

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/363851390>

Cours et Exercices de Géostatistiques | Course and Exercises of Geostatistics

Book · September 2022

CITATIONS

0

READS

6,658

1 author:



Rabah Kechiched

Université Kasdi Merbah Ouarda

86 PUBLICATIONS 335 CITATIONS

[SEE PROFILE](#)

Université Kasdi Merbah-Ouargla

جامعة قاصدي مرباح - ورقلة

Faculté des Hydrocarbures, énergies renouvelable et
Sciences de la Terre et de l'Univers
Département des Sciences de la Terre et de l'Univers
Route de Ghardaia, BP. 511, 30000 Ouargla



كلية المحروقات والطاقات المتتجدة وعلوم الأرض والكون.
قسم: علوم الأرض والكون
طريق غردية ص. ب 511 . 30000 ورقلة

Polykopie

GEOSTATISTIQUE

Cours, Etude de Cas et Exercices Corrigés

Par

Rabah Kechiched

Professeur des Universités en Géologie

Université Kasdi Merbah – Ouragla, Algérie

Destiné aux étudiants de 3^{ème} Année Géologie des Hydrocarbures

(Niveau 1 – Géostatistique linéaire)

Année Universitaire : 2021-2022

Version Décembre 2023

« Si la géostatistique a simplement réussi à donner l'habitude aux gens de distinguer les problèmes différents ou les effets physiques différents, de ne plus confondre un effet de support une avec une erreur d'échantillonnage ou des choses comme ca, [...] eh ! bien je crois que ce serait déjà pas mal, si on était arrivé à ça ! »

Georges Matheron (1987)

Avant-propos

Ce support cours avec des études de cas et exercices, est destiné aux étudiants géologues en troisième année licence de Géologie des Hydrocarbures de formation professionnelle, mais aussi aux étudiants de niveau 1 en géostatistique linéaire. Ce support cours a pour objectif d'initier d'abord les étudiants par le contenu de cette matière en particulier pour les applications de la théorie des variables régionalisées dans le domaine pétrolier. La géostatistique demeure un outil indispensable depuis les premiers phases d'exploration pétrolière jusqu'au le développement et la production.

Ce support cours qui s'associe aussi avec des exercices d'application s'article essentiellement sur huit chapitres sont rédigés selon *le nouveau programme d'harmonisation.* Chaque chapitre présente un exemple d'application dans le domaine pétrolier avec une étude de cas qui traduit réellement l'apport de la méthode géostatistique dans ce domaine. Le cours sera suivi par une série d'exercices de TD et des manipulations avec des études de cas.

SOMMAIRE

CHAPITRE 1. Rappels sur les notions des statistiques descriptives

- 1.1. Analyse de données monovariable
- 1.2. Analyse de données bivariable
- 1.3. Analyse des données multivariées

CHAPITRE 2. Géostatistique – objet, méthodes et domaines d’application

- 2.1 Notions de base sur la géostatistique
- 2.2 Historique des méthodes géostatistiques
- 2.3. Variogramme
- 2.4. Exemple de l’exploitation des résultats de la variographie (domaine pétrolier)

CHAPITRE 3. ESTIMATION GEOSTATISTIQUE – KRIGEAGE

- 3.1 Krigeage ordinaire
- 3.2. Système de krigeage ordinaire

CHAPITRE 4. SIMULATION GEOSTATISTIQUE - SGS

- 4.1 Notions sur la simulation géostatistique
- 4.2. Définition d’après (Journel et Huijbregts, 1978)
- 4.3. Comparaison entre la simulation et estimation
- 4.4 Types de simulation
- 4.5 Méthodes de la simulation géostatistique
- 4.6. Exemples de la simulation géostatistique
- 4.7. Application des simulations conditionnelles

Références Bibliographiques.

Série de TD’s et exercices corrigés

CHAPITRE 1. Rappels sur les notions des statistiques descriptives

1. Notions de l'analyse statistique des données

1.1. Analyse de données monovariable

Ce traitement permet de déterminer les paramètres statistiques de la distribution des variables étudiées (paramètres de position centrale et de dispersion), et de résumer graphiquement les caractéristiques essentielles de la distribution de chaque variable. Cette analyse permet déjà, d'évoquer l'homogénéité ou l'hétérogénéité des données. Le coefficient de variation, l'allure générale de l'histogramme et de la courbe des fréquences cumulées permettraient d'apprécier l'homogénéité ou l'hétérogénéité des données ainsi que la loi de distribution des variables.

1.2. Analyse de données bivariée

Le traitement statistique bivarié consiste à étudier deux variables aléatoires simultanément et de déterminer la relation qui existe entre eux, à l'aide, entre autre du calcul des coefficients de corrélation linéaires simples. Où celles-ci montrent clairement les relations mutuelles qui existent entre les variables. Le coefficient de corrélation est calculé par la formule suivante :

$$\rho = \text{cov}(x, y) / \sigma_x \sigma_y$$

Ou ρ : coefficient de corrélation simple de la population.

x et y : variables ; σ_x et σ_y : écart type des variables x et y .

$\text{Cov}(x, y)$: covariance de variables x et y .

Rappelons que $\text{cov}(x, y) = E(xy) - E(x).E(y)$, il est estimé par :

$$R_{x/y} = \frac{\overline{XY} - \overline{X}\overline{Y}}{S_x . S_y}$$

Avec X et Y : teneurs moyennes des variables X et Y .

S_x et S_y : écarts types de l'échantillon statistique des variables X et Y .

Les coefficients de corrélation linéaires simples peuvent être regroupés dans un tableau récapitulatif appelé "matrice de corrélation". Notons que la valeur d'un coefficient de corrélation varie entre -1 et +1.

La signification de ce dernier dépend du nombre d'individus utilisés pour son calcul. Une façon empirique de calculer la signification c'est de calculer la valeur de :

$$R_{signif} = \frac{2}{\sqrt{n-1}}$$

Où n est le nombre des données utilisées.

1.3. Analyse des données multivariées

1.3. 1. Analyse en Composantes Principales ACP

L'analyse de données multivariée désigne l'ensemble des méthodes, qui permettent de collecter, d'organiser, d'analyser et de présenter les données. Ces méthodes utilisent des techniques géométriques permettant de mettre en évidence des groupes d'individus homogènes, des liens entre individus et variables, des ressemblances ou des différences entre variables et surtout de représenter les données initiales d'un Tableau R (n, p) par un nombre plus petit sans perdre l'information. Soit un tableau (matrice) R (n, p) où n est le nombre de lignes et p le nombre de colonnes.

Tableau. 1 : Exemple de données sur les paramètres pétrophysiques

Ech.	K	Φ	SH	Sw	Vsh
1	50	5	0.5	0.5	10
2	60	5	0.5	0.5	20
3	100	5	1	0	50
Ni				X(ni,pi)	
N	100	10	10		10

a) Représentation géométrique

On peut faire une représentation graphique des données utilisées par un nuage de points lignes (Représentation R^p (et en nuage de point variables) Représentation dans R^n (Fig. 1).

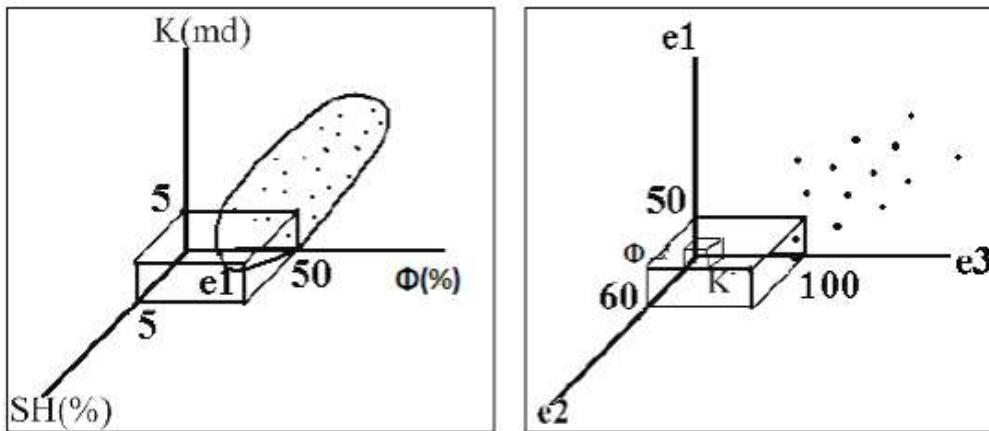


Fig. 1: Représentation géométrique des données : a)- point-lignes R^p et b) point-variables R^n .

Dans R^n , chaque point est défini par n coordonnées. Deux points-lignes apparaissent proche l'un de l'autre, si les deux lignes ont des mesures semblables des variables ou individus et éloignés l'un de l'autre si les mesures sont différentes. En présence d'un nuage de points dans un espace à n dimensions, on cherchera à saisir et comprendre la dispersion des points c'est-à-dire la signification des différents allongements des nuages de points.

L'analyse en Composantes Principales (ACP) est l'une des plus anciennes méthodes de l'analyse factorielle. Elle est utilisée pour l'analyse des données qui sont représentées sous formes de tableaux de n individus et p variables. Soit $R(n, p)$, ce tableau de donnée de terme générale r_{ij} .

En ACP, ce qui importe, c'est la forme du nuage des points par rapport aux axes (droites) qui passent par le centre de gravité du nuage. Il s'agit donc de rechercher les droites qui respectent au mieux l'ensemble des carrés des distances entre couples : C'est celles qui passent par le centre de gravité. D'où la nécessité de procéder à un changement de repère en centrant les variables de la matrice de données initiales. Si de plus les unités des variables sont différentes alors il est préférable de les réduire pour supprimer l'effet unité c'est-à-dire échelle :

$$x_{ij} = \frac{r_{ij} - \bar{r}_j}{\sqrt{n} \cdot \sigma_j}$$

Ainsi après cette transformation (changement), la matrice d'inertie ne serait autre que la matrice de corrélation linéaire simple : $C=XTX$. On appelle inertie du point $N(I)$ par rapport au point P ou l'expression suivant (Fig. 2) :

$$InP(I) = \sum_{i=1}^n m_i \|X_i - P\|^2$$

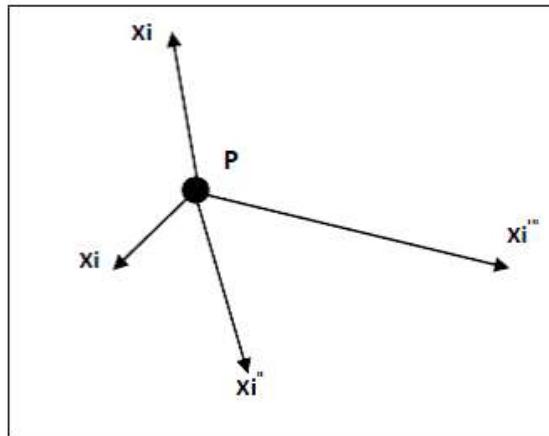


Fig. 2: Représentation géométrique d'inertie des points \$X_i\$ (\$i=1\$ à \$n\$) par rapport au point P.

Après le calcul de C, il est nécessaire de calculer l'inertie du nuage de point dans les principaux axes d'allongement désignés par \$F_i\$ (\$i=1\$ à \$N\$) (Fig. 3) en diagonalisant – c'est à dire en recherchant les premières valeurs propres (variances) et les vecteurs propres correspondants.

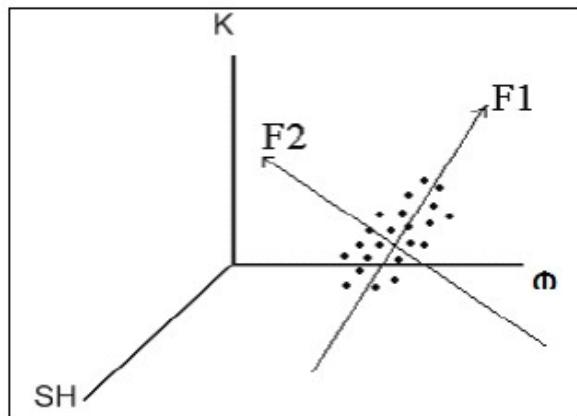


Fig. 3: Représentation d'un nuage de points et les principaux axes d'allongement (Axes factoriels).

La diagonalisation de la matrice C (matrice d'inertie) permet d'obtenir les p vecteurs propres \$U_\alpha\$ et les p valeurs propres \$\lambda_\alpha\$. Le vecteur \$U_\alpha\$ est connu par ces coordonnées dans l'espace \$R^p\$. Les coordonnées des points \$X^i\$ sur les axes factorielles qui sont obtenus par :

$$F_\alpha(i) = X_i U_\alpha$$

$$F_\alpha(i) = \sum_{j=1}^k U_{\alpha j} X_{ij}$$

$F_{\alpha}(i)$ est le facteur de rang α associé à l'individu i : $F \alpha(i)$ est la coordonnée de l'individu (échantillon) i sur l'axe factoriel U . Ce dernier a autant de composantes que de variables. Dans la pratique l'inertie expliquée par l'axe factoriel :

$$T_{\alpha} = \frac{\lambda_{\alpha}}{In(I)} = \frac{\lambda_{\alpha}}{\sum_{\alpha=1}^P \lambda_{\alpha}}$$

Où $In(I)$ est l'inertie totale du nuage de points.

Généralement les résultats d'une ACP sont représentés sous forme graphique (Fig. 4).

Ex : Variances : $F1 = 56,42\%$, $F2 = 17,5\%$, $F3 = 12,8\%$. Variances cumulées =

$F1+F2+F3=86.72\%$: Les 3 facteurs représentent 90.82% de l'information totale.

Tableau. 2 : Exemple de facteurs de charge

Elément	F1	F2	F3
K	-0,48	0,41	0,71
Φ	-0,92	0,24	0,06
So	-0,85	0,26	0,38
Sw	0,85	-0,26	-0,38
Vsh	0,96	0,14	-0,05

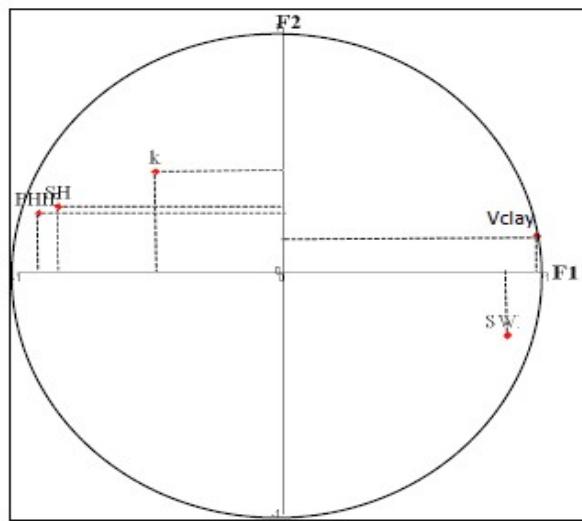


Fig. 4: Représentation des facteurs de charge d'une ACP normée.

Les facteurs des individus sont les coordonnées des individus sur les axes factoriels représentation dans Rp- Leur calcul est effectué à l'aide de la formule suivante : $[Fsc] = [X] \times [F]$: $Fscel = a_{F1}K + b_{F1}\Phi + c_{F1}SH \dots \dots \dots$

CHAPITRE 2. Géostatistique – objet, méthodes et domaines d'application

1. Notions de base sur la géostatistique

1.1. Historique des méthodes géostatistiques

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

Le préfixe "géo" fait référence au domaine des sciences de la terre, qui est historiquement celui où la géostatistique s'est développée. Les champs d'application actuels concernent des domaines très variés.

La désinence "statistique" se réfère à l'utilisation de méthodes probabilistes. L'originalité de la géostatistique par rapport à la statistique classique est la prise en compte de la dépendance entre les observations lorsque celles-ci sont situées dans l'espace. Toutefois, les méthodes géostatistiques ne sont pas exclusivement probabilistes : il existe une branche, connue sous le nom de "géostatistique transitive", qui ne fait pas appel au concept des processus aléatoires (Xavier Emery, 2001).

La géostatistique est connue depuis près de 40 ans dans les mines d'or du Witwatersrand où Daniel Krige proposa une correction statistique à la manière traditionnelle d'estimer la teneur d'un bloc de mineraï à partir d'un nombre limité d'échantillons pris autour du bloc à exploiter. La théorie était formulée 10 ans plus tard par Georges Matheron qui introduisit un outil pour analyser la continuité spatiale des teneurs appelé "le variogramme" et une méthode d'estimation basée sur le variogramme appelée "le krigeage".

L'avenir de la géostatistique semble brillant. Avec toutes les expériences et les développements des trente dernières années, la géostatistique est devenue une alternative possible aux méthodes géométriques traditionnelles de l'estimation des gisements. D'autres disciplines qui utilisent des données distribuées spatialement (pétrole, environnement, hydrologie, océanographie, foresterie) ont commencé à l'adopter.

1.2. Notions de la géostatistique

Matheron (1971), définit la géostatistique de la façon suivante : "la géostatistique est l'application du formalisme des fonctions aléatoires à la reconnaissance et à l'estimation des phénomènes naturels". **Une fonction aléatoire** (F.A) $Z(x)$ est un ensemble de variable aléatoire $Z(x_i)$, définie en chaque point x_i du gisement G:

$$Z(x) = \{Z(x_i), \forall x_i \in G\}$$

Un phénomène minéralisé peut être caractérisé par la répartition dans l'espace d'un certain nombre de grandeurs mesurables, que nous appelons " **variables régionalisées**"(V.R). Ces variables ont une structure d'auto-corrélation qui dépend du module et de la direction du vecteur séparant deux points de mesure. Mathématiquement, une variable régionalisée est une fonction du point x (Ploner & Dutter, 2000). Cette fonction est généralement irrégulière et montre deux aspects complémentaires (Matheron, 1973) :

- Un aspect aléatoire qui explique les irrégularités locales;
- Un aspect structuré qui reflète les tendances du phénomène à grande échelle.

Si au point x_i de l'espace, la V.R $z(x_i)$ est considéré comme valeur unique (valeur vraie) alors la géostatistique étudiera la corrélation spatiale de la V.R et la structure de cette variable dans l'espace. C'est **la géostatistique transitive** (Chauvet, 1999).

Le choix constitutif de la géostatistique minière consiste à interpréter chaque valeur de la variable régionalisée $z(x)$ comme une réalisation particulière d'une variable aléatoire $Z(x_i)$ implanté au point x_i , donc plusieurs réalisations sont possibles. C'est **la géostatistique intrinsèque**.

Ce choix constitutif fondamental conduit à des solutions cohérentes et acceptables pour divers problèmes pratiques miniers posés, tel que le problème de l'estimation des réserves locales et récupérables.

1.3. Variogramme

Soit deux variables aléatoires, $z(x)$ et $z(x+h)$ mesurées en 2 points "x" et "x+h" (Fig. 23), la variabilité entre ces deux points (Fig. 9) se caractérise par une fonction notée $2\gamma(x,h)$ appelée variogramme (Journel et Huijbregts, 1978).

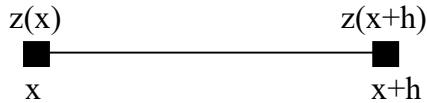


Fig. 9 : Représentation schématique des points de mesure de la variable

Le variogramme théorique est défini comme étant l'espérance quadratique de la variable aléatoire $[z(x) - z(x+h)]$ (Marcotte, 2003) soit:

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

En général, le variogramme $\gamma(x,h)$ (ou semi-variogramme) est une fonction à la fois du point x et du vecteur h (Fig. 10) mais l'estimation de ce variogramme nécessite plusieurs réalisations, or en pratique on ne dispose que d'une seule réalisation, soit le couple $[z(x), z(x+h)]$ de mesure effectuées aux points x et x+h, d'où l'introduction de l'hypothèse intrinsèque: le variogramme ne dépend que du vecteur h en module et en direction et non du point x. En admettant cette hypothèse, le semi-variogramme expérimentale sera estimé par la formule :

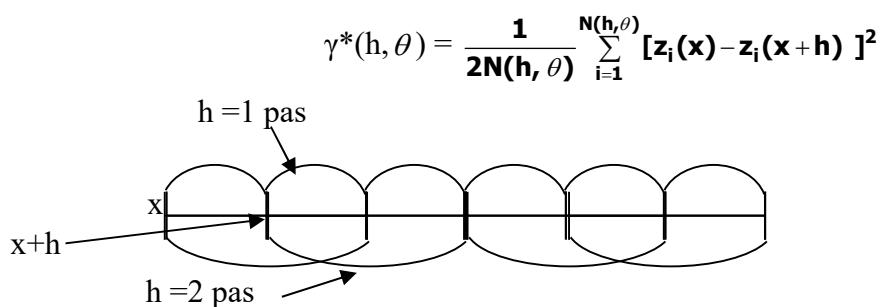


Fig. 10 : Estimation du variogramme

Où $N(h, \theta)$ est le nombre de couples $[z(x), z(x+h)]$ séparé par une distance h suivant une direction θ . Les résultats obtenus peuvent être représentés sous forme graphique (Fig. 8).

a) Propriétés du variogramme directionnel

On fixe une direction θ , en partant de l'origine $h = 0$, $\gamma(x,h) = 0$, en général le variogramme croît avec le module du vecteur h . Cette croissance caractérise le degré de continuité de la variable étudiée.

Généralement, la stabilisation du variogramme est marquée au delà d'une certaine valeur de h notée a appelée portée du variogramme. Ce dernier reste inchangé au-delà de $h = a$ et ne dépend plus de l'interdistance entre les valeurs : les mesures ne sont plus corrélées. La portée "a" donne alors un sens à la notion de zone d'influence d'un échantillon ou d'un sondage. Cette portée n'est pas forcément la même dans toutes les directions de l'espace. Elle permet donc de détecter une éventuelle anisotropie ou isotropie dans la structure de la minéralisation.

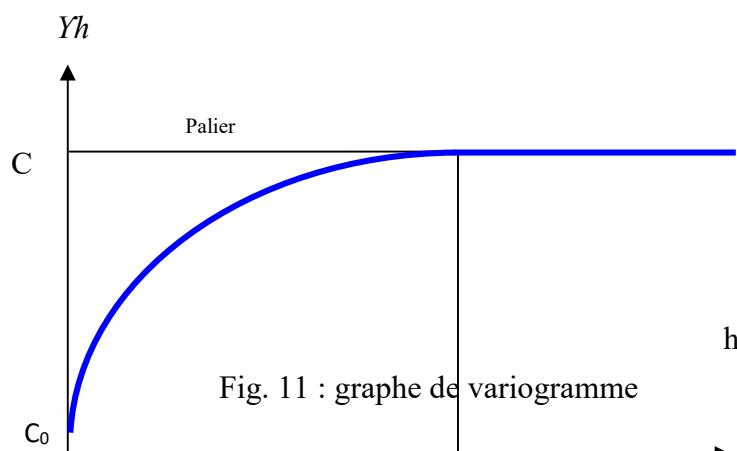


Fig. 11 : graphe de variogramme

Le variogramme croît jusqu'à une c_a : ~~pointe~~ limite appelé palier puis s'aplatit, mais il peut également croître indéfiniment (effet de dérive).

b) Schéma théorique et ajustement des variogrammes

Une fois le variogramme expérimental est calculé, il faut alors déterminer un modèle mathématique qui lui correspond et qui doit être opérationnel et simple à l'emploi : C'est l'ajustement du variogramme expérimental. Les schémas théoriques d'usage courant sont classés en schémas à palier, schémas sans palier et schémas à effet de trous (**Fig. 12**).

- Schéma à palier, présente:

- Un comportement linéaire à l'origine:

✓ Schéma sphérique: $\gamma(h) = \frac{3}{2} \cdot \frac{h}{a} - \frac{1}{2} \cdot \frac{h^3}{a^3} \dots \forall h \in [0, a]$

$$\gamma(h) = 1 \dots \text{pour } h \geq a$$

✓ Schéma exponentiel: $\gamma(h) = 1 - e^{-h/a} \dots \forall h \geq 0$

- Un comportement parabolique à l'origine:

✓ Schéma gaussien: $\gamma(h) = 1 - e^{-h^2/a}$

- Schéma sans palier

- Schéma en puissance: $\gamma(h) = h^\lambda \dots \forall h > 0, \dots 0 < \lambda < 2$
- Schéma logarithmique: $\gamma(h) = a \log(h) \dots \forall h > 0, \dots a = \text{constant} > 0$
- Schéma à effet de trou $\gamma(h) = 1 - \frac{\sin(h)}{h} \dots \forall h > 0, \gamma(h) = \frac{h^2}{6} \dots \forall h \rightarrow \infty$

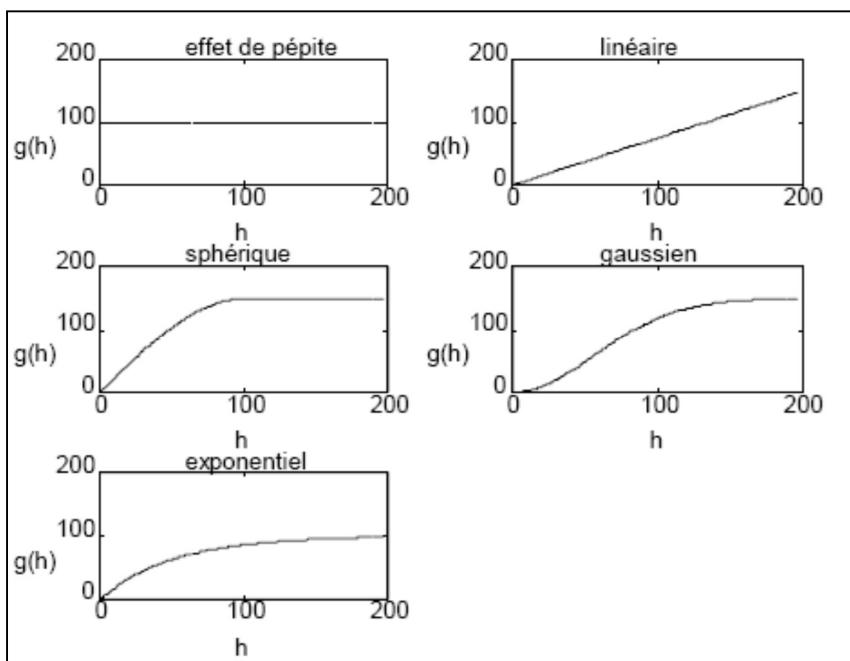


Fig. 12 : Schémas théoriques d'ajustement des variogrammes In (Marcotte, 2003)

c) Isotropie et anisotropie

Le calcul des variogrammes directionnels lorsque le données sont réparties en deux directions 2D. Ceci permet de connaître si la structure de la variabilité est la même dans les différentes directions où non. Les variogrammes présentant les mêmes allures (de paliers et de portées et effet de pépite), on parle de phénomènes **d'isotropie**, sinon le phénomène est dite **anisotrope**.

L'anisotropie est due à des phénomènes génétiques où à des données hétérogènes. Généralement, deux anisotropies sont distinguées.

a) Anisotropie géométrique

Il y a anisotropie géométrique quand les variogrammes présentent la même variabilité globale et en particulier le palier mais ont des portées différentes (Fig. 10).

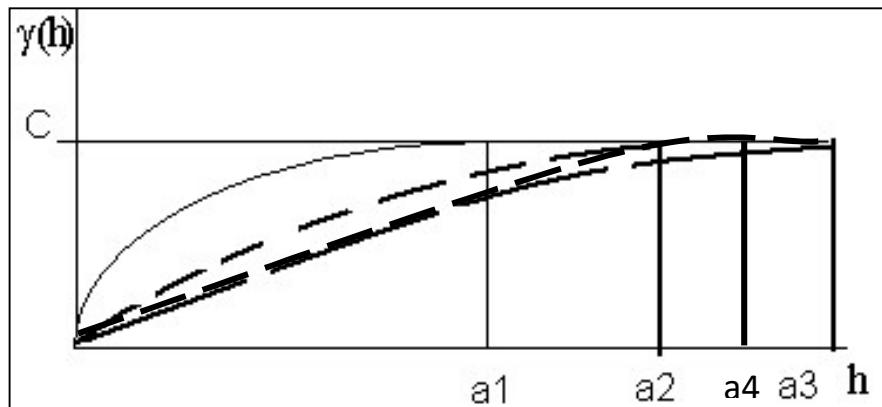


Fig. 13 : Schéma d'une anisotropie géométrique

L'établissement des roses de portées permet l'étude de l'anisotropie en projetant les portées sur une rose. Ce rasasse permet aussi l'identification du sens de la continuité du phénomène étudié (grand portée). Cependant la direction d'aplatissement de l'ellipse représente la direction de la variabilité rapide (Fig. 11).

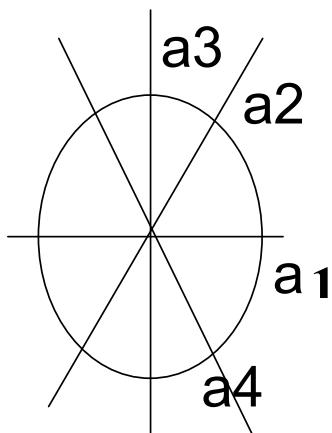


Fig. 14 : Rose des portées d'une anisotropie géométrique

$$\text{Cof. Anis. : } \frac{\text{Petite portée}}{\text{Grande portée}}$$

b) Anisotropie zonale

Cette anisotropie est les plus fréquentes et elle est manifestée par des variogrammes de portées et les paliers différents. Elle correspond généralement à une anisotropie génétique (Fig. 15).

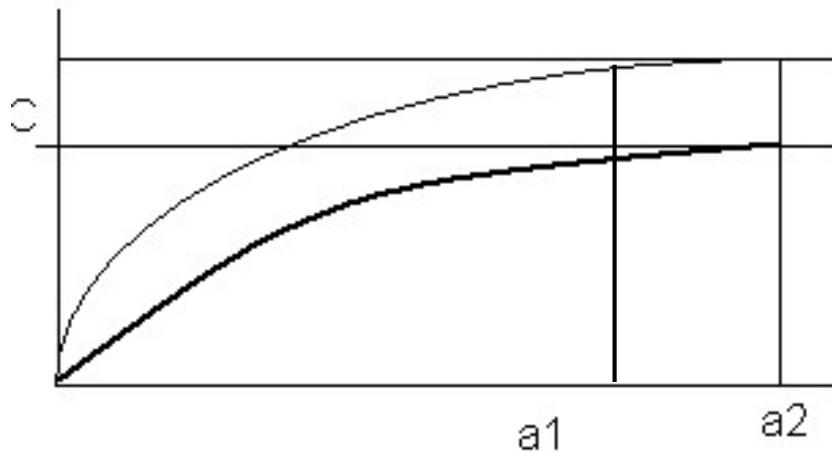


Fig. 15 : Schéma de variogrammes représentants une anisotropie zonale.

2.3. Exemple de l'exploitation des résultats de la variographie (domaine pétrolier)

L'outil géostatistique de variogramme représente un moyen pour étudier la structure de la variabilité régionalisée. Dans cette étude de cas ; c'est la porosité. Les variogrammes directionnels sont construits et ajustés par des modèles de régionalisation sphérique à portée, palier et à effet de pépite. Les roses de portées indiquent le sens d'isotropie et l'anisotropie, les coefficients d'anisotropie donnent le degré de l'anisotropie de la porosité. Ces résultats traduisent les caractéristiques sédimentologiques et pétrophysiques :

- la direction d'isotropie indique la direction de la continuité géologique du phénomène ; (Sens de chenaux par exemple).
- La direction de l'anisotropie qu'est généralement orientée orthogonalement à la direction d'isotropie indique le sens de la discontinuité (grande variabilité).

Ces résultats peuvent être utilisés en exploration et en développement des champs pétroliers. Ceci en optimisant la maille de prospection en fonction des roses des portées.

CHAPITRE 3. ESTIMATION

GEOSTATISTIQUE - KRIGEAGE

La recherche des estimateurs efficace (sans biais et variance d'estimation minimale) a poussé au développement de procédures géostatistiques permettant la construction d'estimateur, sans biais et des variances minimales: Il s'agit des procédures de krigeage. Dans le cas stationnaire, on en reconnaît deux types principaux, selon que la moyenne du processus est connue ou non, si μ est connu le krigeage est ordinaire, si μ est inconnu le krigeage est simple. Ce dernier est, de loin, le plus fréquemment utilisé surtout dans le domaine de la simulation.

1.1. Krigeage ordinaire

L'estimation locale par krigeage ordinaire consiste à trouver le meilleur estimateur linéaire de la valeur moyenne d'une variable régionalisée sur un domaine de petite dimension vis-à-vis des zones de stationnarité du gisement. Le krigeage ordinaire est établi le plus souvent en considérant une F.A qui satisfait aux hypothèses:

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

$$\sigma^2 [z(x) - z(x + h)] = 2 \gamma(h)$$

Le Krigeage est constituée des étapes de **Linéarité**, **d'Autorisation**, **d'Universalité** et **d'Optimalité**. Ces étapes peuvent être dans n'importe quel ordre.

- Linéarité :

$$Z_v^* - m = \sum_{i=1}^N \lambda_i (z_i - m)$$

- Autorisation : il se trouve que nous sommes dans le cadre d'un modèle de stationnarité, de sorte que toutes les combinaisons linéaires sont autorisées. Il n'y a pas de condition d'autorisation.
- Universalité : conditionnellement aux deux contraintes précédentes :

$$E [\lambda z_{xi}^* - z_x] = 0 \iff \sum [\lambda_i z_{xi}^* - z_x] = 0$$

Nous savons que $E[z_x] = m$ (m : la moyenne), mais par hypothèse nous ne connaissons pas m .

Alors pour que : $E[\lambda z^*_{xi} - z_x] = m$ ($\sum_{i=1}^k \lambda_i - 1$) Soit nul,

Il faut et il suffit que : $\sum_{i=1}^k \lambda_i - 1 = 0 \iff \sum \lambda_i = 1$

Telle est la contrainte d'universalité portant sur λ_i .

- Optimalité : il nous reste maintenant à minimiser la variance d'estimation $\sigma^2_E [z_v^* - z_v]$. cette minimisation s'obtient par la technique classique des multiplicateurs de Lagrange. Le formalisme de Lagrange fournit alors le système d'équations linéaires dit "**système de krigeage**" à $n+1$ équations linéaires et $n+1$ inconnus qui contiennent les n pondérateurs λ_i et le paramètre de Lagrange μ

a. **Système de krigeage ordinaire** à $n+1$ équations linéaires et $n+1$ inconnus qui contiennent les n pondérateurs λ_i et le paramètre de Lagrange μ

$$\left\{ \begin{array}{l} \sum_{j=1}^n \lambda_j \gamma(v_i, v_j) + \mu = \gamma(v_i, V) \quad \forall i = 1 \text{ à } n \\ \sum_{i=1}^n \lambda_i = 1 \quad \forall j = 1 \text{ à } n \end{array} \right.$$

La variance d'estimation minimisée ou variance de krigeage s'écrit alors :

$$\sigma^2_K = E \left\{ |z_y^* - z_k|^2 \right\} = 2 \sum_{i=1}^n \lambda_i \gamma(v_i, V) + \mu - \gamma(V, V)$$

Où :

- $\gamma(V, V)$: le variogramme moyen du domaine à estimé intervenant dans la variance de krigeage.
- $\gamma(v_i, V)$: le variogramme moyen entre l'estimé et les estimants.
- $\gamma(v_i, v_j)$: le variogramme moyen entre les estimants.

CHAPITRE 4. SIMULATION

GEOSTATISTIQUE - SGS

1. Notions sur la simulation géostatistique

1.1. Définition d'après (Journel et Huijbregts, 1978)

Considérant par exemple la variable teneur vraie $z_0(x)$ en un point x du gisement. Le modèle probabiliste proposé par la géostatistique consiste à interpréter la répartition spatiale de cette teneur comme une réalisation particulière d'une fonction aléatoire (F.A), $Z(x)$ caractérisée par ses 2 moments, ceux-ci étant estimé à partir des données expérimentales. Ce modèle est adapté aux problèmes pratiques de détermination de diverses dispersions de teneurs puisque l'on démontre que ces dispersions sont caractérisées pour l'essentiel par ces deux premiers moments de la fonction aléatoire $Z(x)$.

Une simulation consiste à tirer une autre réalisation $z_s(x)$ de cette fonction aléatoire. Les deux réalisations ; réelle $z_0(x)$ et simulé $z_s(x)$ sont différentes mais proviennent d'une même fonction aléatoire de $Z(x)$, dont les deux premiers moments sont fixés.

On peut donc, pour l'étude des dispersions des teneurs, travailler indifféremment sur le gisement simulé $\{z_s(x)\}$ ou sur le gisement réel $\{z_0(x)\}$. Le gisement simulé présente l'avantage d'être connu en tous les points et non pas seulement aux seuls points d'implantation des sondages expérimentaux.

- **Conditionnalisation :**

Une F.A. $Z(x)$ possède une infinité de réalisations possibles $\{z_s(x)\}$. On choisira parmi cette infinité de simulation $\{z_{sc}(x)\}$ retrouvent les données expérimentales aux points d'implantation x_i des sondages, c'est-à-dire vérifient :

$$z_{sc}(x_i) = z_0(x_i), \text{ pour chaque sondage } i$$

C'est la conditionnalisation par les données expérimentales ; gisements simulés et gisements réel présenteront aux mêmes endroits les mêmes concentrations de données riches ou de données pauvres.

1.2. Comparaison entre la simulation et estimation

Les gisements simulés passent par les données expérimentales et présentent les mêmes caractéristiques de dispersions que le gisement réel. On peut donc se demander en quoi ces simulations diffèrents d'une estimation ?

1.2.1. L'estimation

Une estimation choisira en chaque point x à fournir un estimateur $z^*_o(x)$ aussi proche que possible de la teneur vraie inconnue $z_0(x)$. Mais ces estimateurs $\{z^*_o(x)\}$ n'ont aucune de reproduire les dispersions vraies $\{z_0(x)\}$; bien au contraire les estimateurs sous-estiment (ou lisse) en générale les dispersions vraies. Le gisement estimé $\{z^*_o(x)\}$ représente une base biaisée dans le mauvais sens pour l'étude des dispersions qui nous intéresse

1.2.2. La simulation

La simulation $\{z_s(x)\}$ retrouve les deux premiers moments expérimentaux des données réelles $\{z_0(x)\}$. C'est-à-dire identifie les principales caractéristiques de dispersions des teneurs vraies. En contre partie, en chaque point x la valeur simulée $z_s(x)$ n'est pas le meilleur estimateur de la $z_0(x)$ que l'on puisse former.

Les objectifs de l'estimation et de la simulation ne sont pas en général compatibles, et une simulation ne serait dispenser des frais de reconnaissance nécessaires à une bonne estimation (Tableau. 17).

Tableau.17 : Comparaison entre estimation et simulation

Caractéristiques à reproduire	Estimation	Simulation
Histogramme	Non (pas obligatoirement)	Oui (obligatoire)
Variogramme	Non (pas obligatoirement)	Oui (obligatoire)
Variance d'estimation	Minimale	Pas forcément la plus petite

1.3. Types de simulation

1.3.1. La simulation non-conditionnelle

Une méthode de simulation **non-conditionnelle** vise à produire des champs montrant la même structure spatiale (variogramme) et le même histogramme que ceux inférés à partir des données observées.

Lorsqu'on applique l'algorithme de simulation correspondant à une méthode donnée, on obtient une réalisation. On peut appliquer l'algorithme autant de fois que l'on veut et obtenir ainsi plusieurs réalisations. Les réalisations sont indépendantes les unes des autres.

1.3.2. La simulation conditionnelle

Une méthode de simulation **conditionnelle** vise les mêmes objectifs que la simulation non-conditionnelle mais, en plus, elle doit assurer que chaque réalisation soit compatible avec les valeurs observées aux points échantillons. Ainsi, si on simule une valeur en un point coïncidant avec une observation, il faut que la valeur simulée soit égale à la valeur réellement observée (Marcotte, 2003)

a) Principe de la simulation conditionnelle

Considérons en chaque point x la teneur réelle $z_0(x)$ est son estimateur de krigeage $z_{OK}^*(x)$ déduit de l'information possible :

$$I : \{z_0(x_i), x_i \in I\}$$

Ces deux valeurs diffèrent par une erreur inconnue :

$$z_0(x) = z_{OK}^*(x) + [z_0(x) - z_{OK}^*(x)]$$

Soit en termes de fonctions aléatoires :

$$Z(x) = Z_K^*(x) + [Z_0(x) - Z_K^*(x)]$$

L'estimateur $z_{OK}^*(x)$ étant un krigeage, les deux F.A. krigeage $Z_K^*(x)$ et erreur de krigeage $[Z(x) - Z_K^*(x)]$ sont orthogonales (propriété de krigeage).

Pour simuler $z_0(x)$, il suffit donc d'ajouter à $z_{OK}^*(x)$ une réalisation quelconque d'une erreur indépendante de z_K^* et isomorphe à l'erreur vraie $[Z(x) - Z_K^*(x)]$. Pour cela :

- On considère une F.A. $Z_S(x)$ indépendante de $Z(x)$ et isomorphe à $Z(x)$. Sur tout réalisation $z_S(x)$ de $Z_S(x)$, la procédure de krigeage $z_{SK}^*(x)$ appliquée sur la même

configuration de données disponibles $\{z_S(x_i), x_i \in I\}$ conduit aux décompositions analogue :

$$z_S(x) = z_{SK}^*(x) + [z_S(x) - z_{SK}^*]$$

$$\text{Soit : } Z_S(x) = Z_{SK}^*(x) + [Z_S(x) - Z_{SK}^*]$$

On remarque alors le deuxième membre de (1) l'erreur vraie inconnue : $[Z_0(x) - Z_{OK}^*(x)]$ Par l'erreur simulée connue $[Z_S(x) - Z_{SK}^*(x)]$

On obtient alors la réalisation connue :

$$z_{SC}(x) = z_{OK}^*(x) + [z_S(x) - z_{SK}^*(x)]$$

Correspondant à la fonction aléatoire :

$$Z_{SC}(x) = Z_{OK}^*(x) + [Z_S(x) - Z_{SK}^*(x)]$$

b) Isomorphisme :

Les F.A. $Z(x)$ étant isomorphe, l'erreur $Z(x)$ et $Z_S(x)$ étant isomorphe, l'erreur $[Z_S(x) - Z_{SK}^*(x)]$ est isomorphe, à l'erreur $[Z_0(x) - Z_{OK}^*(x)]$ et orthogonale à $Z_K^*(x)$. Donc les 2 F.A. $Z(x)$ et $Z_{SC}(x)$ sont également isomorphes, c'est-à-dire qu'elles ont même espérance mathématique et même covariance :

$$E\{Z_{SC}(x)\} = E\{Z(x)\} = m \text{ Supposée nulle ;}$$

$$E\{Z_{SC}(x).Z_{SC}(x')\} = E\{Z(x).Z(x')\}, \forall x, x' \in G$$

c) Conditionnalisation :

Le krigeage fournissant des estimateurs exacts :

$$z_{OK}^*(x_i) = z_0(x_i) \text{ et } z_{SK}^*(x_i) = z_S(x_i), \forall x_i \in I$$

Il vient :

$$z_{SC}(x_i) = z_0(x_i), \forall x_i \in I$$

La réalisation simulée est bien conditionnelle aux données expérimentales.

La figure (Fig. 26) illustre la différence entre le krigeage, les simulations géostatistiques non-conditionnelles et les simulations géostatistiques conditionnelles

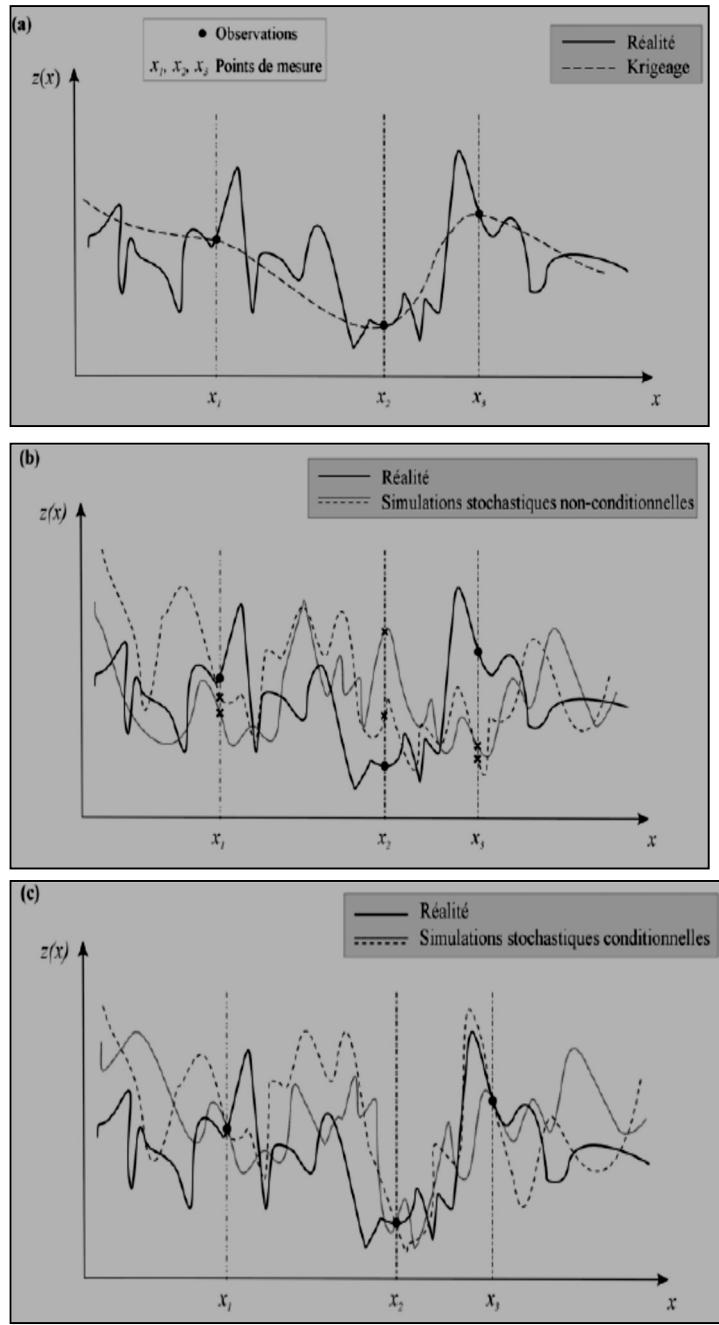


Fig. 26 : Comparaison 1-D des valeurs réelles et des valeurs obtenue. (Journel, 1973).

- (a) par krigage,
- (b) simulations non conditionnelles
- (c) simulations conditionnelles

1.4. Méthodes de la simulation géostatistique

Il existe une plusieurs méthodes de simulation géostatistique, les approches gaussiennes produisent des images d'une variable continue de loi de distribution gaussienne, possédant toutes la même espérance, la même variance et la même fonction de variogramme (ou de covariance). Les approches non-gaussiennes produisent, elles, des images dont la variable ne suit pas une loi de distribution gaussienne. Si on cherche à regrouper ces méthodes, une subdivision possible est la suivante:

1.4.1. Les méthodes gaussiennes : Parmi les méthodes gaussiennes, on peut citer :

- méthodes matricielles basées sur une décomposition de la matrice de covariance ;
- méthode gaussienne séquentielle ;
- moyennes mobiles ;
- bandes tournantes ;
- méthodes autorégressives ;
- méthodes fréquentielles ;

La méthode gaussienne séquentielle est la plus utilisée dans la simulation statique des réservoirs. Selon (P. Kedzierski, 2007), les étapes de cette méthode de simulation sont :

- 1- Choisir (aléatoirement) un nœud de la grille,
- 2- définir un ellipsoïde de recherche indiquant quels points sont à prendre en compte dans la simulation, L'ellipsoïde de recherche est défini à partir du variogramme, et plus précisément à partir des portées de variogrammes dans différentes directions. Il permet de réduire le nombre de points à prendre en compte et par conséquent de réduire les temps de calcul.
- 3- kriger la valeur à cette position,
- 4- à partir de l'estimation du krigeage z^* et de la variance d'estimation σ_e^2 , construire une loi gaussienne de moyenne z^* et de variance σ_e^2 .
- 5- tirer aléatoirement une valeur dans cette distribution (Monte- Carlo) ; cette simulation est dite gaussienne car le tirage de Monte-Carlo se fait dans une loi gaussienne. Cette loi normale est construite à partir des résultats du krigeage au point considéré.
- 6- ajouter la valeur simulée à la liste des valeurs disponibles, on dit que cette simulation qu'elle est séquentielle car on ajoute les valeurs simulées à la liste des valeurs disponibles pour l'étape suivante (manière séquentielle).

Cette simulation est appliquée sur des observations suivant une loi gaussienne, si la distribution est n'est pas gaussienne, Une transformation de la variable en une variable gaussienne Y. Il s'agit d'une anamorphose $Y = F(Z)$. La simulation est effectuée sur Y pour la transformée inverse F^{-1} est appliquée.

1.4.2. Les méthodes non-gaussiennes

- recuit simulé ;
- simulation séquentielle d'indicatrices ;
- méthodes utilisant des champs de probabilité ("p-fields").

La méthode de recuit simulé qui est très utilisée parmi les méthodes non-gaussiennes, est basée sur une technique d'optimisation: le recuit simulé ("simulated annealing"). Cette méthode est d'une grande flexibilité. Brièvement, il s'agit de définir une fonction objectif à optimiser. Cette fonction objectif peut être la différence entre le variogramme expérimental des données simulées et le variogramme théorique que l'on désire reproduire. On peut aussi inclure plusieurs autres termes dans la fonction objective à minimiser pour incorporer diverses contraintes géologiques, le respect d'une courbe tonnage-teneur connue, des relations déterministes, etc. Il existe plusieurs variantes du recuit simulé

1.5. Exemples de la simulation géostatistique

1.5.1. Simulation géostatistique d'une longueur (1D) (In D. Marcotte, 2003)

En vue de disposer un câble de fibre optique sur les fonds marins, on connaît la profondeur de la mer par les points d'observations, on se pose le problème suivant : la profondeur $z(x)$ n'étant connue que dans les points d'observation répartis à peu près chaque 10 Km (Fig. 27), quelle est la longueur d'un câble sous marin qui doit reposer sur le fond de la mer ?

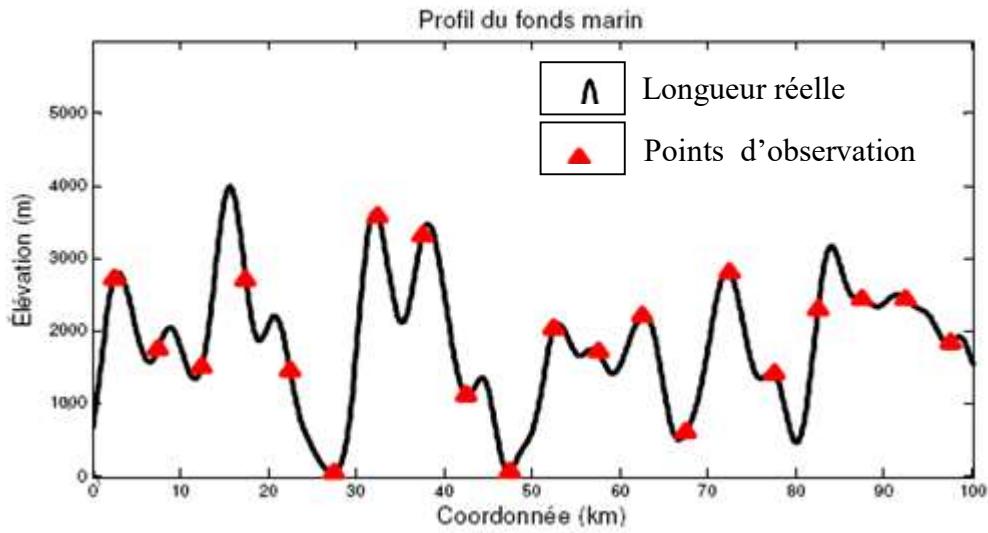


Fig. 27 : La longueur réelle du fond marins et postion des points d'observation.

1.5.1. L'estimation par krigeage

En effectuant le krigeage, on remarque que le krigeage sous estime la longueur du câble avec une longueur de 104,6 km, alors que la longueur réelle est de 110 Km, et par conséquent on va manquer du câble ! (Fig. 28)

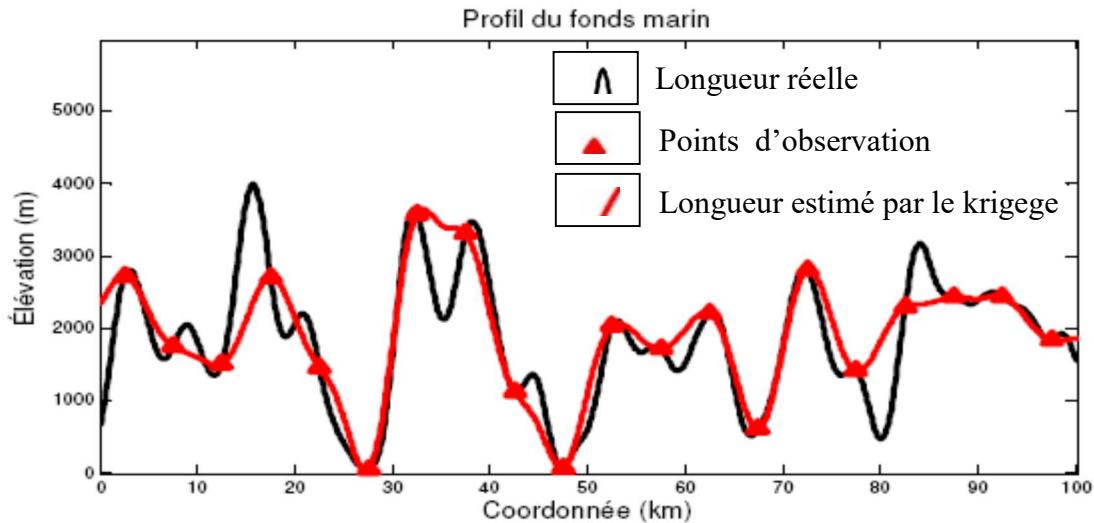
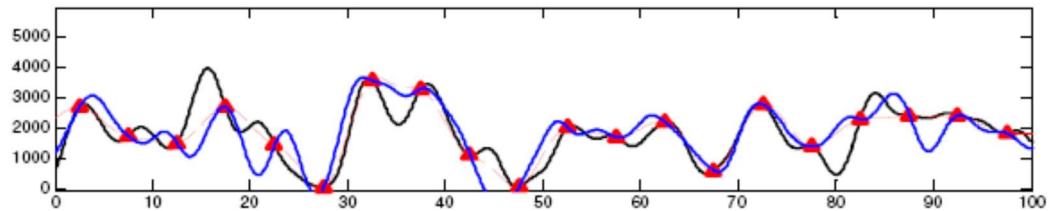


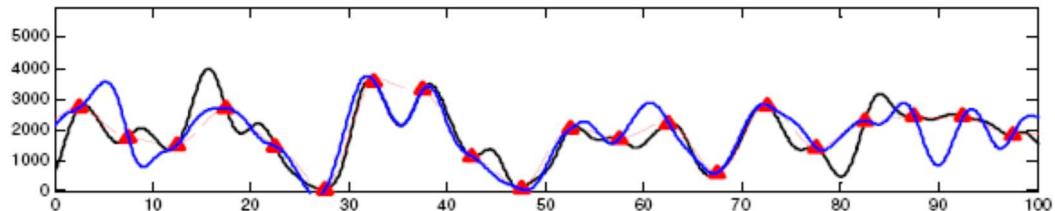
Fig. 28 : La longueur estimé par krigeage en tenant compte les points d'observation.

1.5.2. L'approche de la simulation géostatistique conditionnelle

En effectuant la simulation géostatistique conditionnelle, on remarque que cette approche répond bien à notre problème (Fig. 29).



Réalisation 1



Réalisation 2

Fig. 29 : Simulation conditionnelle de la longueur du câble sous marin

Si on effectue 1000 réalisations, il y'a une possibilité que la moyenne des longueurs simulées égale à la longueur réelle qui est de 110 km (Fig. 30 et 31).

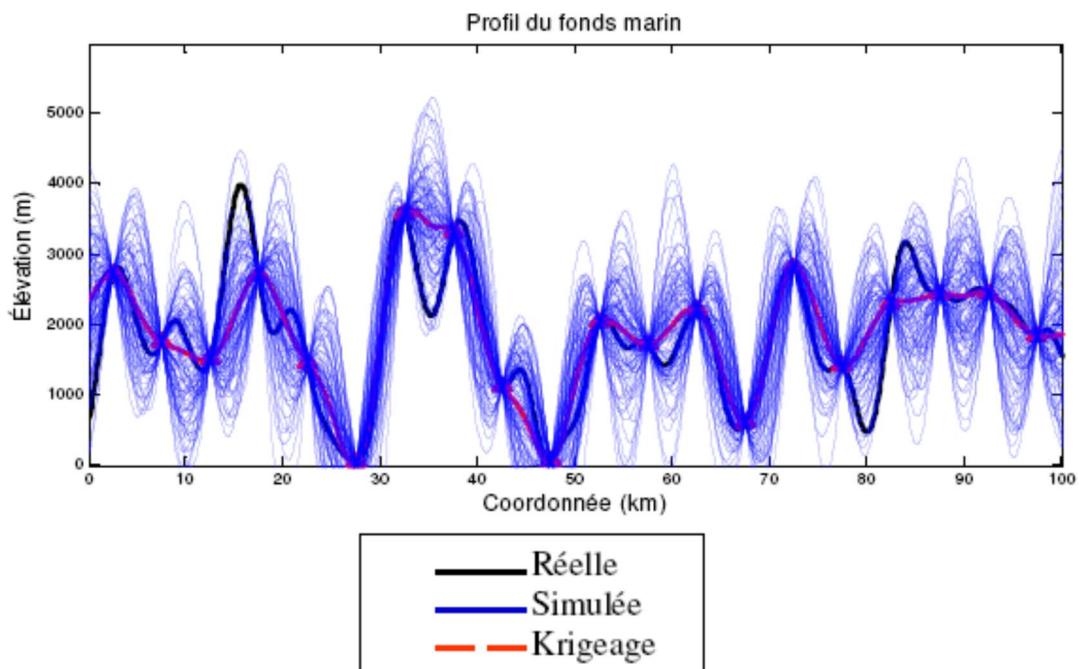


Fig. 30 : Représentation de 1000 réalisations de simulation conditionnelle.

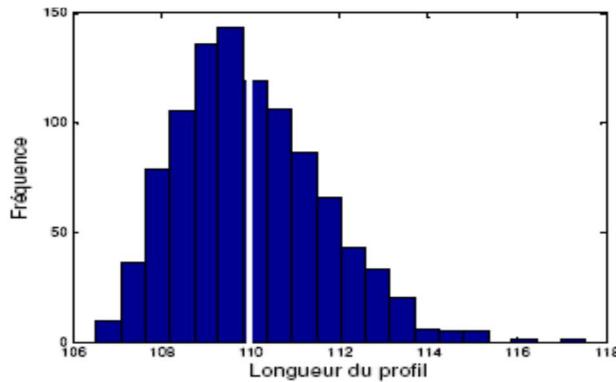


Fig. 31: Histogramme des longueurs simulées (de 1000 réalisations).

1.5.3. Simulation d'un champ de porosité (2D)

Dans cet exemple, la simulation de la porosité est appliquée à deux dimensions (Fig. 32 et 33) montre les points d'observation et la distribution réelle respectivement (Gslib, 2007).

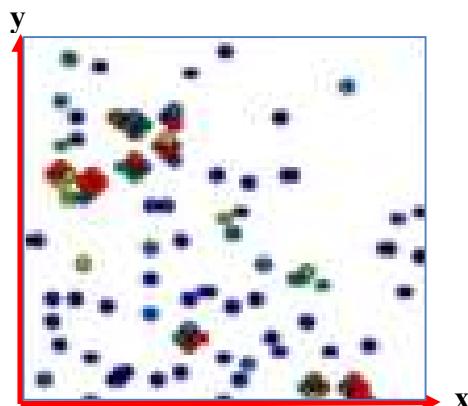


Fig. 32 : Plan d'échantillonnage

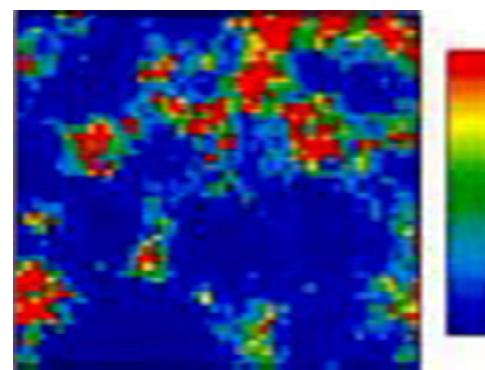


Fig. 33 : La distribution réelle.

A partir des résultats des observations du plan d'échantillonnage (Fig. 32), l'estimation s'effectue par le Krigeage (Fig. 34)

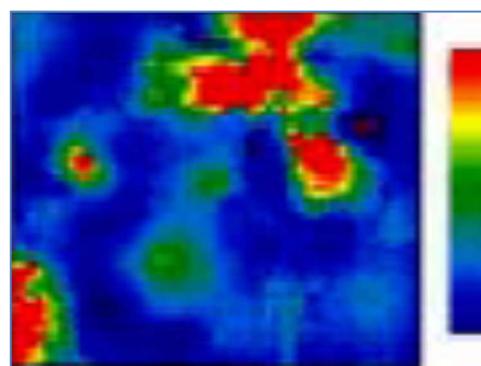


Fig. 34 : Estimation par krigeage

Remarquant que le krigeage (Fig. 34) estime la porosité localement, les points mesurés coïncide avec les points estimés (en faisant la comparaison avec les observations mentionnées sur le plan d'échantillonnage (Fig. 32) parce que le krigeage est un interpolateur exacte, dont les points à estimé correspond au point mesuré). Il est à noté que les variances sont minimes. Si on fait une comparaison avec la distribution réelle (Fig. 33), on constate que la carte obtenue par le krigeage est lissée et on ne visualise que les tendances de la distribution réelle.

En utilisant la simulation géostatistique, on peut avoir plusieurs réalisations de la distribution (Fig. 35 et 36).

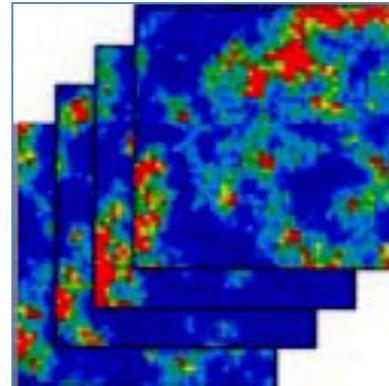
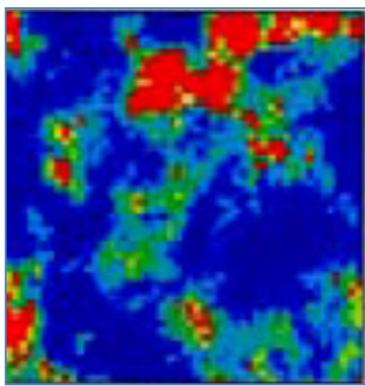


Fig. 35 : carte obtenue par simulation Fig. 36: multiples réalisations de distribution simulées

A partir des réalisations obtenues par la simulation géostatistiques (Fig. 31 et 32), on constate l'absence du lissage avec des valeurs extrêmes importantes. Cette technique donne l'avantage d'obtenir plusieurs réalisations qui répondent au même histogramme et variogramme que la distribution des données initiale (Fig. 32).

1.6. Application des simulations conditionnelles

a) Domaine des mines et pétrole

La simulation conditionnelle est utilisée pour modéliser les précisions sur les estimations des réserves récupérables (i.e. après application d'une teneur de coupure).

Les exemples d'application en pétrole sont les plus nombreux mais aussi les plus complexes. La complexité provient en premier lieu de la complexité géologique des réservoirs (failles, paléochenaux, types de roche, etc.), du peu de données directes (forer un puits coûte

extrêmement cher) et de l'abondance des données indirectes (relevés de sismique réflexion, diagraphies, observations de champs pétroliers comparables, données de production (débits, pressions, etc.), expertise géologique, etc. La plupart des études combinent plusieurs méthodes de simulation (méthodes booléennes, points marqués, simulations géostatistiques).

b) Autres domaines et utilisations possibles

- calculer la précision pour l'estimation de volumes contaminés au delà d'une certaine norme.
- modéliser la variabilité des teneurs observées au concentrateur et étudier l'impact de modifications au plan de minage sur cette variabilité.
- étudier différents scénarios pour le design d'une fosse à ciel ouvert et déterminer les emplacements optimaux pour les équipements. On pourrait envisager la même chose pour les exploitations souterraines mais le problème est beaucoup plus complexe.
- pour une cimenterie, fournir une base de données de "carrières" possibles sur lesquelles on peut tester l'effet d'homogénéisation de différents procédés (sautage, récupération et chargement, broyage, etc.) en vue d'optimiser les modes d'exploitation.
- en environnement, pour l'emmagasinage de déchets nucléaires, fournir des bases de données permettant d'étudier les probabilités d'occurrence de scénarios pessimistes.
- en environnement ou en hydrogéologie, identifier les emplacements névralgiques où des échantillons supplémentaires devraient être prélevés (Denis Marcotte, 2003).

Etude de cas de l'application de la géostatistique dans le domaine pétrolier

1. Variographie

Les variogrammes omnidirectionnels à deux (2) dimensions (2D) ont été construits et ajustées dans la direction 0° avec une tolérance de ± 90° (Fig. 22). Le récapitulatif des résultats est donné au tableau. 12.

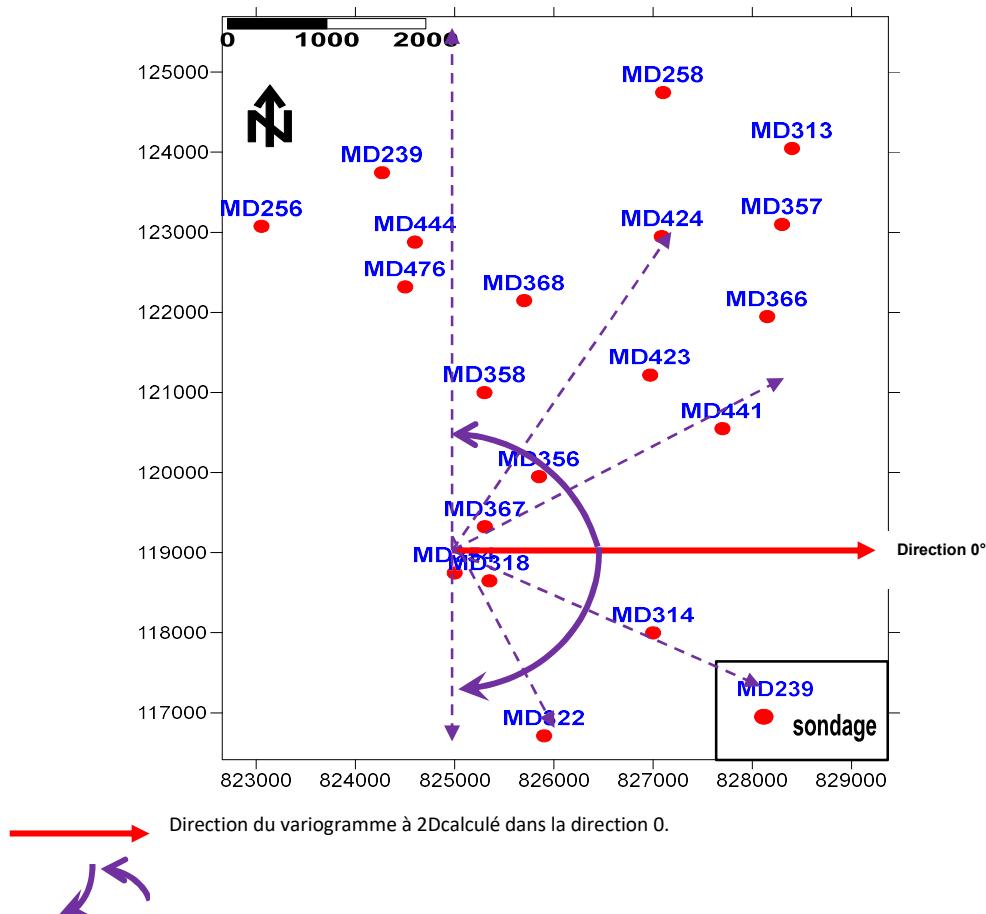


Fig. 22. Procédure de calcul du variogramme.

Tableau. 12 – Récapitulatif des résultats obtenus de l'étude variographique faite sur les données des paramètres pétrophysiques de chaque drain séparément (réservoir R1).

		Réservoir R1 du Cambrien														
Drains du réservoir		Réservoir Ra								Réservoir Ri						
		ID		D2			D3			D5						
Variables régionalisées	Model	Portée (a) (m)	Effet de pépite (C_0) ²	Model	Portée (a) (m)	Effet de pépite (C_0) ²	Model	Portée (a) (m)	Effet de pépite (C_0) ²	Model	Portée (a) (m)	Effet de pépite (C_0) ²	Palier C ($\%_0$) ²			
Porosité (Q)	Sph.	3632	0,19 6	8,62	Sph.	3417	2,24 2	1,67 2	Sph.	4288	2,58	2,3	Sph	3350	2,97	8,14
Saturation en huile (So)	Sph.	3034	3,2	265, 6	Sph.	3100	162, 4	134, 4	Sph.	1950	76	170, 4	Sph	2500	200	75,6
Epaisseur du drain (m)	Sph.	2088	2,03 5	3,90 5	Sph.	2052	0,75	1,67 5	Sph.	2200	2,48	3,6	Sph	1324	2,5	61

Sph. : Modèle sphérique.

2. Estimation géostatistique locale des paramètres petrophysiques par krigeage ordinaire

2.1. Méthodologie de krigeage

Les résultats de l'étude variographique sont utilisés pour l'élaboration du plan de krigeage. La variante de krigeage utilisée pour l'estimation géostatistique locale des paramètres pétrophysiques choisis sont (porosité (\emptyset), saturation en huile (S_o) et l'épaisseur (Epai)) des blocs 100*100 m, Ceci, consiste à un krigeage de chaque drain séparément pour assurer la stationnarité des données.

2.2. Krigeage de chaque drain séparément

a) Plan de krigeage

Le krigeage ordinaire à deux dimensions (2D) des paramètres pétrophysiques (porosité (\emptyset), saturation en huile (S_o) et l'épaisseur (Epai)) par bloc a été effectuée dans chaque drain. Les blocs ont les dimensions 100m x 100 m Le nombre de mailles en X est de 54 et il est de 81 en Y (Fig. 23).

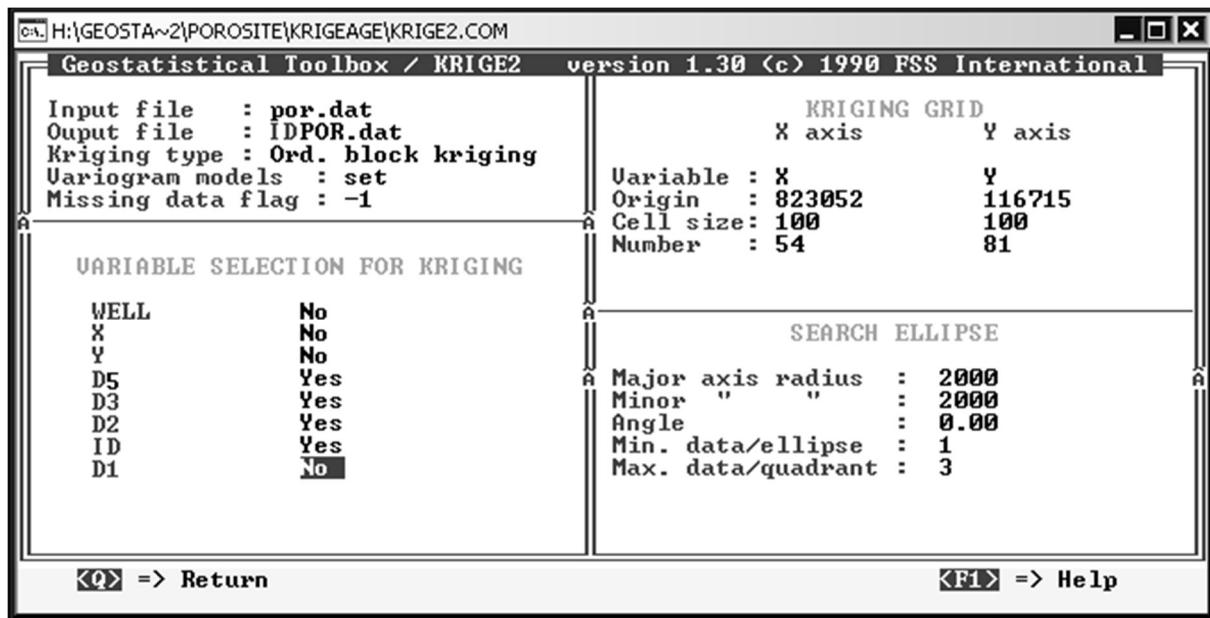


Fig. 23 : Exemple du plan de krigeage de la porosité (\emptyset) dans les drains du réservoir à l'aide du logiciel Geostatistical toolbox.

b) Statistiques des résultats de krigeage de chaque drain

Une étude statistique de chaque drain a été établie, afin de déterminer les paramètres statistiques des blocs krigés (Tableau. 13):

- *Drain D5 (réservoir Ri)*: Ce drain est caractérisé par une porosité (\emptyset) faible, elle est en moyenne dans les blocs de 4,95 %, La saturation en huile (S_o) des blocs krigé varie de 16, 90 à 43, 98%. L'épaisseur moyenne des blocs krigés est de 25,67 m.
- Drain D3 (réservoir Ra) : Ce drain qui est transition vers D5, ce caractérise par une faible porosité (\emptyset), elle est en moyenne de 6,38 %. La saturation en huile (S_o) moyenne dans les blocs krigés est aussi faible, elle est de 29,65%. L'épaisseur de ce drain varie de 12,45 à 18,81 m, elle est en moyenne dans les blocs krigés de 16,24m.
- *Drain D2 (réservoir Ra)* : Le drain D2 est caractérisé par une porosité (\emptyset) assez élevée, elle varie de 5,12 à 9,23 %, dont la moyenne dans les blocs krigés de 7,30 %. Une saturation en huile (S_o) élevée qui varie de 19,02 à 62,37 % , avec une moyenne 36, 36 %. L'épaisseur dans les blocs krigées est 24,79 m.
- *Drain ID (réservoir Ra)* : Ce drain est caractérisé par une porosité (\emptyset) varie de 3,37 à 10,81 %. La moyenne est de 7,33 %. Cette valeur de la porosité (\emptyset) est la plus élevée

dans tous drains. La saturation en huile (S_o) est aussi, élevée, elle varie de 14,96 à 60,80 % dont la moyenne 40,49 %. L'épaisseur moyenne des blocs krigés est de 28,24 m.

Tableau. 13 – Statistique élémentaire des paramètres pétrophysiques des blocs krigées 100*100 m de la zone-17.

Paramètre pétrophysique krigé			N. Bloc	Min (%)	Max (%)	\bar{x}	S^2 (unité) ²	
Réservoir R1	Ri	D5	Porosité (Ø) (%)	3187	1,31	7,40	4,95	2,40
			Saturation en huile (S_o) (%)	3187	16,90	43,98	28,36	53,43
			Epaisseur (Epai) (m)	3187	1,70	51,70	25,67	189,34
	Ra	D3	Porosité (Ø)(%)	3187	4,21	7,89	6,38	0,42
			Saturation en huile (S_o)	3187	14	48,86	29,65	76,21
			Epaisseur (Epai)	3187	12,45	18,81	16,24	2,11
	Ra	D2	Porosité (Ø)(%)	3187	5,12	9,23	7,30	1,15
			Saturation en huile (S_o)(%)	3187	19,02	62,37	36,36	52,26
			Epaisseur (Epai)(m)	3187	22,55	26,97	24,79	0,61
	ID	ID	Porosité (Ø) (%)	3187	3,37	10,81	7,33	1,78
			Saturation en huile (S_o) (%)	3187	14,96	60,80	40,49	72,95
			Epaisseur (Epai) (m)	3187	25,59	30,71	28,24	0,93

b) Variances de Krigeage

L'étude statistique des variances de krigeage des drains a montré que le maximum des variances a été enregistré dans le krigeage saturation en huile (S_o), elle dépasse 100 (%)² en moyenne dans la plus part les drains. Les variances moyennes de porosités (\varnothing) sont faibles dans les deux drains D2 et D3. Par contre, elles sont relativement élevée dans D5 et ID. Les variances de krigeages moyennes des épaisseurs sont assez faibles varie de 1, 27 % dans le drain D2 à 3,02 (m)² dans le drain ID. Une variance de krigeage de l'épaisseur élevée a été observée dans le drain D5, elle est en moyenne de 51,40 (²m). Il est a noté, que les valeurs élevées de variances indiquent une erreur élevée de l'estimation. Cette erreur, est peut être due à la masse de donnée faible ou bien une hétérogénéité de données.

b) Classification des blocs krigés

Une classification des résultats par intervalle des paramètres pétrophysiques a été faite pour chacun des drains (Tableau. 14). Elle montre :

- *Drain D5 (réservoir Ri)* : Ce drain contient 1590 blocs ayant une porosité (\emptyset) qui varie de 4 à 6 % représentent 49,9 % de l'ensemble de blocs. On note que 1780 blocs ayant une saturation en huile (S_o) varient de 20 à 30 %, soit 55,9 % de l'ensemble des blocs du drain.
- *Drain D3 (réservoir Ra)* : Ce drain est caractérisé par la présence de 2242 blocs ayant une porosité (\emptyset) qui varie de 6 à 8 % représentent 70,3 % de l'ensemble de blocs. 1362 blocs sont caractérisés par une saturation en huile (S_o) qui varie de 30 à 40 %, soit 42,7 % de l'ensemble des blocs du drain. Il est à noté, que tous les blocs ayant des épaisseurs varient de 10 à 20 m.
- *Drain D2 (réservoir Ra)* : Dans ce drain, 1775 blocs ayant une porosité (\emptyset) varient de 6 à 8 % représentent 55,7 % de l'ensemble des blocs. La saturation de 1323 blocs varie de 30 à 40 % soit, 41,5 % de l'ensemble de blocs de ce drain. On note que tous les blocs ayant des épaisseurs varient de 20 à 30 m.
- *Drain ID (réservoir Ra)* :

Ce drain est caractérisé par la présence de 1548 blocs ayant une porosité (\emptyset) qui varient de 6 à 8 % représentent une proportion de 48,6 % de l'ensemble des blocs. Il est caractérisé par la présence de 1479 blocs dont la saturation en huile (S_o) varie de 40 à 50 % soit, 46,4 % de l'ensemble des blocs. La quasi-totalité des blocs ont des épaisseurs varient de 20 à 30 m, avec 3034 blocs, soit, 95,2 % de l'effectif des blocs.

Tableau. 14- Classification des résultats de krigeage de chacun de drain par intervalles – Réservoir Ri + Ra.

Réservoir	Réservoir R1 du Cambrien																
	Réservoir Ra													Réservoir Ri			
	Drain				ID			D2			D3			D5			
Résultats krigeage	de	à	Fi %	N.Bloc	de	à	Fi %	N.Bloc	de	à	Fi %	N.Bloc	de	à	Fi %	N.Bloc	
Porosité(Ø)	0	2	0	0	0	2	0	0	0	2	0	0	0	2	7,5	240	
	2	4	0,7	23	2	4	0	0	2	4	0	0	2	4	13,6	433	
	4	6	17,6	562	4	6	14,1	449	4	6	29,7	945	4	6	49,9	1590	
	6	8	48,6	1548	6	8	55,7	1775	6	8	70,3	2242	6	8	29	924	
	8	10	31	989	8	10	30,2	963	8	10	0	0	8	10	0	0	
	10	12	2	65	10	12	0	0	10	12	0	0	10	12	0	0	
Saturation en huile (So)	14	20	1,8	56	14	20	1	33	14	20	23,1	736	14	20	8,5	272	
	20	30	11,8	376	20	30	24,2	772	20	30	22,7	723	20	30	55,9	1780	
	30	40	28,3	903	30	40	41,5	1323	30	40	42,7	1362	30	40	27,2	867	
	40	50	46,4	1479	40	50	29,6	942	40	50	11,5	366	40	50	8,4	268	
	50	63	11,7	373	50	63	3,7	117	50	63	0	0	50	63	0	0	
Epaisseur (Epai)	0	10	0	0	0	10	0	0	0	10	0	0	0	0	10	16,5	527
	10	20	0	0	10	20	0	0	10	20	100	3187	10	20	23,6	752	
	20	30	95,2	3034	20	30	100	3187	20	30	0	0	20	30	23,3	741	
	30	40	4,8	153	30	40	0	0	30	40	0	0	30	40	16	511	
	40	52	0	0	40	52	0	0	40	52	0	0	40	52	20,6	656	

3. Estimation des quantités initiales de l'huile en place

3.1. Méthodologie de l'estimation locale

Les quantités initiales en place en huile ont été estimées. La méthode utilisée consiste estimer ces quantités dans chaque bloc des drains désignés. Etant donnée, la relation d'estimation volumétrique d'un réservoir comme suit :

$$\text{Volume d'huile initialement en place} = V \phi(So).$$

Les volumes ainsi, estimés par cette méthode sont d'unité (m^3)³. Il suffit de convertir le volume en barils en multipliant $V * 6,28981$.

L'estimation des quantités sera effectuée localement, c'est-à-dire ; dans les blocs, en utilisant les résultats de krigeage.

$$\text{Volume huile en place BLOC} = V_{Bloc^*} \phi_{Bloc^*} So_{Bloc^*},$$

Où :

V_{Bloc} = Surface X épaisseur krigée du bloc (surface d'un bloc = $100 * 100 m$) ;

ϕ_{Bloc^*} : La porosité (ϕ) krigée du bloc ;

So_{Bloc^*} : La saturation en huile (So) krigée du bloc.

La quantité en huile dans un drain donnée est égale à la somme des volumes d'huile des blocs :

$$\text{Volume huile en place du drain} = \sum_{Bloc\ 1}^{Bloc\ N} (V_{Bloc^*} \phi_{Bloc^*} So_{Bloc^*})$$

N : nombre de blocs constitue le drain de réservoir.

3.2. Quantités d'huile en place

Les quantités d'huile initiales en place ont estimé pour chaque drain séparément (Tableau. 15). Les résultats obtenus montrent que les drains ID + D2 sont les drains les plus

prometteurs. Le drain ID est caractérisé par des quantités initiales en place de l'huile de l'ordre de 171 millions de barils, le drain D2 avec des quantités de 134 millions de barils. Le drain D3 est caractérisé par des quantités de 62 millions de barils. Les quantités de l'huile dans le drain D5 ont été estimées à l'ordre de 71 millions de barils. Total des quantités dans tous les drains, objet de l'estimation sont de l'ordre de 440 millions de barils.

Tableau. 15- Estimation des réserves des drains de réservoir Ri + Ra.

Réservoir	Drains	Nombre de Blocs	Min(%)	Max(%)	\bar{x} (%)	Volume initiale de l'huile en place (bbl)	
Réservoir R1	Ri	D5	3187	1547,57	52165,05	22287,26	
	Ra	D3	3187	6465,16	37625,04	19723,14	
		D2	3187	15915,09	74228,94	42310,36	
		ID	3187	8741,59	106393,48	53782,82	
Total des quantités initiales de l'huile en place (tous les drains) Réservoir R1=							
440.136.067,55 bbl							

2.3.3. Classification des blocs en fonction des quantités d'huile en place

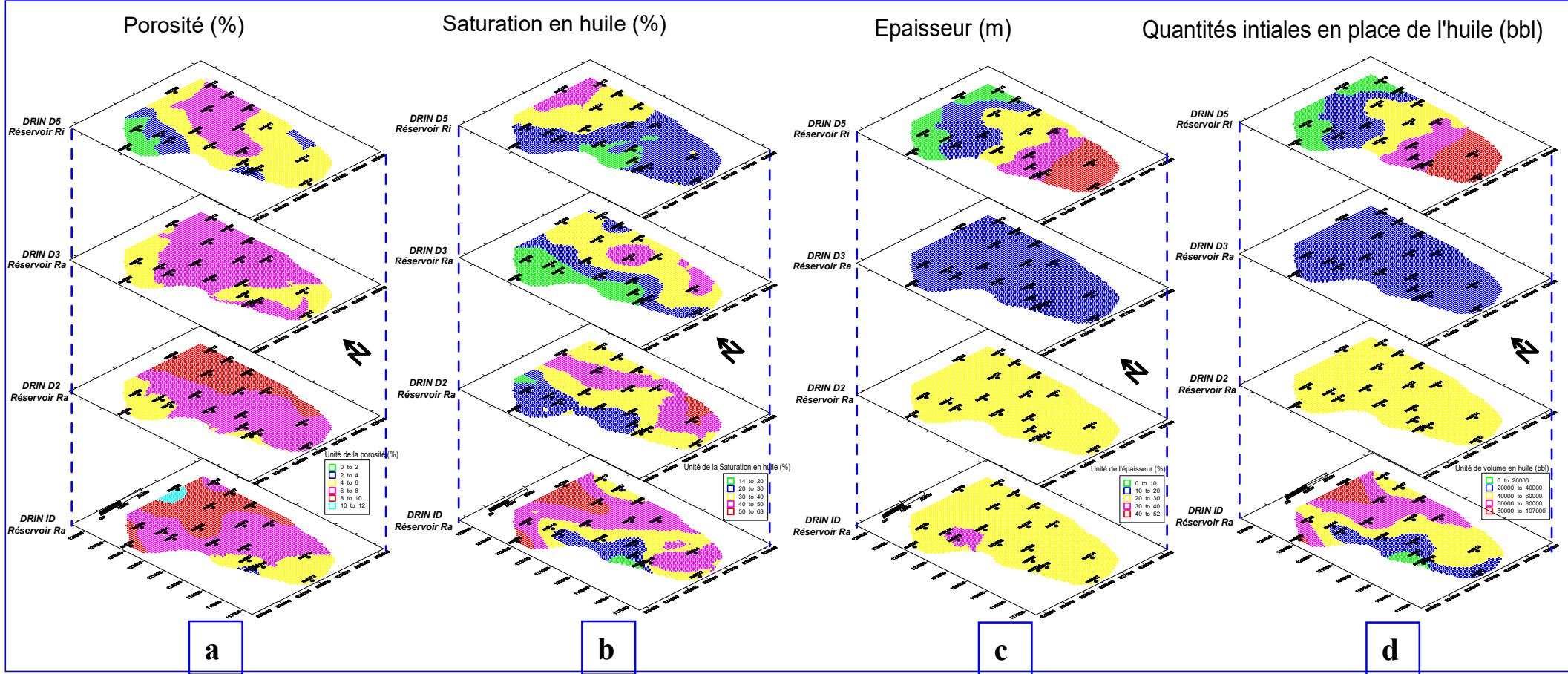
Une classification des quantités initiales de l'huile en place a été faite, afin de connaître la répartition des quantités au sein des blocs (Tableau. 16). Il montre :

- Drain ID : La majorité des blocs de ce drain, avec 1318 blocs, soit 41,4 % ayant une quantité de 40000 à 60000 barils / bloc ;
- Drain D2 : La majorité des blocs de ce drain, avec 1477 blocs, soit 46,3 % ayant une quantité de 40000 à 60000 barils / bloc;
- Drain D3 : La majorité des blocs de ce drain, avec 1633 blocs, soit 51,2 % ayant une quantité de 20000 à 40000 barils / bloc ;
- Drain D5 : La majorité des blocs de ce drain, avec 1613 blocs, soit 50,6 % ayant une quantité de 20000 à 40000 barils / bloc.

Tableau. 12 – Paramètres statistiques des quantités initiales de l'huile en place dans les drains du réservoir Ra et Ri.

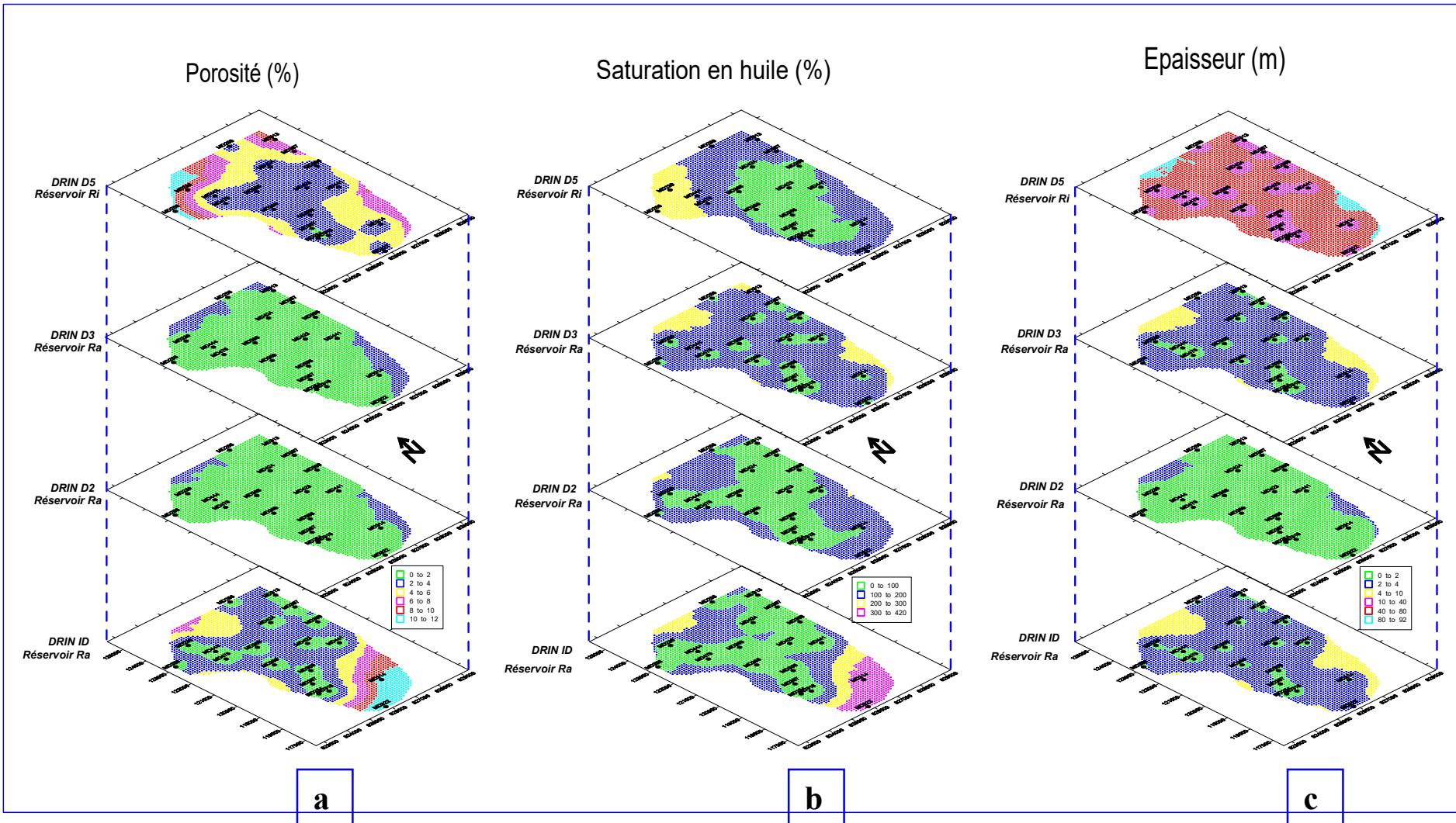
Tableau. 16 – Classification des quantités initiales de l'huile en place par intervalles des blocs – Réservoir Ri et Ra – zone 17.

Réservoir	Réservoir R1 du Cambrien															
	Réservoir Ra							Réservoir Ri								
Drains	ID				D2				D3				D5			
Statistiques	Min	Max	%	Nb Blok	Min	Max	%	Nb Blok	Min	Max	%	Nb Blok	Min	Max	%	Nb Blok
	0	20000	2,8	89	0	20000	1,7	53	0	20000	48,8	1554	0	20000	39,9	1273
	20000	40000	18,9	602	20000	40000	43	1370	20000	40000	51,2	1633	20000	40000	50,6	1613
	40000	60000	41,4	1318	40000	60000	46,3	1477	40000	60000	0	0	40000	60000	9,4	301
	60000	80000	30,3	967	60000	80000	9	287	60000	80000	0	0	60000	80000	0	0
	80000	107000	6,6	211	80000	107000	0	0	80000	107000	0	0	80000	107000	0	0



- Cartographie de porosité dans les blocs krigés 100*100 m.
- Cartographie de la saturation en huile dans les blocs krigés 100*100 m.
- Cartographie de l'épaisseur dans les blocs krigés 100*100 m.
- Cartographie des quantités initiales en place de l'huile dans les blocs 100*100 m.

Fig.– Cartographie des résultats de krigage par drain du réservoir R1 de la zone-17 (Champ de Hassi Messaoud).



- Cartographie de porosité dans les blocs krigés 100*100 m.
- Cartographie de la saturation en huile dans les blocs krigés 100*100 m.
- Cartographie de l'épaisseur dans les blocs krigés 100*100 m.

Fig. – Cartographie des variances des résultats de krigeage dans les blocs par drain du réservoir R1 de la zone-17 (Champ de Hassi Messaoud).

Références Bibliographiques

Deutsch & Journel 1992 :pp. 209-213, GSLIB. Geostatistical Software Library and users guide. Oxford University Press, New York, USA, 340 p.

Chauvet. (1999) : Aide mémoire de géostatistique linéaire application du domaine minier. Les presses de l'école des mines de Paris, 14p.

Gslib (2007): Gaussian Simulation for Porosity Modeling, Reservoir Modeling with GSLIB, presentation ppt, 17 dispositives.

Goovaerts 1997 : Geostatistics for natural resources evaluation. Oxford University Press, New York, USA, 483 p.

Howarth & Earle 1979 : Application of a generalized power transformation to geochemical data. Mathematical Geology, 11 : 45-62

Isaaks & Srivastava 1989 : An introduction to applied geostatistics. Oxford University Press, New York, USA, 561 p.

Joseph & Bhaumik 1997 : Improved estimation of the Box-Cox transform parameter and its application to hydrogeochemical data. Mathematical Geology, 29 : 963- 976.

Journel A. G. (1978): Cours de géostatistique, 3^{eme} année. Tome II, institut national polytechnique de Lorraine, France, 122p.

Journel A. G. & Huijbregts (1978) : Mining geostatistics. Academic Press, London, U. K, 600p.

Kechiched, R. (2007) : Estimation des réserves locales du gisement de phosphate de Bled El Hadba par la procédure géostatistique de krigeage ordinaire, Mémoire d'ingénieur, UBMA, 70 p.

Kedzierski, P.(2007): Révisions de géostatistiques, résumé publié sur internet, 7p.

Matheron, G. (1971) : La théorie des variables régionalisées et ces applications. Les cahiers de CMM5. Paris ENSMP, 212p.

Matheron, G. (1973): The Intrinsic Random Function and their application. In Advances in Applied Probability. Dec .1973, n°3, pp439 – 468.

Marcotte D. “Géostatistique minière”. Ecole polytechnique, Montréal, 1997, 230 p.

Marcotte D. (2003): Géostatistique minière. Cours publiés sur internet, Ecole polytechnique, Lorraine, France, 230p.

Morino, A “Analyse en Composantes Principales”. Edition CISIA. Ceresta, France, 1988, 198 p.

Ploner A. & Dutter R. (2000): New directions in geostatistics. In : Journal of statistical planning and inference, edition Elsevier, pp 499-509.

Sokal & Rohlf 1995. Revue dans Hoyle 1973 pp. 413-422). Biometry. The principles and practice of statistics in biological research. Third edition. Freeman, New York, USA, 887 p.

SONATRACH (2005), “La stratigraphie du champ Hassi Messaoud”, Rapport interne, Division Exploration. 2005, 35p.

SONATRACH (2002) “subdivision des zones dans champ Hassi Mesaoud”, Rapport interne, Division Production, 2002, 25p.

Stoline 1991: An examination of the lognormal and Box and Cox family of transformations in putting environmental data. Environmetrics, 2 : 85-106 p.

Xavier, Emery (2001): Géostatistiques linéaire, Ecole des mines de Paris, 405 p.

TDS ET EXERCICES

TD1. Travaux de l'analyse statistique par logiciel

STATISTICA

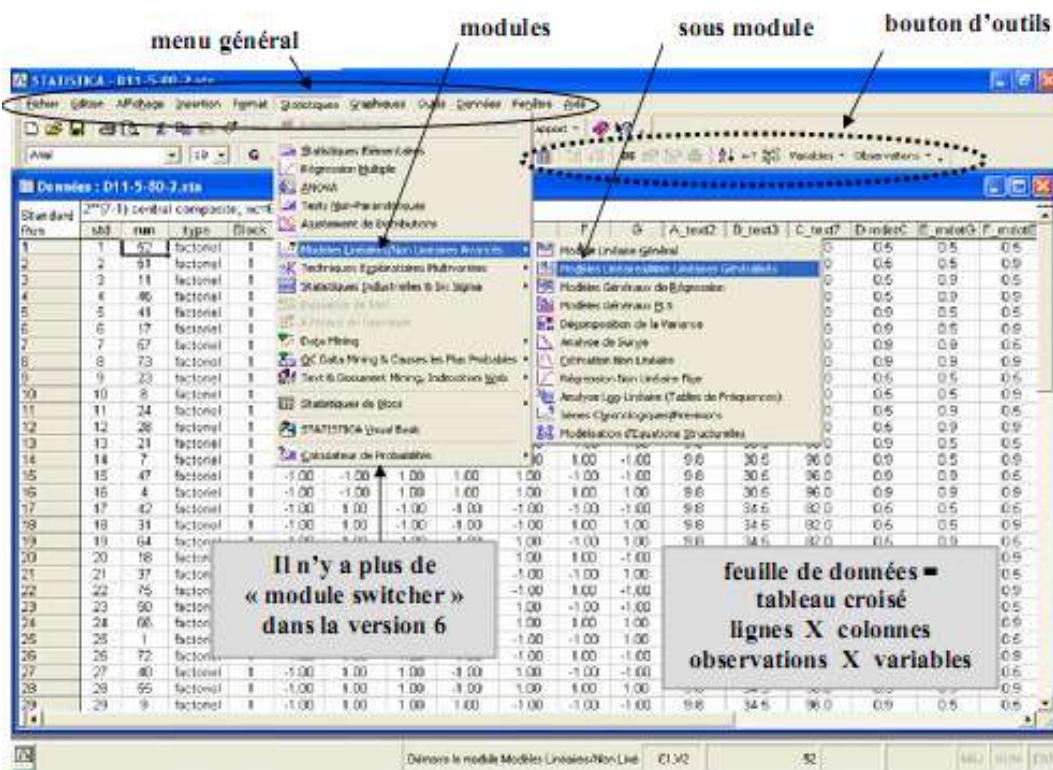
« Dans ce document la version DEMO (Version d'essai) a été utilisée dans le but de montrer les manipulations et la partie pratique du module de « Application de la géostatistique dans le domaine pétrolier). L'auteur décline la responsabilité de toute utilisation sans licence.

1. STATISTIQUE : APPLICATION PAR LOGICIEL – STATISTICA 6.61

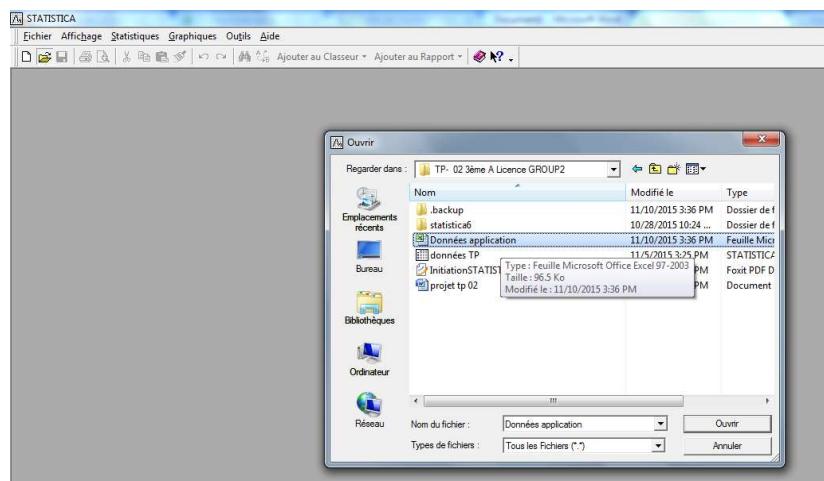
1.1. Fichier de données :

Plusieurs méthodes peuvent être utilisées pour la préparation du fichier des données parmi lesquelles :

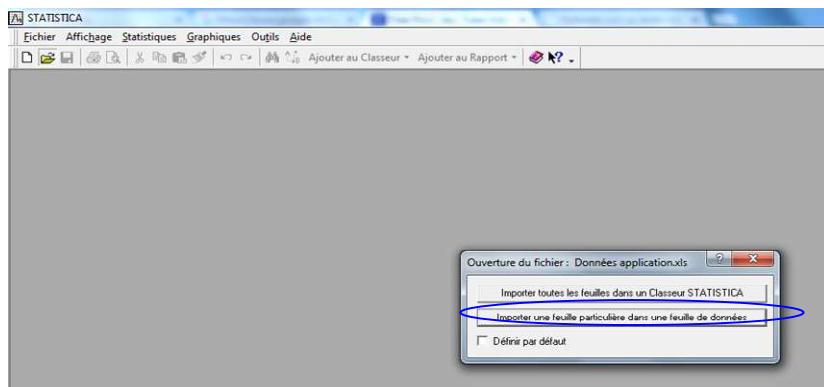
- Saisir les données manuellement puis enregistrer sous format *.sta



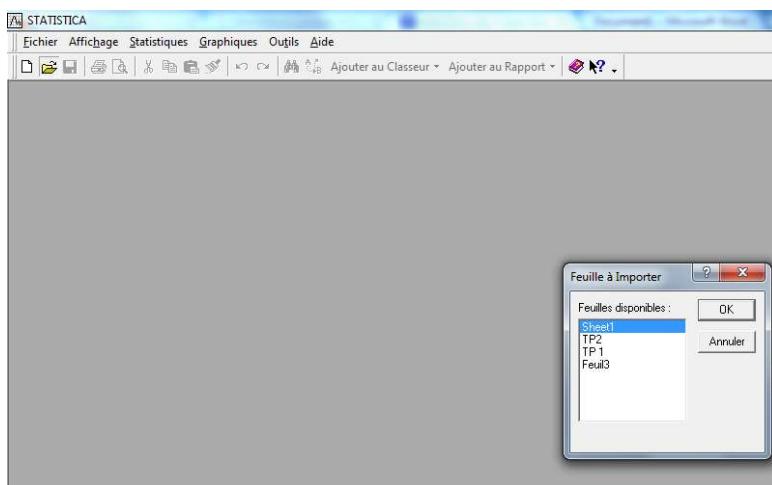
- Importer à partir un fichier Excel des donnée en tableau (variables : Colonnes ; Observations : lignes)



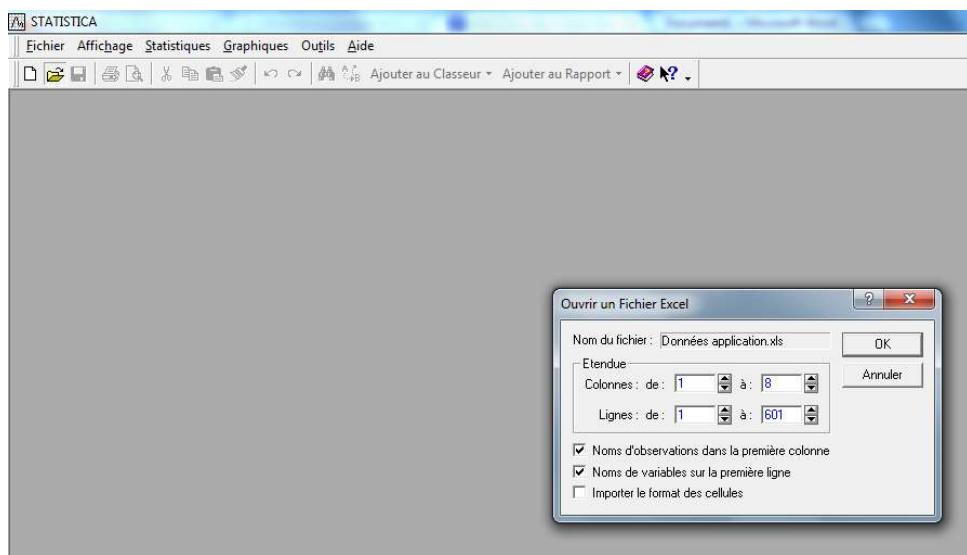
- Importer une feuille particulière :



- Choisir la feuille de données :



- Déclarer les noms des d'observations et variables



- Spécifier les valeurs manquantes (Codes VM) (cliquer sur une variable puis faire toutes spécifications pour les autres variables)

	4 MgO	5 Fe2O3	6 CaO	7 SiO2
9	17.48	0.37	34.86	
8	9.4	0.51	37.38	
3	7.2	0.52	38.06	
	9.24	0.33	40.87	
5	7.86	0.7	42.06	
5	10.4	0.44	41.46	
5	8.98	0.31	40.5	
3	7.7	0.35	41.34	
7	9.04	0.39	40.36	
1	7.8	0.43	41.46	
3	8.56	0.4	40.56	
2	6.29	0.56	42.26	
	7	0.31	42.4	
9	5.92	0.45	43.62	
1	6.76	0.41	42.7	
2	5.24	0.54	44.34	
7	4.42	0.55	46.24	
9	8.72	0.41	40.42	
9	3.96	0.83	44.38	
6	3.94	0.69	44.7	
2	1.41	0.49	48.56	
2	3.96	0.53	44.22	
2	5.36	0.44	43.56	
3	3.04	0.82	44.66	
9	5.4	0.4	43.7	
5	3.8	0.51	46.12	
4	1.12	0.45	48.2	

Variable 6

Nom : CaO Type : Double OK
 Code des VM : -1 Annuler
 Longueur : 8
 Format d'Affichage
 Général Toutes Spécif.
 Nombre Valeurs-Texte
 Date
 Heure
 Scientifique
 Monétaire
 Pourcentage
 Fraction
 Personnalisée Valeurs/Stats...
 Description détaillée (étiquette ou formule avec Fonction) : Guide des fonctions
 Etiquettes : saisissez tout texte personnalisé. Formules : utilisez les noms de variables ou leur numéro (v1, v2, ...); v0 représente les numéros d'observations.
 Exemples : (a) = mean(v1:v3; sqrt(v7); AGE) (b) = v1+v2; commentaires (saisissez tout type de commentaire après la formule)

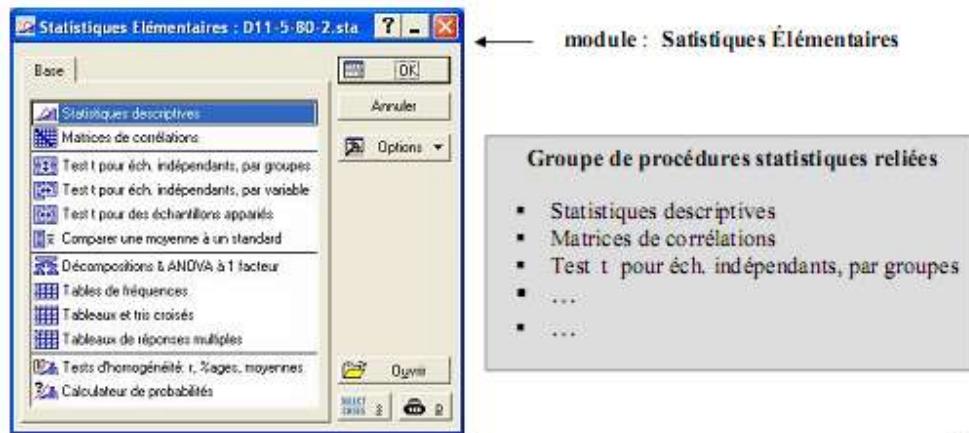
- Enregistrer sous – Format *.sta

N.B. Si vous importez toute un classeur, l'enregistrement se fait en format *.stw

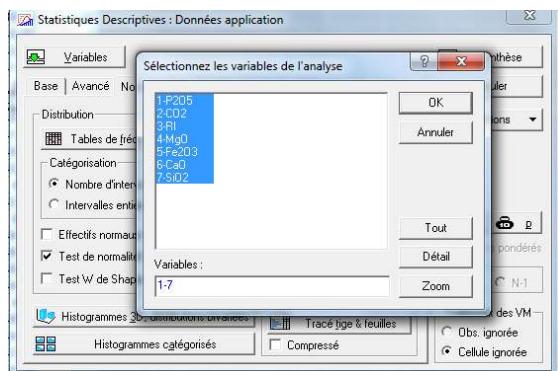
1.2. Traitements statistiques

1.2.1. Tableau 1statistique

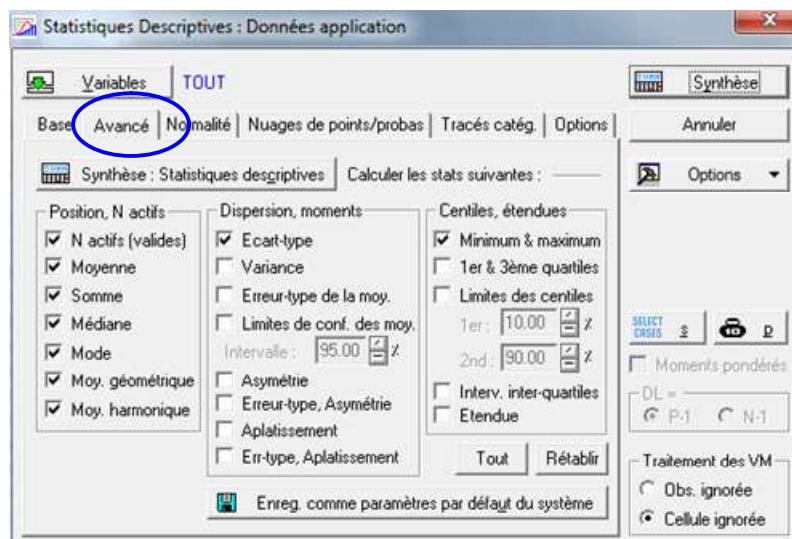
a) Aller au module Statistique puis sous-module Statistiques élémentaires :



- Préciser les variables



- Préciser les paramètres statistiques élémentaires (dans l'onglet : avancé et puis synthèse)



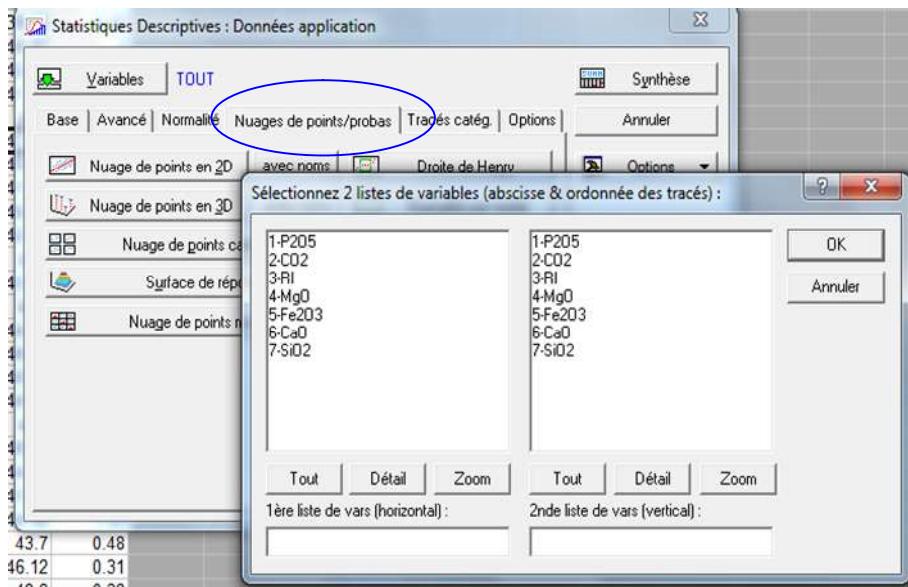
- Copier le tableau (copier avec noms (Vars/Obs.)

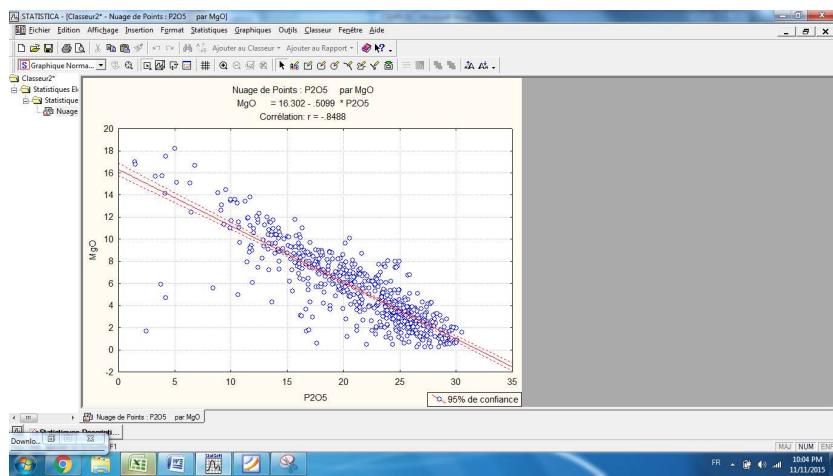
STATISTICA - [Classeur1* - Statistiques Descriptives : Données application]

The screenshot shows the STATISTICA software interface. A data grid is displayed with columns: Moyenne, Harmonique, Médiane, Mode, Effectif, Somme, Minimum, Maximum, and Ecart-type. Below the grid, a context menu is open, showing options like 'Collage Spéciale...', 'Copier', 'Coller avec Noms (View/Obj...)', 'Coller', 'Coller Spéciale...', 'Remplacer...', 'Rechercher/Remplacer...', 'Atteindre...', 'Liens DDE...', 'Urgence...', 'Objet', and 'Capture d'Ecran'. The status bar at the bottom shows 'C1/V1' and '600'.

1.2.2. Nuage de points X/Y par exemple P₂O₅/MgO

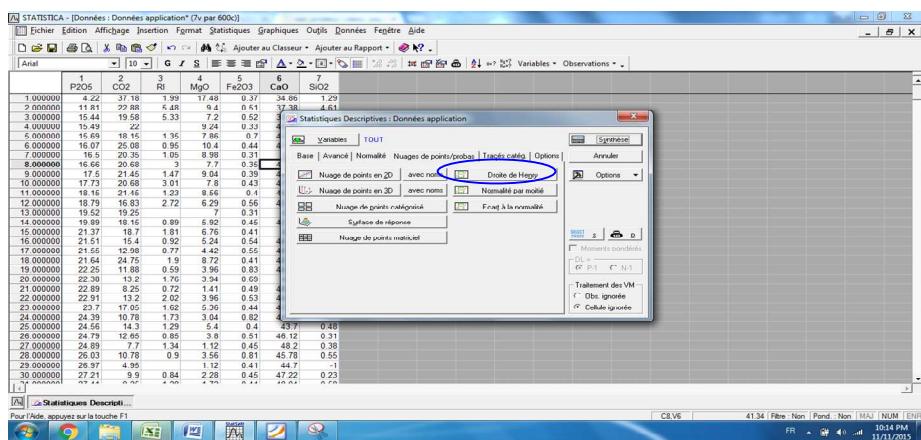
- Sur l'onglet nuages de points / probas
- Choisir P₂O₅ versus MgO puis OK

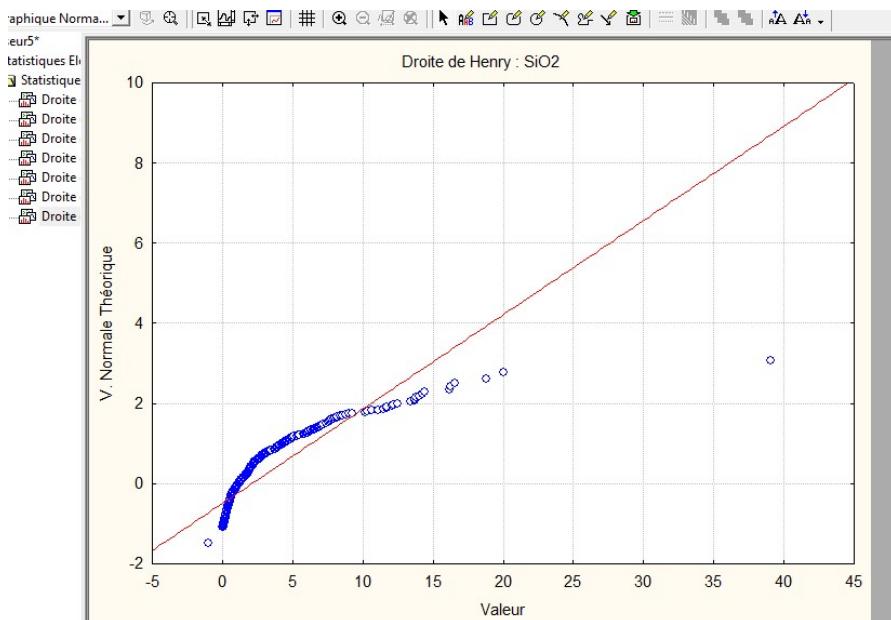




1.2.3. Graphe de probabilités (droite d'Henry)

Sur l'onglet droit d'Henry

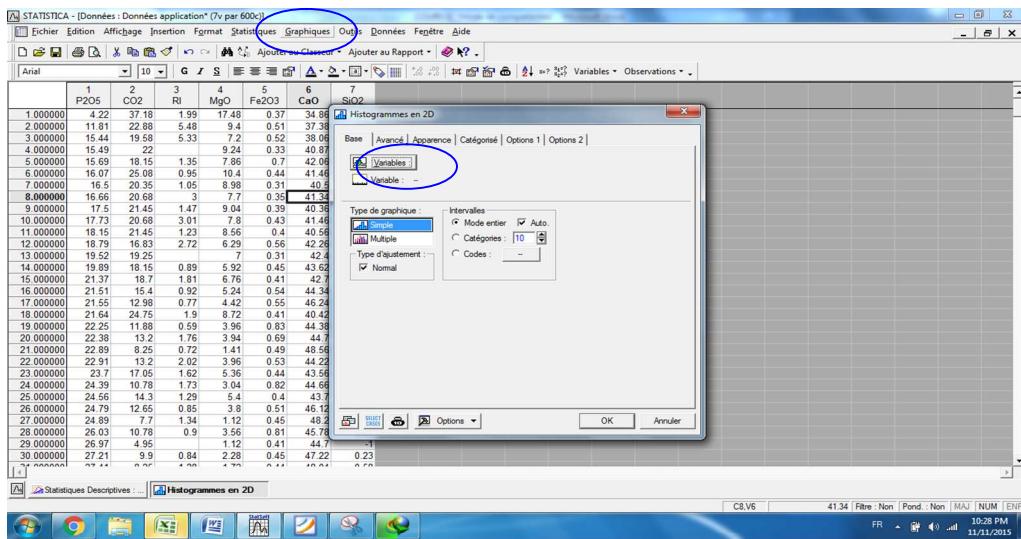


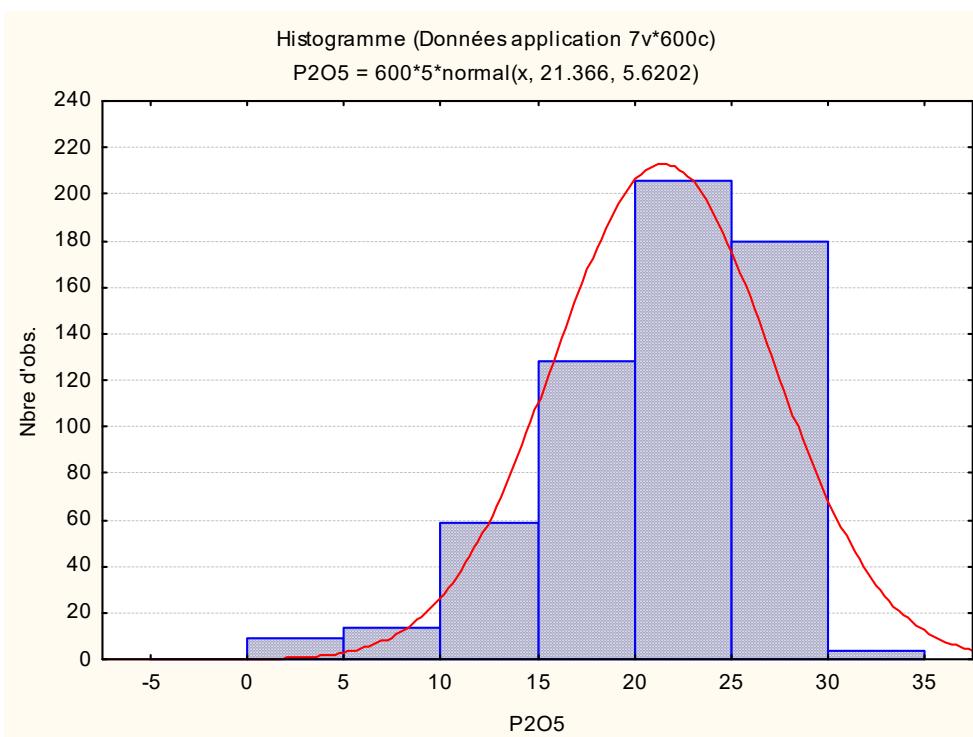


1.2.4 Histogrammes des fréquences relatives

Sur le module graphique puis préciser la variable puis OK

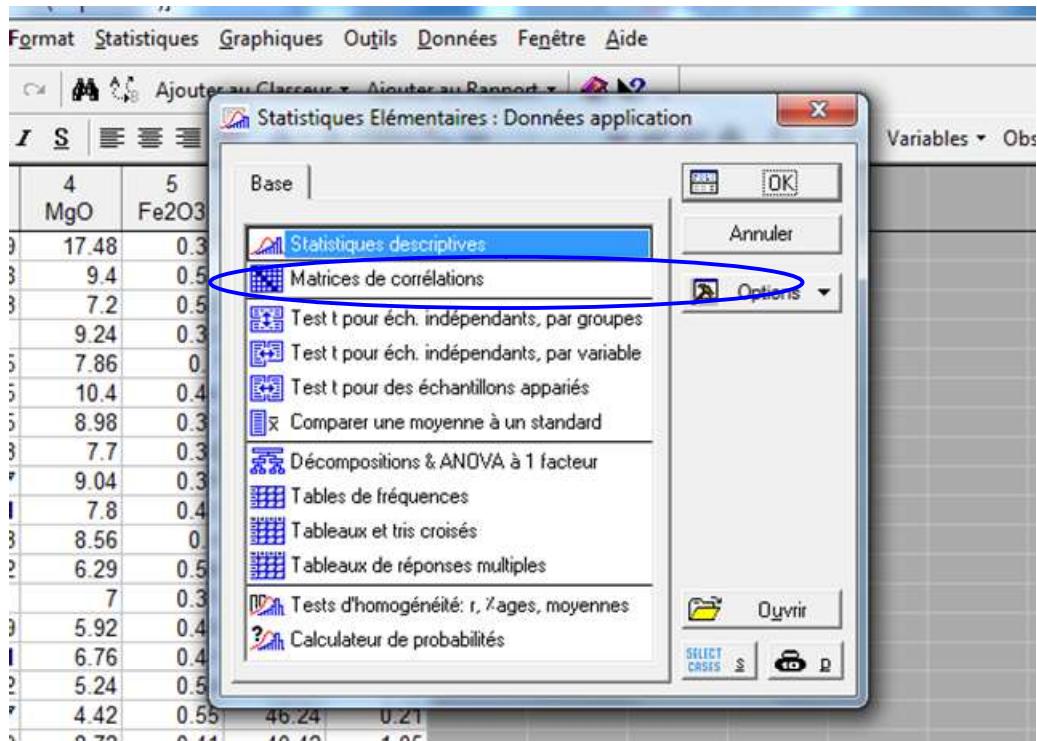
Pour le nombre de catégories (nombre de classes il faut toujours appliquer la relation $n = \text{racine carrée du } N$ (Nombre total d'observations))





1.2.5. Coefficient de corrélation

- Dans le sous-module Statistique élémentaire, matrice de corrélation



2- spécifier la liste 1 et la liste 2 des variables puis OK puis Synthèse

N.B : il faut choisir le même nombre de variables autant dans la liste 1 que dans la liste 2

The screenshot shows the STATISTICA software interface. A data grid is visible in the background with columns labeled 1 NS, 2 DARIN, 3 Nech, 4 Prof (m), 5 K (mD), 6 Q (%), 7 So (%), 8 Sw (%), and 9 Vsh (%). A cursor is positioned over the 9 column header. In the foreground, a dialog box titled "Matrices de Corrélation : DONNES" is open. It contains two tabs: "1 liste de variables" (selected) and "2 listes (matrice rectang.)". Under "1 liste de variables", the variable "K (mD)\Vsh" is listed. Under "2nde liste", the variable "K (mD)\Vsh" is also listed. Below these tabs are buttons for "Synthèse" (highlighted), "Annuler", and "Options". The "Options" tab includes a note: "(Seules les matrices carrées avec une liste unique peuvent être sauvegardées)". At the bottom of the dialog box are buttons for "Nuage matriciel des variables sélectionnées", "SELECT CHSSE", "Moments pondérés", "DL =", "P-1" (radio button selected), "N-1", "Traitement des VM", "Obs. ignorée" (radio button selected), and "Cellule ignorée".

- Une fois sectionner synthèse, la matrice de corrélation est affichée.

STATISTICA - [Classeur1* - Corrélations (DONNES)]

Fichier Edition Affichage Insertion Format Statistiques Graphiques Outils Données Classeur Fenêtre Aide

Ajouter au Classeur Ajouter au Rapport ?

Arial 10 G I S

Classeur1* Corrélations (DONNES)

Statistiques Elémentaires (DONNES) N significatives marquées à p < .05000

Corrélation N=827 (Observations à VM ignorées)

Variable	K (mD)	Q (%)	So (%)	Sw (%)	Vsh (%)
K (mD)	1.00	0.34	0.21	-0.07	-0.09
Q (%)	0.34	1.00	0.45	-0.32	-0.04
So (%)	0.21	0.45	1.00	-0.20	-0.09
Sw (%)	-0.07	-0.32	-0.20	1.00	-0.09
Vsh (%)	-0.09	-0.04	-0.09	-0.09	1.00

1.2.6. Analyse en Composantes Principales (ACP°)

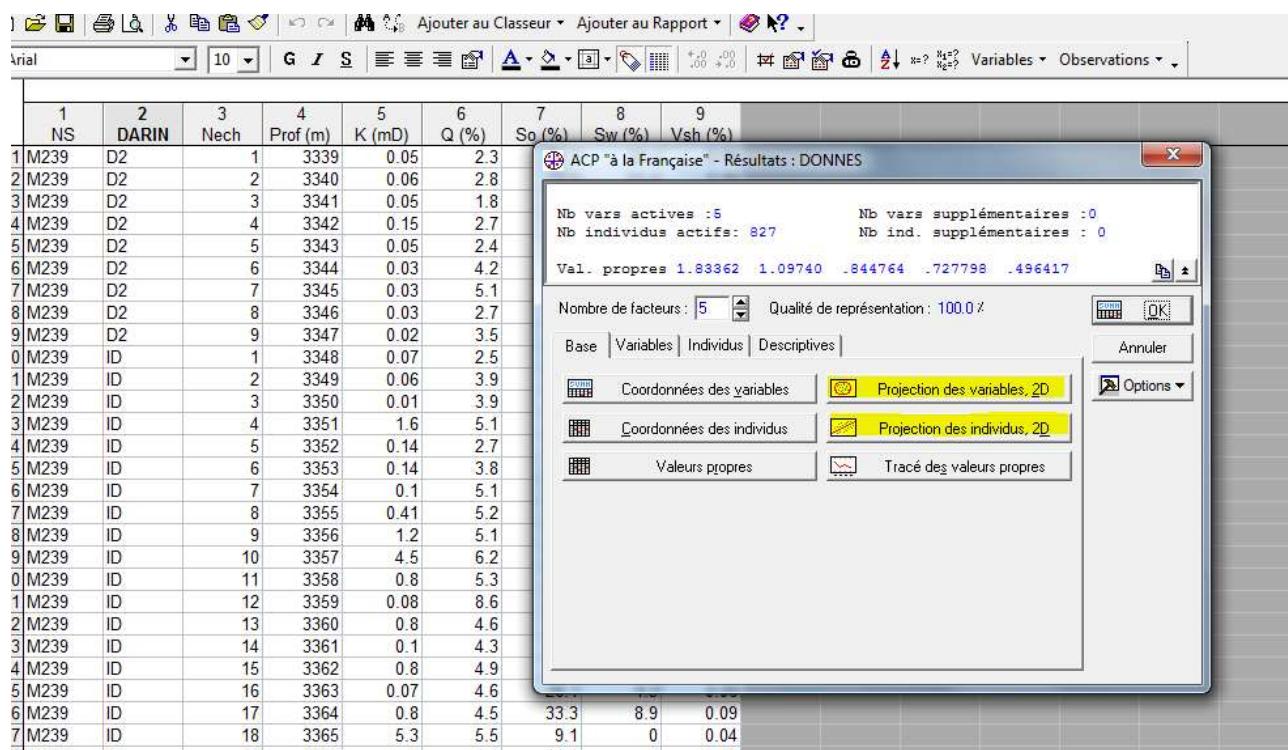
- Sous l'onglet de **Statistiques** et dans la rubrique de **techniques exploratoires multivariées**, Aller à ACP « à la française » où analyse factoriel

The screenshot shows the STATISTICA software interface. The menu bar includes Fichier, Edition, Affichage, Insertion, Format, Statistiques, Graphiques, Outils, Données, Fenêtre, and Aide. The toolbar contains icons for opening files, saving, printing, and various statistical tools. A context menu is open over the 'Techniques Exploratoires Multivariées' option in the Statistiques menu, listing options such as 'Calculateur de Probabilités', 'Statistiques de Blocs', 'STATISTICA Visual Basic', and several multivariate analysis methods including 'ACP à la Française', 'Analyse Canonique', 'Fiabilité et Analyse d'Echelle', 'Arbres de Décision', 'Analyse des Correspondances', 'Analyse de Proximité', and 'Analyse Discriminante Générale'. The main window displays a data grid with columns labeled 1(NS), 2(DARIN), 3(Nech), 4(Prof(m)), 5(K(mD)), 6(Q(%)), 7(So(%)), 8(Sw(%)), and 9(Vsh(%)). The status bar at the bottom right shows the date and time.

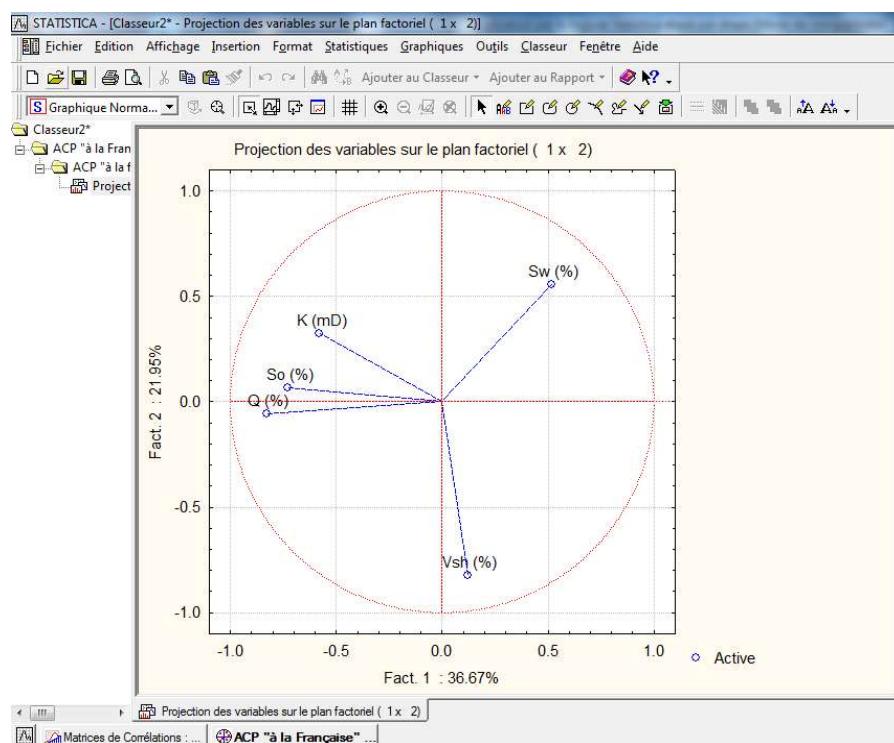
- Sélectionner les variables dans la première liste :

The screenshot shows the STATISTICA software interface with a dialog box titled 'Sélectionnez les vars actives & supplémentaires, avec individus actifs, et catégorielle illustrative'. The dialog lists variables in four columns: 1-NS, 2-DARIN, 3-Nech, 4-Prof(m); 5-K(mD), 6-Q(%), 7-So(%), 8-Sw(%); 9-Vsh. Some variables are highlighted in blue, indicating they are selected. Below the dialog, a data grid displays the same variables with numerical values. The status bar at the bottom right shows the date and time.

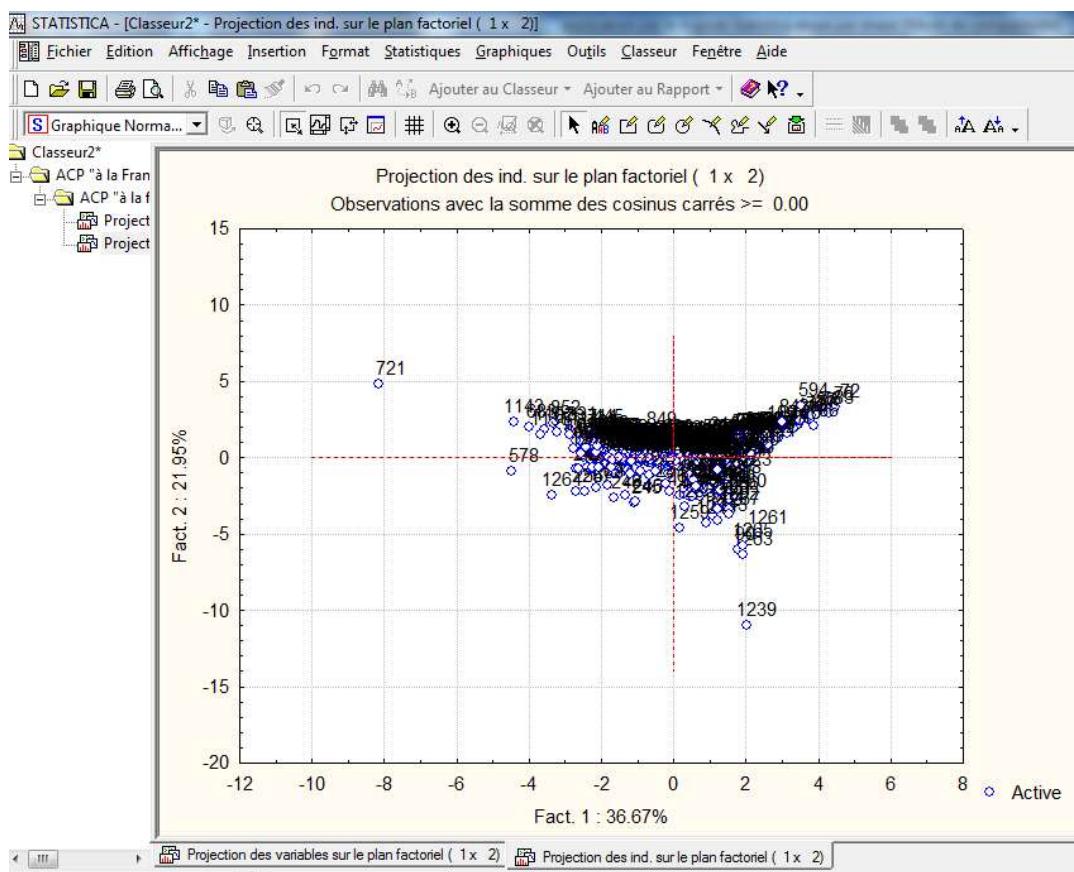
- Une fois sélectionner les variables, OK, les calculs seront affiché



a) projection des variables à 2D : Aller dans cette boîte de dialogue et sélectionner projection des variables à 2D et sélectionner les deux facteurs exprimant le maximum de l'information généralement (F1 versus F2)



a) projection des individus à 2D : Aller sans cette boîte de dialogue et sélectionner projection des individus à 2D et sélectionner les deux facteurs exprimant le maximum de l'information généralement (F1 versus F2).



TD2 : loi de gausse.

Il convient de rappeler que la fonction de répartition d'une variable gaussienne centrée réduite U (moyenne nulle et variance égale à 1) (Fig. 1) est donnée par la fonction de densité $f(x)$ est de la forme :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}x^2}$$

Il est noté $N(0,1)$, et sa fonction de répartition est :

$$F(x) = \frac{1}{\sqrt{2\pi}} \int e^{\frac{-1}{2}x^2} dx$$

Le graphe $N(0,1)$ est une courbe symétrique par rapport à Oy

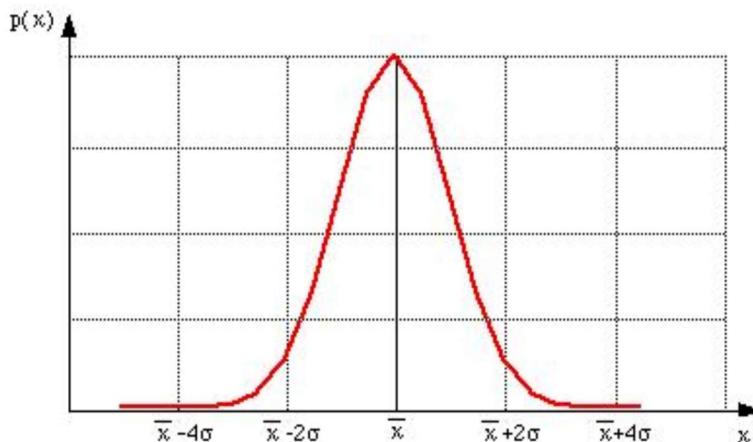
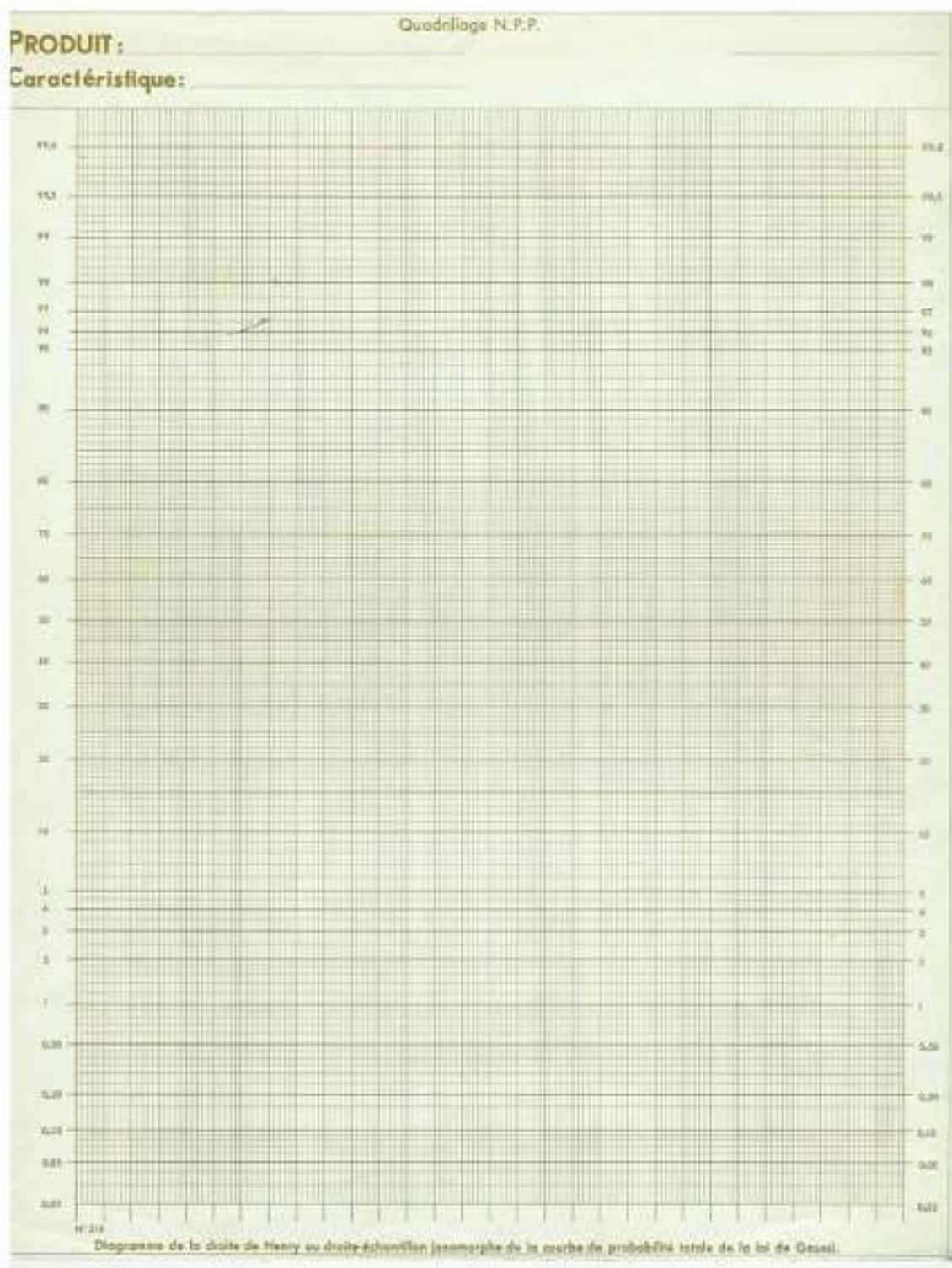


Fig.1 : Fonction de répartition d'une variable gaussienne réduite U (Moyenne =0, écart-type=1)

4.16.1 Loi normale N(0,1)

Fonction de répartition N(0,1)										
x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
Quantiles particuliers										
$x = F^{-1}(p)$	0.5	0.6	0.7	0.8	0.9	0.95	0.975	0.99	0.995	0.999
	0.0000	0.2533	0.5244	0.8416	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902

Papier Gaussien



TD3 :

Enoncé :

A partir du fichier ci-joint :

Il est demandé d'effectuer les opérations statistiques suivantes :

1. Effectuer le Tableau statistique (table des fréquences) pour chaque élément.
2. Calculer les paramètres de position centrale et les paramètres de dispersion.
3. Tracer l'histogramme de fréquences relatives, déterminer le mode.
4. Tracer l'histogramme de fréquences cumulées, déterminer la médiane.

Exercice : 01

Enoncé :

A partir du fichier ci-joint (donné en support électronique) :

Il est demandé d'effectuer les opérations statistiques suivantes :

1. Effectuer le Tableau statistique (table des fréquences) pour chaque élément.
2. Calculer les paramètres de position centrale et les paramètres de dispersion.
3. Tracer l'histogramme de fréquences relatives, déterminer le mode.
4. Tracer l'histogramme de fréquences cumulées, déterminer la médiane.
5. Tracer sur papier gaussien la droite d'Henry, commenter

Solution

Traitements statistiques des données

Les résultats des analyses chimiques des 260 échantillons de Fe_2O_3 ont subi un traitement statistique

1. Effectuer le Tableau statistique (table des fréquences) pour chaque élément.
- 1- Le tableau statistique des teneurs en Fe_2O_3

X_i	n_i	n_i cumulé	Fréquence relative n_i/N	Fréquence cumulé	$(X_i * n_i)/N$	$(X_i - \bar{X})^2$
15,25	4	4	0,0154	0,0154	0,2346	42,1
15,5	70	74	0,2692	0,2846	4,1731	38,913
16	39	113	0,15	0,4346	2,4	32,925
17	26	139	0,1	0,5346	1,7	22,449
18	17	156	0,0654	0,6	1,1769	13,973
19	13	169	0,05	0,65	0,95	7,4966
20	15	184	0,0577	0,7077	1,1538	3,0206
21	14	198	0,0538	0,7615	1,1308	0,5446
21,5	1	199	0,0038	0,7654	0,0827	0,0566
23	13	212	0,05	0,8154	1,15	1,5926
23,5	1	213	0,0038	0,8192	0,0904	3,1046
24	1	214	0,0038	0,8231	0,0923	5,1166
25	3	217	0,0115	0,8346	0,2885	10,641
30	7	224	0,0269	0,8615	0,8077	68,261
35	17	241	0,0654	0,9269	2,2885	175,88
45	7	248	0,0269	0,9538	1,2115	541,12
55	5	253	0,0192	0,9731	1,0577	1106,4
65	7	260	0,0269	1	1,75	1871,6
		260			X=21,738	3945,1942

2. Calculer les paramètres de position centrale et les paramètres de dispersion.

Paramètres d'une Série Statistique

Une série statistique peut se caractériser par 2 grands types de paramètres:

Le tableau statistique en Fe_2O_3 résume certaine paramètre statistique

Xmax	X min	étendu	Variance S2	Ecart type S	Coeff de variation Cv
65	15,25	49,75	15.17	3.89	17.92

2 - 1 - Paramètres de position centrale

Les paramètres de position centrale ont été calculés. Ces paramètres permettent de quantifier la tendance centrale des valeurs d'une série statistique. Les principaux paramètres de tendance centrale sont le mode, la médiane et la moyenne arithmétique. Ces paramètres ont été déterminés et calculés à partir du tableau 1

I)- Le mode

Il est défini comme étant la valeur de la variable aléatoire qui a l'effectif le plus élevé.

L'histogramme montre la présence des 2 modes qui sont :

$$Mo_1 = 15.5 \text{ ppm} \quad Mo_2 = 35 \text{ ppm}$$

2)- La médiane

La médiane est la valeur de la variable qui correspond à un effectif cumulé de 50%.

Médiane = 17 ppm

3) - Le moyenne arithmétique

En présence de données discrètes, la moyenne est calculée à l'aide de la formule suivante :

$$\bar{X} = M = \sum_{i=1}^N \frac{X_i}{N}$$

La moyenne de Fe₂O₃ est de : 21,738 ppm

4) - les quartiles d'une série statistique ?

-détermination des quartiles et des déciles : d'après la courbe cumulative des fréquences

- Soit une série statistique X de taille n .

Le premier quartile Q_1 est la plus petite valeur de la série telle qu'au moins 25 % des données soient inférieures ou égales à Q_1 .

Le troisième quartile Q_3 est la plus petite valeur de la série telle qu'au moins 75 % des données soient inférieures ou égales à Q_3 .

$Q_1 = 15.5 \text{ ppm}$ $Q_2 = 17 \text{ ppm}$ $Q_3 = 21 \text{ ppm}$

2 - 2 - paramètres de dispersion

Les paramètres de dispersion permettent de quantifier la dispersion des valeurs de la série statistique. Les principaux paramètres sont l'étendue, la variance, l'écart type et le coefficient de variation.

1) L'étendue : c'est la différence entre les valeurs extrêmes de la série statistique ordonnée :

$$E = x_{\max} - x_{\min} = 49,75 \text{ ppm}$$

2)- La variance S^2

Dans le cas de variable discrète - dans notre cas, elle est calculée à l'aide de la formule suivante :

$$S^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N} = \sum_{i=1}^n \frac{x_i^2}{N} - \bar{x}^2 = 15.17 \text{ ppm}^2$$

3)- L'écart type

C'est la racine carrée de la variance

$$\sigma = \sqrt{S^2} = 3.89 \text{ ppm}$$

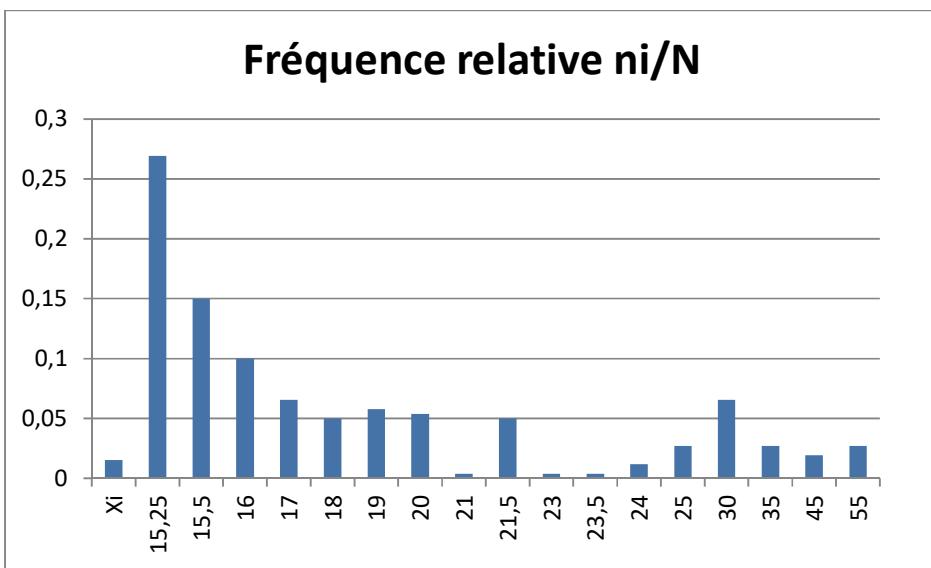
4)- Coefficient de variation C_v

C'est le rapport écart type sur la moyenne multiplié par 100

$$C_v = \frac{s}{\bar{x}} \cdot 100 = 17.92$$

3. Tracer l'histogramme de fréquences relatives, déterminer le mode.

3 - l'histogramme. On obtient un histogramme en plaçant les X_i par ordre croissant sur l'axe des abscisses et la fréquence relative sur l'axe des ordonnées



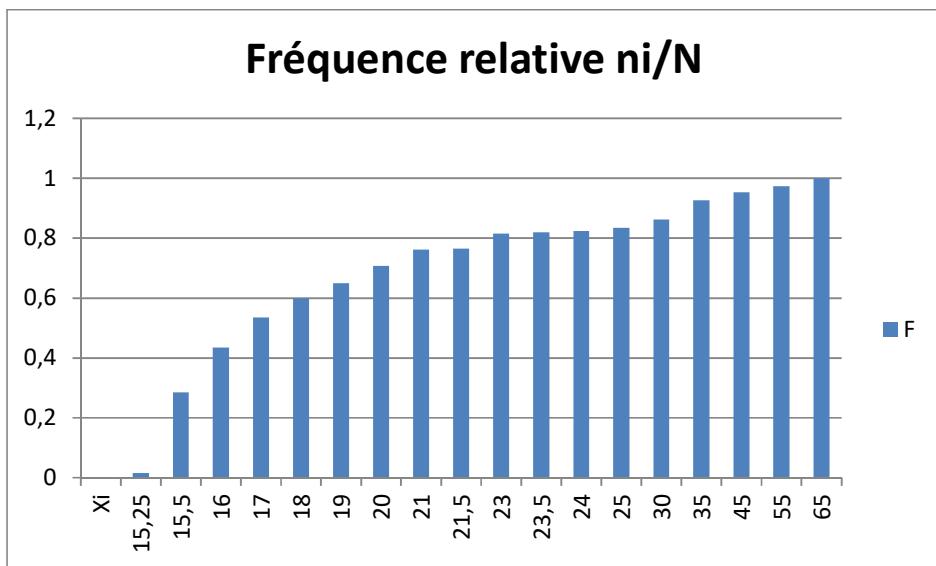
L'histogramme de Fe2O3 présente 2 modes qui sont : 15.5 ppm et 35 ppm

On a vu deux mode qui représenté l'association de deux population

4. Tracer l'histogramme de fréquences cumulées, déterminer la médiane.

La médiane est la valeur de la variable qui correspond à un effectif cumulé de 50%.

Médiane = 17 ppm



TD- 4 : Coefficient de corrélation

Exercice : 02

Enoncé :

Il est demandé d'effectuer les opérations suivantes :

Projetez le nuage de point dans un diagramme binaire (Pb, Zn) ?

Calculez le coefficient de corrélation simple $r_{Pb, Zn}$ et vérifiez sa signification ?

Etablissez la formule de la régression linéaire simple ?

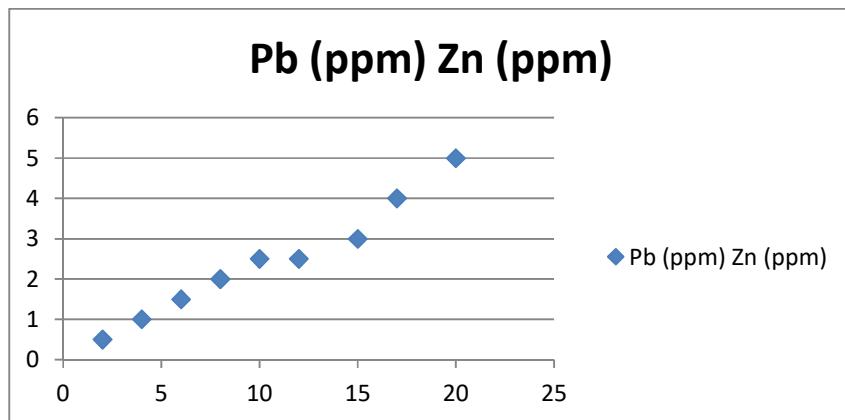
N°	1	2	3	4	5	6	7	8	9
Pb (ppm)	15	20	10	8	6	2	17	4	12
Zn (ppm)	3	5	2.5	2	1.5	0.5	4	1	2.5

Solution

1. Projetez le nuage de point dans un diagramme binaire (Pb, Zn) ?

Pb (ppm)	Zn (ppm)	$pb * Zn$ (ppm^2)
15	3	45
20	5	100
10	2,5	25
8	2	16
6	1,5	9
2	0,5	1
17	4	68
4	1	4
12	2,5	30
10,44	2,44	33,11

Tableau des données se Pb,Zn.



NUAGE DE POINT

2. Calculez le coefficient de corrélation simple $r_{\text{Pb}, \text{Zn}}$ et vérifiez sa dignification ?

Le coefficient de corrélation est calculé par la formule suivante :

$$r_{xy} = \text{cov}(xy)/S_x \cdot S_y$$

Où:

r : Coefficient de corrélation simple de la population ;

x et y : Variables ;

S_x et S_y : Écart type des variables x et y ;

$\text{Cov}(x,y)$: Covariance de variables x et y . Il est à rappeler que :

$$\text{cov}(x,y) = E(xy) - E(x)E(y)$$

$$S_{\text{Pb}} = 6,09 \text{ ppm}$$

$$S_{\text{Zn}} = 1,42 \text{ ppm}$$

$$\text{Cov}(\text{Pb}, \text{Zn}) = 7,58 \text{ ppm}^2$$

$$\text{Donc : } r_{\text{Pb}, \text{Zn}} = 0.87 \text{ ppm}$$

La signification de ce dernier dépend du nombre d'individus utilisés pour son calcul.

Une façon empirique de calculer la signification c'est de calculer la valeur de :

$$R_{\text{signif}} = \frac{2}{\sqrt{n-1}}$$

Où n est le nombre des données utilisées.

$$R_{\text{signif}} = 2 \sqrt{(9-1)} = 0.71$$

La signification du coefficient de corrélation dépend du nombre de données utilisées pour son calcul. Généralement ce dernier n'est significatif que s'il satisfait aux conditions d'approximation suivants :

$$r_{xy} > +\frac{2}{\sqrt{n-1}} \quad \text{ou bien} \quad r_{xy} < -\frac{2}{\sqrt{n-1}}$$

On a $r_{Pb,Zn} > R_{\text{signif}}$

Alors le coefficient de corrélation est significatif

3. Etablissez la formule de la régression linéaire simple ?

La régression linéaire simple par la méthode des moindres carrés

Les valeurs optimales d'ajustement des paramètres de la droite $Y=aX+b$ pour le critère des moindres carrés sont données par les relations :

$$a = \text{Cov}(X, Y) / (\sigma_X)^2 \quad \text{où} \quad (\sigma_X)^2 : \text{le variance de Pb} : (\sigma_X)^2 = 37,03 \text{ ppm}^2$$

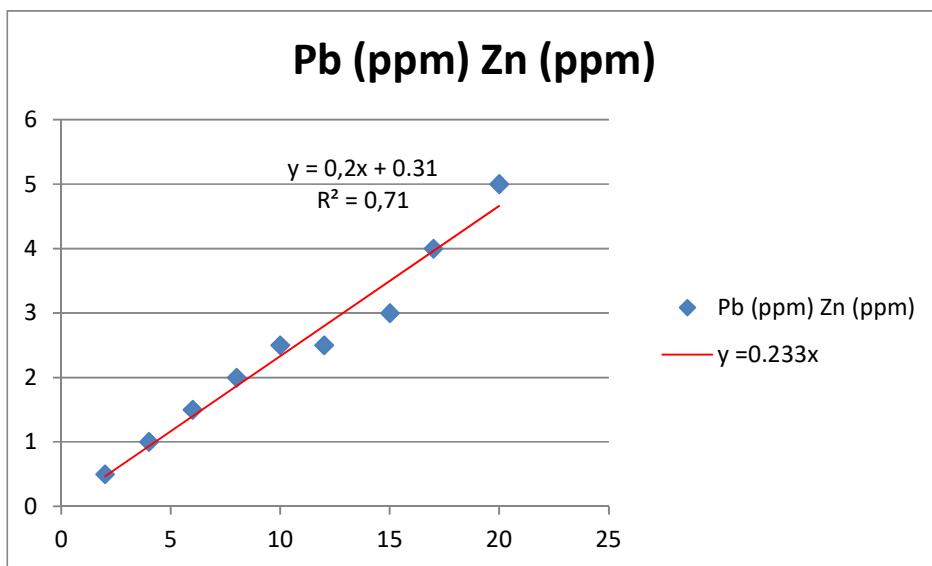
$$a = 0,20$$

$$b = m(Y) - a \cdot m(X)$$

$$b = m(Zn) - a * m(Pb)$$

$$b = 0,30628027$$

$$y = 0,2x + 0,31$$



Droite de régression linéaire simple

TD5: AJUSTEMENT DES VARIOGRAMMES

Exercice : 03

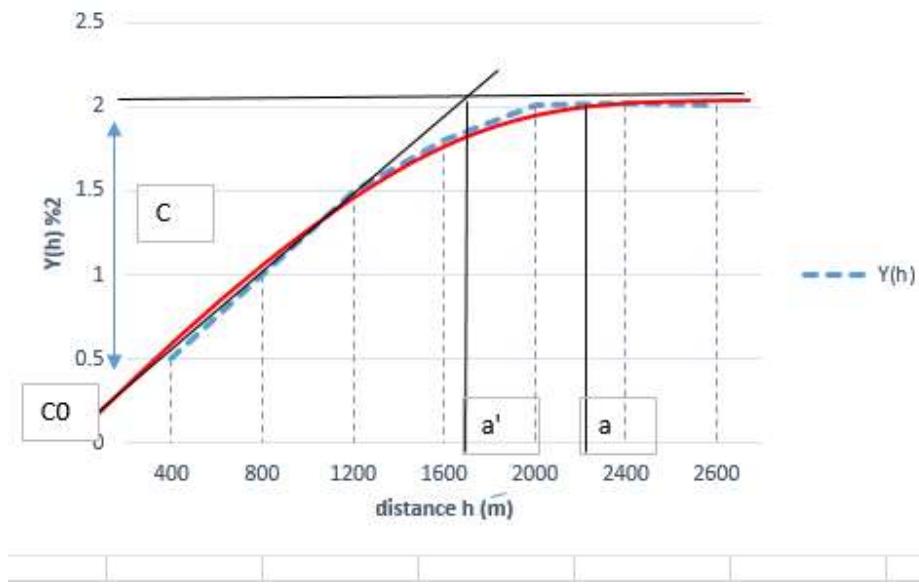
Les résultats du variogramme de la porosité (Unité de variable régionalisée : %) ; sont donnés au tableau ci – dessous :

N° de pas	Distance (h) (m)	Couples	$\gamma^*(h) (\%)^2$
1	400	69	0.5
2	800	152	1
3	1200	118	1.5
4	1600	66	1.8
5	2000	64	2.01
6	2400	50	2.02
7	2600	40	2.01

1. Tracer, le variogramme du tableau, en utilisant une échelle appropriée.
2. Ajuster ce variogramme par un model de régionalisation ? Justifier ce choix.
3. Ecrire la formule du variogramme ajusté (les paramètres de variogramme).
- 4- Interpréter le variogramme sur le plan géologique (géologique).

Solution

1. tracer le variogramme.



- Variogramme expérimentale directionnel de la porosité ajustée à l'aide d'un schéma sphérique
- 2. l'ajustement de variogramme se fait par un model sphérique.

- Justification :

Dans la pratique il existe plusieurs méthodes d'ajustement, cependant l'ajustement « à la main » est la méthode la plus simple et la plus juste.

Il faudra compte de :

- L'existence ou non du palier
- De l'effet de pépite et du palier expérimental
- Du comportement à l'origine et de la tangente à l'origine dans le cas de comportement linéaire pour la proposition du schéma théorique.
- Le schéma sphérique est un schéma à palier et le comportement de variogramme à l'origine est linéaire.

La projection de l'intersection de la tangente à l'origine avec la ligne de palier engendre sur l'axe des x une valeur: $a = 2/3 a'$.

3. La formule du variogramme ajusté (les paramètres de variogramme).

Le model d'ajustement est le model sphérique, de formule suivent :

$$\gamma(h) = C_0 + C \left(\frac{3}{2} \frac{h}{a} - \frac{1}{2} \frac{h^3}{a^3} \right)$$

Les paramètres de variogramme tracée:

- **Effet de pépite $C_0 = 0.25 \text{ \%}^2$**
- **Palier $C = 2.1 - 0.25 = 1.85 \text{ \%}^2$**
- **Portée $a = 2200 \text{ m}$**

4- Interpréter le variogramme sur le plan géologique (géologique).

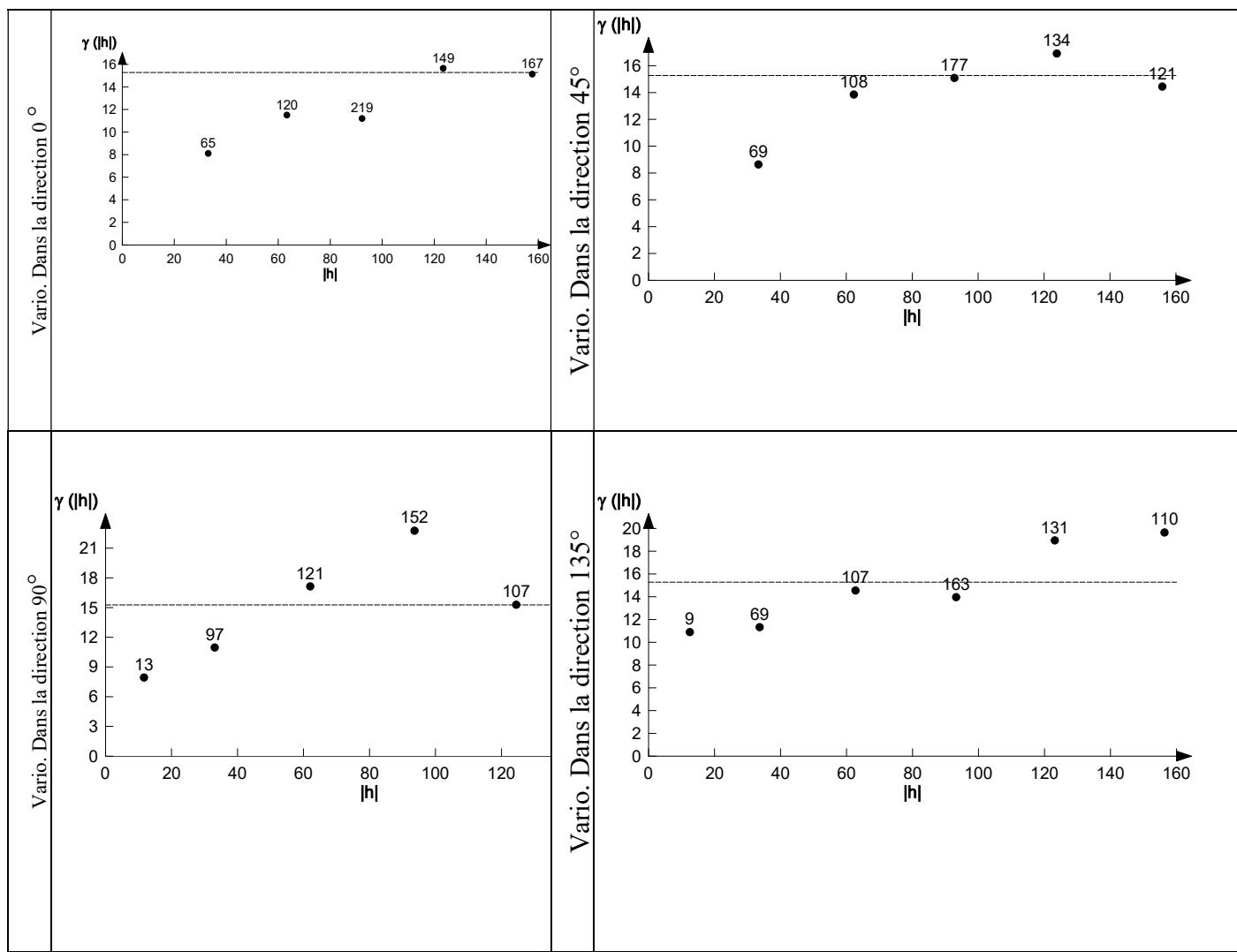
Le variogramme indique que il y a de corrélation entre les échantillons séparés au fur et à mesure que la distance augmente, $400 \text{ m} < h < 2000 \text{ m}$, lorsque le variogramme est atteint sa limite supérieure c'est à dire son palier $c = 2.1\%^2$, il n'y a plus de corrélation entre les échantillons séparés par cette distance h , $h = 2200$: cette distance critique est appelée portée du variogramme, qui fournit une définition plus précise de la notion de zone d'influence.

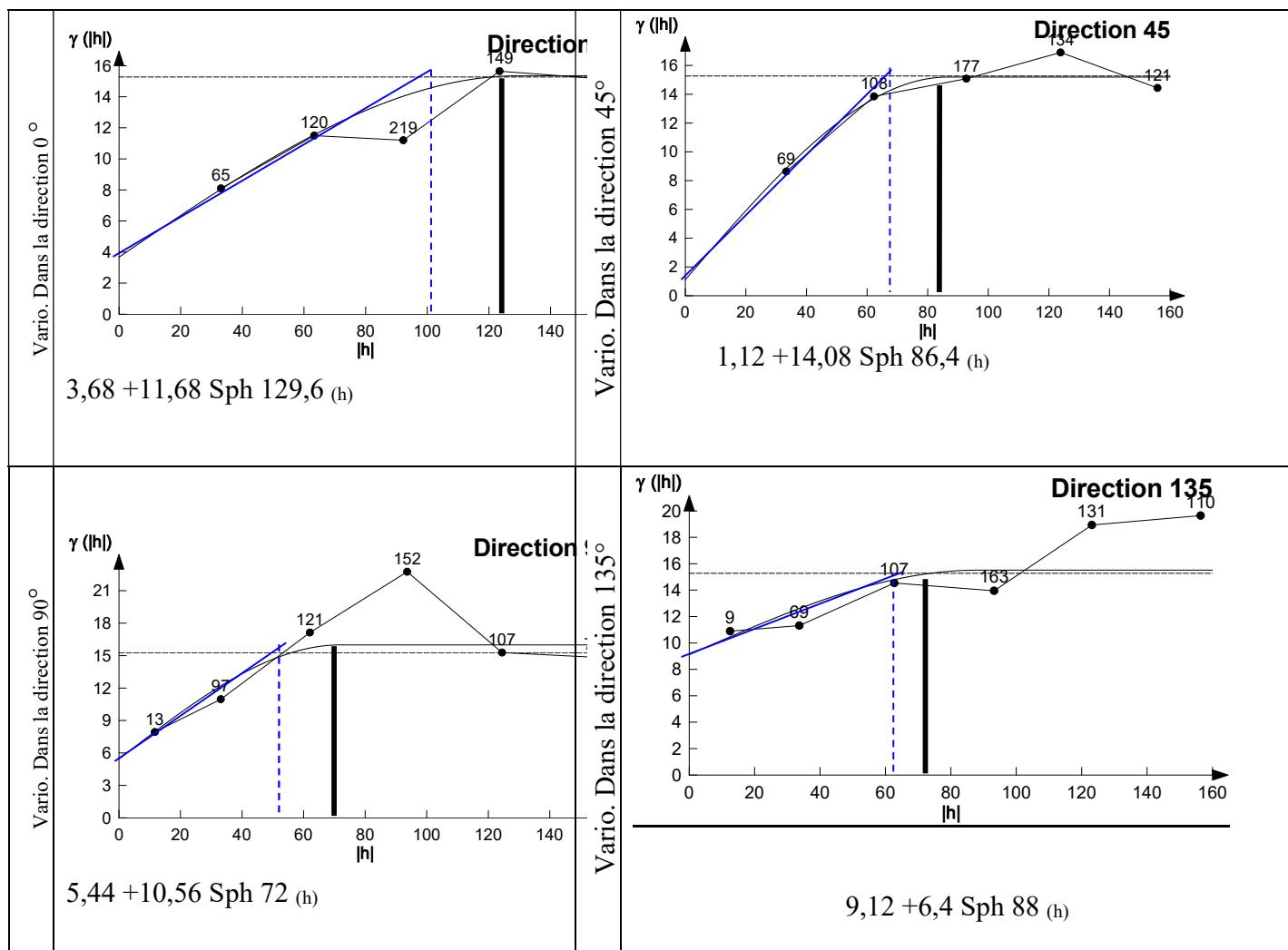
TD5- Variographie et anisotropie

Le logiciel VARIOWIN 2.21 a été utilisé pour calculer les variogrammes directionnels à deux dimensions (2D), sur les données du Cadmium (Cd) (ppm).

Il est demandé de :

1. Ajuster les variogrammes par des schémas de régionalisation, justifiez votre choix.
2. Tracer la rose de portées et calculer le coefficient d'anisotropie, Commentez-les.

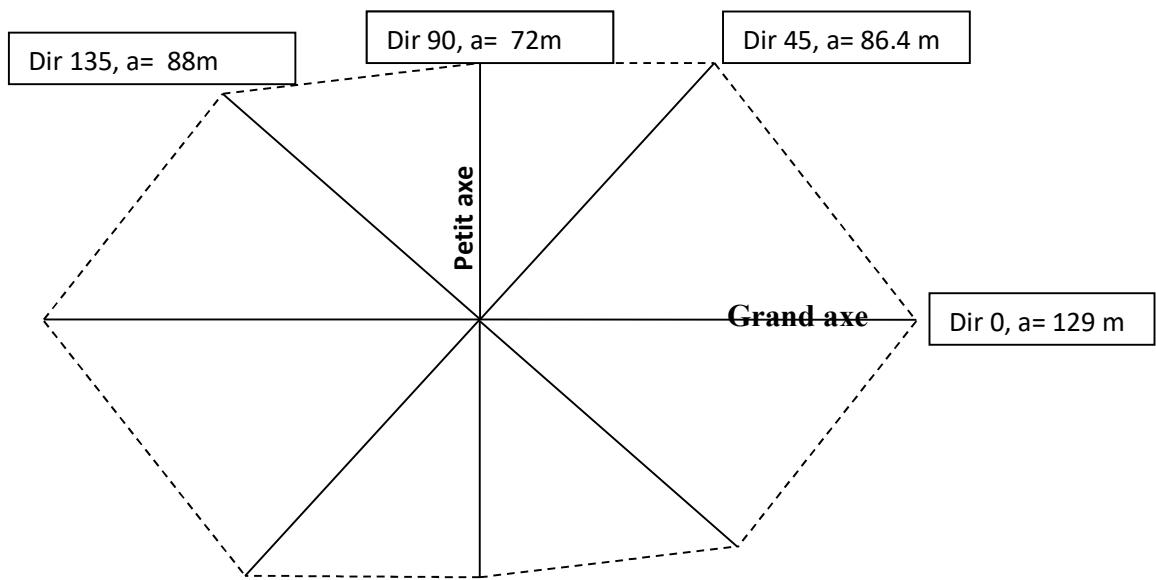




Justification

- Présence de palier
- Linéaire à l'origine
- La projection de l'intersection de la tangente à l'origine avec le palier (à) représente plus de 2/3 de la porté, donc le modèle pour tous les variogrammes est sphérique.

Rose des portées



Coefficient d'anisotropie:

$$C = aP/aG$$

$$C = 72/129 = 0,55$$

Commentaire:

- Une anisotropie a été mise en évidence, l'axe de continuité des teneurs en Cd est orienté dans la direction 0° , caractérisée par une portée de 129. L'axe de la grande variabilité est dans la direction 90° caractérisé par une portée de 72 m.
- La maille de prospection est alors, adaptée selon les portées.