridge : Cambridge University Press.
- Priest, S. (1993). *Discontinuity Analysis for Rock Engineering*, London : Chapman and Hall.
- Renard, D. et P. Ruffo (1993). Depth, dip and gradient. A. Soares (Ed.), *Geostatistics Troia '92*, Volume 1, Dordrecht, p. 167–178. Kluwer.
- Ribeiro Jr, P. et P. Diggle (1999). Bayesian inference in Gaussian model-based geostatistics. Technical Report ST-99-08, Department of Mathematics and Statistics, Lancaster University, Lancaster.
- Rivoirard, J. (1984). *Le comportement des poids de krigeage*. Thèse de docteur-ingénieur en sciences et techniques minières - option géostatistique, Ecole Nationale Supérieure des Mines de Paris, Fontainebleau.
- Roig, J. et M. Faure (2000). La tectonique cisailante polyphasée du Sud Limousin (Massif central français) et son interprétation dans un modèle d'évolution polycyclique de la chaîne hercynienne. *Bulletin de la Société géologique de France* 171(3), 295–307.
- Roig, J., M. Faure, et C. Truffert (1998). Folding and granite emplacement inferred from structural, strain, TEM and gravimetric analyses : the case study of the Tulle

- antiform, SW French Massif Central. *Journal of Structural Geology* 20(9/10), 1169–1189.
- Schneider, S. (2002). *Pilotage automatique de la construction de modèles géologiques surfaciques*. Thèse de docteur en informatique, Ecole des Mines de Saint-Etienne et Université Jean-Monnet, Saint-Etienne.
- Sénégas, J. (2002). *Méthodes de Monte-Carlo en vision stéréoscopique. Application à l'étude de modèles numériques de terrain*. Thèse de docteur en géostatistique, Ecole Nationale Supérieure des Mines de Paris, Fontainebleau.
- Tacher, L. et A. Parriaux (1997). Calcul et représentation de l'incertitude associée aux modèles géologiques. *Colloque modélisation du sous-sol*, Documents du BRGM 274, p. 108–111. Editions du BRGM.
- Turk, G. et J. O'Brien (2002). Modelling with implicit surfaces that interpolate. *ACM Transactions on Graphics* 21(4), 855–873.
- van den Boogaart, K. et H. Schaeben (2002). Kriging of regionalized directions, axes and orientations I. Directions and axes. *Mathematical Geology* 34(5), 479–503.
- Vieira, M. (1999). *Estimation bayésienne par des méthodes MCMC - Application à la surveillance des moteurs asynchrones*. Thèse de docteur en sciences pour l'ingénieur, Faculté des Science de Nice - Sophia-Antipolis, Nice.
- Wackernagel, H. (2003). *Multivariate Geostatistics* (3rd, completely revised ed.), Berlin : Springer.
- Young, D. (1987). Random vectors and spatial analysis by geostatistics for geotechnical applications. *Mathematical Geology* 19(6), 467–479.