



REPUBLIQUE DU CAMEROUN

REPUBLIC OF CAMEROON

**Paix – Travail - Patrie
Peace – Work – Fatherland**

**UNIVERSITE DE NGAOUNDERE
THE UNIVERSITY OF NGAOUNDERE**

**ECOLE DE GEOLOGIE ET D'EXPLOITATION MINIERE
SCHOOL OF GEOLOGY AND MINING ENGINEERING**



Mémoire de Master

Sur le thème :

Application des simulations géostatistiques aux traitements des données géoélectriques : cas de la simulation séquentielle gaussienne

En vue de l'obtention d'un diplôme de Master Recherche en Science de l'Ingénieur

Spécialité : Géostatistiques Appliquées à l'Industrie Minière

Par :

KABE MOUKETE Eric Bruno

12A085EM

Devant le jury composé de :

ABDOUL Wahabou, Professeur, EGEM

MAMBOU NGUEYEP Luc Leroy, Chargé de Cours, EGEM

MEYING Arsène, Chargé de Cours, EGEM

Président du jury

Examineur

Rapporteur

Année académique 2018 / 2019

DEDICACE

A ma défunte mère BAKANE Régine et Bruno mon frère aîné,
des piliers de mes édifices.

REMERCIEMENTS

C'est avec un grand plaisir que je réserve ces quelques lignes en guise de reconnaissance et de gratitude à l'égard de tous ceux qui ont contribué de près ou de loin à l'élaboration de ce travail.

Ma première pensée va vers le bon Dieu, qui me soutient face aux obstacles.

Je tiens à remercier :

Le Directeur de l'Ecole de Géologie et d'Exploitation Minière, Pr. ABDOUL Wahabou pour le cadre de formation mis à la disposition des étudiants ;

Le précédent Directeur Pr. NGOUNOUNO Ismaïla, pour les dispositions prises lors de l'ouverture du Master Recherche ;

Mon encadrant Dr. MEYING Arsène, pour la disponibilité, les qualités intellectuelles, et le professionnalisme qu'il a su apporter à ce travail ;

Ma famille, pour l'assistance morale et financière inconditionnelles qu'elle démontre à mon égard ;

Dr. MAMBOU et Dr. ABDOURAMANI, pour les conseils éclairés qui ont été d'une grande utilité pour mon développement personnel.

M. MFENJOU Martin Luther, M. MBOUNJA BESSEME Fritz et M. SEINI, ainsi que tous les autres enseignants, pour leurs contributions à ma formation ;

Mes camarades et amis Achemetyaye, Medjo, Mbeba, Njikam Souleyman, Boroh, Seunkole Bila, Bikai, Sontsa, tous les étudiants de la spécialité Géostatistiques Appliquées à l'Industrie Minière, et ceux de la spécialité Géophysique Minière, pour les échanges de connaissance et les moments agréables qu'on a partagé.

Mes frères et sœurs Blandine, Nicole, Moïse, Bruno, Jeannine, Charlotte et Ebenezer, pour m'avoir toujours accompagné dans mes entreprises.

Merci à tous.

Table des matières

DEDICACE	i
REMERCIEMENTS	ii
LISTES DES FIGURES	vi
LISTE DES TABLEAUX.....	viii
RESUME	ix
ABSTRACT.....	x
INTRODUCTION GENERALE	1
Chapitre 1 : GENERALITES.....	4
1.1 Les méthodes géostatistiques	4
1.1.1 Analyse et modélisation de la structure spatiale des données	5
1.1.1.1 Fonction de covariance.....	5
1.1.1.2 Analyse variographique.....	5
1.1.2 Les méthodes univariées.....	7
1.1.2.1 Le krigeage ordinaire	7
1.1.2.2 Le krigeage simple	9
1.1.3 Les simulations géostatistiques	9
1.1.3.1 Les simulations non-conditionnelles.....	13
1.1.3.2 Les simulations conditionnelles	14
1.2 Approche probabiliste de la théorie inverse discrète	15
1.2.1 Mesures bruitées et variables aléatoires	16
1.2.2 Distributions Gaussiennes et test des hypothèses de statistiques gaussiennes	18
1.3 Application des notions de théorie inverse à la prospection électrique	20
1.3.1 Principe de la méthode de courant continu.....	20

1.3.2	Inversion des données géoélectriques.....	21
•	Approche non-linéaire.....	21
•	Approche linéaire	22
Chapitre 2 : ALGORITHME DES SIMULATIONS SEQUENTIELLES GAUSSIENNES.....		23
2.1	Algorithme de simulation séquentielle gaussienne.....	23
2.1.1	Principe des simulations séquentielles gaussiennes	23
2.1.2	Conception d'un algorithme de simulation séquentielle gaussienne.....	25
2.1.2.1	Algorithme de simulation : cas à une dimension (1D).....	25
2.1.2.2	Algorithme de simulation : cas à deux dimensions (2D)	27
2.2	Analyse variographique automatique et semi-automatique.....	29
2.2.1	Principe et algorithme de la modélisation par la méthode des moindres carrés.....	30
2.2.1.1	Cas particulier d'un ajustement linéaire.....	30
2.2.1.2	Algorithme de la méthode pour $Q = 1$	31
2.2.2	Analyse variographique intelligente à l'aide des algorithmes génétiques.....	33
2.2.2.1	Formulation de la fonction objective	34
2.2.2.2	Paramétrage du programme génétique.....	35
2.3	Réalisation du logiciel ResSimSeq2D	37
2.3.1	Choix de l'environnement de programmation	37
2.3.2	Les fonctionnalités de ResSimSeq2D	37
2.3.2.1	Les applications	38
•	La description statistique.....	38
•	La description géostatistique.....	38
•	Le système de gestion des fichiers	39
2.3.2.2	Les séquences d'exécution de chaque tâche	40
2.3.3	Validation des résultats de ResSimSeq2D	41

2.3.3.1	Les distributions statistiques	41
2.3.3.2	Les variogrammes	42
2.3.3.3	Les pseudo-sections	42
Chapitre 3 : Etude d'un cas pratique d'inversion des données géoélectriques à l'aide de ResSimSeq2D		45
3.1	Origine des données	45
3.2	Descriptions statistique et géostatistique des résistivités mesurées	48
3.2.1	Les distributions statistiques.....	48
3.2.2	Les variogrammes initiaux	49
3.3	Inversion des données de l'expérimentation.....	51
3.3.1	Les courbes de sondage (simulation 1D).....	51
3.3.2	Inversion 2D	52
3.4	Interprétations des résultats et discussions	54
3.4.1	Interprétations des courbes de sondages et pseudo section	54
3.4.2	Les valeurs extrêmes	56
3.4.3	L'influence du voisinage sur la qualité de la simulation	57
3.4.4	L'analyse variographique	59
CONCLUSION GENERALE.....		60
REFERENCE BIBLIOGRAPHIQUE		xi

LISTES DES FIGURES

Figure 1: Exemple de fonction de covariance.....	5
Figure 2: Modèle sphérique.	6
Figure 3: Modèle exponentiel.	7
Figure 4: La partie grise donne la probabilité pour qu'une réalisation soit comprise entre d et $d + \partial d$	17
Figure 5: Le point de maximum de vraisemblance dML de la distribution $P(d)$ donne la réalisation qui a la plus grande probabilité d'occurrence. En général, cette valeur est différente de la moyenne des mesures d	17
Figure 6 : Distribution gaussienne centrée réduite(A), et de variance 2 (B).....	18
Figure 7 : Configuration d'un voisinage de calcul	28
Figure 8: Algorithme de la méthode par simulation séquentielle gaussienne.....	29
Figure 9 : Croisement entre deux chromosomes.....	34
Figure 10 : Mutations d'une série de gène sur chaque chromosome	34
Figure 11 : Algorithme d'exécution du programme génétique.....	36
Figure 12 : Boite de dialogue pour l'application pour la description statistique.	38
Figure 13 : Boite de dialogue pour l'analyse variographique.....	39
Figure 14 : Histogramme des résistivités sur Prosys (A) et sur ResSimSeq2D (B)	41
Figure 15 : Variogramme réalisé sur ArcGis (A) et sur ResSimSeq2D (B).....	42
Figure 16 : Pseudo section sur Prosys (A), sur ResSimSeq2D (B)	43
Figure 17 : Pseudo section sur Res2DInv pour le même profil d'essai	43
Figure 18 : Carte de localisation du site de l'expérimentation (à refaire)	47
Figure 19 : Matériels utilisés lors de l'expérimentation.....	47
Figure 20 : Les histogrammes et distributions pour chaque profil.....	48
Figure 21 : Exemple de profil représenté par des points de mesures de pa	49
Figure 22 : Les variogrammes calculés pour chacun des profils	50
Figure 23 : Courbe des sondages du profil 4.	51
Figure 24 : Liste des variogramme de la simulation 1D pour le profil 4	51
Figure 25 : Autres courbes de sondages	52
Figure 26 : Pseudo section des résistivités pour le profil 8	52

Figure 27 : Pseudo section des résistivités pour le profil 4	53
Figure 28 : Pseudo section des résistivités pour le profil 2	53
Figure 29 : Pseudo section des résistivités pour le profile 6, avec des données présentant des valeurs extrêmes.....	54
Figure 30 : Continuité des couches de terrains peu résistives, sur les profils 04, 08 et 02	56
Figure 31 : Histogramme du profil 05 avec valeurs extrêmes (A) et après suppressions des valeurs extrêmes	57
Figure 32 : Pseudo section du profil 05 après suppression des valeurs extrêmes.....	58
Figure 33 : Pseudo section du profile 06 après suppression des valeurs extrêmes	58

LISTE DES TABLEAUX

<i>Tableau 1: Les classes essentielles définies lors de la programmation du logiciel.</i>	40
<i>Tableau 2: Procédure d'implémentation des tâches principales.</i>	40
<i>Tableau 3 : Listes des outils utilisés lors de la mesure.</i>	45
<i>Tableau 4 : Informations de la courbe du profil 04</i>	55
<i>Tableau 5 : Informations de la courbe du profil 05</i>	55
<i>Tableau 6 : Informations de la courbe du profil 06</i>	55

RESUME

Les travaux présentés dans ce document portent sur la conception, la réalisation et l'implémentation d'un algorithme de simulation séquentielle gaussienne, pour résoudre des problèmes inverses gaussiens non-linéaires. Le document revient en entrée de jeu sur les fondements théoriques des méthodes d'inversions et des méthodes de simulations géostatistiques, pour ensuite relater les phases de conception, réalisation et application de cet algorithme à travers le logiciel ResSimSeq2D.

Le logiciel ResSimSeq2D a permis le traitement par inversion au sens géostatistique des données prélevées dans la localité de Djallo (environ 5 *km* de Meiganga). A travers ce logiciel il a été possible d'obtenir des modèles 1D de courbe de sondage et des modèles 2D de pseudo section. Les interprétations des courbes de sondages ont montré que le sous-sol du site d'expérimentation est composé d'environ quatre couche de terrain de résistivité et d'épaisseur distincte. A partir des pseudo section il a été possible d'apprécier la continuité des probables nappes aquifères dans le sol.

Malgré les résultats appréciables obtenus à travers le logiciel ResSimSeq2D, il a fait preuve de certaines limites quant à sa vitesse d'exécution lors des différents tests de la validation, qui ont cependant été légèrement amélioré, quoi que des meilleures améliorations sont encore envisageables.

Mots clés : Inversion gaussienne, simulations séquentielles, résistivités, krigeage ordinaire, algorithme génétique.

ABSTRACT

The aim of this essay is to present all the step converging to design, realize and apply a sequential gaussian program, to solve a non-linear inverse problem. In the present document, one will look at some theoretical basic concept discrete inverse theory and on basic concepts about geostatistical modeling, then follow the flow with those others parts of the essay, which presents all the step to achieve while designing and coding algorithms to produce a software named ResSimSeq2D.

ResSimSeq2D allows one to admin all the data collected during a tomography's survey. Those data were collected in Djallo, a small town nearby Meiganga (about 5 *km*). Using this software, one might be able to produce 1D models of vertical electric sounding, and 2D models of profile's section. While interpreting 1D models, one came out with the conclusion saying that on the studied field there is probably four different layers with distinct thickness and resistivity. One use 2D models to appreciate the spatial variability of the layer of interest.

Despite the good quality of all the data obtained, ResSimSeq2D shows some weak, about how he deals with data. One tried to solve and decrease that weakness, but they still better options to have a new version of ResSimSeq2D.

Keywords : Gaussian inversion, Sequential gaussian simulation, resistivity, Ordinary krigage, genetic programing.

INTRODUCTION GENERALE

Il est intéressant de porter un regard particulier sur l'usage des méthodes géostatistiques dans le domaine de la géophysique, où de nombreux auteurs s'en servent pour réaliser des programmes d'inversion des données géophysiques. Même si dans l'état de l'art il apparaît (cependant rarement) des avis contre-favorable quant à l'utilisation des méthodes de simulations séquentielles gaussiennes, comme c'est le cas dans les travaux de *Shahrokh Paravarzar et al. (2015)*, les simulations séquentielles gaussiennes sont apparues très tôt comme étant une alternative à la résolution des problèmes inverses ; c'est le cas par exemple des travaux de *Hansen et Klaus (2008)*, qui ont proposé une solution aux problèmes inverses linéaires par application des simulations séquentielles gaussiennes. Cette approche des problèmes inverses au sens stochastique est de plus en plus adoptée durant ces dernières années. En effet, les premières écoles de pensée liées aux méthodes de la théorie inverse discrète, abordaient l'inversion des données sur un aspect probabiliste (*William Menke, 1989*), où il fallait déterminer la distribution de probabilité des données mesurées et ensuite estimer les paramètres du modèle calculé qui permettent d'obtenir une distribution des données calculées très proche de celle des données mesurées.

Des travaux comme ceux de *Jim Yeh et al. (2002)*, par exemple, présentent une approche géostatistique par cokrigage, pour réaliser une inversion des données géoélectriques mesurées sur le terrain. Dans ces travaux, ils partent du principe qu'au sein d'une formation géologique étudiée, la différence de potentielle saurait expliquer les conductivités électriques, et vis-versa. A partir d'une fonction de covariance généralisée qu'ils vont remodeler itérativement pour pouvoir tenir compte de l'aspect non-linéaire de la conductivité électrique dans l'espace, ils parviennent à implémenter une modélisation par cokrigage et obtiennent à la fin des modèles 2D des approximations de résistivités de terrain.

D'autres comme ceux de *L. Ouadif et al. (2014)*, montrent qu'il est possible de réaliser des cartes de résistivité et autres modèles 2D des résistivités mesurées sur le terrain, à partir d'une

modélisation géostatistique à l'aide du krigeage ordinaire, qui nécessite au préalable une analyse variographique.

Contexte

Un des points essentiels qu'il faudra retenir ici est que, les phénomènes géoélectriques sont non-linéaires (cela est démontrable par les lois qui les gouvernent). De nombreuses méthodes classiques tiennent compte de cet aspect en ayant recours à des développements de Taylor, ou autre méthode de reparamétrage et linéarisation d'un problème non-linéaire. Cependant, en plus d'une bonne prise en compte de la structure spatiale des données mesurées, les méthodes géostatistiques de simulation séquentielle gaussienne se présentent comme étant une alternative plutôt pertinente, quant à l'entreprise d'une modélisation inverse qui saurait tenir compte de la non-linéarité des résistivités tout en apportant le maximum possible de véracité à la réponse du modèle inverse.

C'est dans cette idée que le travail qui sera présenté dans ce document, vient s'inscrire dans l'école de pensée portant sur le développement des méthodes d'inversions Gaussiennes non-linéaires, basées sur des concepts d'études de fonctions aléatoires et variables régionalisées de la géostatistique, ceci dans une situation d'application des méthodes d'inversion lors du traitement des données de résistivité mesurées à l'issue d'une campagne de prospection géoélectriques.

Problématique

Ce travail devrait permettre à la communauté scientifique et acteurs du domaine des géosciences, de répondre aux questions telles que « les simulations séquentielles gaussiennes ne sont-elles pas appropriées pour des inversions gaussiennes ? », ou encore, « les données de résistivités mesurées lors d'une campagne, ne suivent-elles pas une loi gaussienne pour permettre une utilisation des simulations séquentielles gaussiennes ? », ou alors « pourquoi devoir opter pour des méthodes lourdes d'hypothèses, s'il est possible d'en choisir une plus simple à mettre en œuvre et qui demande moins d'hypothèses ? ». C'est ainsi que le travail présenté dans ce document, viendra contribuer à la réponse au problème central de la prise en compte de la non-linéarité des phénomènes étudiés dans le domaine de la géophysique et des géosciences de manière générale, lors de l'application des méthodes d'inversions sur les données mesurées et traitées en vue d'une future interprétation.

Objectifs

L'objectif général visé dans ce travail est de proposer un algorithme de simulation séquentielle gaussienne, à utiliser pour réaliser une inversion gaussienne non-linéaire suivant une approche géostatistique. Plus spécifiquement il s'agit de :

- Rédiger un algorithme d'analyse variographique automatique à l'aide du principe des algorithmes génétiques ;
- Rédiger un algorithme de simulation séquentielle gaussienne comprenant l'analyse variographique automatique et une configuration de voisinage de calcul ;
- Concevoir et réaliser une plateforme logicielle pour mieux manipuler les données mesurées, et visualiser les résultats.

La présentation du travail reporté dans ce document sera structurée en quatre chapitres. Le premier chapitre qui portera sur les généralités des notions et outils suscités, le second qui portera sur la présentation des algorithmes de simulations qui ont été conçus, le troisième pour présenter la réalisation du logiciel et la validation de ses résultats, puis viendra un quatrième chapitre portant sur l'application du logiciel conçu à l'inversion des données issue d'une campagne de mesure géoélectriques.

Chapitre 1 : GENERALITES

Ce chapitre est consacré à la présentation des aspects théoriques des notions et outils de la géostatistique, ainsi que de l'approche probabiliste de la théorie inverse discrète et les méthodes d'inversion couramment utilisées pour réaliser des inversions des données géoélectriques.

De nos jours, de nombreux domaines des géosciences font de plus en plus appel aux méthodes géostatistiques, lors de l'analyse et du traitement des données relevant de la mesure d'une activité ou d'un phénomène naturel. Les raisons de ce penchant progressif sont multiples, mais nous pouvons entre autres citer le fait que la plupart des méthodes géostatistiques développées pour la modélisation des phénomènes naturels sont sans biais, et ne nécessitent pas une connaissance exhaustive du phénomène étudié ou de ses réalisations. Une autre raison qui viendrait s'ajouter aux dernières citées est qu'en générale, les méthodes géostatistiques ont la capacité, du fait de la prise en compte de la structure spatiale du phénomène étudié, de pouvoir relativement se rapprocher d'une représentation réaliste du dit phénomène. Les méthodes de la géostatistique sont nombreuses, allant des méthodes de type univariable (**krigeage simple et ordinaire, krigeage intrinsèque d'ordre k et universel**), aux méthodes de simulation (simulation des **processus de Weiner-Levy, simulation séquentielle gaussienne...**), ceci en passant par les méthodes multivariées (les **cokrigeages**). L'intérêt portera plus tard sur l'usage des méthodes de simulations. Les simulations géostatistiques peuvent être rangées en deux grandes familles, les simulations non-conditionnelles et les simulations conditionnelles. Par la suite l'intérêt sera particulièrement tourné vers les méthodes de simulations conditionnelles, plus spécifiquement les simulations séquentielles gaussiennes.

1.1 Les méthodes géostatistiques

La géostatistique dispose d'une grande variété de méthodes pouvant être utilisée pour la modélisation des variables régionalisées et des fonctions aléatoires. Ces méthodes peuvent être rangées en deux grands groupes de méthodes à savoir les méthodes linéaires et les méthodes non-linéaires. Pour chaque famille on distingue des méthodes univariées, comme le krigeage simple et ordinaire, les krigeages universel et intrinsèque d'ordre k, les krigeages d'indicateur et disjonctif, et des méthodes multivariées comme le cokrigeage simple, cokrigeage d'indicateur etc... Toutefois dans ce chapitre, il ne sera pas abordé les notions multivariées.

1.1.1 Analyse et modélisation de la structure spatiale des données

En amont de toute étude géostatistique figure très souvent une analyse variographique, qui consiste à attribuer un modèle de variogramme aux observations connues. En effet le variogramme est l'outil de description de la variable régionalisée par excellence.

1.1.1.1 Fonction de covariance

Soit $z(x)$ la variable régionalisée étudiée comme réalisation d'une fonction aléatoire $Z(x)$ stationnaire d'ordre deux. Dans ce cas, la fonction de covariance existe et possède les propriétés suivantes (Emery Xavier, 2001 ; Nicolas B. et al, 1955) :

$$C(h) = cov[Z(x+h), Z(x)] \text{ ne dépend que de } h ; \quad (1)$$

$$C(0) = var[Z(x)] = \sigma^2 \text{ constante ;} \quad (2)$$

$$C(h) = C(-h) ; \quad (3)$$

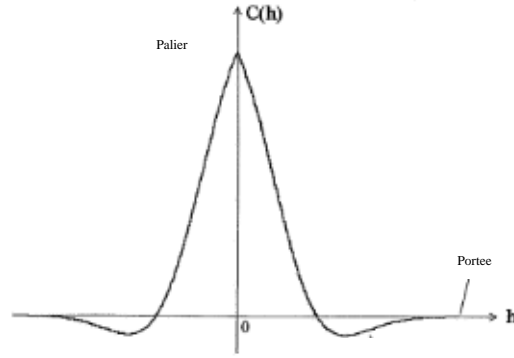


Figure 1: Exemple de fonction de covariance

1.1.1.2 Analyse variographique

Le variogramme est défini par la relation suivante :

$$\gamma(h) = \left(\frac{1}{2}\right) * E\{[Z(x) - Z(x+h)]^2\} ; \quad (4)$$

Cette relation donne la valeur théorique du variogramme ; cependant dans la pratique on ne dispose que du variogramme expérimental, à partir duquel il faudra ajuster un modèle théorique. L'estimateur du variogramme est donné par :

$$\gamma^*(h) = \left(\frac{1}{2 * |N(h)|} \right) * \sum_{N(h)} [Z(x) - Z(x + h)]^2 \quad (5)$$

où $N(h) = \{(x_i, x_j) \mid x_i - x_j = h\}$ et $|N(h)| = \text{Card}(N(h))$.

Après avoir calculé le variogramme expérimental, il sera ajusté à un modèle prédéfini. Il existe plusieurs modèles prédéfinis, deux d'entre eux sont présentées en guise d'exemple :

- Modèle sphérique (avec palier)

$$\begin{cases} \gamma(h) = Co + C * \left(1.5 * \left(\frac{h}{a} \right) - 0.5 * \left(\frac{h}{a} \right)^3 \right), & si \ h \leq a \\ \gamma(h) = Co + C, & si \ h > a \end{cases} \quad (6)$$

où a est la portée ; Co l'effet pépité ; $Co + C$ est le palier ; avec $Co = 0$;

- Modèle exponentiel (avec palier pratique)

$$\gamma(h) = Co + C * \left(1 - \exp\left(\frac{-h}{a}\right) \right) \quad (7)$$

avec $Co = 0$;

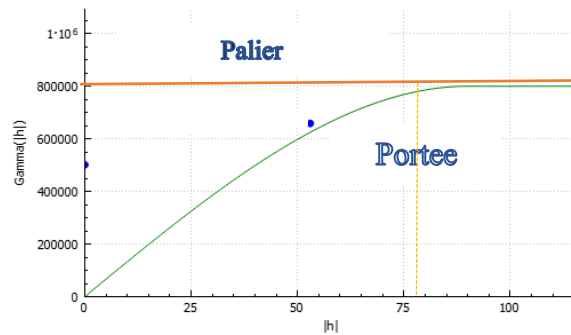


Figure 2: Modèle sphérique.

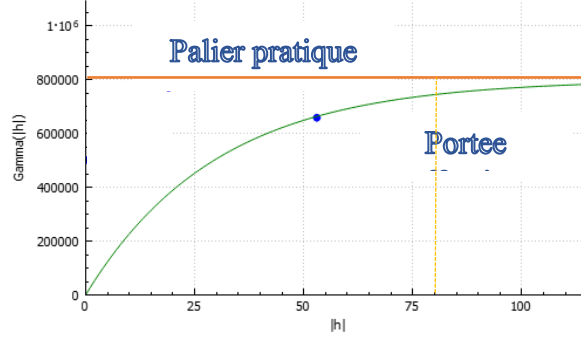


Figure 3: *Modèle exponentiel.*

1.1.2 Les méthodes univariables

Dans le domaine de la géostatistique on retrouve plusieurs outils utiles à l'estimation des valeurs d'un phénomène variant dans un espace géographique ; de telle sorte qu'il existe une certaine ambiguïté sur les critères de choix de l'outil adéquat face à un problème posé. Toutefois, ici se trouve une présentation des plus essentiels, autour desquelles s'articulent les autres. Il s'agira notamment du **krigeage ordinaire**, et du **krigeage simple**.

1.1.2.1 Le krigeage ordinaire

Encore appelé krigeage à moyenne inconnue, il correspond à la méthode d'estimation ponctuelle adéquate lorsque la moyenne du phénomène estimé est inconnue. En général, il est appliqué aux nœuds d'une maille régulière formée, pour évaluer (estimer) le paramètre étudié dans toute la zone d'étude, on parle alors de krigeage ponctuel (*Emery Xavier, 2001*).

Son application nécessite au préalable une **analyse variographique (modéliser un variogramme théorique à partir du variogramme expérimentale calculé)** ; étape qui apporte dans un sens une certaine complexité à la tâche, cependant tout en apportant dans un autre sens un degré de réalité appréciable à l'estimation.

Supposons que l'on veuille estimer un bloc v centré au point x_0 . Notons Z_v la vraie valeur (inconnue) de ce bloc et Z_v^* l'estimateur que l'on obtient.

On veut minimiser :

$$Var[Z_v - Z_v^*] = Var[Z_v] + Var[Z_v^*] - 2Cov[Z_v, Z_v^*] \quad (8)$$

Substituant l'expression de l'estimateur dans cette équation, on obtient :

$$\mathbf{Var}[\mathbf{Z}_v - \mathbf{Z}_v^*] = \mathbf{Var}[\mathbf{Z}_v] + \sum \sum \lambda_i \lambda_j \mathbf{Cov}[\mathbf{Z}_i, \mathbf{Z}_j] - 2 \sum \lambda_i \mathbf{Cov}[\mathbf{Z}_v, \mathbf{Z}_i] \quad (9)$$

Pour que l'estimateur soit sans biais, il faut que :

$$\sum_{i=1}^n \lambda_i = 1 ;$$

En minimisant cette variance d'estimation on obtient la relation suivante (*Emery Xavier 2001*) :

$$\left\{ \begin{array}{l} \sum_{j=1}^n \lambda_j \mathbf{Cov}[\mathbf{Z}_i, \mathbf{Z}_j] + \mu = \mathbf{Cov}[\mathbf{Z}_v, \mathbf{Z}_i], \quad \forall i = 1 \dots n \\ \sum_{j=1}^n \lambda_j = 1 \end{array} \right. \quad (10)$$

Ainsi se définit le système de krigeage ordinaire, avec pour variance la grandeur définit par :

$$\sigma_k^2 = \mathbf{Var}[\mathbf{Z}_v] - \sum_{i=1}^n \lambda_i \mathbf{Cov}[\mathbf{Z}_v, \mathbf{Z}_i] - \mu \quad (11)$$

Il est possible de définir le variogramme en fonction du covariogramme (de la covariance en quelque sorte), et cette relation peut être employée ici pour obtenir un système de krigeage ordinaire en fonction du variogramme :

$$\left\{ \begin{array}{l} \sum_{j=1}^n \lambda_j \gamma(x_i, x_j) - \mu = \bar{\gamma}(v, x_i), \forall i = 1 \dots n \\ \sum_{j=1}^n \lambda_j = 1 \end{array} \right. \quad (12)$$

Ayant à ce moment pour variance la grandeur définit par :

$$\sigma_k^2 = \sum_{i=1}^n \lambda_i \bar{\gamma}(v, x_i) + \bar{\gamma}(v, v) - \mu \quad (13)$$

Il est intéressant de visualiser le système de krigeage ordinaire sous forme matricielle :

$$\mathbf{K}_0 \lambda_0 = \mathbf{k}_0 \quad (14)$$

où

$$K_0 = \begin{bmatrix} \sigma^2 & Cov(Z_1, Z_2) & \dots & Cov(Z_1, Z_n) & 1 \\ Cov(Z_2, Z_1) & \sigma^2 & & Cov(Z_2, Z_n) & 1 \\ \vdots & & \ddots & \vdots & \\ Cov(Z_n, Z_1) & Cov(Z_n, Z_2) & \dots & \sigma^2 & 1 \\ 1 & 1 & & 1 & 0 \end{bmatrix};$$

$$k_0 = \begin{bmatrix} Cov(Z_1, Z_v) \\ Cov(Z_2, Z_v) \\ \vdots \\ Cov(Z_n, Z_v) \\ 1 \end{bmatrix}; \lambda_0 = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_n \\ \mu \end{bmatrix};$$

Cette forme est intéressante parce que celle-ci présente un aspect (matriciel) suivant lequel une méthode de résolution algébrique peut être implémentée pour calculer les différents éléments $(\lambda_1, \lambda_2, \dots, \lambda_n)$.

1.1.2.2 Le krigeage simple

Encore appelé krigeage à moyenne connue, il correspond à la méthode d'estimation ponctuelle adéquate lorsque la moyenne du phénomène estimé est connue de façon générale (empirique). Son contexte d'application est similaire à celui du krigeage ordinaire.

En effet en posant $Z^* = Y^* - m$ et $Z = Y - m$ on retrouvera les mêmes relations que celle régissant l'application du krigeage ordinaire.

1.1.3 Les simulations géostatistiques

Les méthodes géostatistiques ont été développées progressivement, les unes complétant et/ou comblant les lacunes des précédentes. Elles peuvent être réparties en deux types de méthodes. Les méthodes linéaires, qui sont les plus anciennes et qui ont été développées pour des cas de modélisation se rapportant à des résolutions d'équations linéaires (et des estimateurs définis sous forme de combinaisons linéaires), et les méthodes non-linéaires, pour la modélisation des phénomènes dénotant une relation non-linéaire avec les vecteurs de l'espace. Les méthodes de simulation s'inscrivent dans la deuxième famille (celle des méthodes non-linéaire). Elles viennent en effet s'ajouter aux précédentes méthodes non-linéaires de modélisation par espérance conditionnelle, par krigeage disjonctif (sur modèle mosaïque et modèle isofactoriel), en apportant comme atout le fait qu'elles ne nécessitent pas de fortes hypothèses pour être appliquées, ce qui n'est pas le cas des précédentes. *Jean-Paul CHILES (1995)*, dans ses travaux, présente les

simulations séquentielles gaussiennes comme étant une méthode appartenant à la famille des simulations conditionnelles. Il expliqua que lors d'une simulation séquentielle gaussienne, l'estimation d'une nouvelle réalisation est conditionnelle aux réalisations précédentes obtenues par la même simulation et aussi aux données initiales disponible pour décrire le problème posé. Des travaux réalisés durant ces dernières années montrent que les méthodes de simulations peuvent être considérées comme étant les méthodes les plus en vue dans la recherche scientifique en géostatistique fondamentale et appliquée ; il suffit de rentrer dans l'état de l'art du développement et amélioration des méthodes géostatistiques pour le confirmer.

Comme premier exemple, les travaux de *Masoomeh Delbari et al. (2009)*, mettent à profit les méthodes de simulations séquentielles gaussiennes, pour cartographier la distribution spatiale des teneurs en eau des sols d'un site d'étude en Autriche. L'étude a été réalisée à partir de 157 échantillons de sol et à partir des simulations ils ont été capable d'obtenir des résultats qui ont des histogrammes et semi-variogrammes qui épousent convenablement ceux des données mesurées. *Omid Asghari et al. (2009)*, parviennent à la conclusion que même si les méthodes de krigeage ordinaire sont de bon estimateur, les simulations offrent encore plus de qualité aux prédictions de la variable étudiée, ceci par le biais de ses multiples exécutions qui peuvent permettre d'établir un modèle d'incertitude de la variable étudiée.

Nasser M. et Omid A. (2012), se sont servis des méthodes de simulations séquentielles gaussiennes, pour évaluer la distribution spatiale des valeurs en RQD, obtenues à partir d'analyse des échantillons de forages prélevés sur un site d'étude (dépôt minéral de phosphate en Gazestan au centre de l'Iran) en vue d'en déduire les zones de faille possible sur le site d'étude. A l'issue de leurs travaux ils ont pu établir une carte de distribution des valeurs de RQD, en concluant sur le fait que, des valeurs de $RQD < 20\%$ correspondent à des zones de faille (probablement) et les zones de $RQD > 50\%$ représentent les endroits où le massif est encore sain.

Mingkai Qu et al. (2012), reviennent sur le fait qu'à cause de leurs effets de lissage, les méthodes de krigeage sont progressivement délaissées lorsqu'il s'agit d'estimer la distribution spatiale d'un phénomène, au profit des méthodes de simulations séquentielles gaussiennes qui savent mieux se rapprocher de la réalité. C'est pour cette raison qu'ils ont fait usage de ces dernières, pour simuler la distribution spatiale du nickel contenu dans un site d'étude dans la localité de Wuhan en Chine. Le but était à partir de ces simulations, de produire les modèles d'incertitudes qui ont été utilisés

par la suite pour évaluer les zones à risque sanitaire, et dégager des axes de proposition de mesure d'atténuation ou de correction. Les mêmes auteurs réalisent en 2014 des travaux similaires pour évaluer la contamination cette fois ci en métaux lourds, d'un autre site de la même localité, en se servant toujours des méthodes de simulations séquentielles gaussiennes, pour cartographier les zones aux risques écologiques potentiels. Il en est de même dans les travaux de *Albuquerque et al. (2013)*, où figure l'usage des méthodes de simulations séquentielles gaussiennes, pour modéliser la distribution spatiale de l'uranium contenu dans un cours d'eau transfrontalier entre le Portugal et l'Espagne. Après avoir réalisés une centaine de simulation à l'aide du logiciel Space-stat ils ont pu apprécier la variabilité spatiale de l'uranium et proposer de meilleurs axes de décision quant aux stratégies de corrections et/ou atténuations envisageables.

Fatemeh Soltani et al. (2013), ont quant à eux aussi mis en comparaison les méthodes de simulation séquentielle gaussienne et de krigeage ordinaire, lors de l'évaluation de la distribution spatiale en cuivre dans un dépôt minéralisé de la localité de Sungun (Nord-Ouest de l'Iran). Et à l'issue de leurs travaux, ils parviennent à mieux apprécier la distribution spatiale et le modèle d'incertitude des teneurs en cuivre à l'aide des simulations séquentielles car grâce à elles, il a été possible de quantifier les probabilités pour que dans l'espace ces teneurs prennent certaines valeurs.

D'autres travaux, comme ceux de *Pham T. D. (2015)*, relatent de l'usage des simulations gaussiennes comme filtre de signal lors d'un exercice de traitement d'image, leurs permettant ainsi de mesurer la variance et la variabilité du signal, ainsi que le bruit du signal, ceci sans avoir besoin d'une méthode d'apprentissage quelconque, contrairement à la méthode BLUE basée sur la théorie d'estimation statistiques, qui nécessite au préalable une méthode d'apprentissage pour dégager la structure spatiale du signal étudiée.

Il est possible de mentionner aussi les travaux de *Shuaipu Zhang et al. (2017)*, dans lesquels les auteurs devaient pouvoir prédire l'humidité du sol, en vue d'en déduire son évolution au cours de la variation des saisons, et en déduire des modèles de déshumidification. Pour cela, ils ont dû faire usage des méthodes géostatistiques de simulations séquentielles gaussiennes, appliquées à des données prises par intervalle de temps sur une surface d'étude, en justifiant leurs choix par le fait que ces méthodes offrent en plus d'une bonne prédiction spatiale, la capacité de déterminer les distributions par seuil d'humidité fixé en tenant compte de la probabilité que cette humidité prenne une valeur donnée. Ainsi, ils ont pu établir des cartes de distribution spatiale de l'humidité dans

leur zone d'étude, sur lesquelles ils pouvaient s'appuyer pour prédire l'irrigation naturelle de leur zone d'étude en fonction des saisons. Ici encore il est mis en évidence l'efficacité des méthodes de simulations séquentielles gaussiennes, en comparaison avec d'autres méthodes géostatistiques qui auraient pu produire aussi des résultats comparables.

Gabor Szatmari et László Pásztor (2018) ont mis à l'épreuve et en comparaison des diverses méthodes de cartographie digitale des sols notamment le krigeage universel, la simulation séquentielle gaussienne, la méthode de la forêt aléatoire associée au krigeage et la méthode de la forêt de régression des quantiles. Il s'agissait de voir lesquelles de ces méthodes (géostatistique d'une part, et du Machine Learning ou méthode d'apprentissage d'autre part) sont plus appropriées à l'exercice encouru. A l'issue de cette étude qui consistait à produire des modèles de distribution spatiale et d'en comparer leurs modèles d'incertitudes relativement à la prédiction spatiale, il en ressort que les simulations séquentielles gaussiennes (géostatistiques) et la méthode de la forêt de régression des quantiles (méthode d'apprentissage) sont des meilleurs outils de cartographie digitale des sols, car ces dernières ont présenté les meilleurs modèles d'incertitudes, contrairement aux deux autres qui ont tendance à surestimer les grandeurs mesurées sur le terrain.

Sabrine Methanie et al. (2019) ont réalisé des travaux dont l'objectif était de déterminer si les méthodes géostatistiques peuvent être convenablement utilisées pour estimer le volume de sol contaminés par des agents chimiques organiques et inorganiques, et comparer la capacité de chaque méthode à être utilisée dans un exercice de prédiction spatiale. Dans ce travail les auteurs comparent des méthodes d'interpolation de l'inverse des distances, polygonale, du krigeage ordinaire, et simulation séquentielle gaussienne, lors d'un exercice d'estimation de volume de sol contaminé. Pour les comparer, ils estiment la distribution spatiale des contaminants à partir de chacune de ces méthodes et apprécient les différents résultats obtenus, les uns par rapport aux autres. Les données utilisées pour ce faire ont été prélevées sur un site situé dans le Michigan (USA), contaminé à l'arsenic, au cuivre, au chrome, au polychlorophénole, aux dioxines et polychlorobenzo-furannes. A l'issue de leurs travaux, ils arrivent à la conclusion que la méthode de krigeage ordinaire est un meilleur estimateur comparé aux méthodes polygonale et d'inverse des distances, mais qu'elle a tendance à vouloir se rapprocher de la moyenne de la distribution (caractère lisse de la méthode), c'est pourquoi les simulations séquentielles gaussiennes, qui corrigent ce caractère de lissage sont d'avantage utilisées.

Les méthodes de simulations géostatistiques sont nécessaires pour des problèmes impliquant des transformations non-linéaires des grandeurs mesurées. Elles peuvent être regroupées en deux familles. Les simulations conditionnelles d'une part, pour lesquelles la prédiction obtenue par simulation doit suivre les valeurs issues de l'expérimentation, et les simulations non-conditionnelles, qui ne sont appelées à aucune opération de conditionnement. Dans les lignes suivantes, il sera présenté chacune de ces deux familles.

1.1.3.1 Les simulations non-conditionnelles

Une méthode de simulation non-conditionnelle vise à reproduire des champs montrant les mêmes structures spatiale (variogramme) et les mêmes histogrammes que les valeurs mesurées lors de l'expérimentation (*Richard Webster, 2007*). Les simulations non-conditionnelles peuvent être gaussiennes ou non-gaussiennes.

- **Méthode matricielle (gaussienne)**

En supposant qu'il faudra simuler n points montrant une covariance $C(h)$, il faudra construire la matrice de covariance $K_{n \times n}$ (identique à la matrice utilisée pour un krigeage ordinaire ou simple). Ensuite il suffira de décomposer cette matrice par la méthode de factorisation LU , avec $L = U^T$ (méthode de Cholesky), puis tirer n valeurs $y_i; i = 1, \dots, n$, aléatoirement d'une loi normale centrée réduite, et poser $z = Ly$; z pourra être alors considérée comme une réalisation de Z ayant le bon covariogramme. En effet il est démontrable que :

$$E[zz^T] = LU = K. \quad (15)$$

- **Méthode des bandes tournantes**

La méthode des bandes tournantes (*G. Matheron, 1973 ; Journel, 1974*), est la plus ancienne des méthodes géostatistiques utilisées pour simuler des données des processus aléatoires autocorrélés dans l'espace à trois dimensions. Elle permet d'obtenir des réalisations indépendantes suivant une direction, en partant d'un point centré sur le volume d'intérêt (région de l'espace étudié). Pour son application, il faudrait au préalable connaître la fonction de covariance $C^1(h)$, correspondant à la fonction de covariance $C^3(h)$ dans l'espace à trois dimensions.

Une documentation plus détaillée fournira plus amples informations sur les méthodes non-conditionnelles, car ce document se limite juste à une simple présentation de cette famille de simulation.

1.1.3.2 Les simulations conditionnelles

Les méthodes conditionnelles, en plus du variogramme et de l'histogramme des données mesurées, doit reproduire les valeurs de l'expérimentation aux différents points de mesure.

- **Méthode matricielle (avec conditionnement sur les données mesurées)**

(Richard Webster, 2007) Dans le cas d'une application de cette méthode associée à un conditionnement sur les données mesurées, il faut considérer N observations et T points non évalués. La matrice symétrique de covariance sera par conséquent de dimension $N + T$, et elle sera de la forme :

$$\begin{bmatrix} C_{NN} & C_{NT} \\ C_{TN} & C_{TT} \end{bmatrix} = \begin{bmatrix} L_{NN} & 0 \\ L_{TN} & L_{TT} \end{bmatrix} \begin{bmatrix} U_{NN} & U_{NT} \\ 0 & U_{TT} \end{bmatrix} \quad (16)$$

Les T valeurs sont tirées dans la loi normale centrée réduite, et forment le vecteur de données \mathbf{g} , et les N observations de conditionnement doivent suivre une loi gaussienne de distribution et forment le vecteur \mathbf{z} , de sorte que le vecteur \mathbf{y} suivent aussi une distribution gaussienne et soit donné par la relation suivante :

$$\mathbf{y} = \left[L_{TN}(L_{NN})^{-1} + L_{TT}\mathbf{g} \right] \quad (17)$$

Il peut être aisément démontré que la réalisation \mathbf{y} sera conditionnée par les observations \mathbf{z} , de même matrice de covariance aussi

.

- **Méthode du recuit simulé**

(Richard Webster, 2007) La méthode du recuit simulé est une méthode appartenant à une famille d'algorithme destinée par définition à résoudre les problèmes d'optimisation. Elle est basée sur le principe général de relaxation stochastique décrit par Kirkpatrick (Kirkpatrick et al., 1983). La méthode a été réadaptée au domaine de la géostatistique par Deutsch et Journel (Deutsch et Journel, 1992), et dans ce cas, la fonction objective à minimiser représente l'écart entre le

variogrammes des observations et le variogramme des réalisations par simulation, donnée par la relation suivante :

$$G_i = \sum_{m=1}^M [\gamma_i^*(h)_m - \gamma(h)_m]^2. \quad (18)$$

La procédure de simulation est résumée dans les étapes suivantes :

- **Etape 1 :** Calculer le variogramme des données mesurées et générées suivant la loi de distribution des données de l'expérimentation, T valeurs considérées comme une réalisation par simulation ;
- **Etape 2 :** Calculer la première valeur de G

$$G_i = \sum_{m=1}^M [\gamma_i^*(h)_m - \gamma(h)_m]^2 \quad (19)$$

- **Etape 3 :** Utiliser un algorithme de relaxation sur les valeurs $z(x_i)$ observées pour avoir la meilleure configuration de la réalisation à l'itération i ;
- **Etape 4 :** Retenir le variogramme précédent comme étant le variogramme courant et reprendre le tirage d'une nouvelle réalisation, jusqu'à ce que G_i soit relativement petit.

Plus tard dans ce document il sera fait mention d'une troisième méthode conditionnelle, celle des simulations séquentielles gaussiennes, qui est au centre des travaux réalisés.

1.2 Approche probabiliste de la théorie inverse discrète

La théorie inverse (ou l'inversion ou résolution du problème inverse) est un ensemble structuré de techniques d'analyse mathématique, appliqué sur des données généralement issues d'une expérimentation, dans le but d'en extraire une information utile sur le phénomène physique étudié. Ces informations seront dans la suite considérées comme étant les paramètres d'un modèle, qu'il faut obtenir par inversion des mesures de la réalité (*William Menke, 1989*). Il existe trois écoles de pensée concernant la théorie inverse. Il s'agit de :

- L'approche probabiliste (plus ancienne), pour laquelle les données mesurées et les paramètres du modèle sont considérés comme étant des variables aléatoires, et la tâche la plus grande à gérer est de déterminer la loi de distribution qu'ils suivent. Cette approche conduit naturellement à une analyse d'erreur et test d'hypothèse ;

- L'approche déterministe ;
- L'approche continue.

Le problème à résoudre connu sous le nom de problème inverse est une forme algébrique donnée par :

- $d = Gm$, dans le cas des problèmes inverses linéaires ;
- $d = g(m)$, dans le cas des problèmes inverses non-linéaires

Les méthodes d'inversion basées sur les concepts géostatistiques sont de la première école de pensée, seulement qu'à la place des principes purement statistiques, ces méthodes font usage d'outils de la géostatistique descriptive. Toutefois, il est important de revenir sur les concepts de base de l'approche probabiliste, ce qui constitue l'objet de ce chapitre.

Il est important de rappeler que dans la suite de ce chapitre, on désigne par fonction de distribution la fonction de densité d'une loi de distribution et cette dernière se confond par abus de langage, à la fonction de répartition qui est celle fournissant les véritables valeurs de probabilité par intervalle d'étude.

1.2.1 Mesures bruitées et variables aléatoires

Les mesures expérimentales sont pour la plupart du temps accompagnées d'incertitude sur la mesure, ou encore bruitées. Il a été par conséquent introduit dans la théorie inverse, des notions de probabilité et de variables aléatoires, pour mieux représenter l'aspect perturbé d'une mesure effectuée en un point. Chaque variable aléatoire possède des propriétés précises, qui peuvent être évaluées et utilisées pour décrire la variable aléatoire. Ces propriétés se résument dans la loi de distribution de probabilité de la variable aléatoire.

La distribution de probabilité d'une variable aléatoire (loi de probabilité), représente une fonction qui évalue la probabilité pour qu'une réalisation de la variable ait une valeur dans le voisinage d'une mesure d_i données. En effet la probabilité pour avoir une réalisation entre d_i et $d_i + \partial d$ est donnée par la relation $P(d)\partial d$. Les distributions de probabilité sont des bons moyens pour décrire les variables aléatoires, mais sont représentées par des fonctions continues, qui dans la plupart des cas sont compliqué à manipuler. **En général la valeur la plus recherchée quand il faut faire appel à cette approche, c'est la valeur de réalisation avec la plus grande probabilité, elle**

fournit une certaine description globale sur les données initiales. Puisque $P(d)$ est une probabilité (relativement) sur d , on peut écrire :

$$\int_{-\infty}^{+\infty} P(d) \partial d = 1 \quad (20)$$

Lorsque la distribution n'est pas symétrique, le point de maximum de vraisemblance n'est plus un bon descripteur des mesures effectuées. Dans cette situation, la moyenne ou espérance de la variable aléatoire (**v.a**) est un meilleur descripteur des données mesurées. Cette grandeur est :

$$E(d) = \int_{-\infty}^{+\infty} d \cdot P(d) \partial d \quad (21)$$

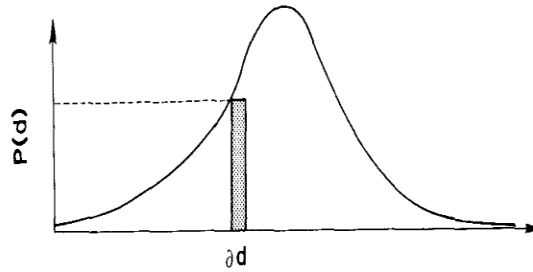


Figure 4: La partie grise donne la probabilité pour qu'une réalisation soit comprise entre d et $d + \partial d$.

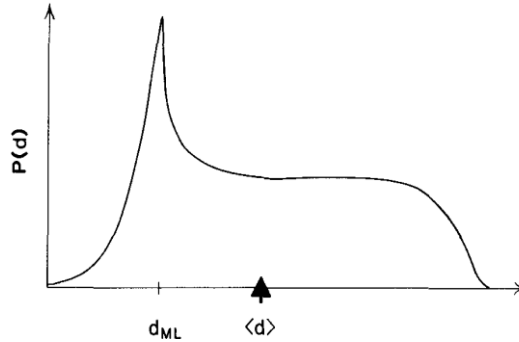


Figure 5: Le point de maximum de vraisemblance d_{ML} de la distribution $P(d)$ donne la réalisation qui a la plus grande probabilité d'occurrence. En général, cette valeur est différente de la moyenne des mesures d .

Un autre descripteur de distribution est connu sur le nom de 'largeur global' de la distribution. Une mesure quantitative de cette largeur correspond à l'aire en dessous de la fonction de distribution. En prenant comme fonction la parabole $(d - E(d))^2$, il est possible d'obtenir cette largeur, à partir de son espérance vaut la variance de la distribution et est donnée par :

$$\sigma^2 = \int_{-\infty}^{+\infty} (d - E(d))^2 P(d) \partial d \quad (22)$$

σ est la mesure la mesure de la largeur global de la distribution

1.2.2 Distributions Gaussiennes et test des hypothèses de statistiques gaussiennes

Extirper la loi de distribution d'une variable aléatoire donnée peut être un exercice assez particulier et compliquée, mais dans la majeure partie des cas, les données ont tendance à suivre une simple distribution gaussienne de loi donnée par :

$$P(d) = \frac{1}{\left(2\pi^{\frac{1}{2}}\right)\sigma} \exp \left[-\frac{(d - E[d])^2}{2\sigma^2} \right] \quad (23)$$

Elle a pour moyenne $E[d]$ et pour variance σ^2 . Les distributions gaussiennes sont les plus communes car elles expliquent beaucoup de phénomène du hasard (aléatoire), comme par exemple le concept des bruits aléatoires (provenant de plusieurs possible source différente) dont la distribution se rapproche d'une loi de distribution gaussienne.

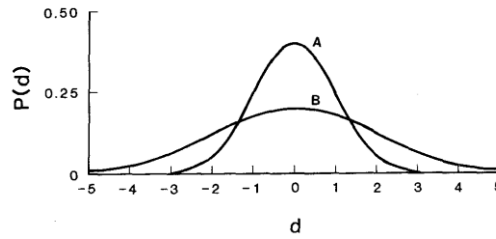


Figure 6 : Distribution gaussienne centrée réduite(A), et de variance 2 (B).

Généralement l'hypothèse voulant que les données mesurées suivent une distribution gaussienne est raisonnable ; néanmoins, il est important d'avoir certains moyens de tester cette hypothèse.

Tout d'abord, soit un ensemble v de variable aléatoire x_i , suivant une loi gaussienne centrée réduite. Soit ensuite la variable aléatoire :

$$\chi^2 = \sum_{i=1}^{Card(v)} x_i^2 \quad (24)$$

Cette variable aléatoire suit une loi de χ^2 à $card(v)$ degré de liberté. C'est une distribution unimodal de moyenne $Card(v)$ et de variance $2Card(v)$, et avec une loi de distribution fonctionnelle de la forme :

$$P(\chi^2, Card(v)) = \frac{[\chi^2]^{\frac{Card(v)-2}{2}} \exp\left[-\frac{\chi^2}{2}\right]}{2^{\frac{Card(v)}{2}} \Gamma\left(\frac{Card(v)}{2}\right)} \quad (25)$$

où Γ est la fonction gamma.

Ensuite, nous allons supposer que nous disposons d'une méthode de résolution du problème inverse. Assumons en plus que le modèle soit explicite, alors il est possible de calculer la variation entre les données mesurées et les paramètres estimés du modèle, à partir de $e = d - g(m^{est})$. La question est la suivante, cette erreur suit elle une loi gaussienne indépendante à variance uniforme ?

Pour répondre à cette question, il faut premièrement construire l'histogramme des erreurs, en s'assurant que chaque intervalle de l'histogramme ait la même largeur. On obtient ainsi un histogramme à unités d'aires, et chaque aire A_i de chaque intervalle est notée. On compare ces valeurs de A_i aux A'_i obtenues en utilisant une loi gaussienne de moyenne $E[e]$ et de variance $Var(e)$. La grandeur de comparaison est la suivante :

$$X^2 = \sum_{i=1}^p \frac{(A'_i - A_i)^2}{A'_i} \quad (26)$$

Pour des valeurs de cette grandeur proche de 0, on peut conclure que l'erreur suit une loi gaussienne. De la même manière, cette grandeur peut servir à évaluer l'hypothèse sur les données mesurées. En règle générale, lorsque $X_{obs}^2 > 5\%$, on ne devrait plus conclure sur une hypothèse gaussienne.

Il est démontrable que X^2 , peu importe la loi des x_i , suit une loi de χ^2 . Ainsi cette approche peut être utilisée pour tester d'autres hypothèses en plus d'une hypothèse gaussienne.

1.3 Application des notions de théorie inverse à la prospection électrique

La prospection électrique est une technique d'exploration de surface et subsurface basée sur l'étude des propriétés électriques des formations géologiques.

Elle puise en générale ses outils d'application, d'analyse et traitement dans les méthodes géophysiques électriques. Ces méthodes peuvent être regroupées en deux familles de méthodes, celles dites méthodes actives (méthode de courant continue, méthode du potentiel spontané...), et celles dites passives (méthode de la polarisation provoquée...). Une littérature plus appropriée sur la différenciation de chacune d'elles est fournie en référence (*H. Robert Bruger, 1992 ; Philip Kearey et al., 2002*). La méthode d'intérêt dans la suite sera celle de courant continu, et la grandeur mesurée et étudiée sera les résistivités des formations géologiques.

1.3.1 Principe de la méthode de courant continu

La méthode consiste à injecter à partir d'une ou deux électrodes un courant continu dans le sol, et à l'aide des autres électrodes, mesurer la différence de potentiel dans le champ de potentiel électrique créé par le passage du courant. L'objectif de cette approche est de pouvoir déterminer la distribution spatiale des résistivités des formations géologiques, à partir des mesures de différence de potentiel effectuées dans l'espace étudié.

La mesure des résistivités peut être réalisée en un point (ou encore point par point), ou suivant un profil de mesure orienté dans une direction précise. Fondamentalement, la différence de potentiel en un point ou entre deux électrodes, est liée à la résistivité mesurée par la relation :

$$V_r = \frac{\rho I}{2\pi r} ; \quad (27)$$

et pour des cas généraux, suivant le schéma de mesure (la configuration des électrodes de mesures), on a :

$$\nabla V = \frac{\rho I}{K} ; \quad (28)$$

K étant le facteur géométrique associé au schéma de prise des données. Il a été fourni dans la bibliographie des références pouvant donner plus de détails sur les principes et fondement de la

méthode des courants continus. Mais toutefois, il était nécessaire d'aborder de façon platonique les bases de cette méthode.

1.3.2 Inversion des données géoélectriques

Bien avant les méthodes basées sur les concepts géostatistiques, il existe des méthodes dites classique d'inversion de données géoélectriques.

En générale il existe deux approches pour étudier les problèmes inverse pour des données géoélectriques. Ces problèmes sont non-linéaires et par conséquent, une première approche est de générer des solutions itératives pour le problème $d = g(m)$, où m représente les paramètres du modèle et d les données mesurées sur le terrain, et une deuxième approche consiste à reparamétriser le problème en le linéarisant.

- **Approche non-linéaire**

L'exercice consiste à estimer m^{est} pour pouvoir calculer la prédiction donnée par la relation :

$$d^{cal} = g(m^{est}). \quad (29)$$

Il est possible de définir l'écart $\nabla d = d^{obs} - d^{cal}$, entre les données calculées et les données mesurées sur le terrain. Alors il faudra rechercher une valeur de m^{est} qui minimise cet écart.

Dans des problèmes de la forme explicite $f(x) = d - g(m) = 0$, et si les données et les paramètres à priori du modèle sont indépendantes, m^{est} s'obtient par une formule itérative donnée par (William Menke, 1984) :

$$\begin{cases} m_{n+1}^{est} = E[m] + G_n^{-g} \{d - g(m_n^{est}) + G_n[m_n^{est} - E[m]]\} \\ G_n^{-g} = \{[cov\ m]^{-1} + G_n^T [cov\ d]^{-1} G_n^T\}^{-1} G_n^T [cov\ d]^{-1} \end{cases} \quad (30)$$

où $[G_n]_{ij} = \left[\frac{\partial g_i}{\partial m_j} \right]_{i=1, \dots, N, j=1, \dots, M}$ (encore appelée matrice Hessienne) est évaluée en m_n^{est} , $E[m]$ (valeur

probable) et $[cov\ m]$ (covariance) sont des informations a priori sur les paramètres du modèle et sont disponible lorsque la distribution des paramètres est connue.

- **Approche linéaire**

Tout comme dans la sous-section précédente, il est toujours question de trouver m^{est} qui minimise l'écart ∇d . Mais seulement cette fois-ci, à l'aide d'un développement de Taylor sur $g(m)$ on peut écrire (*William Menke, 1984*) :

$$g(m) \cong g(m_n^{est}) + \nabla g[m - m_n^{est}] = g(m_n^{est}) + G[m - m^{est}]. \quad (31)$$

En posant $\nabla m = [m - m^{est}]$, alors on peut écrire :

$$\begin{cases} G_n \nabla m_{n+1} = d - g(m_n^{est}) \\ m_{n+1}^{est} = m_n^{est} + \nabla m_{n+1} \end{cases} \quad (32)$$

Il est important de rappeler que ceci est une présentation générale des méthodes de résolution des problèmes inverses non-linéaires au sens de la norme L_2 , qui sont les méthodes classiques les plus retrouvées dans l'état de l'art de la recherche traitant des méthodes d'inversion sur les données géoélectriques.

En revanche dans la suite du document, il figure l'implémentation, étape par étape d'une méthode d'inversion basée sur les concepts géostatistiques, qui consistera à utiliser les simulations séquentielles gaussiennes pour produire une distribution spatiale des résistivités calculées (par inversion). Les chapitres de la partie suivante font office de la présentation de cette implémentation.

Chapitre 2 : ALGORITHME DES SIMULATIONS ET MATERIEL DE REALISATION DE ResSimSeq2D

Dans ce chapitre figure la présentation des algorithmes qui implémentent les simulations séquentielles gaussiennes, ces algorithmes qui pour finir seront traduits en un langage de programmation pour concevoir un programme de simulations géostatistiques proprement dit. Il s'agit des algorithmes de simulations séquentielles, des algorithmes de modélisation automatique des variogrammes isotropes à l'aide de programme génétique intelligent et aussi entre les lignes, des algorithmes pour des fonctions de second plan tout aussi important, comme l'algorithme de recherche du voisinage de calcul.

2.1 Algorithme de simulation séquentielle gaussienne

Les simulations séquentielles gaussiennes appartiennent à la famille des méthodes de simulations conditionnelles (*Jean-Paul CHILES, 1995*). Elles ont été utilisées dans les travaux présentés dans ce document pour développer une méthode d'inversion pour des problèmes non-linéaires. En effet le but de l'inversion de façon générale, est de pouvoir fournir des paramètres d'un modèle pouvant expliquer les mesures effectuées, qui permettront de produire des prédictions plus régulières que les données mesurées, et que l'on pourra interpréter pour en tirer des informations importantes sur le phénomène physique étudié. Mais comme cela a été mentionné dans le chapitre 1, il peut aussi s'agir de prédire par inversion une distribution de valeurs calculées qui se rapprocherait le mieux possible de celle des données mesurées lors de l'expérimentation. C'est ainsi qu'à l'aide des simulations séquentielles gaussiennes, il faudra produire des valeurs calculées ayant des distributions proches de celles des données mesurées.

Ce chapitre abordera le principe de la méthode de simulation géostatistique séquentielle gaussienne, ensuite viendront des sections traitant de la rédaction des algorithmes qui gouvernent la simulation implémentée.

2.1.1 Principe des simulations séquentielles gaussiennes

(*Jean-Paul CHILES, 1995*) Soit à simuler $Z_i = Z(x_i)$ aux points $x_i, i = 1, \dots, N$. Le principe est le suivant ; il faut tirer Z_1 dans sa loi a priori, puis Z_2 dans sa loi conditionnelle à Z_1 , ensuite Z_3 dans sa loi conditionnelle à Z_1 et Z_2 , etc... Il faudrait au préalable se rassurer que les données

mesurées suivent une loi gaussienne, et de cette façon la loi de Z_i conditionnellement à Z_1, \dots, Z_{i-1} sera gaussienne, de moyenne Z_i^* estimée par krigeage, et de variance σ_{Ki}^2 correspondant à la variance de krigeage obtenue en estimant Z_i^* . Les données de conditionnement comprennent tout d'abord les données mesurées lors de l'expérimentation, mais aussi les données des points précédemment simulés.

La simulation séquentielle gaussienne par du principe que pour le premier point à simuler, l'erreur d'estimation par krigeage suit une loi normale de moyenne nulle et de variance $\sigma_K^2(x_0)$. Rappelons que pour une variable y qui suit une loi normale de paramètre m_y et de variance σ_y^2 , la variable $Y = \frac{y-m_y}{\sigma_y}$ suit la loi normale centrée réduite.

Pour implémenter la simulation il faut suivre les étapes suivantes :

- **Etape 1 :** Tracer l'histogramme des données mesurées pour se rassurer qu'elles suivent une loi gaussienne (et en cas de doute il faudra réaliser un test d'hypothèse) ;
- **Etape 2 :** Calculer le variogramme expérimental des données et modéliser leur variogramme théorique (et isotrope) ;
- **Etape 3 :** Configurer la maille de calcul, et sélectionner un nœud de la maille auquel il va falloir réaliser une simulation ;
- **Etape 4 :** Définir un voisinage de calcul ;
- **Etape 5 :** Calculer Z_i^* et σ_{Ki}^2 à l'aide d'une méthode de krigeage ordinaire pour estimer les paramètres de la distribution de Z_i , ensuite on pose $Z_i = Z_i^* + U \sigma_{Ki}$, où U est une réalisation tirée dans la table de la loi normale centrée réduite ; cette nouvelle valeur de Z_i sera insérée dans la liste des données de conditionnement pour la prochaine simulation à effectuer, les étapes 1 à 5 sont répétées pour parcourir toute la grille créée ;
- **Etape 6 :** Dans la mesure où les données auraient été transformées initialement pour permettre l'application des simulations gaussiennes, il faudra appliquer la transformation réciproque sur les réalisations par simulation.

- **Cas particulier du processus de Wiener-Levy** (*Jean-Paul CHILES, 1995*)

Il s'agit d'un cas particulier à une dimension, pour lequel la méthode séquentielle est facilement utilisable. Le processus à modéliser est considéré comme étant gaussien à accroissement

indépendant, et de variogramme $\gamma(h) = \omega|h|$, avec $\omega > 0$. Sa simulation en un ensemble de point $x_i, i = 0, \dots, N$, à partir d'une valeur initiale a priori $Z_o = Z(x_0)$ peut être itérativement obtenue par la relation suivante :

$$Z(x_i) = Z(x_{i-1}) + U_i \quad (33)$$

où U_i est une variable aléatoire gaussienne centrée et de variance $2\omega(x_i - x_{i-1})$, indépendantes des variables aléatoires U_j précédentes.

2.1.2 Conception d'un algorithme de simulation séquentielle gaussienne

L'algorithme va être établi pour le cas 1D, correspondant à l'inversion des données de sondages électriques verticaux, et 2D pour l'inversion sur les données de tomographie.

2.1.2.1 Algorithme de simulation : cas à une dimension (1D)

Pour cet algorithme il existe deux possibilités ; la première étant de suivre le schéma général d'implémentation d'une simulation séquentielle, et la deuxième étant de procéder par la simulation des processus de Wiener-Levy. Les deux algorithmes ont été conçus mais nous avons retenu la première approche pour des raisons qui seront abordées plus tard dans ce document. Néanmoins, dans cette section figure la présentation des deux algorithmes.

a- Approche suivant le principe général des simulations séquentielles

```
/* Etape 1 */
/* Au préalable il faut initialiser les variables listeClasse et probaClasse; et on va considérer que
chaque variable utilisée a été bien déclarée */
//Calcul des probabilités de chaque classe
pasClasse <- (max(Rho) - min(Rho))/nbreClasse ; /* nbreClasse correspond au nombre de classe définit
par l'utilisateur */
minRho <- min(Rho) ; // pour contenir la plus petite valeur du tableau des Rho

listeClasse[0] <- 0.0 ;
Pour (i <- allant de 1 à nbreClasse) faire
    listeClasse[i] = (minRho + i*pasClasse) ;
    nbreElt <- 0 ;
    Pour (j <- allant de 1 à taille(Rho)) faire
        Si (Rho[j] >= listeClasse[i]&&(Rho[j] < listeClasse[i+1]) )
            nbreElt += 1 ;// compte le nombre d'element par classe
        finSi
    probaClasse[i] <- nbreElt/taille(Rho) ;
    finPour
finPour
// L'histogramme recherché sera le diagramme en bande de probaClasse = f(listeClasse)

/* Etape 2 */
```

```

Ecrire(" Etape 2 : modélisation du variogramme. M | A ") ; // Manuellement ou modélisation automatique
Lire(ChoixMethd)
Si( ChoixMethd == M ) Alors
    //Calcul du variogramme expérimental
    listeEquidt <- strideComputing(x) ; //Calcul la liste des équidistances possibles
    varioValeur <- varioComputing(listeEquidt, Rho) ; //Pour calculer le variogramme expérimental
    Repeter
        Ecrire("Entrer les parametres du modeles theorique : a , Co, et C") ;
        Lire(a, Co, C) ;
        ecartVarioGram <- ecartVarioComputing(varioValeur, a, Co, C) ; /* Pour calculer l'écart
entre le modèle théorique et le modèle expérimental */
        Jusqu'a(CndArret) // CndArret = ecartVarioGram relativement petit
Sinon
    //Calcul du variogramme expérimental
    listeEquidt <- strideComputing(x) ; //Calcul la liste des équidistances possibles
    varioValeur <- varioComputing(listeEquidt, Rho) ; //Pour calculer le variogramme expérimental
    varioTheoriqValeur <- GA_VarioComputing(varioValeur) ;
finSi

/* Etape 3 & 4 */
listeNoeud <- gridCreation(x) ; // Pour créer les nœuds de simulation
voisinagePoint <- neighborConfig(x , 150) ; /* Le système de krigeage est efficace pour moins de 150 point
d'observations */

/* Etape 5 */
K <- covarianceMatrice(x, a, Co, C) ;
krigCoef <- LU_Facto(K) ;
Zi <- krigeageComputing(krigCoef) ; // Fonction de calcul par krigeage
varZi <- krigVariance(krigCoef, a, Co, C) ;
Ui <- drawRandGauss(Zi, varZi) ;
Zi <- Zi*Ui + varZi ;

Si( i < N)
    goto LABEL ;
finSi

```

b- Cas de simulation des mouvements brownien ou processus de Wiener-Levy

Dans ce cas présent le voisinage sera tout simplement les deux point x_i et x_{i+1} tel que, pour un point x à simuler, on a : $x_i < x < x_{i+1}$. La moyenne de la distribution est alors obtenue par :

$$Z_i^* = \frac{((x_{i+1}-x)Z(x_i)+(x-x_i)Z(x_{i+1}))}{x_{i+1}-x_i} \quad \text{et} \quad \sigma_K^2 = \frac{2\omega(x-x_i)(x_{i+1}-x)}{x_{i+1}-x_i}$$

L'algorithme de la première étape reste identique à celui rédigé pour le cas précédent. A la deuxième étape l'algorithme de modélisation automatique du cas précédent, qui utilise une méthode de modélisation par algorithme génétique (nous y reviendrons dans le chapitre suivant), sera remplacé par un algorithme d'ajustement linéaire (puisque pour les processus Wiener-Levy le variogramme est linéaire). On aura l'algorithme suivant :

```

/* Etape 2 */
Ecrire(" Etape 2 : modélisation du variogramme. M | A ") ; // Manuellement ou modélisation automatique
Lire(ChoixMethd)

```

```

Si( ChoixMethd == M ) Alors
  //Calcul du variogramme expérimental
  listeEquidt <- strideComputing(x) ; //Calcul la liste des équidistances possibles
  varioValeur <- varioComputing(listeEquidt, Rho) ; //Pour calculer le variogramme expérimental
  Repeter
    Ecrire("Entrer les paramètres du modèles théorique :  $\omega$ ") ;
    Lire( $\omega$ ) ; //Le seul type de modèle disponible c'est le modèle linéaire
    ecartVariogram <- ecartVarioComputing(varioValeur,  $\omega$ ) ; /* Pour calculer l'écart entre le
modèle théorique et le modèle expérimental */
    Jusqu'a(CndArret) // CndArret = ecartVariogram relativement petit
Sinon
  //Calcul du variogramme expérimental
  listeEquidt <- strideComputing(x) ; //Calcul la liste des équidistances possibles
  varioValeur <- varioComputing(listeEquidt, Rho) ; //Pour calculer le variogramme expérimental
  varioTheoriqValeur <- MCO_Computing(varioValeur) ; //Ajustement par moindres carrés ordinaires
finSi

```

La troisième étape ne changera pas, et la quatrième consistera tout simplement à rechercher les deux points les plus proches de part et d'autre du nœud de calcul. L'algorithme de la cinquième étape sera à l'image de celui rédigé dans le cas précédent, à la différence que la fonction de calcul par krigeage sera remplacée par la nouvelle expression de l'estimation de la moyenne, et le tirage sera différent.

2.1.2.2 Algorithme de simulation : cas à deux dimensions (2D)

Les 06 étapes à suivre pour implémenter une simulation séquentielle s'applique aussi bien en une dimension qu'en deux dimensions (de l'espace). Cependant, jusqu'à présent dans la littérature consacrée au développement des méthodes séquentielles gaussiennes, il ne figure pas un cas particulier de simplification semblable à celui du processus de Wiener-Levy à une dimension. L'algorithme pour les cas à deux dimensions reste identique à celui du cas général à une dimension. La différence va principalement se trouver au niveau des opérations de calcul des distances entre deux points, et aussi au niveau de la gestion du voisinage de calcul.

Le voisinage du calcul de Z_i^* est un facteur très déterminant pour la simulation. Il est préférable de choisir comme voisinage en 2D, un espace centré en x , x étant le point de calcul. La forme géométrique la plus simple pouvant répondre à cette condition est celle d'un cercle.

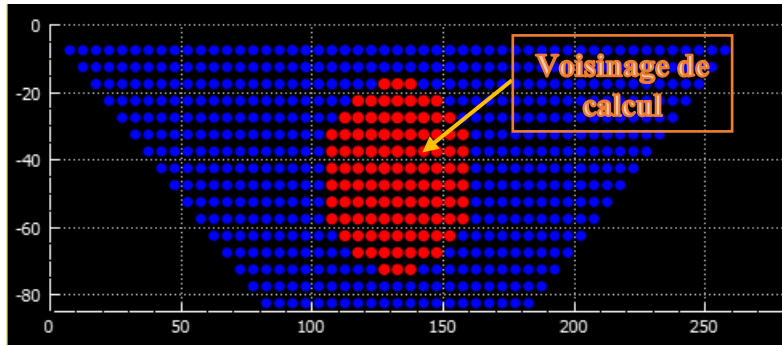


Figure 7 : Configuration d'un voisinage de calcul

Toutefois, en configurant le voisinage il ne faut pas oublier que pour plus de 150 points d'observation, l'estimation par krigeage n'est plus une opération stable (Jean-Paul CHILES, 1995 ; Xavier EMERY, 2001). Pour implémenter proprement la sélection du voisinage, l'algorithme suivant a été utilisé :

```
/* Voisinage */
voisinageCoord[2][] ;

P0[2] ; //Les coordonnées du nœud de calcul
P0[0] <- Xo ;
P0[1] <- Yo ;

Pour (i <- 1 allant à taille(x)) faire
    P1[2] ;
    P1[0] <- x[k] ;
    P1[1] <- y[k] ;
    tempDist <- distanceC(P0, P1);

    Si(tempDist <= lagMax) Alors
        voisinageCoord[1][i] <- (x[k]) ;
        voisinageCoord[2][i] <- (y[k]) ;
    finSi
finPour
```

L'analyse variographique représente aussi une étape très importante pour la simulation séquentielle gaussienne. En effet, puisqu'à chaque nouvelle réalisation la valeur simulée précédemment est enregistrée dans la liste des données de conditionnement, il apparaît clairement que le variogramme des données de conditionnement doit subir une mise à jour. Il faudrait par conséquent que l'algorithme de simulation puisse automatiquement calculer et modéliser un nouveau variogramme à chaque nouvelle réalisation, pour éviter que cette tâche soit faite manuellement. Le processus automatique de l'analyse et la modélisation variographique sera abordé dans le chapitre suivant.

L'algorithme de simulation séquentielle gaussienne peut finalement être résumé par l'algorithme suivant :

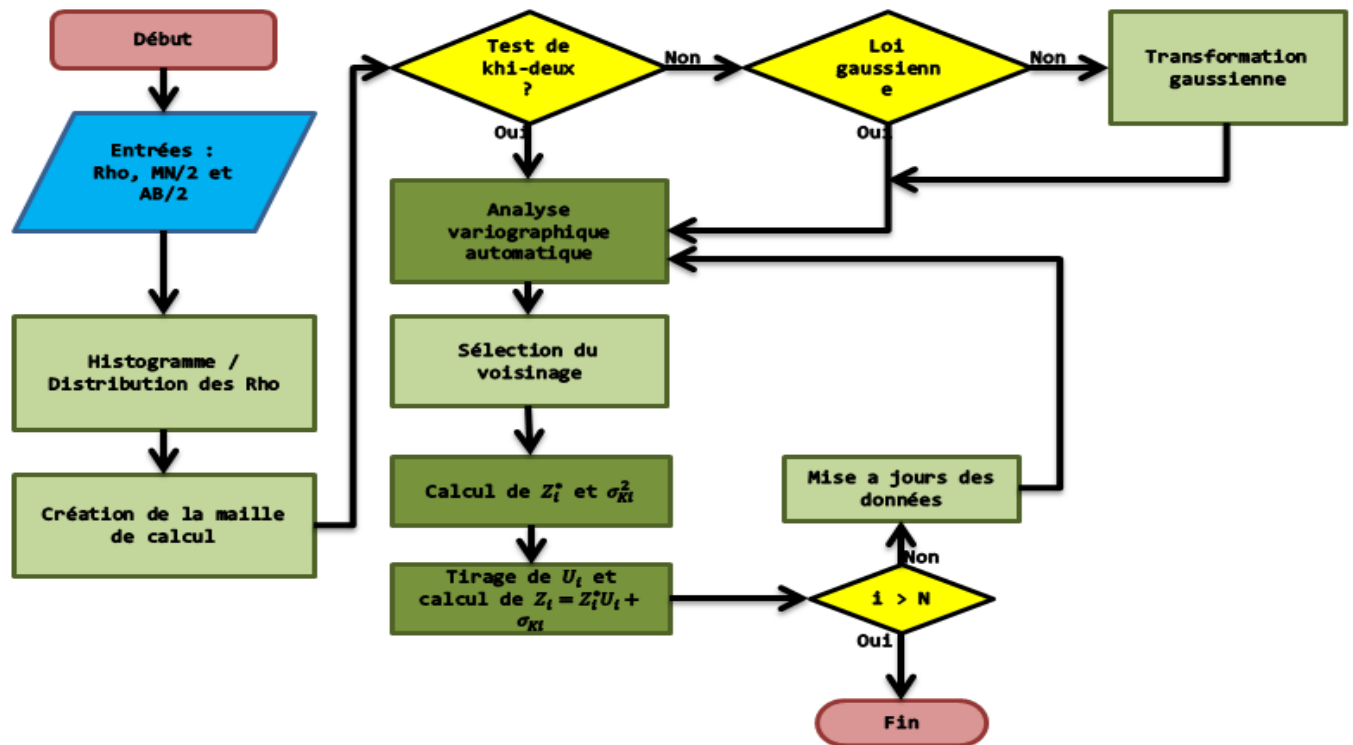


Figure 8: Algorithme de la méthode par simulation séquentielle gaussienne

2.2 Analyse variographique automatique et semi-automatique

L'analyse variographique est considérée comme étant la clé de toute étude géostatistique. En effet, pour être en mesure de produire un modèle qui possèdera la même structure spatiale que les données mesurées, il faudrait être capable de faire ressortir le modèle de variogramme de ces données. L'exercice consistant à obtenir ce modèle a été introduit et présenté dans le chapitre des généralités sur les méthodes géostatistiques. Seulement, comme il en a été fait mention dans le chapitre précédent, pour les algorithmes de simulations, l'analyse variographique devient très couteuse lorsqu'elle se fait manuellement. C'est pourquoi il a fallu au cours des travaux qui ont été réalisés, penser à dresser des algorithmes de modélisation automatique des variogrammes. Dans la littérature traitant de ce sujet, il apparait régulièrement deux approches. La première porte sur l'usage de la méthode des moindres carrés pour produire des modèles de variogramme

théoriques, comme c'est le cas par exemple dans les travaux de *Desassis et Renard (2012)*, et la deuxième approche couramment retrouvée porte sur l'usage des algorithmes génétiques pour réaliser la même tâche, c'est le cas par exemple des travaux de *Zhang et al. (2018)*.

2.2.1 Principe et algorithme de la modélisation par la méthode des moindres carrés

Le variogramme expérimental peut être considéré comme étant un vecteur de n valeurs numériques $\gamma = [\gamma_1, \gamma_2, \dots, \gamma_n]$, calculées pour n equidistances $h = [h_1, h_2, \dots, h_n]$. L'exercice consiste à écrire le modèle théorique sous la forme d'un polynôme de degré Q , donné par :

$$\gamma^*(h) = \sum_{q=0}^Q A_q ||h||^q, \quad (34)$$

et par la suite rechercher les valeurs de A_q qui minimise l'écart :

$$Err = \sum_{i=1}^n er_i^2 = \sum_{i=1}^n (\gamma_i^* - \gamma_i)^2. \quad (35)$$

Afin d'obtenir les paramètres $A = [A_1, A_2, \dots, A_Q]$, on cherche les racines des dérivées premières de Err , en prenant les A_q comme étant les variables inconnues.

2.2.1.1 Cas particulier d'un ajustement linéaire

Si le variogramme théorique est supposé linéaire, alors on peut écrire :

$$\gamma^*(h) = A_0 + A_1 ||h||, \quad (36)$$

et il est possible de déterminer A_0 et A_1 par la méthode des moindres carrés. L'écart quadratique Err peut être réécrit comme suit :

$$Err = \sum_{i=1}^n er_i^2 = \sum_{i=1}^n (A_0 + A_1 h_i - \gamma_i)^2 \quad (37)$$

Les premières dérivées de Err sont :

$$\begin{cases} \frac{\partial Err}{\partial A_0} = 2 \sum_{i=1}^n (A_0 + A_1 h_i - \gamma_i) \\ \frac{\partial Err}{\partial A_1} = 2 \sum_{i=1}^n h_i (A_0 + A_1 h_i - \gamma_i) \end{cases} ; \quad (38)$$

et en recherchant leurs racines on obtient le système suivant :

$$\begin{cases} nA_0 + A_1 \sum_{i=1}^n h_i = \sum_{i=1}^n \gamma_i \\ A_0 \sum_{i=1}^n h_i + A_1 \sum_{i=1}^n h_i^2 = \sum_{i=1}^n h_i \gamma_i \end{cases}$$

Ce système peut être écrit sous la forme matricielle $G\lambda = Y$, où :

$$G = \begin{bmatrix} n & \sum_{i=1}^n h_i \\ \sum_{i=1}^n h_i & \sum_{i=1}^n h_i^2 \end{bmatrix} ; \quad \lambda = [A_0, A_1]^T \text{ et } Y = \begin{bmatrix} \sum_{i=1}^n \gamma_i & \sum_{i=1}^n h_i \gamma_i \end{bmatrix}^T$$

Étant donné que $\det(G) \neq 0$, en posant $\lambda = [G^T G]^{-1} G^T Y$ (ou à partir de tout autre méthode de résolution du système formé) on obtient les valeurs de A_0 et A_1 .

2.2.1.2 Algorithme de la méthode pour $Q = 1$

```
/* Moindres Carrés Ordinaires de degré 1 */
G[2][2]; Y[2]; Y[1] <- 0.0; Y[2] <- 0.0;
cA <- 0.0; B <- 0.0;
Pour (i <- 0 allant à taille(h)) faire
    A <- A + h[i]*h[i] ;
    B <- B + h[i] ;
    coefB[0] <- coefB[0] + (h[i])*(g[i]) ;
    coefB[1] <- coefB[1] + g[i] ;
finPour

G[0][0] <- taille(h); G[0][1] <- B;
G[1][0] <- B; G[1][1] <- A ;

//Factorisation LU
U <- G ;
L <- factoLU(U);

sY[taille(Y)] ;
sY[0] <- Y[0] ;
Pour( i <- 1 allant à taille(Y)) faire
    sS <- 0.0;
```

```

    Pour( l <- 1 allant à i) faire
        sS <- sS + (L[i][j])*sY[j];
    finPour
    sY[i] <- Y[i] - sS;
finPour

nn <- taille(sY);
sX[nn];
sX[nn-1] <- sY[nn-1] / U[nn-1][nn-1];
Pour( k <- (nn-2) allant à 0) faire
    sS <- 0.0;
    Pour( l <- (k+1) allant à nn) faire
        sS <- U[k][l]*sX[l];
    finPour
    sX[k] <- (sY[k] - sS)/U[k][k];
    k <- k - 1;
finPour
coefMCO <- sX ;

//Modèle de variogramme théorique obtenu
MCO_VarioVal[taille(h)] ;
Pour(k <- 0 allant à taille(h)) faire
    MCO_VarioVal[k] <- h[k]*coefMCO[1] + coefMCO[0] ;
finPour

```

Les algorithmes pour des cas de degré deux et trois ont aussi été implémenté, et suivent la même structure. Des essais sur des données issues de différentes expérimentation montre que le modèle linéaire n'est pas approprié pour ajuster les variogrammes de résistivité, et c'est la raison pour laquelle les processus de Wiener-Levy n'ont pas été retenus pour les cas de simulation 1D.

Après la modélisation par la méthode des moindres carrés, il faut présenter la deuxième approche, basée sur l'usage des algorithmes génétiques pour obtenir à l'aide d'un programme intelligent le modèle de variogramme théorique.

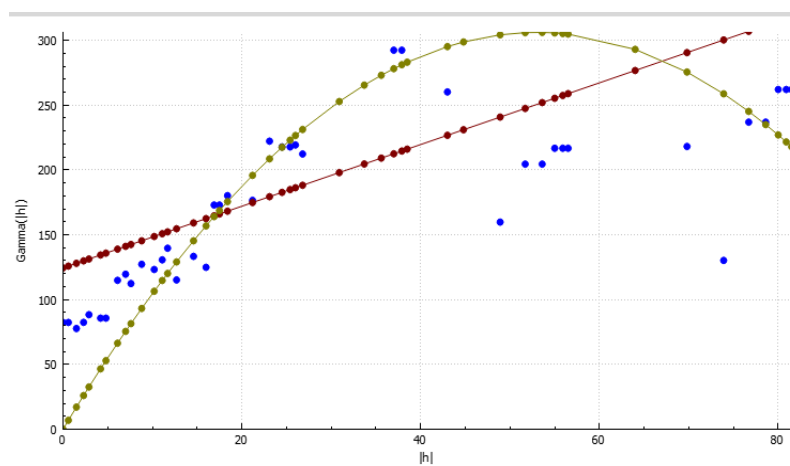


Figure 4.1 : Variogramme expérimental en bleu (en pointillés), modèle linéaire obtenu par moindres carrés en rouge (degré 1), et en couleur de bronze (degré 2)

2.2.2 Analyse variographique intelligente à l'aide des algorithmes génétiques

Le principe des algorithmes génétiques a été proposé pour la première fois par John Holland (*K.F. Man et al., 2001*). Il est inspiré du mécanisme de sélection naturelle, pour lequel des individus plus fort résistent et survivent mieux dans un environnement devenant de plus en plus hostile et compétitif.

Les méthodes d'algorithme génétique présument que la solution probable à un problème posé est individuelle et peut être représentée par un ensemble de paramètres. Ces paramètres sont à l'image des gènes d'un chromosome, et peuvent être codés (ou structurés) en une chaîne de caractères du système binaire (pour le codage binaire) ou tout autre chiffage (codage réel, codage alphanumérique...). La qualité d'une solution est évaluée à partir d'une valeur positive appelée '*fitness value*'. Cette valeur devrait être liée à la fonction objective à optimiser par application des algorithmes génétiques.

L'implémentation des méthodes de programmation génétique passe par l'usage des opérateurs génétiques tels que la sélection, les mutations et les croisements. Des meilleures spécifications quant à la définition et l'usage de chacun de ces opérateurs sont fournies dans la documentation en référence (*K.F. Man et al., 2001 ; Zhanglin Li, 2018*), mais il est utile de présenter sommairement leurs fonctionnalités. Initialement, le programme génétique génère une liste de solution à évaluer pour résoudre le problème posé ; cette liste est appelée population initiale :

- **Sélection** : la sélection consiste à tirer un nombre d'individu dans la population initiale, elle peut être effectuée suivant un schéma précis, à partir d'un critère de sélection, ou alors tout simplement de façon aléatoire ;
- **Croisements** : les croisements sont des opérations d'échange de gènes entre deux chromosomes, l'échange peut s'effectuer point par point, suivant une série de point, mais échanger toute l'information génétique contenues dans chaque chromosome n'est pas envisageable ;
- **Mutations** : les mutations correspondent à l'altération (la transformation) d'un gène ou d'une série de gènes pour un chromosome donné.

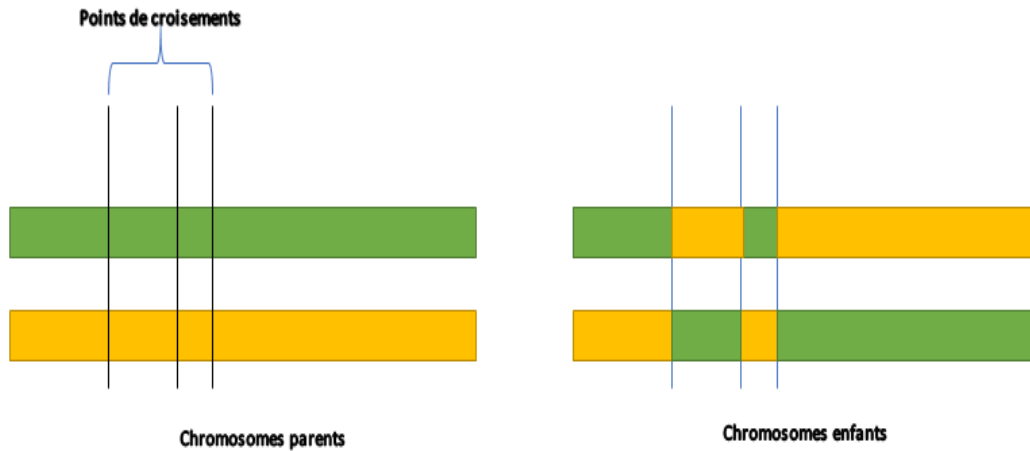


Figure 9 : Croisement entre deux chromosomes

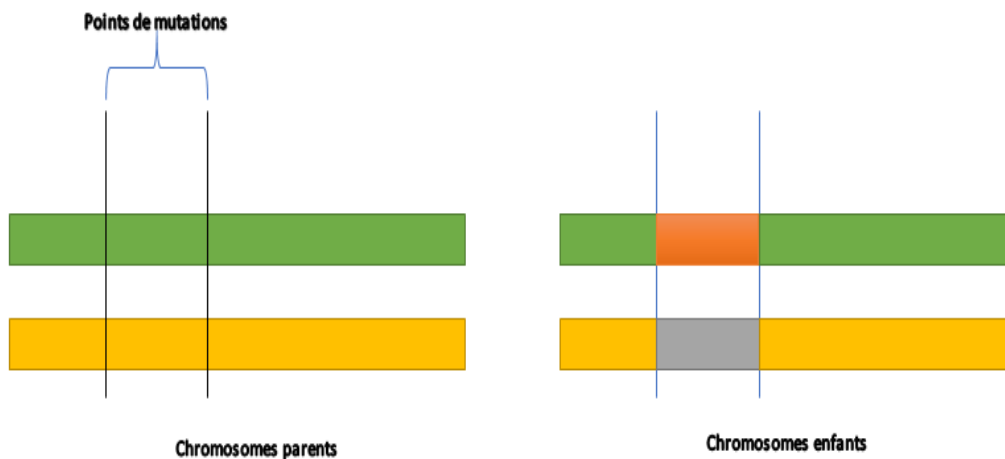


Figure 10 : Mutations d'une série de gène sur chaque chromosome

2.2.2.1 Formulation de la fonction objective

La fonction objective d'un programme génétique doit pouvoir permettre au programme de mesurer la qualité d'une solution (d'un chromosome) en l'évaluant par une fonction d'évaluation (*fitness function*) et en comparant la valeur de fitness à une valeur objective (celle que l'on souhaiterait réaliser). En générale cette valeur objective est donnée sous la forme d'une grandeur à minimiser (ou maximiser, en fonction du problème).

Le modèle de variogramme expérimental est représenté par $g^*(h_i) = g_i, i = 1, \dots, n$. Le but de la méthode est de déterminer le modèle (sphérique, exponentiel ou gaussien), ainsi que les paramètres du variogramme isotrope, qui permettront de mieux ajuster le modèle expérimental. Il suffit par conséquent de considérer le modèle théorique comme étant une fonction donnée par :

$$g_{GA}^*(h_i) = f(mdl_{type}, range, sill, nugget), \quad (39)$$

ensuite définir l'écart entre le modèle expérimental et le modèle théorique par :

$$Er_i = g_{GA}^*(h_i) - g^*(h_i), \quad (40)$$

et rechercher à l'aide d'un programme génétique le chromosome $mdl_{type}.range.sill.nugget$, qui minimise la variance de Eri .

2.2.2.2 Paramétrage du programme génétique

Il faut dans cette section définir le codage utilisé, le type de sélection, le schéma de croisement et de permutation, la fonction d'évaluation, et pour finir le critère de sélection des meilleures solutions.

- **Le codage :** le codage utilisé c'est le codage réel. Le chromosome est représenté dans le programme par 1 – 2,5 – 4.8 – 10 par exemple, où le premier élément correspond à un modèle de variogramme, le second correspond à la portée, le troisième au palier et le dernier à l'effet pépité ;
- **La sélection :** la sélection va être aléatoire ;
- **Le croisement :** le croisement va suivre le schéma de croisement uniforme (*K.F. Man et al., 2001*) ;
- **La mutation :** la mutation va consister à tirer deux positions du chromosome et à transformer les gènes concernés ;
- **La fonction d'évaluation :** en fonction du type de modèle sélectionné le programme va utiliser la formule du modèle théorique et les paramètres de la solution pour l'évaluer ;
- **Les meilleures solutions :** après chaque itération le programme sélectionne les chromosomes qui ont présenté la plus petite variance de l'écart avec le modèle expérimental.

Pour réaliser les croisements et mutations sur les chromosomes sélectionnés, il faut définir une probabilité de croisement et de mutation. Le calcul de ces valeurs est l'objet d'une discussion quant

aux considérations à observer lors de l'exercice encouru (*K.F. Man et al., 2001*). Toutefois dans la littérature on trouve généralement des probabilités de croisement proche de 70 à 80%, et des probabilités de mutation environnant les 40%. Ce sont ces valeurs qui ont été utilisées dans le programme génétique.

L'implémentation du programme génétique peut être représentée par le diagramme suivant :

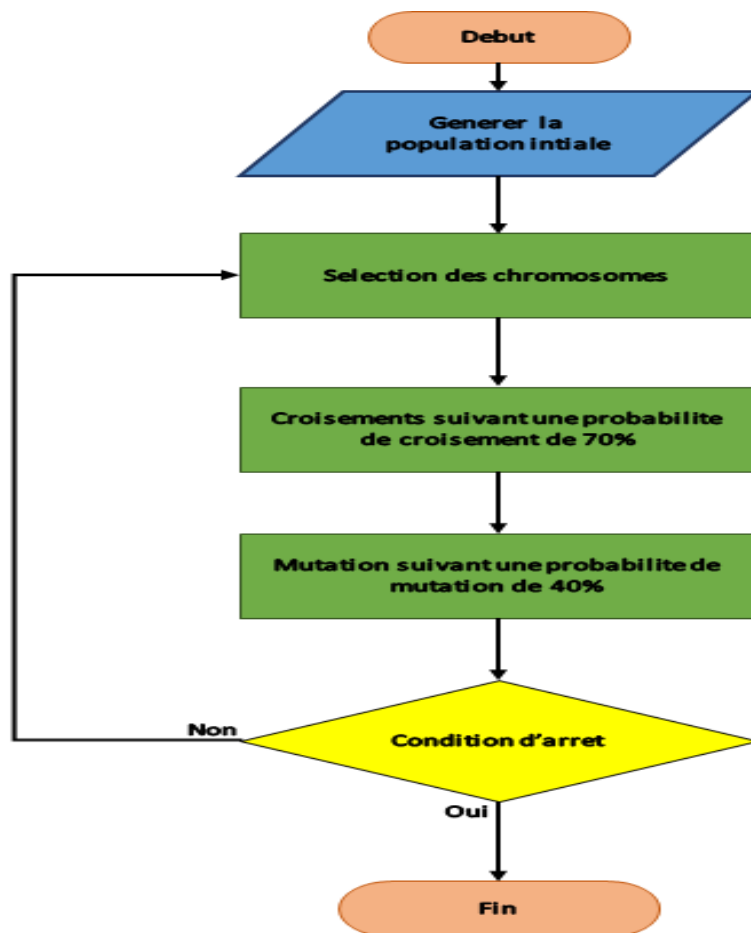


Figure 11 : Algorithme d'exécution du programme génétique

Les différents algorithmes présentés dans ce chapitre ont été traduits en un langage de programmation dans le but de réaliser une plateforme logicielle, permettant une meilleure prise en charge des données et une visualisation des résultats. Dans le chapitre suivant figure la présentation des étapes de la réalisation du programme ResSimSeq2D.

2.3 Réalisation du logiciel ResSimSeq2D

Ce chapitre présente les différents éléments qui ont contribué au développement de la plateforme de traitement des données nommée ResSimSeq2D. Il s'agit d'un logiciel démontrant de multiples applications qui seront présentées dans les prochaines lignes. Ce sera aussi l'occasion de revenir sur la technique programmation utilisée pour rendre effectif ce logiciel.

2.3.1 Choix de l'environnement de programmation

L'IDE (Integrated Development Environment – Environnement de Développement Intégré) choisit pour développer ResSimSeq2D est **Qt Creator**. C'est un environnement de développement intégré multiplateforme faisant partie de la framework Qt. Il est orienté pour la programmation en C++. Il intègre directement dans l'interface un débogueur, un outil de création d'interfaces graphiques, des outils pour la publication des codes ainsi que la documentation Qt.

Le fait que cet environnement regroupe plusieurs outils a fortement nourrit la raison pour laquelle il a été choisi. En plus de cela il faut ajouter que le langage qu'il utilise (C++) montre un bon avantage quant à la matérialisation numérique de certain objet tel que les profils mesures... En effet le langage de programmation C++ est orienté objet, et ainsi offre la possibilité de créer des classes particulières d'objet et des fonctions membres ou méthodes qui commandent l'interaction des objets de cette classe avec l'ensemble du logiciel ; on parle d'encapsulation. C'est ainsi qu'il a été possible par exemple de modéliser une classe d'objet nommée **GeoelecProfile**, pour regrouper en une seule variable, toute l'information disponible sur les activités et les mesures réalisées pour un profil donné dans l'étude du cas pratique.

2.3.2 Les fonctionnalités de ResSimSeq2D

Le traitement des données électriques suivant une approche géostatistique doit passer par plusieurs étapes, à savoir principalement :

- la description statistique (application A_1) et géostatistique des données (A_2) ;
- la définition d'une dimension d'étude et d'analyse des résistivités ;
- la simulation séquentielle gaussienne des données (A_3 pour le cas 1D, et A_4 pour 2D) ;

Pour permettre à l'utilisateur de mieux administrer chaque étape du traitement de ses données, le logiciel dispose de boîtes de dialogue de configuration et d'opération, il offre aussi l'opportunité à l'utilisateur de sauvegarder des résultats dans des fichiers qu'il peut réutiliser par la suite en cas de nécessité.

2.3.2.1 Les applications

Le logiciel réalise plusieurs tâches. Ici sera présenté juste quelques-unes d'entre elles qui sont considérées comme principales, à partir desquelles il est possible de l'apprécier.

- **La description statistique**

Il s'agit ici de produire la distribution statistique et l'histogramme des résistivités mesurées sur le terrain, et aussi, de retrouver certaines grandeurs permettant de décrire la distribution, comme la moyenne arithmétique, la largeur globale (écart-type), ou encore le point de maximum de vraisemblance

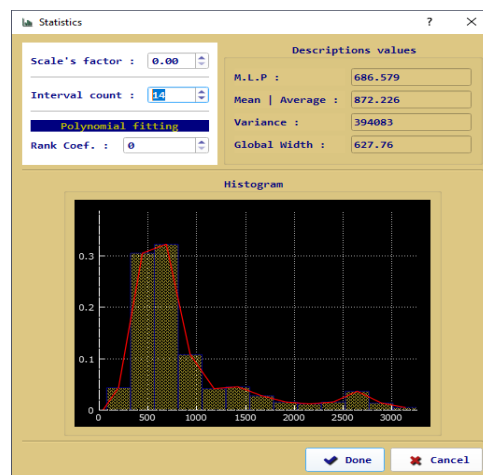


Figure 12 : Boîte de dialogue pour l'application pour la description statistique.

- **La description géostatistique**

L'outil de description géostatistique des résistivités c'est le variogramme. Le logiciel est capable de calculer le variogramme initiale des résistivités mesurées pour un profil donné. Il demande au préalable certaines configuration, qu'il va utiliser par la suite pour calculer le modèle expérimental, et laisse le choix à l'utilisateur d'ajuster son modèle expérimental soit par algorithmes génétiques, soit manuellement.

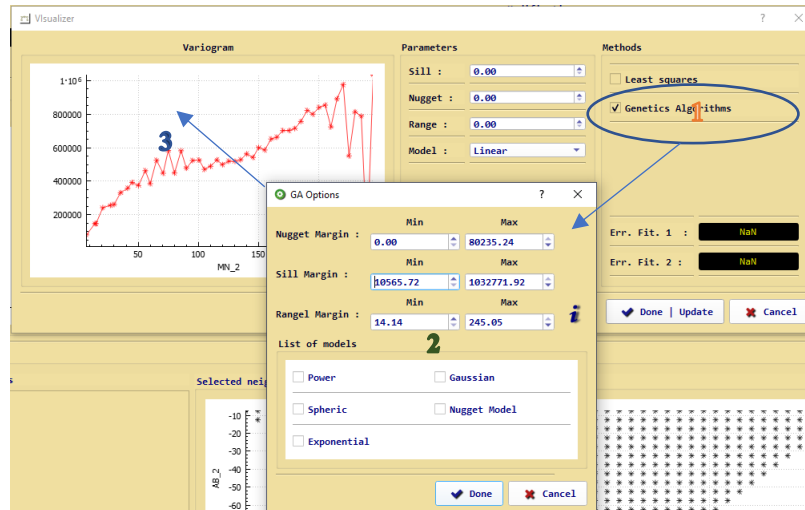


Figure 13 : Boite de dialogue pour l'analyse variographique

Le choix de la dimension de travail (1D ou 2D) est directement lié aux options proposées dans les menus. En effet lorsque l'utilisateur décide de travailler en mode traitement de données de sondages, il a choisi de travailler en 1D, et quand il travaille avec les options de réalisation de tomographie il choisit de travailler en 2D.

Les simulations vont de pair avec le choix de la dimension de travail, de sorte que, pour l'espace à une dimension le logiciel produit par simulation une courbe de sondage, et pour le cas 2D, il produit une pseudo section des résistivités.

• Le système de gestion des fichiers

Le logiciel propose une gestion des fichiers de sauvegarde. Cette application montre un très grand avantage lors des simulations car en effet, à partir de ces fichiers l'utilisateur peut comparer après chaque itération les paramètres du nouveau variogramme avec ceux des variogrammes précédents. Cette fonctionnalité montre encore un autre avantage qui est celui d'éviter à l'utilisateur de devoir redéfinir par exemple les paramètres de configuration du voisinage de calcul après chaque itération.

Ces différentes tâches mettent en interaction avec le logiciel, des objets particuliers créés pour permettre une meilleure manipulation des données initiales. Le tableau suivant fournit les types d'objet les plus importants, créés lors de la réalisation du logiciel.



Tableau 1: Les classes essentielles définies lors de la programmation du logiciel.

Type ou classe d'objet	Utilité dans la programmation du logiciel
<i>GeoelecProfile</i>	Pour regrouper en une seule variable, toute l'information disponible sur les activités et mesures réalisées sur un profil.
<i>GA_Population</i>	Pour les opérations de programmation génétique, un objet de cette classe sauvegarde les différents chromosomes ayant survécu à la sélection naturelle.
<i>Neighborhood</i>	Pour permettre la configuration et la gestion des sommets du graphes considérés comme étant les éléments du voisinage de calcul.

2.3.2.2 Les séquences d'exécution de chaque tâche

Les différentes procédures de réalisation des tâches citées précédemment, sont présentées dans le tableau ci-après (**Tableau 2**). Et par la suite viendra un organigramme regroupant toutes les fonctionnalités du logiciel.

Tableau 2: Procédure d'implémentation des tâches principales.

Tâche	Procédure
A_1	Importation des données → Configuration des champs de la base de données → Sélection du profil à analyser → Cliquer sur l'icône  .
A_2	Importation des données → Configuration des champs de la base de données → Sélection du profil à analyser → Sur la page 'Variogram' du logiciel, cliquer sur  .
A_3	Importation des données → Configuration des champs de la base de données → Sélection du profil à analyser → Dans le menu 'VES...'

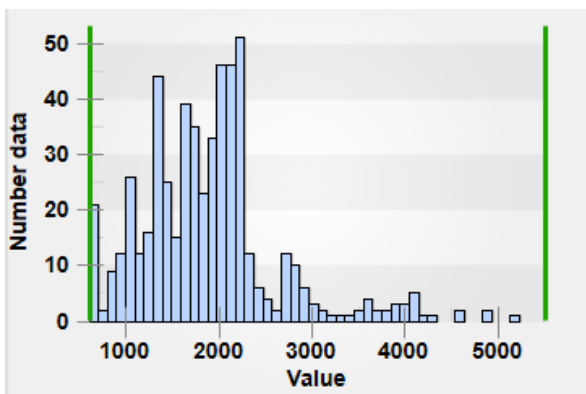
	choisir ‘ Configuration ’ → Définir les paramètres et enregistrer → Dans le menu ‘ VES... ’ choisir ‘ Run inversion ’
A_4	Importation des données → Configuration des champs de la base de données → Sélection du profile à analyser → Dans le menu ‘ Tomography ’ choisir ‘ Configuration ’ → Définir les paramètres et enregistrer → Dans le menu ‘ Tomography ’ choisir ‘ Run inversion ’

2.3.3 Validation des résultats de ResSimSeq2D

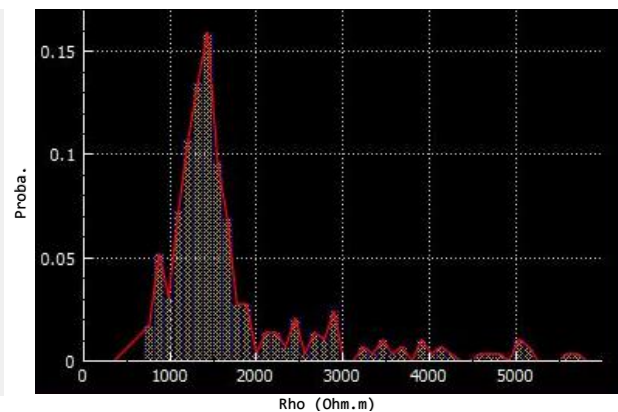
La validation des résultats du logiciel ResSimSeq2D consistera dans les sections suivantes, à comparer les distributions, variogrammes et pseudo-section obtenus à partir de ce logiciel, à ceux obtenus en utilisant des logiciels de référence tel que Prosys de Iris Instruments, ou encore ArcGis et Res2DInv.

2.3.3.1 Les distributions statistiques

A partir du logiciel Proxys, il est possible de produire la distribution statistique des résistivités mesurées le long d’un profil. Les figures suivantes présentes les résultats obtenus d’une part sur ResSimSeq2D, et d’autre part su Proxys :



(A)



(B)

Figure 14 : Histogramme des résistivités sur Prosys (A) et sur ResSimSeq2D (B)

Et il apparaît que les distributions obtenues par ResSimSeq2D sont valides, en comparaison avec celles obtenues sur Prosys.

2.3.3.2 Les variogrammes

Le calcul du variogramme expérimental n'est pas une fonction fréquemment retrouvée dans les logiciels d'inversion classique, c'est pourquoi les résultats de ResSimSeq2D ont été comparés à ceux obtenus sur ArcGis. Ces résultats sont représentés sur les figure suivante :

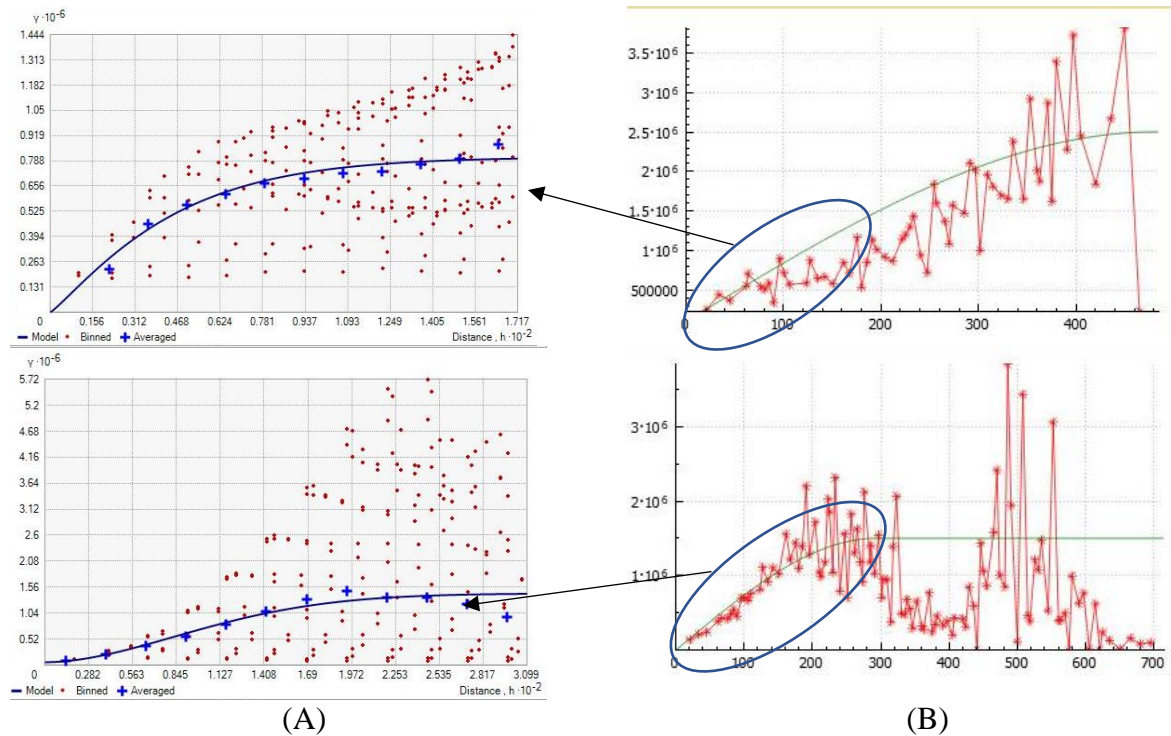


Figure 15 : Variogramme réalisé sur ArcGis (A) et sur ResSimSeq2D (B)

Les deux résultats ne sont pas représentés à la même échelle, mais on peut lire pratiquement les mêmes valeurs de palier et portée pour variogramme obtenu sur ResSimSeq2D.

2.3.3.3 Les pseudo-sections

Ici la comparaison peut se faire en même temps en ResSimSeq2D, Prosys et Res2Dinv, en ceci que le modèle de pseudo section devrait pouvoir être proche de celui visualiser sur proxys lors du prétraitement des données, et en même temps il devrait être relativement proche du résultat obtenu sur Res2Dinv qui implémente une inversion 2D à l'aide des moindres carrés pondérés. Les résultats sont présentés dans les figures suivantes :

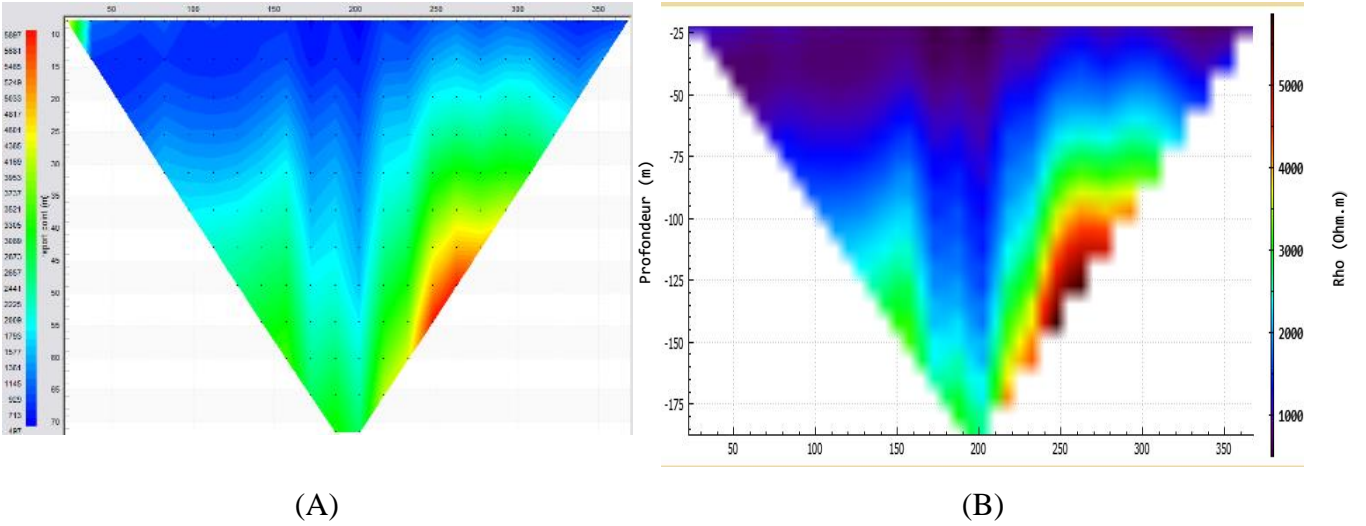


Figure 16 : Pseudo section sur Prosys (A), sur ResSimSeq2D (B)

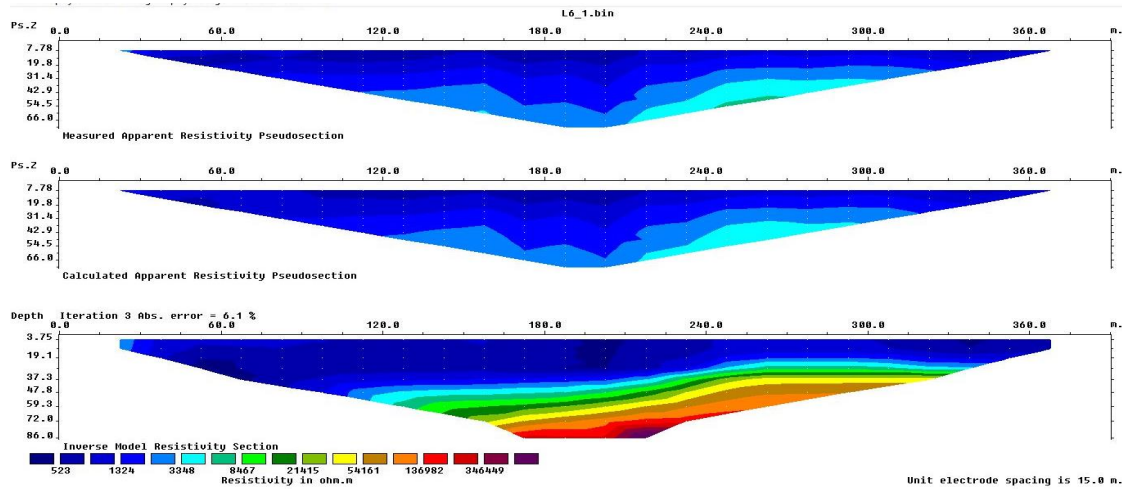


Figure 17 : Pseudo section sur Res2DInv pour le même profil d'essai

En effet, le résultat sur Prosys correspond à la distribution spatiale de résistivité que le modèle inverse devrait pouvoir reproduire. Il apparaît clairement que ResSimSeq2D y parvient, cela démontre de la fiabilité de son processus d'inversion 2D et si son résultat est relativement différent de celui obtenu sur Res2DInv, c'est dû au simple fait que ce dernier logiciel fait usage d'une méthode déterministe d'inversion tandis que le logiciel ResSimSeq2D se sert d'une approche probabiliste ayant la capacité de mieux se rapprocher de la réalité.

De cet exercice il en ressort que, le logiciel ResSimSeq2D produit des résultats qui démontrent un degré raisonnable de véracité et de réalisme, et peut par conséquent être utilisé dans un cas réel d'étude des données mesurées lors d'une campagne géophysique.

Une fois le logiciel présenté, il faut passer à l'application sur un cas pratique d'étude des résistivités mesurées sur le terrain, ce qui fera l'objet du prochain chapitre.

Chapitre 3 : Etude d'un cas pratique d'inversion des données géoélectriques à l'aide de ResSimSeq2D

Dans ce chapitre sera présentée l'application des simulations séquentielles gaussiennes sur les résistivités mesurées lors d'une expérimentation, pour produire des modèles de distribution spatiale et statistique des prédictions par inversion, à travers le logiciel ResSimSeq2D.

3.1 Origine des données

L'expérimentation consistait à mesurer les résistivités du sous-sol d'un site situé dans la localité de Djallo à environ 5 kilomètres du centre-ville de Meiganga, dans la région de l'Adamaoua (Cameroun). Meiganga est la principale ville du département de Mbéré. Cette zone d'étude, avec environ 1129 km^2 de surface, se situe entre les longitudes $14^{\circ}00'$ et $14^{\circ}25'$ et les latitudes $06^{\circ}28'$ et $06^{\circ}43'$ (Figure 2), avec des altitudes allant de 920 à 1200 m (Meying Arsene et al., 2018).

Les mesures ont été effectuées suivant 09 profils orientés dans la direction N150E, pour pouvoir recouper les éléments structuraux majeurs, qui suivent la direction N60E. Les mesures sont effectuées suivant la configuration d'électrode de type Schlumberger (quadripôle), avec deux électrodes d'injections du courant continu, et deux autres à travers lesquelles le resistivimètre Syscal Junior (de Iris Instrument, voir figure 19), mesure les différences de potentiel. L'objet de l'expérimentation est de localiser les probables nappes d'eau présente dans le sous-sol du site. L'ensemble du matériel nécessaire ayant été utilisé pour les mesures est représenté dans l'image de la figure 19, dont les désignations par outils sont répertoriées dans le tableau suivant :

Tableau 3 : Listes des outils utilisés lors de la mesure.

Etiquettes	Désignations	Utilités
A	Resistivimètre	Il permet d'injecter le courant dans le sol, et mesure la

		différence de potentiel dans le champ électrique qui se crée.
B	Récepteur GPS	Il permet de mesurer des coordonnées géographiques nécessaires lors de la mesure des données.
C	Boussole avec clinomètre	Pour les mesures de directions lors de l'ouverture des layons.
D	Bobine conductrice	Elles permettent la circulation du courant électrique provenant du resistivimètre, pour que ce dernier puisse traverser chaque électrode fixée dans le sol.
E	Electrodes de mesures et d'injections	Elles permettent en même temps d'injecter dans le sol le courant provenant du resistivimètre, et mesurer la différence de potentiel pour évaluer les valeurs de résistivités.
F	Marteau	Pour des besoins de prélèvement d'échantillon si nécessaire et occasionnellement il est utile lors de la fixation des électrodes dans le sol.

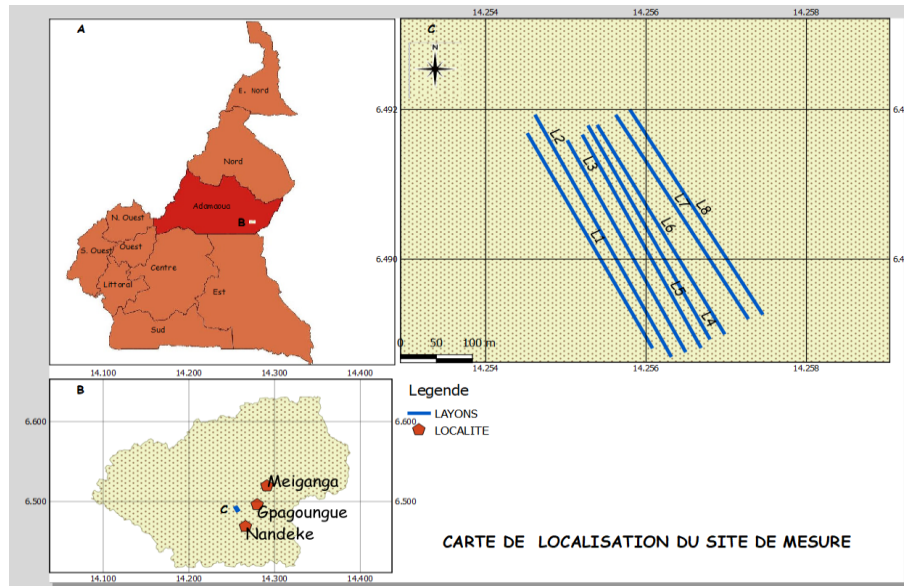


Figure 18 : Carte de localisation du site de l'expérimentation (à refaire)

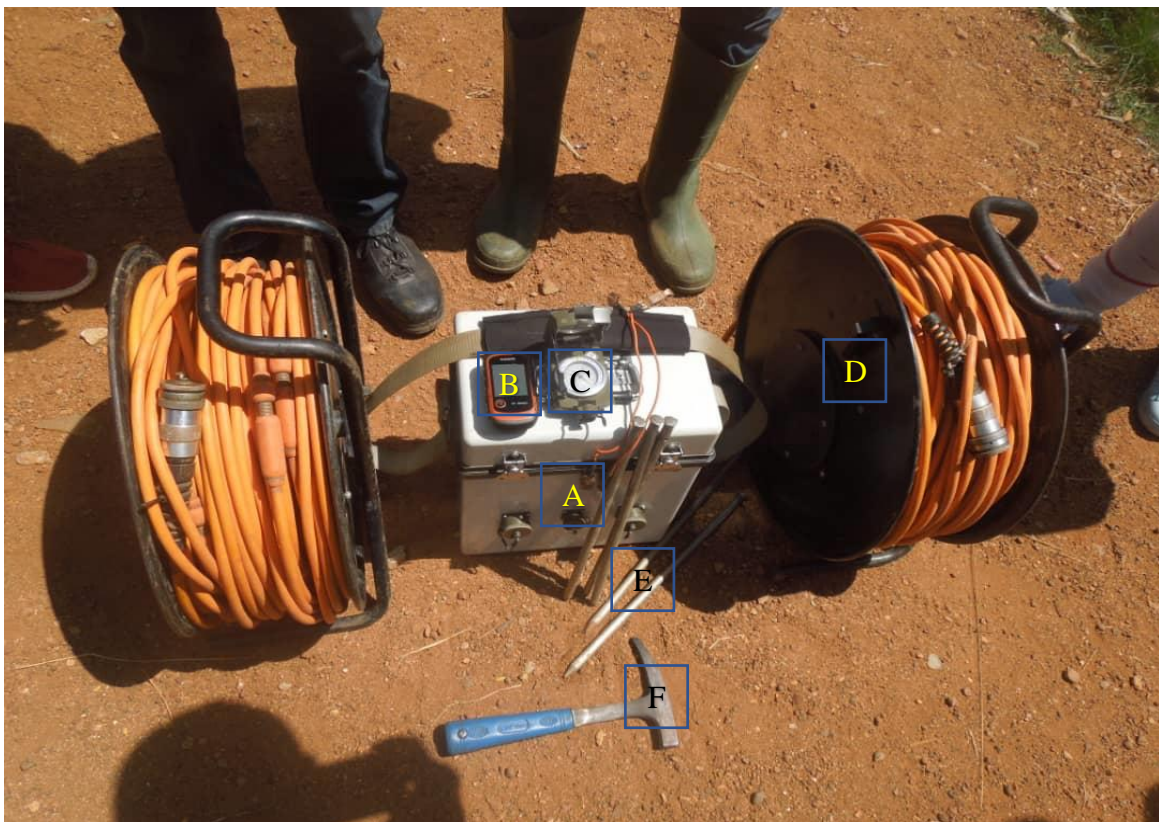


Figure 19 : Matériels utilisés lors de l'expérimentation

3.2 Descriptions statistique et géostatistique des résistivités mesurées

Dans cette section, il est question de dégager la distribution des résistivités mesurées sur chaque profil, afin de voir les profils pour lesquels il est possible d'appliquer la simulation gaussienne sans transformation sur la distribution initiale des données. Ensuite il faudrait aussi produire les modèles de variogramme initiaux pour chaque profil, afin d'en dégager quelques observations qui pourront servir par la suite dans l'application des simulations.

3.2.1 Les distributions statistiques

A partir des données mesurées sur le terrain, le logiciel a produit les histogrammes suivants :

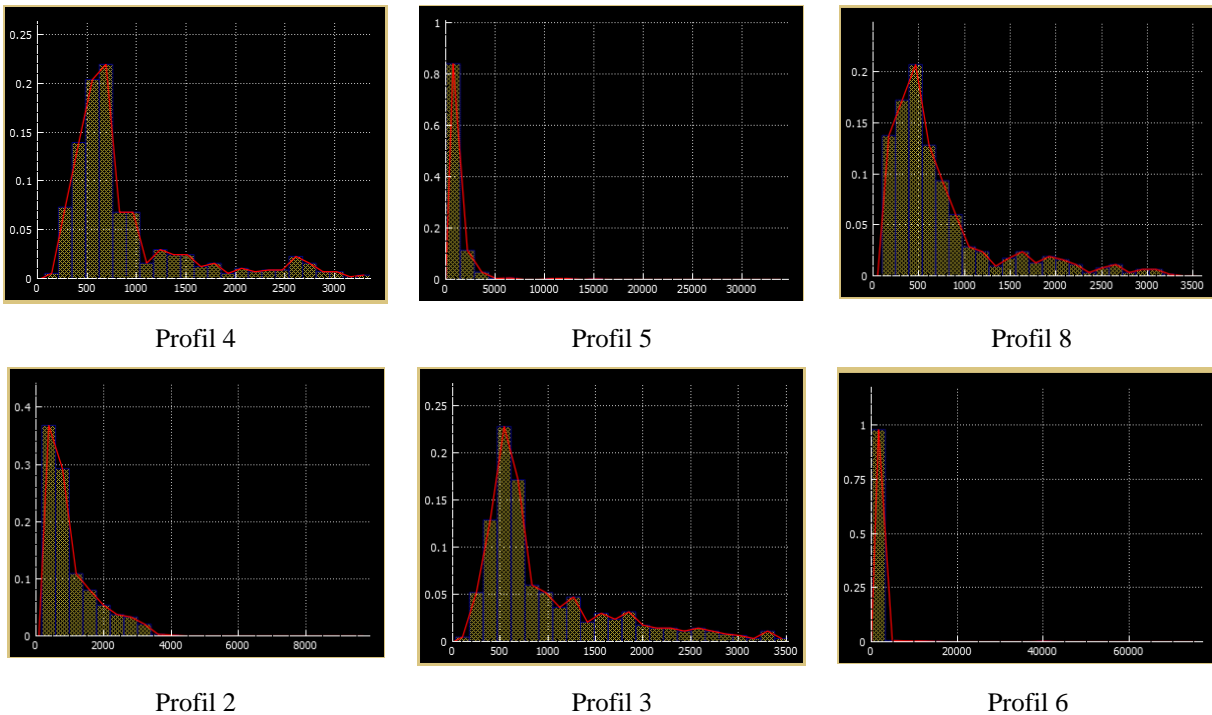


Figure 20 : Les histogrammes et distributions pour chaque profil

Au vu des histogrammes obtenus, on peut admettre que les résistivités suivent une loi gaussienne, sauf pour certains cas (Profil 5, Profil 6) où la présence de valeurs très grande (des valeurs extrêmes) devant la valeur moyenne masque la distribution des données. Ce problème de valeur extrême fera l'objet d'une analyse plus tard dans ce document, car en effet, d'un point de vue géostatistique il serait préférable de ne pas l'intégrer dans les données à traiter, pourtant d'un point

de vue physique ne pas l'inclure serait considéré comme passer à côté d'une information très pertinente.

3.2.2 Les variogrammes initiaux

Considérons un profil orienté dans une direction le long duquel des mesures ont été réalisées à l'aide du resistivimètre. Alors il est possible de considérer les résistivités comme étant des réalisations d'une variable régionalisée (variable aléatoire dont la réalisation dépend de la position dans l'espace), $\rho_a(\frac{AB}{2}, \frac{MN}{2})$, où $\frac{AB}{2}$ et $\frac{MN}{2}$ sont les coordonnées des points de réalisation dans l'espace de mesure

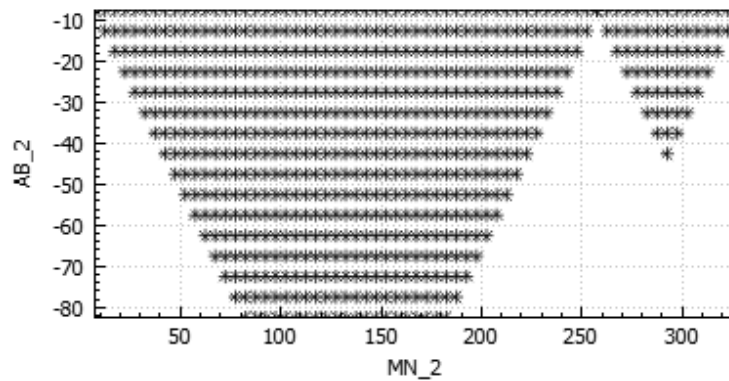
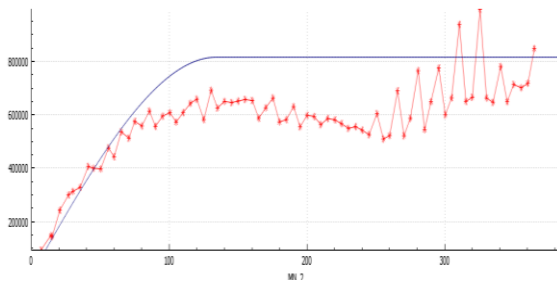
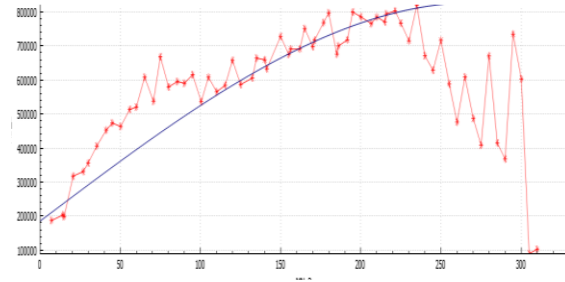


Figure 21 : Exemple de profil représenté par des points de mesures de ρ_a

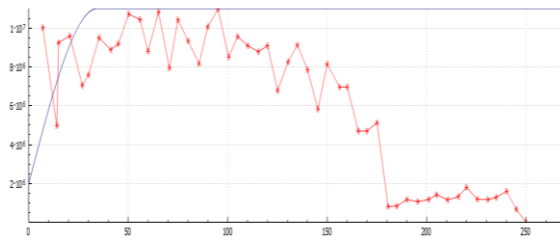
Il est donc possible de calculer le variogramme de chaque profil à partir de cette représentation, et le programme conçu fourni les résultats suivants :



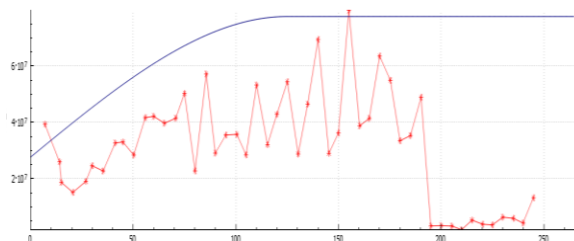
Variogramme profil 2



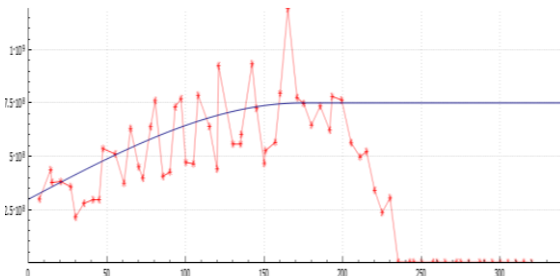
Variogramme profil 3



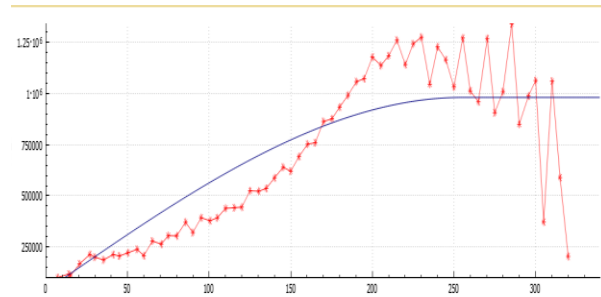
Variogramme profil 5



Variogramme profil 6



Variogramme profil 7



Variogramme profil 8

Figure 22 : Les variogrammes calculés pour chacun des profils

Le premier constat observable est que à plus de 200 mètres le variogramme a du mal à schématiser la structure spatiale des résistivités mesurées, cette information sera très utiles lorsqu'il faudra définir les paramètres du voisinage d'un point de calcul, puisque si un point renseigne très peu sur le calcul de la valeur probable en un autre pourquoi ne pas l'exclure du calcul et réduire ainsi le cout de l'opération (il faut rappeler que le nombre de points utilisés est très déterminant pour ce type d'exercice).

La seconde observation est que, les variogrammes des profils 5 et 6 sont plus irrégulier que les autres. En effet la présence d'une valeur très grande par rapport aux autres, influence fortement le modèle de variogramme de chacun de ces profils. Une fois de plus il faudra encore discuter sur la question de maintenir ou retirer les valeurs extrêmes de la base de données (ici ce sera pour permettre une meilleure appréciation de la variabilité spatiale des résistivités).

Une troisième observation est que, les variogrammes n'ont pas été calculés suivant une direction. En effet, les résistivités ont plus de chance de varier suivant la verticale que l'horizontale, et de ce fait il n'est plus nécessaire de rechercher la direction dans laquelle le phénomène étudié présente une meilleure continuité. L'objectif du modèle omnidirectionnel est de produire directement un modèle isotrope de variogramme qui peut être utilisé pour réaliser des prédictions par simulation.

3.3 Inversion des données de l'expérimentation

3.3.1 Les courbes de sondage (simulation 1D)

A partir des résistivités mesurées pour un profil donnée, et suivant un point d'étude choisis, le logiciel implémente les simulations séquentielles pour produire des courbes de sondage qui peuvent être interprétées pour obtenir des informations sur le site d'expérimentation. Les résultats sont présentés dans les figures suivantes, toujours pour nos 06 profils cibles de départ.

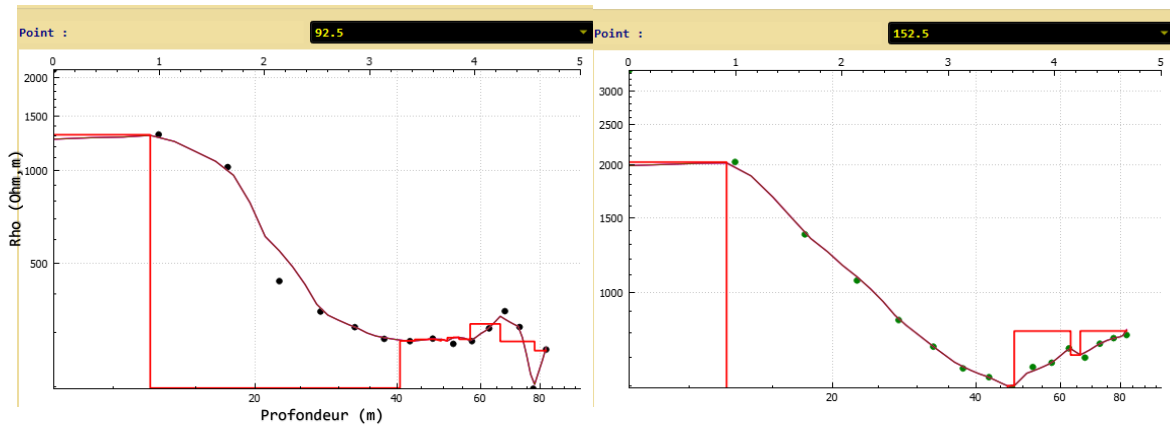


Figure 23 : Courbe des sondages du profile 4.

Les courbes de sondages montrent que les zones de plus faibles résistivités sont comprises entre 40 et 50 mètres de profondeur, et il est aussi possible de constater ces courbes mettent en évidence l'alternance de 03 couches de formations d'épaisseur et résistivités distinctes.

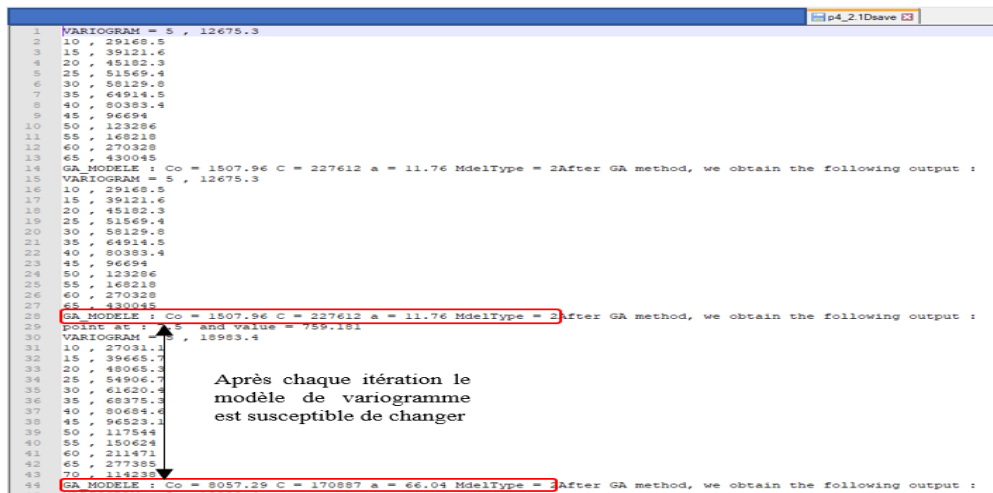
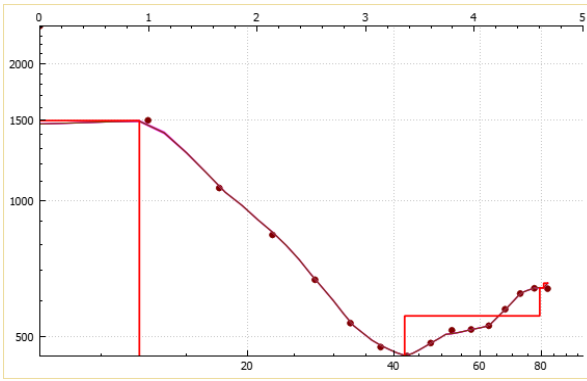
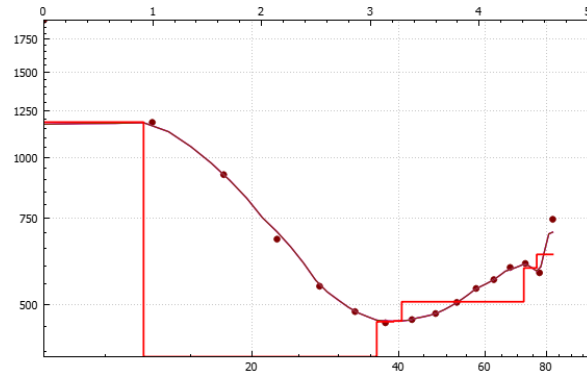


Figure 24 : Liste des variogramme de la simulation 1D pour le profil 4



Courbe de sondage du profil 5



Courbe de sondage du profil 8

Figure 25 : Autres courbes de sondages

3.3.2 Inversion 2D

Dans le cas 2D, il s'agit de produire les pseudo sections des résistivités pour chaque profil de mesures. Les résultats obtenus sur ResSimSeq2D sont présentés dans les figures suivantes :

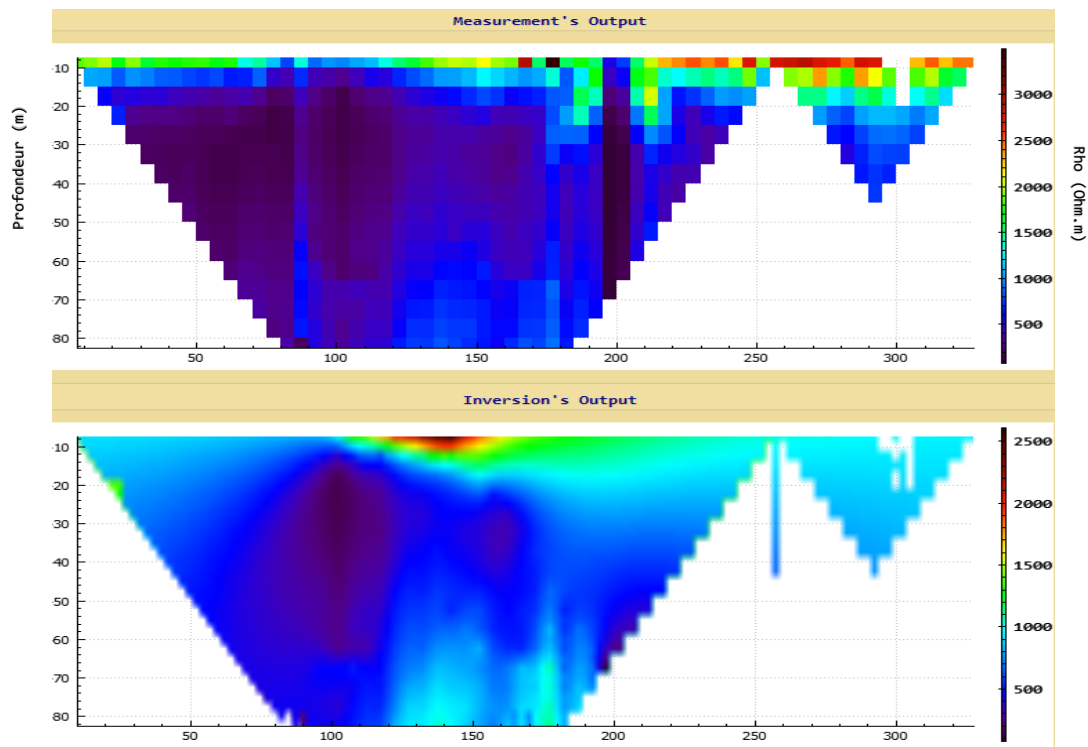


Figure 26 : Pseudo section des résistivités pour le profil 8

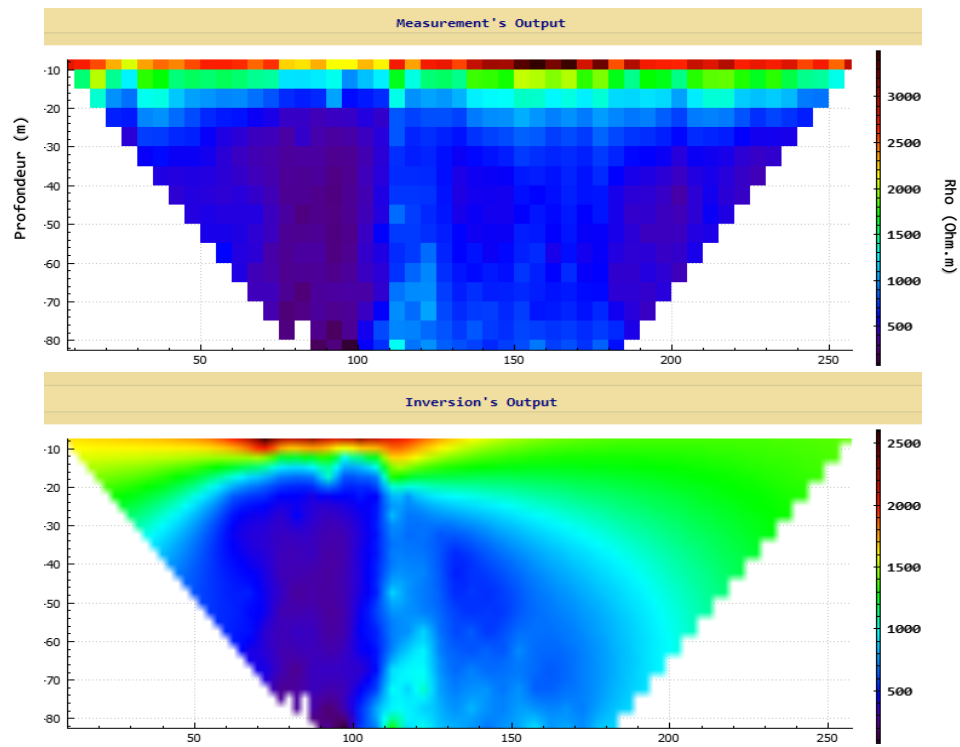


Figure 27 : *Pseudo section des résistivités pour le profil 4*

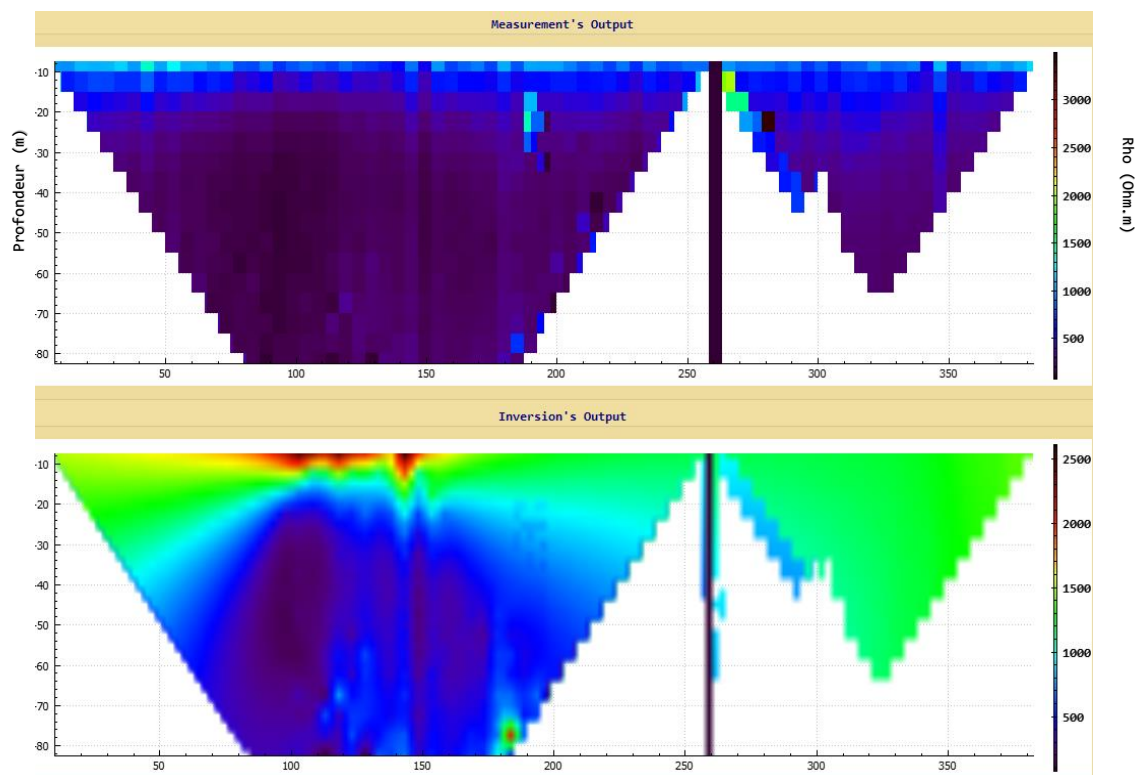


Figure 28 : *Pseudo section des résistivités pour le profil 2*

Sur ces figures, la pseudo section au-dessus représente la distribution spatiale de des résistivités mesurées, et en dessous, la distribution spatiale des résistivités calculées par inversion. Il est possible à partir de ces résultats de faire certains constats.

Tout d’abord, même dans le cas des simulations 2D, la présence des valeurs extrêmes continue de masquer la distribution des résistivités mesurées (Figure 29), ce problème ne dépend pas des dimensions de l’espace, mais il est directement lié à la nature et la compréhension physique d’une mesure des résistivités.

Néanmoins les résultats de l’inversion 2D révèlent aussi que, les plus faibles résistivités sont situées à 40 mètres de profondeur, probablement entre 40 et 55 mètres.

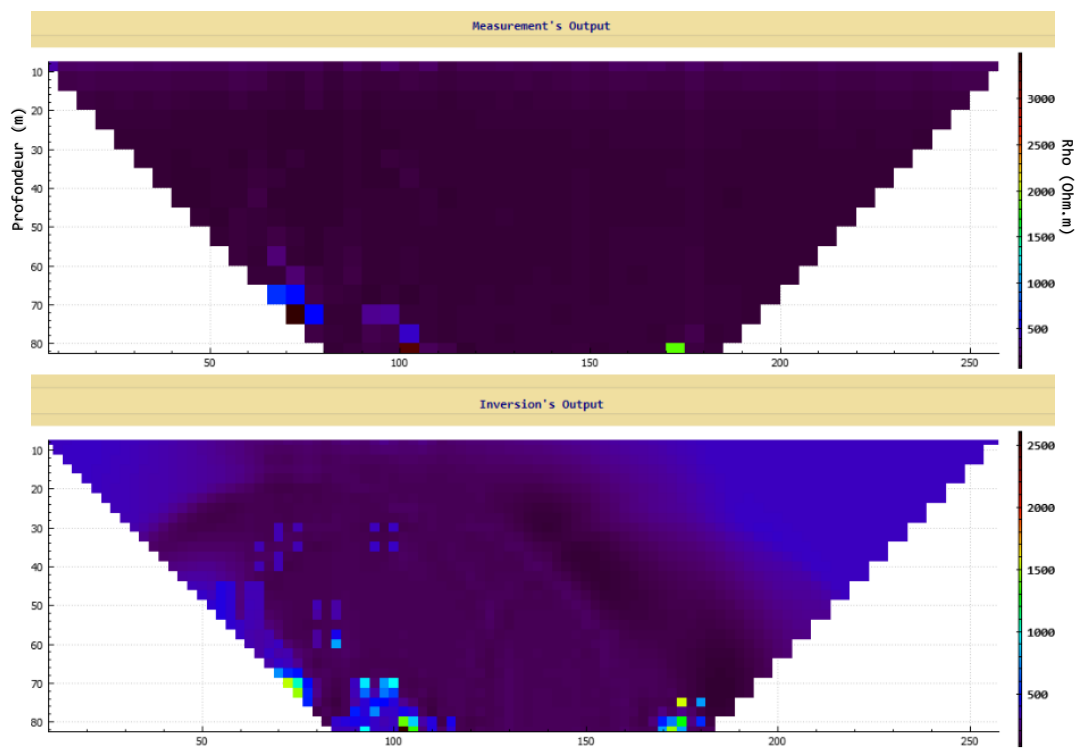


Figure 29 : Pseudo section des résistivités pour le profil 6, avec des données présentant des valeurs extrêmes

3.4 Interprétations des résultats et discussions

3.4.1 Interprétations des courbes de sondages et pseudo section

Les courbes de sondages ont pour objectifs en générale de dégager les différents horizons traversés par le courant électrique perpendiculairement à la surface du sol. Dans les tableaux suivants,

pour les courbes présentées plus haut, seront répertoriées les informations relatives aux couches de terrain explorées :

Tableau 4 : Informations de la courbe du profil 04

Index	1	2	3
Profondeur (m)	0	12,5	44,5
Epaisseur (m)	12,5	32	
Résistivité ($\Omega \cdot m$)	2000	600	800

Tableau 5 : Informations de la courbe du profil 05

Index	1	2	3	4
Profondeur (m)	0	12,5	42,5	80
Epaisseur (m)	12,5	30	37,5	
Résistivité ($\Omega \cdot m$)	1500	400	600	650

Tableau 6 : Informations de la courbe du profil 06

Index	1	2	3	4
Profondeur (m)	0	12,5	36	75
Epaisseur (m)	12,5	23,5	39	
Résistivité ($\Omega \cdot m$)	1200	306	500	650

Il est possible d'affirmer que le site étudié présente une alternance de trois à quatre couches de terrain. Une étude géologique de la région pourrait contribuer à confirmer ou infirmer cette affirmation, mais dans le souci de rester dans la problématique traitée dans ce document, les travaux réalisés n'iront pas aborder cet aspect.

Toutefois, à partir des profils présentés, il est possible de remarquer une certaine continuité de la couche située à environ 40 mètres de profondeur.

A partir des pseudo sections réalisées, il est possible d'apprécier d'avantage la continuité des résistivités dans la direction perpendiculaire à N150E. La figure qui suit montre que les formations ayant résisté au passage du courant, ont tendance à suivre une certaine continuité dans l'espace a trois dimensions :

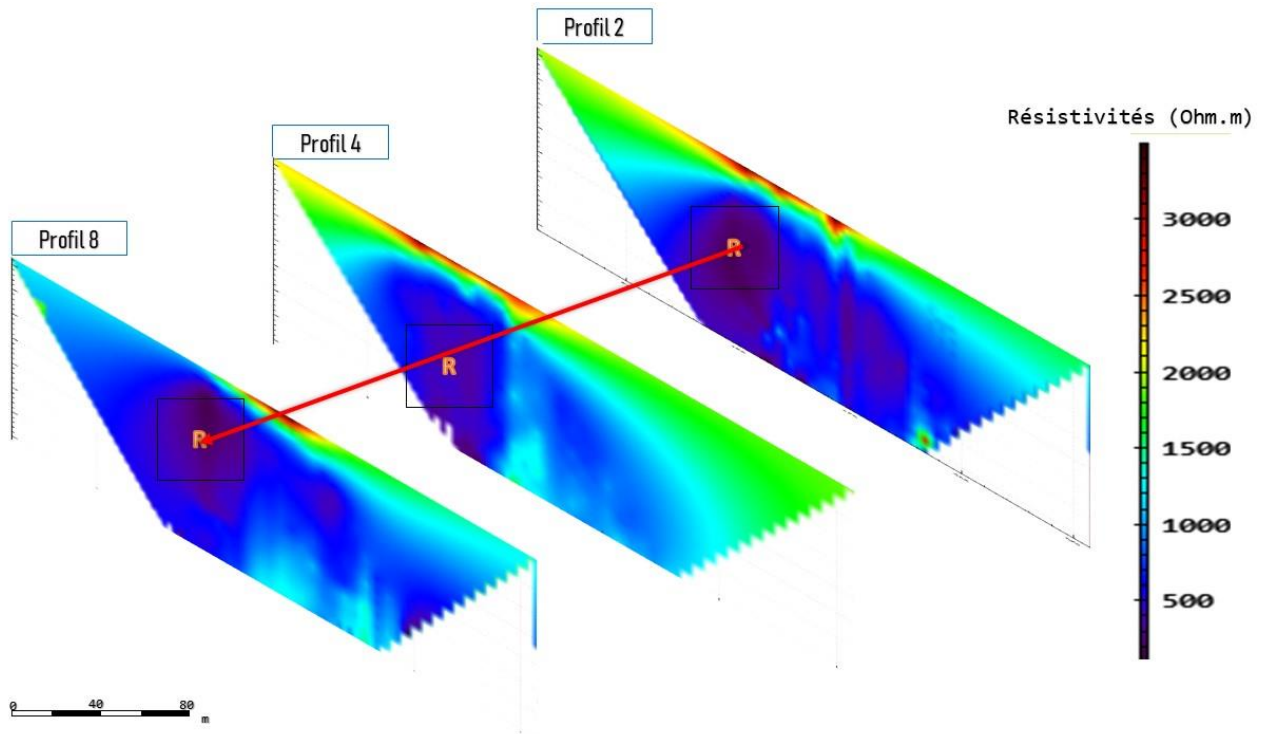


Figure 30 : Continuité des couches de terrains peu résistives, sur les profils 04, 08 et 02

Le point **R** repéré sur le profil 02 a été retrouvé sur les profils 04 et 08, une preuve que la formation responsable des faibles résistivités à cet endroit est continue, du moins du profil 02 au profil 08.

3.4.2 Les valeurs extrêmes

Dans une section plus haut, il a été souligné le fait que certaines mesures contiennent des valeurs qui sont d'une échelle de grandeur plus importante que le reste. En effet ce genre de valeur a tendance à masquer la distribution des données. Dans la pratique, lorsque dans un ensemble de données se trouve une valeur supérieure à cinq fois l'écart-type des données mesurées, cette valeur peut être considéré comme extrême (William Menke, 1984). En éliminant les valeurs extrêmes du

profil 05 par exemple, sa distribution se laisse mieux apprécier comme le montre la figure suivant ce paragraphe (Figure 32 et Figure 33). Cependant, cette valeur de forte résistivité pourrait très bien aussi représenter un corps très compact enfouis dans le sol, ou encore un vide sec tel que les dolines. *Matheron G. (1978)* dans son cours intitulé «**TECHNICAL PARAMETRIZATION OF ORE RESERVES**» parle de l'effet d'information, pour se référer à l'influence qu'a une information sur un problème posé. Ainsi pour pouvoir mieux apprécier les distributions des données mesurées sur les profils, les valeurs extrêmes ont été retiré de la base (ceci uniquement pour les profils qui en avaient).

3.4.3 L'influence du voisinage sur la qualité de la simulation

Un facteur qui s'est montré très déterminant lors des essais et de l'application des simulations à la modélisation inverse 2D des résistivités, est la taille du voisinage. Car quand cette taille augmente, la précision pour une réalisation en un point va augmenter aussi. Le voisinage glissant apporte plus de réalisme aux simulations, cependant en étant couteux en temps d'exécution. L'algorithme présenté dans le chapitre 2, conçu pour la création du voisinage, ne forme pas des voisinages de plus de 150 points, pour que la matrice de krigeage soit stable. En plus cet algorithme ne travaille pas avec le voisinage glissant, mais plutôt des blocs de voisinages qui se déploient suivant les régions de l'espace d'étude.

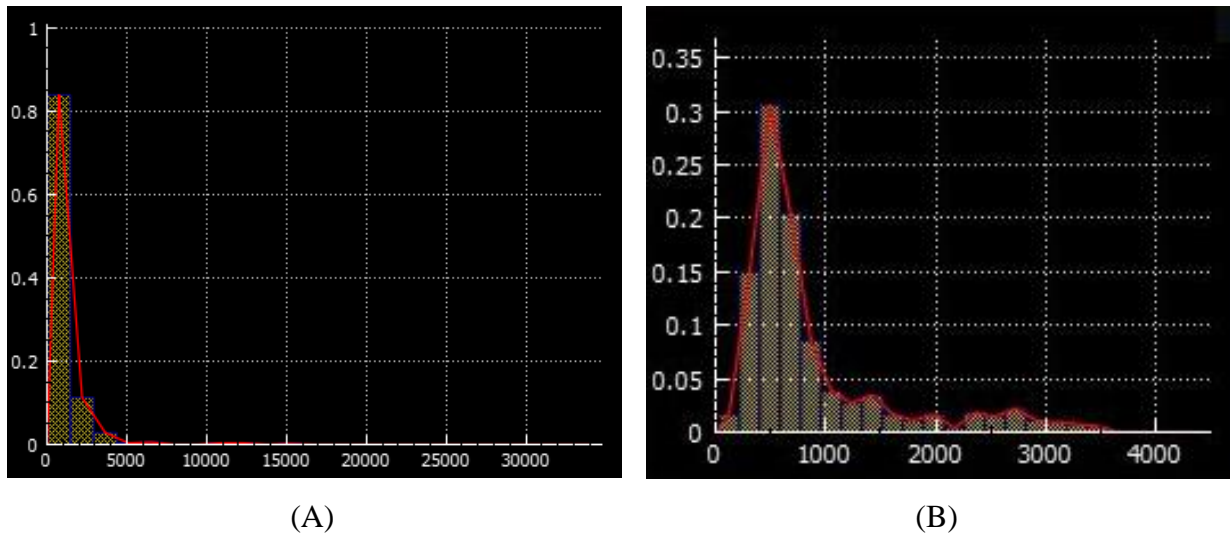


Figure 31 : Histogramme du profil 05 avec valeurs extrêmes (A) et après suppressions des valeurs extrêmes

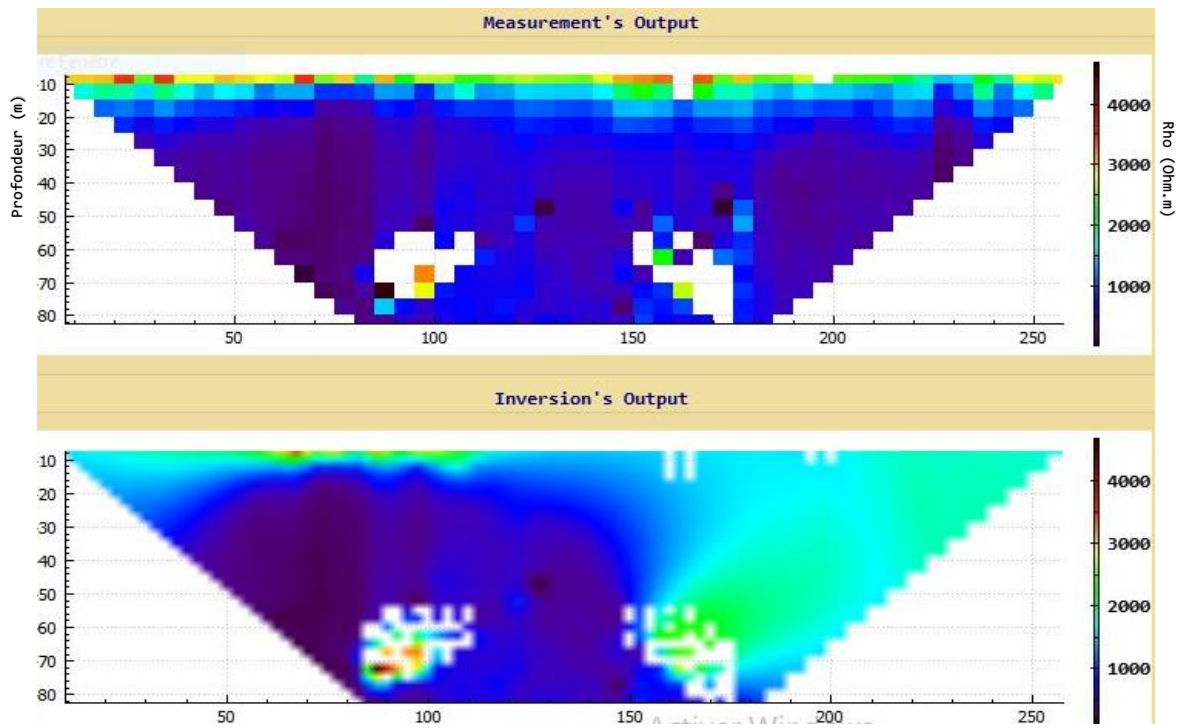


Figure 32 : Pseudo section du profil 05 après suppression des valeurs extrêmes

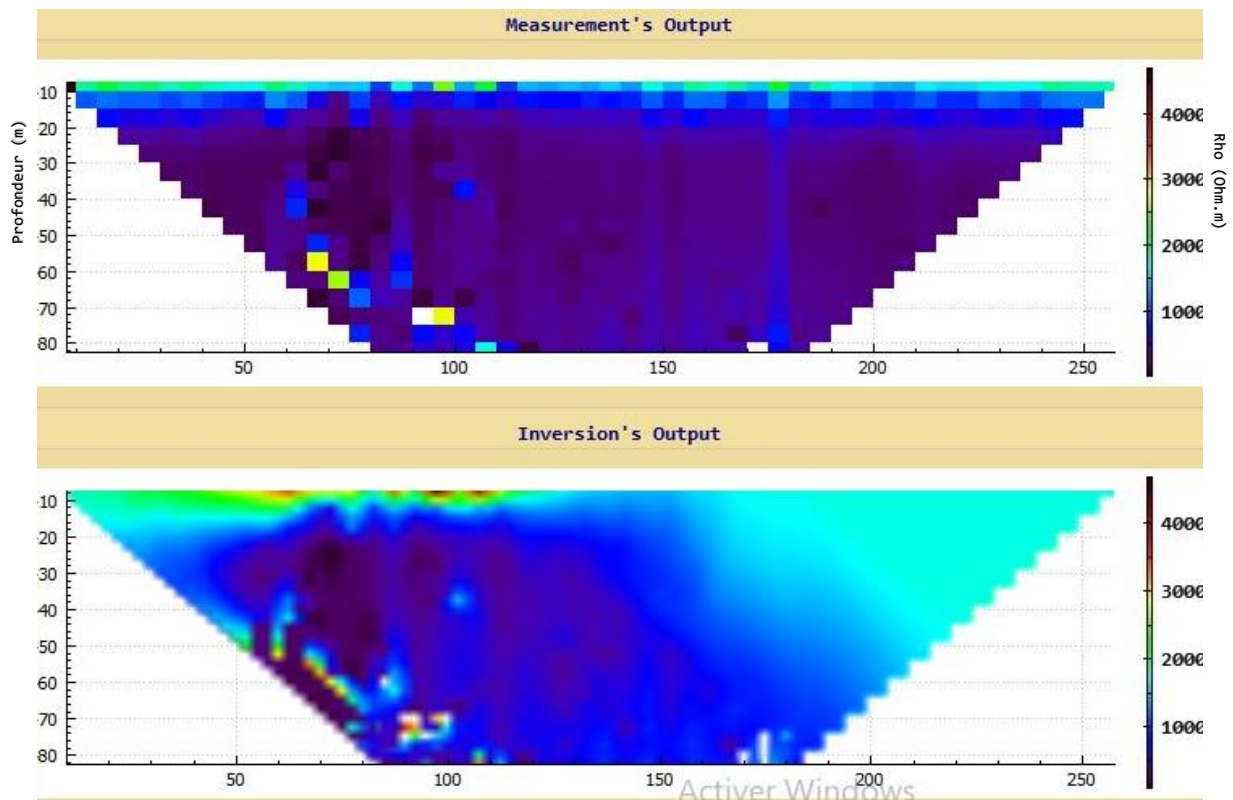


Figure 33 : Pseudo section du profile 06 après suppression des valeurs extrêmes

3.4.4 L'analyse variographique

Elle revient, car ayant démontrée d'une certaine complexité quant à sa mise en œuvre automatique. En effet il aura fallu, pour un voisinage de type glissant, recalculer un nouveau variogramme à chaque nouvelle réalisation obtenue. Cette approche a eu l'avantage de fournir des résultats très proches de la réalité, cependant elle est très couteuse en capacité pour l'ordinateur, et en plus elle est aussi de plus en plus lente quand le nombre de données de conditionnement augmente. Une astuce a été de se servir de l'analyse variographique par bloc de voisinage, de sorte que, le logiciel recherche un nouveau variogramme quand on est passé d'un bloc de voisinage à l'autre.

Il faut revenir sur un point tout aussi essentiel. Les simulations ont été implémentées de façons itératives, le nombre d'itération étant défini par l'utilisateur. Le modèle de pseudo section représente tout simplement la moyenne arithmétique des réalisations en chaque point pour chaque itération. Il fallait revenir sur cet aspect itératif pour préciser que, lorsque l'algorithme génétique devant modéliser le variogramme prend beaucoup d'itérations pour le faire, il affecte le processus d'inversion ; ceci constitue, malgré les résultats appréciables obtenu, une limite au jeune algorithme conçu. Une approche pour y remédier serait de faire usage d'une méthode d'apprentissage plus rapide et directe (qui n'aurait pas besoin d'une structure itérative) pour obtenir le modèle de variogramme isotrope à utiliser dans les simulations.

CONCLUSION GENERALE

Il a été question dans le cadre des travaux présentés dans ce document, de produire un algorithme d'inversion utilisant les simulations séquentielles gaussiennes. Les différentes phases de conception, réalisation et implémentation de cet algorithme ont été présentées dans les différentes parties qui ont constitué ce document. Les atouts de cet algorithme ont été mis en évidence lors de la validation du logiciel ResSimSeq2D, qui est sa plateforme d'implémentation, et aussi lors de l'utilisation de ResSimSeq2D dans un cas pratique d'inversion des données géoélectriques, prélevées lors d'une campagne visant à déceler des zones probables de nappes aquifères. Le programme a pu réaliser des inversion 1D et 2D à l'aide des simulations séquentielles gaussiennes, et les résultats ont permis de dire que le sous-sol est constitué d'environ quatre formation d'épaisseur et de résistivité distincte. Les résultats ont aussi permis d'évaluer raisonnablement la continuité de la zone de basse résistivité, afin de mieux apprécier l'extension de la probable nappe aquifère. Malgré la qualité appréciable des résultats obtenus, ResSimSeq2D présente des limites dont les améliorations sont envisageables dans un cadre démontrant de meilleures échéances de faisabilité, notamment en entreprenant des méthodes d'apprentissage pour l'analyse variographique automatique.

De cette façon, ce document aura démontré la pertinence et les intérêts à utiliser les simulations séquentielles gaussiennes, lors des inversions gaussiennes non-linéaires abordées suivant une approche géostatistique, et aura contribuer ainsi à répondre au problème central de la problématique de départ.

REFERENCE BIBLIOGRAPHIQUE

A.

Abrahamsen P., Hauge R. and Kolbjørnsen O., Geostatistics Oslo 2012, *Quantitative geology and geostatistics*, 2012, Vol. **XIV**, p**559**.

Albuquerque M., Antunes I., Seco M., Roque N. M., Sanz G., “Sequential Gaussian Simulation of Uranium Spatial Distribution – a Transboundary Watershed Case Study”, *ScienceDirect : Procedia Earth and Planetary Science* 2014, Vol. **8**, p**5**.

B.

Boudour Z., *Détermination de l'épaisseur du mort-terrain par inversion des données électromagnétiques aéroportées*, Université de Montréal, Ecole Polytechnique de Montréal, 2011, p**106**.

Burger H. R., *Exploration Geophysics of The Shallow Subsurface*, Edited, Prentice Hall, New Jersey, 1992, p**490**.

C.

Chiles J.-P., « Quelques méthodes de simulation de fonctions aléatoires intrinsèques », *Cahiers de Géostatistique*, 1995, Vol. **5**, p**16**.

D.

Dagnelie P., *Analyse Statistiques A Plusieurs Variables*, Edited by L. P. A. D. GEMBLoux, A.S.B.L, Belgique, 1986, p**360**.

Desassis N., Renard D., “Automatic variogram modeling by iterative least squares”, 2012, p**23**.

E.

Emery X., *Géostatistique Linéaire*, Edited, France, 2001, p**405**.

F.

Fatemeh Soltani, Peyman Afzal, Omid Asghari, “Sequential Gaussian Simulation in the Sungun Cu Porphyry Deposit and Comparing the Stationary Reproduction with Ordinary Kriging”, *Universal Journal of Geoscience*, 2013, Vol. **2**, p**8**.

G.

Gábor Szatmári, László Pásztor, “Comparison of various uncertainty modelling approaches based on geostatistics and machine learning algorithms”, *GEODERMA*, 2018, p**337**.

J.

Jacques Dubois, Michel Diamant, Jean-Pascal Cogne, *Géophysique : cours et exercices*, Edited, Dunod, Paris, France, 2011, p**120**.

Jim Yeh T.-C., Liu S., Glass R. J., Baker K., Brainard J. R., Alumbaugh D., “A geostatistically based inverse model for electrical resistivity surveys and its application to vadose zone hydrology”, *WATER RESOURCES RESEARCH*, 2002, Vol. **38**(12), p**14**.

K.

Kirkpatrick S., Gelatt C. D., Vecchi M. P., “Optimization by Simulated Annealing”, *SCIENCE*, 1983, Volume **220**(No. 4598), p**11**.

M.

Man K.F., Tang K.S. and Kwong S., *Genetic Algorithms : Concepts and Designs*, Edited by M. A. J. Michael J. Grimble, Springer, London, 2001, p**346**.

Marescot Laurent, “A weighted least-squares inversion algorithm: application to geophysical frequency-domain electromagnetic data”, *Bull. Soc. vaud. Sc. nat.*, 2003, Vol. **88**(0037-9603), p**24**.

Masoomah Delbari, Peyman Afrasiab, Willibald Loiskland, “Using sequential Gaussian simulation to assess the field-scale spatial uncertainty of soil water content”, *Catena*, 2009, Vol. **79**, p**7**.

Matheron G., Technical parametrization of ore reserves, 1978, p**110**.

Mejer Hansen T., Klaus Mosegaard, “VISIM: Sequential simulation for linear inverse problems”, *Computers & Geosciences*, 2007, Vol. **34**, p24.

Meying Arsène , Bidichael Wahile W. E., Gouet Daniel, Ndougsa-Mbarga Théophile , Kuiate Kelian and Ngoh Jean D., “Hydrogeophysical Investigation for Groundwater Resources from Electrical Resistivity Tomography and Self-Potential Data in the Méiganga Area, Adamawa, Cameroon”, *Hindawi : International Journal of Geophysics*, 2018, p15.

Mingkai Qu, Weidong Li, Chuanrong Zhang, “Assessing the risk cost in delineating soil nickel contamination using sequential Gaussian simulation and transfer functions”, *Ecological Informatics*, 2012, Vol. **13**, p17.

Mingkai Qu, Weidong Li, Chuanrong Zhang, “Spatial Distribution and Uncertainty Assessment of Potential Ecological Risks of Heavy Metals in Soil Using Sequential Gaussian Simulation”, *Human and Ecological Risk Assessment: An International Journal*, 2014, Vol. **20(3)**, p17.

N.

Nasser Madani E., Omid Asghari, “Fault detection in 3D by sequential Gaussian simulation of Rock Quality Designation (RQD)”, *Arab J Geosci*, 2012, p11.

Nicolas Bez, Jacques Rivoirard, Jean-Charles Poulard, « Le covariogramme : un outil structural », *Session II : Analyse de l'information*, 1995.

O.

Omid Asghari, Fatemeh Soltani, Hassan Bakhshandeh A., “The Comparison Between Sequential Gaussian Simulation (SGS) of Choghart Ore Deposit and Geostatistical Estimation Through Ordinary Kriging”, *Australian Journal of Basic and Applied Sciences*, 2009, Vol. **3**(ISSN 1991-8178), p12.

Ouadif L., Bahi L., Baba K., “Using geostatistical method to delimit a water bearing formations”, *J. Mater. Environ. Sci*, 2014, Vol. **6**(No. 3), p7.

P.

Penz S., *Modélisation et inversion des données électriques en courant continu : vers une prise en compte efficace de la topographie*, Ecole nationale supérieure des mines de Paris, 2012, p181.

Pham Tuam D., “Estimating Parameters of Optimal Average and Adaptive Wiener Filters for Image Restoration with Sequential Gaussian Simulation”, *IEEE SIGNAL PROCESSING LETTERS*, 2015, Vol. **22**(11), p**12**.

Philip Kearey, Micheael Brooks, Ian Hill, *An Introduction to Geophysical Exploration*, Edited, Blackwell Publishing, Ames, Iowa, 2002, Vol. 3, p**281**.

R.

Richard Webster, Margaret Oliver A., *Geostatistics for Environmental Scientists*, Edited by John Wiley & Sons Ltd, United Kingdom, 2007, p**333**.

S.

Sabrina Metahni, Lucie Coudert, Erwan Gloaguen, Karima Guemiza, Guy Mercier, Jean-Francois Blais, “Comparison of different interpolation methods and sequential Gaussian simulation to estimate volumes of soil contaminated by As, Cr, Cu, PCP and dioxins/furans”, *Environmental Pollution*, 2019, p**255-266**.

Shahrokh Paravarzar, Xavier Emery, Nasser Madani, “Comparing sequential Gaussian and turning bands algorithms for cosimulating grades in multi-element deposits”, *Comptes Rendus Geoscience*, 2015, p**347**.

Z.

Zhanglin Li, Xialin Zhang, Keith C. Clarke, Gang Liu, Rui Zhu, “An automatic variogram modeling method with high reliability fitness and estimates”, *Computers and Geosciences*, 2018, p**39**.