

SIMULATION DU MODÈLE GAUSSIEN SEUILLÉ CONDITIONNÉ PAR DES CONTRAINTES DE CONNEXITÉ.

Denis ALLARD
Centre de Géostatistique
Ecole des Mines de Paris
35, rue Saint-Honoré
77 305 Fontainebleau, FRANCE

RÉSUMÉ : Un des paramètres essentiels des écoulements de fluide dans les réservoirs pétroliers hétérogènes est la connexité de la phase perméable. Aujourd'hui, les simulations conditionnelles des faciès géologiques ne prennent pas en compte l'information disponible concernant la connexité entre points, qui peut provenir, par exemple, des tests de puits ou de l'interprétation géologique. Un algorithme de simulation du modèle gaussien seuillé qui respecte des relations de connexité de la forme « x est connecté à x' » ou « x n'est pas connecté à x' » est présenté. Des exemples illustrent la méthode, en montrant ses limites et soulèvent le problème de l'adéquation entre le modèle de fonction aléatoire gaussienne (choisi *a priori*) et les contraintes qui le conditionnent.

ABSTRACT : One of the main parameter in oil and gas reservoir exploitation is the connectivity of the permeable phase. Generally, conditional simulations of the geological lithofacies do not take into account the available information about connectivity that can be provided by well-test or geological interpretation for instance. Using the truncated gaussian model, a simulation algorithm is proposed which respects connectivity constraints such as " x is connected to x' " and " x is not connected to x' ". Practical cases will illustrate the algorithm, its limits and the problem of the consistency between the model of the gaussian random function and the constraints of the conditional simulations.

1 Introduction

Un des objectifs majeurs de la géostatistique pétrolière est de proposer des modèles de réservoirs hétérogènes qui parviennent à reproduire l'hétérogénéité de celui-ci en intégrant le maximum de données et d'informations disponibles. À cette fin plusieurs modèles stochastiques ont été développés parmi lesquels on peut citer —sans que la liste ne soit exhaustive— le modèle gaussien seuillé [13], la simulation séquentielle d'indicateurs [1] et le schéma booléen [11]. Ce dernier ne permet pas à ce jour de conditionnement satisfaisant. La simulation séquentielle d'indicateurs permet d'honorer des données sous forme de valeurs exactes ou d'appartenance à un intervalle. Le modèle gaussien seuillé permet aujourd'hui des simulations conditionnées par des valeurs et/ou des inégalités, éventuellement bruitées [6] et régularisées [5]. Par une généralisation au cas multivariable, une cosimulation [14, 8] parvient à prendre en compte l'information apportée par une autre variable

mieux échantillonnée, comme l'impédance acoustique. Finalement, ces différents modèles intègrent des quantités d'informations de plus en plus importantes. Toutefois, un des paramètres essentiels en vue des écoulements de fluides, la connexité de la phase perméable ou des très hautes perméabilités, n'est jamais pris en compte. Une tentative a été faite en ce sens [12], mais elle se limitait à un aspect géométrique particulier et local. Plus récemment, un algorithme de simulation respectant les courbes issues de tests de puits par une méthode de recuit simulé a été présenté [3].

Dans cet article nous proposons une méthode de simulation du modèle gaussien seuillé, conditionné par des contraintes de connexité du type « x est connecté à x' » et/ou « x n'est pas connecté à x' ». L'algorithme est dérivé de l'échantillonneur de Gibbs, introduit en 1984 par les frères Geman [7].

La connexité de l'ensemble aléatoire représentant la phase perméable dépend très fortement du modèle choisi et de la grille sur laquelle celui-ci sera simulé. Aussi, après avoir brièvement rappelé quelques définitions concernant le modèle gaussien seuillé et la connexité sur une grille régulière, nous montrerons au paragraphe 3 l'influence du choix du modèle et du paramètre de portée sur certaines caractéristiques de connexité. Nous présenterons ensuite formellement l'algorithme de simulation au paragraphe 4. Avant de conclure, nous consacrerons le dernier paragraphe à la présentation de quelques simulations, et au problème posé par la cohérence entre le modèle —choisi *a priori*— et les contraintes

2 Notations et définitions

Sur la grille carrée $G = \{1, \dots, n\} \times \{1, \dots, n\}$, chaque site $x = (i, j)_{1 \leq i, j \leq n}$ possède 4 plus proches voisins : $(i-1, j)$, $(i+1, j)$, $(i, j-1)$, $(i, j+1)$, les sites appartenant aux bords (resp. aux coins) ne possédant quant à eux que 3 (resp. 2) plus proches voisins. Soit A un ensemble aléatoire stationnaire sur G de proportion $p = P(x \in A)$ et de covariance $C(h) = P(x \in A \text{ et } x+h \in A) - p^2$. Introduisons quelques définitions nécessaires à l'étude de la connexité sur l'ensemble aléatoire A .

Définition 1 Deux sites x et x' de A sont dits connectés, et on note $x \leftrightarrow x'$, s'il existe un chemin reliant x à x' entièrement contenu dans A . Un chemin est une séquence de sites $x_1, \dots, x_n$ où $(x_i, x_{i+1})_{1 \leq i \leq n-1}$ sont des couples de plus proches voisins.

Définition 2 La fonction de connexité, τ , est la probabilité que deux sites de l'ensemble aléatoire A soient connectés :

$$\tau(x, x') = P(x \leftrightarrow x' \mid x, x' \in A) \quad x, x' \in G$$

Enfin, on introduit l'ensemble des sites connectés à un site donné :

Définition 3 La composante connexe en x_0 , notée C_{x_0} est l'ensemble des sites de A connectés à x_0 :

$$C_{x_0} = \{x \in A : x \leftrightarrow x_0\}$$

Les composantes connexes forment une partition de A : ce sont les plus grands sous-ensembles connexes de A . On notera par la suite $|\cdot|$ le cardinal d'un sous-ensemble de G . En particulier, $|G| = n^2$ et $E\{|A|\} = pn^2$.

Les contraintes imposées peuvent être de 4 types : $(x_i)_{i \in I} \subset A$, $(x_j)_{j \in J} \subset A^c$, $(x_l \leftrightarrow x'_l)_{l \in L}$ et enfin $(x_m \nleftrightarrow x'_m)_{m \in M}$ où I , J , K et L sont des ensembles d'indices. Une méthode de simulation des deux premiers types de contraintes est présentée dans [6]. Par la suite, nous ne considérerons plus que les deux derniers ensembles de contraintes : $\Gamma_1 = (x_l \leftrightarrow x'_l)_{l \in L}$ et $\Gamma_0 = (x_m \nleftrightarrow x'_m)_{m \in M}$. On notera $\Gamma = \Gamma_0 \cup \Gamma_1$. Il faut remarquer que ces contraintes ne sont pas locales mais font intervenir tout le champ de la simulation, le chemin reliant deux sites quelconques pouvant passer par tout site de G . Notons $\mathcal{A}(\Gamma)$ l'ensemble des sous-ensembles de G qui satisfont aux contraintes de connexité Γ . Nous n'aborderons pas le problème de la cohérence de ces contraintes, et nous supposerons dans la suite de cet article qu'il existe des solutions satisfaisant Γ : $\mathcal{A}(\Gamma) \neq \emptyset$. Le problème posé est la simulation d'un ensemble aléatoire $A \in \mathcal{A}(\Gamma)$ de proportion p et de covariance $C(h)$. À cette fin, nous allons utiliser le modèle gaussien seuillé.

Sur G on associe le vecteur aléatoire gaussien $\mathbf{Y} = (Y_x)_{x \in G}$ de moyenne $E\{Y_x\} = 0$, $x \in G$ et de covariance $\rho(h) = E\{Y_x \cdot Y_{x+h}\}$, $x, x+h \in G$. On notera $\mathbf{y} = (y_x)_{x \in G}$ une réalisation de $\mathbf{Y}$. On définit l'ensemble aléatoire A comme l'ensemble des sites dont la coordonnée du vecteur gaussien associé dépasse un certain seuil y_c :

$$A = \{x \in G : Y_x \geq y_c\}$$

A est un ensemble aléatoire stationnaire, et on rappelle que [13]

$$\begin{aligned} p &= 1 - G_1(y_c) \\ C(h) &= g_1(y_c) \sum_{n \geq 1} \frac{H_{n-1}^2(y_c) \rho^n(h)}{n} \end{aligned}$$

où g_1 et G_1 sont respectivement la densité et la fonction de répartition gaussienne univariée, et où H_n est le polynôme d'Hermite normalisé de degré n .

3 Etude non conditionnelle de la connexité

Avant de mettre en œuvre l'algorithme de simulation conditionnée par des contraintes de connexité, nous allons montrer comment évoluent le nombre de composantes connexes, N , et la fonction de connexité, en fonction des modèles et de leurs paramètres.

Le nombre de composantes connexes

Partant de $p = 0$ où $N = 0$, faisons augmenter p . Dans un premier temps, de nouvelles composantes connexes apparaissent dans les nombreux espaces vides, et $E\{N\}$ augmente; mais peu à peu, des composantes connexes jusqu'alors distinctes sont mises en contact, tendant à diminuer $E\{N\}$. Ces deux effets antagonistes trouvent un équilibre pour une valeur particulière de p où $E\{N\}$ est maximum après laquelle le nombre de composantes connexes diminue jusqu'à $N = 1$ pour $p = 1$ (figure 1).

Lorsque le paramètre de portée de la covariance augmente, la régularité de l'ensemble aléatoire défini au-dessus de la coupure augmente aussi. Les composantes connexes sont alors plus grandes et moins nombreuses pour un même seuil de coupure. L'influence de ce paramètre est très importante, comme cela peut être vu à la figure 1. L'influence du modèle est moins importante et plus complexe. Deux modèles différents de même portée intégrale $\xi = \sum_{h \in \mathbb{Z}^2} C(h)$ sont comparés à la même figure 1. Les composantes connexes

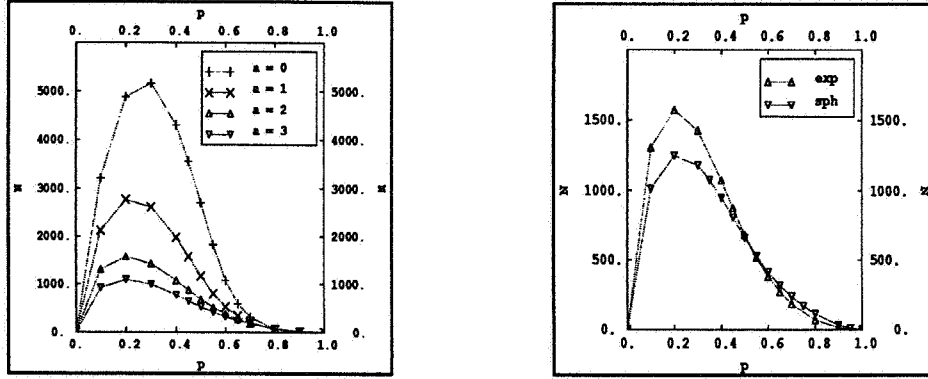


FIG. 1 - Nombre de composantes connexes en fonction de la proportion (moyenne sur 250 simulations sur une grille 200×200 noeuds). A gauche: Covariance exponentielle factorisée de paramètre a . A droite: Covariance exponentielle factorisée et sphérique isotrope de même portée intégrale $\xi = 16$.

sont moins nombreuses dans le modèle sphérique aux faibles proportions, mais un peu plus nombreuses aux fortes proportions. Ces différences sont toutefois peu importantes en comparaison d'une variation du paramètre de portée.

La fonction de connexité

Pour une distance h fixée, on reporte la probabilité $P_p(x \leftrightarrow x+h \mid x, x+h \in A)$ en fonction de la proportion p (figure 2). L'existence d'un seuil, séparant le domaine où $\tau(h) \simeq 0$ et où $\tau(h) \gg 0$, apparaît clairement. La courbe est légèrement déplacée vers les proportions plus importantes lorsque la distance $|h|$ augmente, vers les proportions plus faibles lorsque le paramètre de portée augmente.

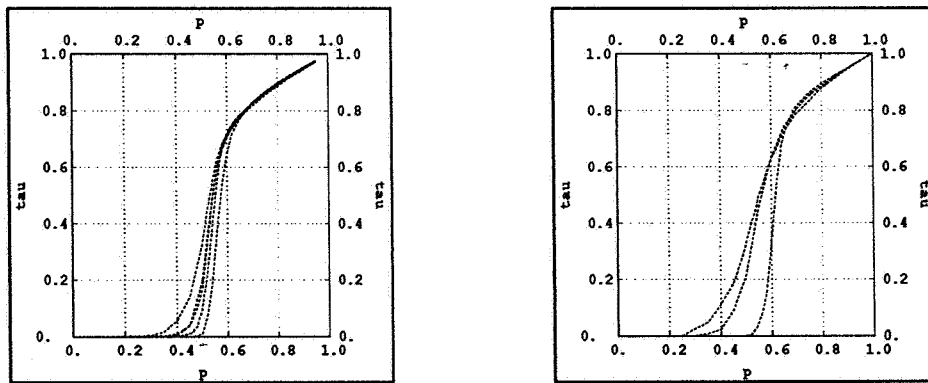


FIG. 2 - Fonction de connexité τ en fonction de la proportion p (Covariance exponentielle factorisée). A gauche: $a = 2$ et $h = (d, d)$, avec $d = 10, 20, 40, 80, 160$ de gauche à droite. A droite: $h = (20, 20)$ avec $a = 0, 2, 4$ de droite à gauche.

Si l'ensemble aléatoire A est stationnaire avec une portée finie (ie. $\exists R < \infty$ tel que si $|h| > R$, alors $C(h) = 0$), on peut montrer que la fonction $\theta(p) = \lim_{|h| \rightarrow \infty} P_p(x \leftrightarrow x+h \mid x, x+h \in A)$ est telle que

$$\theta(p) \begin{cases} = 0 & \text{si } p < p_c \\ > 0 & \text{si } p > p_c \end{cases}$$

En théorie de la percolation, p_c est appelé *seuil de percolation*. Il sépare le phénomène en deux phases [9] : lorsque $p < p_c$, toutes les composantes connexes sont finies; au contraire, lorsque $p > p_c$, il existe une composante connexe infinie sur $\mathbb{Z}^d$, et elle est unique. Nous la noterons C_∞ .

On constate expérimentalement que la fonction de connexité $\tau(h)$ est isotrope pour la norme euclidienne $|h|$ malgré la non-isotropie de la covariance exponentielle factorisée et de la grille carrée. Il n'existe pas de justification théorique générale de ce résultat. Toutefois, pour la percolation de liens non corrélée on établit [2] l'approximation suivante :

$$h \in \mathbb{Q}^2, nh \in \mathbb{Z}^2, \quad \tau(nh) \simeq e^{-n \cdot \varphi(p) \cdot g(h)} \quad \text{pour } n \text{ grand}$$

où $g(h)$ peut être prolongée sur $\mathbb{R}^2$ par la norme euclidienne.

De façon plus générale, si A est stationnaire et ergodique, on montre que la fonction de connexité suit une décroissance exponentielle dans l'argument de laquelle intervient une fonction $\varphi(p)$ dont on peut montrer a) qu'elle est continue et non croissante sur $[0, 1]$, b) que $\varphi(p) \rightarrow \infty$ pour $p \rightarrow 0$, c) que $\varphi(p) = 0$ pour $p > p_c$.

La fonction de connexité expérimentale est reportée à la figure 3. Pour de faibles proportions, $\tau(h)$ décroît rapidement avec $|h|$ pour tendre vers 0 : dans ce cas en effet, les composantes connexes sont de petites tailles et deux sites éloignés n'ont qu'une probabilité infime d'être connectés. Lorsque p augmente, la décroissance en $|h|$ est plus lente, mais on observe toujours que $\tau(h)$ tend vers 0 lorsque $|h|$ tend vers l'infini, et ce jusqu'au seuil de percolation p_c , au-delà duquel la fonction $\varphi(p)$ est identiquement nulle. Dans ce cas, $\tau(h)$ décroît puis atteint un palier lorsque $|h|$ tend vers l'infini. En effet, aux courtes distances deux sites connectés de A peuvent appartenir soit à une petite composante connexe, soit à la composante connexe infinie. Mais lorsque $|h| \rightarrow \infty$, deux sites ne sont connectés que s'ils appartiennent tous deux à la composante connexe infinie et $\tau(h) = P(x \leftrightarrow x+h | x, x+h \in A) \simeq P(x \in C_\infty \text{ et } x+h \in C_\infty | x, x+h \in X) \simeq (P(x \in C_\infty)/p)^2$.

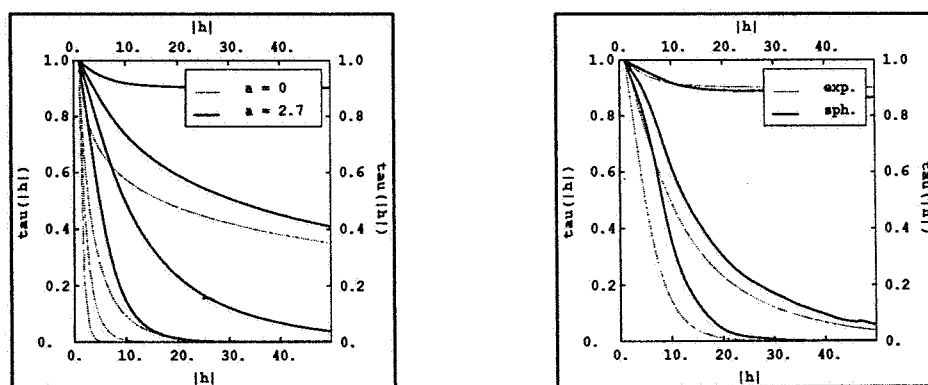


FIG. 3 - Fonction de connexité $\tau(|h|)$ A gauche: cov. exp. fact. pour deux valeurs du paramètre de portée (de bas en haut : $p = 0.2, 0.4, 0.5, 0.6$). A droite: cov. exp. fact. et cov. sph. de même portée intégrale $\xi = 29$ (de bas en haut : $p = 0.2, 0.4, 0.6$).

Sur cette même figure, on observe que $\tau(h)$ augmente à $|h|$ et à p fixés avec le paramètre de portée a . Pour deux modèles de covariance ayant même portée intégrale, la probabilité

de connexion est plus importante dans le modèle sphérique que dans le modèle exponentiel factorisé lorsque la proportion est inférieure au seuil de percolation. La situation s'inverse au-delà de p_c .

4 Algorithme de simulation

Présentation formelle

Soit B l'ensemble des vecteurs $\mathbf{y} \in \mathbb{R}^{n^2}$ vérifiant les conditions C . Une simulation conditionnelle de $\mathbf{Y}$ est une réalisation de la loi conditionnelle qui respecte ces contraintes. Dans le modèle gaussien, la densité f à simuler est donc :

$$\mathbf{y} \in \mathbb{R}^{|G|} \quad f(\mathbf{y}) = \frac{g(\mathbf{y})\mathbf{1}_B(\mathbf{y})}{\int_{\mathbb{R}^{n^2}} g(\mathbf{y})\mathbf{1}_B(\mathbf{y}) d\mathbf{y}}$$

où g est la densité multigaussienne et $\mathbf{1}$ la fonction indicatrice : $\mathbf{1}_B(\mathbf{y}) = 1$ si $\mathbf{y} \in B$, $\mathbf{1}_B(\mathbf{y}) = 0$ sinon. L'ensemble B est de géométrie complexe; en particulier, il ne se factorise pas en un produit d'intervalles, les contraintes n'étant plus ponctuelles. Conditionnellement à $(Y_x = y_x)_{x \neq x_0}$, l'ensemble $\tilde{B}_{x_0}$ auquel la composante Y_{x_0} doit appartenir est un intervalle de $\mathbb{R}$ d'une des trois formes suivantes : $] -\infty, y_c[$, $[y_c, +\infty[$ ou $\mathbb{R}$.

Il n'existe pas de méthode permettant la simulation directe de $\mathbf{Y} \in B$ selon la loi f . Une solution consiste à mettre en œuvre une méthode itérative: partant d'un vecteur $\mathbf{y}_0 \in B$ quelconque, on en modifie ses coordonnées au cours des itérations de manière à aboutir, au bout d'un certain temps, à un vecteur acceptable dans le cadre du modèle. La méthode utilisée ici est l'échantillonneur de Gibbs, introduit en 1984 par les frères GEMAN [7] dont on trouvera une mise en œuvre géostatistique dans [6, 5].

Soit $\{x_1, x_2, \dots, x_k, k \in \mathbb{N}\}$, un schéma de visite de G que l'on supposera périodique, de période $|G| = n^2$. La k -ième transition de $\mathbf{Y}^k$ à $\mathbf{Y}^{k+1}$ est définie de la façon suivante (on note $\tilde{\mathbf{Y}}_{x_k}^k$ le vecteur $\mathbf{Y}^k$ privé de la composante correspondant au site x_k):

$$Y_x^{k+1} = \begin{cases} Y_x^k & \text{si } x \neq x_k \\ Y_{x_k}^{KS} + \sigma_{KS}(x_k)U_k & \text{si } x = x_k \end{cases}$$

où $Y_{x_k}^{KS} = Y_{x_k}^{KS}(\tilde{\mathbf{Y}}_{x_k}^k)$ est le krigeage simple du site x_k par les autres sites de G , $\sigma_{KS}^2(x_k)$ est la variance de ce krigeage et U_k est une variable gaussienne éventuellement seuillée, telle que $Y_{x_k}^{k+1} \in \tilde{B}_{x_k}$. Les $\mathbf{Y}^k$ forment une chaîne de Markov, la loi de $\mathbf{Y}^k$ ne dépendant que de l'état précédent, $\mathbf{Y}^{k-1}$, et étant indépendante de tous les autres. La probabilité de transition

$$P_k(\mathbf{y}^k, F) = P(\mathbf{Y}^{k+1} \in F \mid \mathbf{Y}^k = \mathbf{y}^k) = \int_F dP_k(\mathbf{y}^k, \mathbf{y}^{k+1})$$

s'exprime à l'aide de la densité conditionnelle $f(\cdot \mid \tilde{\mathbf{y}}_{x_k})$:

$$dP_k(\mathbf{y}, \mathbf{v}) = \delta_{\tilde{\mathbf{y}}_{x_k}^k}(\tilde{\mathbf{v}}_{x_k}) f(v_{x_k} \mid \tilde{\mathbf{y}}_{x_k}) dv_{x_k}$$

avec

$$f(v_{x_k} | \bar{y}_{x_k}) = \frac{g_1\left(\frac{v_{x_k} - y_{x_k}^{KS}}{\sigma_{KS}(x_k)}\right) \mathbf{1}_{\tilde{B}_{x_k}}(v_k)}{\int_{\mathbb{R}} g_1\left(\frac{v_{x_k} - y_{x_k}^{KS}}{\sigma_{KS}(x_k)}\right) \mathbf{1}_{\tilde{B}_{x_k}}(v_k) dv_k}$$

On peut montrer [6] que f —loi de la gaussienne tronquée sur B — est la densité d'une probabilité stationnaire de P_k , c'est à dire

$$\forall F \in \mathcal{B}_B \quad \int_B f(y) P_k(y, F) dy = \int_F f(y) dy$$

où $\mathcal{B}_B$ est la tribu des boréliens de B .

Cette densité est unique si B est compact et connexe [4] ou si $B = \mathbb{R}^{n^2}$ [10]. Si $B \neq \mathbb{R}^{n^2}$ est connexe mais non compact, il n'y a pas à notre connaissance de résultat concernant l'unicité de la loi stationnaire. Moyennant une approximation acceptable, on peut toutefois s'affranchir de cette restriction, en considérant une troncature de la loi gaussienne sur l'ensemble $B \cap [-a, a]^{n^2}$ où a est une constante positive suffisamment grande. Cela revient à modifier le modèle —on ne considère plus le modèle gaussien, mais le modèle gaussien tronqué sur $[-a, a]^{n^2}$. En pratique, $a > 10$ n'entraîne aucun changement perceptible du modèle et n'a aucune influence sur la mise en œuvre de l'algorithme.

Il n'est pas toujours possible de déterminer si l'ensemble B des vecteurs qui satisfont à des conditions Γ données est connexe ou non. Toutefois, il existe quelques cas simples pour lesquels B est connexe :

- Si $\Gamma = \Gamma_1$, c'est à dire si les contraintes se réduisent à des conditions de connexité uniquement, alors $[y_c, +\infty]^{n^2} \subset B$ et B est connexe.
- De même, si $\Gamma = \Gamma_0$, c'est à dire si les contraintes se réduisent à des conditions de non-connexité seulement, alors $]-\infty, y_c]^{n^2} \subset B$ et B est encore connexe.
- Si $\Gamma = \Gamma_0 \cup \Gamma_1$, mais que les relations de *non-connexité* ne font intervenir que des relations entre deux composantes connexes dont une au moins n'a qu'un site conditionnant, alors B est toujours connexe. Un point conditionnant est un point x_m ou x'_m , $m \in M$ qui apparaissent dans la définition de Γ_0 (cf section 2).

Dans ces trois cas simples, l'échantillonneur de Gibbs échantillonne bien l'ensemble des solutions possibles.

Dans les autres cas

Revenons à l'ensemble $\mathcal{A}(\Gamma)$ des sous-ensembles de G qui satisfont aux relations de connexité Γ . Nous dirons que deux sous-ensembles $\mathbf{x}$ et $\mathbf{x}' \in \mathcal{A}$ sont en relation, notée $\mathbf{x} \rightleftharpoons \mathbf{x}'$ s'il est possible d'atteindre $\mathbf{x}'$ à partir de $\mathbf{x}$ (et inversement) en un nombre fini de transformations élémentaires. La transformation élémentaire au site quelconque $x \in G$, notée t_x , appliquée à $\mathbf{x} \in \mathcal{A}(\Gamma)$ est :

$$t_x(\mathbf{x}) = \begin{cases} \begin{cases} \mathbf{x} + \{x\} & \text{si } x \notin \mathbf{x} \\ \mathbf{x} \setminus \{x\} & \text{si } x \in \mathbf{x} \end{cases} \\ \mathbf{x} & \text{sinon} \end{cases} \quad \text{pour autant que } t_x(\mathbf{x}) \in \mathcal{A}(\Gamma)$$

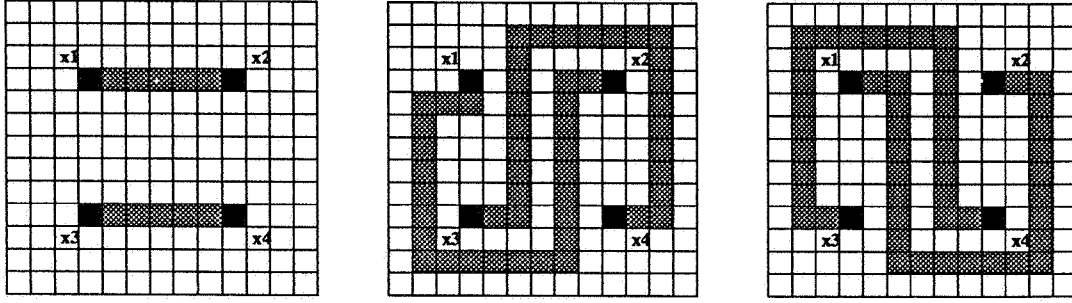


FIG. 4 - Les conditions sont : $(x_1 \leftrightarrow x_2)$; $(x_3 \leftrightarrow x_4)$; $(x_1 \not\leftrightarrow x_3)$. Les trois solutions appartiennent à des classes d'équivalence différentes.

Cette transformation définit une dynamique déterministe correspondant à la dynamique de l'échantillonneur de Gibbs. La relation $\rightleftharpoons$ est une relation d'équivalence, et les classes d'équivalences associées forment autant de classes de solutions. Il est important de souligner que deux ensembles x_1 et x_2 appartenant à des classes distinctes ne sont pas accessibles l'un à l'autre par des transformations élémentaires. Un exemple en est donné à la figure 4 où 4 sites conditionnants $\{x_1, \dots, x_4\}$ sont représentés. Les conditions sont les suivantes : $(x_1 \leftrightarrow x_2)$; $(x_3 \leftrightarrow x_4)$; $(x_1 \not\leftrightarrow x_3)$. Il existe 3 classes de solutions illustrées ci-dessous, et il est impossible de passer de l'une à l'autre par des transformations élémentaires sans rompre une des trois relations de connexité.

Plus généralement, à chaque classe $A_i(\Gamma)$ correspond une classe B_i de vecteurs $\mathbf{Y}$. La conséquence sur l'algorithme de simulation est que, partant d'un vecteur $\mathbf{y}_0 \in B$, l'échantillonneur de Gibbs ne parcourra que les solutions appartenant à la classe d'équivalence de $\mathbf{y}_0$ pour la relation $\rightleftharpoons$, en ignorant toutes les solutions appartenant aux autres classes. Ainsi, si B se partitionne en n_c classes disjointes, $(B_i)_{1 \leq i \leq n_c}$, notons $g(B_i)$ la mesure de l'ensemble B_i par rapport à la densité multigaussienne g ($g(\mathbb{R}^{n^2}) = 1$):

$$g(B_i) = \int_{B_i} g(\mathbf{y}) d\mathbf{y}$$

La densité f de la gaussienne tronquée sur B s'écrit alors

$$f(\mathbf{y}) = \frac{g(\mathbf{y})\mathbf{1}_B(\mathbf{y})}{g(B)} = \frac{g(\mathbf{y})\sum_i \mathbf{1}_{B_i}(\mathbf{y})}{\sum_i g(B_i)} = \sum_i \frac{g(B_i)}{\sum_i g(B_i)} f_i(\mathbf{y})$$

où

$$f_i(\mathbf{y}) = \frac{g(\mathbf{y})\mathbf{1}_{B_i}(\mathbf{y})}{g(B_i)}$$

est la densité de la gaussienne tronquée sur B_i . Celle-ci est la densité d'une loi stationnaire pour la probabilité de transition P_k . f apparaît comme la somme des f_i , pondérée par le système de probabilité engendré par les mesures $g(B_i)$.

En résumé, dans le cas où il y a plusieurs classes de solutions, l'échantillonneur de Gibbs n'explorera pas l'ensemble de toutes les solutions B , mais uniquement celles appartenant à la même classe que l'image initiale, B_i . De plus, la loi simulée sera f_i , gaussienne tronquée sur la classe B_i , et non f , gaussienne tronquée sur B .

Description algorithmique

Dans tous les cas de figure, l'algorithme à mettre en oeuvre est le suivant :

- 1) Construire $\mathbf{x}_0 \in \mathcal{A}(\Gamma)$, puis $\mathbf{y}_0 \in B$.
- 2) Parcourir G selon un schéma de visite $(x_k)_{k \in \mathbb{N}}$ périodique de période $|G| = n^2$; à l'itération k :
 - a) calculer la krigeage simple $Y_{x_k}^{KS}$ au site x_k , et sa variance de krigeage, $\sigma_{KS}^2(x_k)$;
 - b) déterminer à quel intervalle $(]-\infty, y_c[$, $[y_c, +\infty[$ ou $\mathbb{R}$) doit appartenir la coordonnée Y_{x_k} du vecteur $\mathbf{Y}$ afin de respecter les contraintes Γ ;
 - c) tirer une gaussienne indépendante U_k appartenant à l'intervalle $]-\infty, u_c[$, $[u_c, +\infty[$ ou $\mathbb{R}$, selon a), où $u_c = \frac{y_c - Y_{x_k}^{KS}}{\sigma_{KS}(x_k)}$;
 - d) assigner la valeur $Y_{x_k}^{KS} + \sigma_{KS}(x_k)U_k \rightarrow Y_{x_k}$.
- 3) Arrêter les itérations.

Remarque : En pratique, on choisira pour $\mathbf{y}_0$ un vecteur gaussien non corrélé construit à partir de $\mathbf{x}_0$.

5 Mise en oeuvre pratique

A l'aide d'exemples de simulations conditionnelles, nous allons aborder les problèmes de l'adéquation entre le modèle gaussien inféré et les contraintes de connexité, de la vitesse de convergence de l'algorithme, ainsi que les conséquences pratiques entraînées par l'existence de plusieurs classes de solutions. Dans la suite de cet article, nous utiliserons le modèle de covariance exponentiel factorisé afin d'accélérer l'étape de krigeage.

Influence du conditionnement

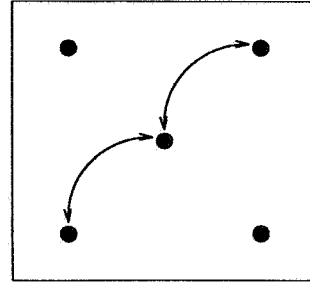
Dans ce paragraphe, nous allons supposer que les conditions de connexité sont telles qu'il y a une classe de solutions unique, B . Dans ce cas, quel que soit le vecteur de départ $\mathbf{y}_0 \in B$, les $(\mathbf{Y}^k)_{k \in \mathbb{N}}$ convergent en loi vers $\mathbf{Y}$, vecteur gaussien tronqué sur B , de densité $f(\mathbf{y}) = g(\mathbf{y})\mathbf{1}_B(\mathbf{y})/g(B)$. Ce vecteur aléatoire n'est plus gaussien; en particulier, ses deux premiers moments ne caractérisent pas la loi entière, et ceux-ci ne sont pas explicitement calculables. Une façon de rendre le vecteur $\mathbf{Y} \in B$ gaussien serait de randomiser les contraintes par un échantillonnage sur des simulations non conditionnelles. Nous n'avons pas réalisé de tests en ce sens, car notre objectif est précisément d'étudier l'évolution du modèle à simuler en fonction des contraintes. La fonction de connexité, τ , introduite au paragraphe 3, permet de faire quelques prévisions qualitatives.

Si Γ se réduit à une condition unique de connexité : " $x \leftrightarrow x'$ ", la mesure de l'ensemble des solutions est alors la fonction de connexité τ , $g(B) = P(x \leftrightarrow x') = \tau(x' - x)$ dont nous avons vu expérimentalement (cf figure 3) que, quasi nulle si $p < 0.4$, elle croît très rapidement autour de $p = 0.5$. Lorsque plusieurs contraintes de connexité et de non-connexité sont simultanément présentes, la probabilité que celles-ci soient vérifiées dans le modèle gaussien *a priori*, $P(\Gamma \text{ est vérifié}) = g(B)$, présente un maximum¹, pour une

1. En effet, $g(B) = 0$ si $p = 0$ et si $p = 1$, alors que $g(B) > 0$ si $p \in]0, 1[$.

proportion assurant un compromis entre les contraintes de connexion et de non-connexion. Expérimentalement, il apparaît que cette proportion est proche du seuil de percolation p_c et que c'est également autour de cette valeur que les différentes caractéristiques de connexité : nombre de composantes connexes, taille et forme de celles-ci, etc., ont la plus grande variabilité. On peut donc prévoir que le modèle simulé, f , sera différent selon que l'on est en-deçà, autour ou au-delà du seuil de percolation.

Voyons cela sur un exemple. Pour les conditions de connexité schématiquement représentées ici, —notées Γ' —, on réalise des simulations pour trois valeurs de la proportion théorique : inférieure, proche, et supérieure à p_c ($p = 0.2$, $p = 0.4$, $p = 0.7$). Les grandeurs expérimentales suivantes sont calculées en fonction du nombre de balayages, k : $\sigma^{*2}(k)$, $m^*(k)$, $\rho^*(k)$ et $p^*(k)$ respectivement égales à la variance, à l'espérance, à la covariance entre plus proches voisins et à la proportion expérimentale du vecteur $(y^k)_{k \in \mathbb{N}}$. Ces résultats sont reportés à la figure 5.



Si la proportion est proche du seuil de percolation, les paramètres du modèle *a priori* sont bien respectés. Au contraire, si la proportion est petite par rapport au seuil de percolation, la probabilité *a priori* des contraintes de connexion de Γ' est très faible. Afin de répondre aux sollicitations des contraintes, le nombre de sites au-dessus de la coupure doit augmenter ($p^* \geq p$), entraînant que $E\{(Y_x)^k\} \geq 0$. Inversement, si la proportion théorique est grande par rapport au seuil de percolation, c'est la probabilité *a priori* des contraintes de non-connexion de Γ' qui est très faible, et dans ce cas, $p^* \leq p$ et $E\{(Y_x)^k\} \leq 0$. La covariance à la 100-ième itération est bien reproduite pour $p = 0.4$ et $p = 0.7$. En revanche, le palier est supérieur à 1 lorsque $p = 0.2$. L'histogramme est décalé dans le sens indiqué par l'espérance : vers les plus fortes valeurs lorsque $p = 0.2$, vers les plus faibles valeurs lorsque $p = 0.7$. D'autres exemples pourraient être présentés. Tous vont dans le même sens : plus la configuration des contraintes de connexité est improbable dans le modèle gaussien *a priori*, plus les statistiques calculées sur les simulations conditionnelles s'écartent de celles du modèle. En fait, l'inférence du modèle (covariance et proportion) devrait intégrer les informations de connexité éventuellement disponibles qui conditionneront les simulations.

Vitesse de convergence

Lors de la mise en oeuvre de l'algorithme de simulation se pose également le problème de la vitesse à laquelle les $(Y^k)_{k \in \mathbb{N}}$ convergent vers la loi limite Y . On observe (par exemple à la figure 5) que les 4 grandeurs expérimentales présentées au paragraphe précédent se stabilisent ensemble au cours des itérations. La vitesse de stabilisation de la variance expérimentale —dorénavant choisie comme indicateur de la convergence— dépend du modèle de covariance de la gaussienne et des contraintes de simulation : proportion p et contraintes de connexité. Le premier facteur est très important. Pour une simulation non conditionnelle (modèle de covariance exponentielle factorisée de paramètre a), le nombre k de balayages nécessaires pour obtenir une bonne stabilisation de la variance expérimentale est reporté au tableau 1 :

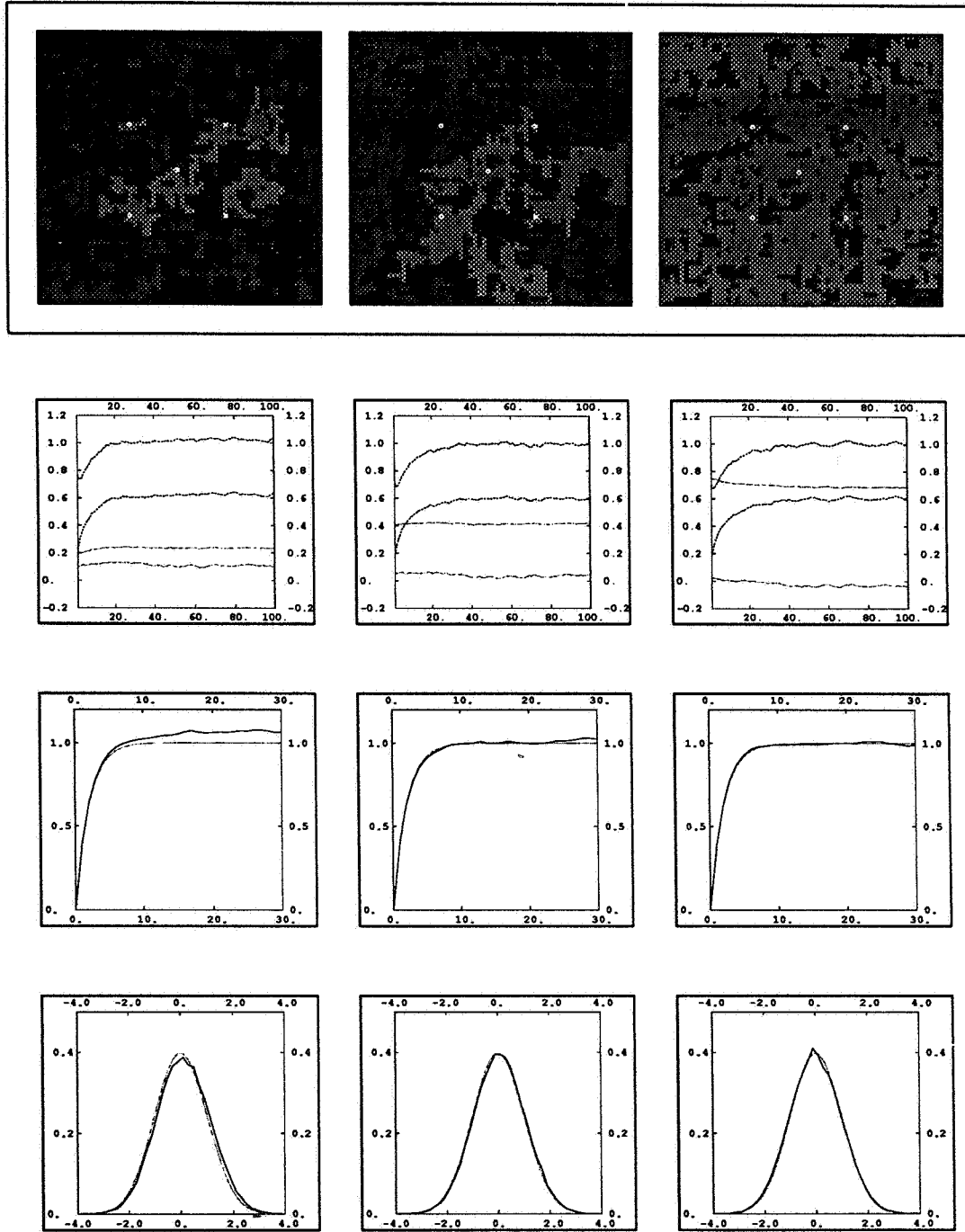


FIG. 5 - *En haut*: 3 simulations conditionnelles pour les contraintes Γ' avec $p = 0.2$; $p = 0.4$ et $p = 0.7$. Les pores sont en noir; les grains sont en gris (en gris clair, les 3 composantes connexes conditionnantes). En blanc sont figurés les 5 sites conditionnants. *Puis*, moyennés sur 12 simulations: les statistiques élémentaires ($\sigma^2(k)$, $m^*(k)$, $\rho^*(k)$ et $p^*(k)$) en fonction du nombre de balayages, le variogramme et l'histogramme à la 100-ième itération.

Tableau 1

a	$\rho = e^{-1/a}$	k
2.0	0.605	40
3.0	0.715	100
4.0	0.779	> 400

Ce tableau confirme que l'échantillonneur de Gibbs est un très mauvais (car très lent) algorithme de simulations non conditionnelles.

Mais on observe également que plus les contraintes C sont improbables, plus le conditionnement est fort et plus l'algorithme converge rapidement. Cet effet est nettement perceptible à la figure 6 où la variance expérimentale est représentée en fonction du nombre de balayages pour les trois valeurs de la proportion ci-dessus, ainsi que pour une simulation non conditionnelle.

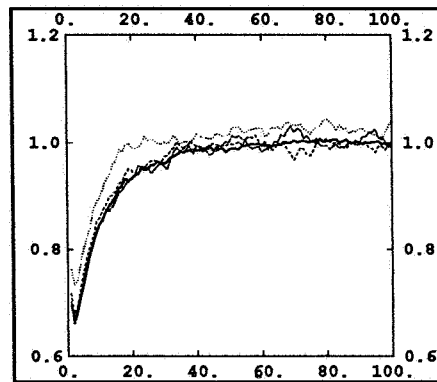


FIG. 6 - Variance expérimentale en fonction du nombre de balayages. Influence de la proportion sur des simulations conditionnées par les contraintes Γ' : $\cdots$ $p = 0.2$; $---$ $p=0.4$; $-.-.-$ $p=0.7$; $—$ non conditionnelle.

Plusieurs classes de solutions

S'il existe plusieurs classes de solutions $(B_i)_{1 \leq i \leq n_c}$ de probabilités respectives $g(B_i)$, l'image de départ conditionnera la classe à laquelle les simulations appartiendront. Afin de simuler la loi f , il faudrait pouvoir tirer une image de départ dans la classe i selon le système de probabilité $g(B_i)/g(B)$. Celui-ci n'est pas calculable, et cette démarche n'est pas possible. Reprenons l'exemple de la figure 4. Deux simulations conditionnelles sont présentées à la figure 7, correspondant aux deux premières classes de solutions. Les statistiques, moyennées sur 12 simulations, sont proches du modèle *a priori* pour la première classe, au contraire de la seconde classe, plus improbable dans le modèle gaussien.

Les contraintes de connexité telles que nous les avons présentées ne spécifient pas toujours suffisamment le modèle à simuler, lorsque l'algorithme de simulation utilisé est l'échantillonneur de Gibbs. Une part est laissée à l'arbitraire dans le choix de l'image de départ, ou dans les poids accordés à chacune s'il y en a plusieurs. Cet arbitraire est en contrepartie une souplesse inattendue, car il laisse à l'utilisateur la possibilité d'intégrer des interprétations géologiques qualitatives.

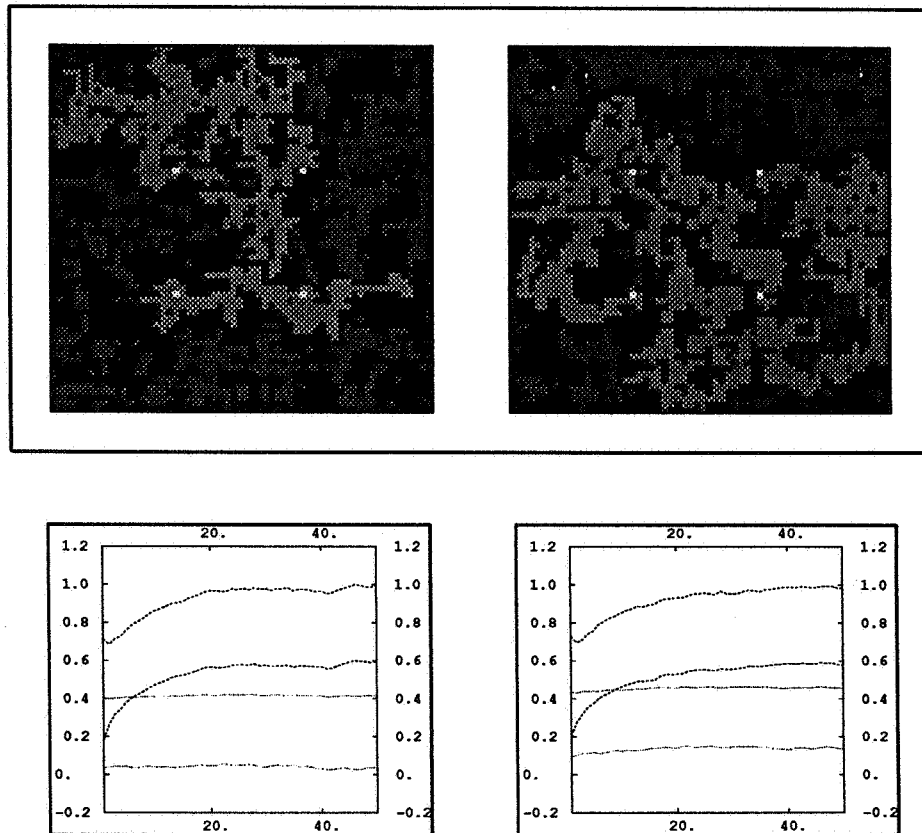


FIG. 7 - Simulations conditionnelles et statistiques élémentaires ($\sigma^2(k)$, $m^*(k)$, $\rho^*(k)$ et $p^*(k)$) moyennées sur 12 simulations pour deux classes de solutions des mêmes conditions de connexité.

6 Conclusion

L'échantillonneur de Gibbs est un algorithme puissant et souple permettant de réaliser des simulations conditionnelles qui prennent en compte des contraintes complexes. La simulation de lithofaciès géologiques peut ainsi être conditionnée par des contraintes de connexité. La loi simulée sera d'autant plus différente du modèle *a priori* que le conditionnement sera fort. Aussi, les statistiques sur les simulations pourront-elles s'écarter de façon notable de celles du modèle *a priori*. Afin de respecter celles-ci, il conviendrait de tenir compte des contraintes que l'on cherche à respecter dans l'inférence du modèle. Il peut arriver que le conditionnement fasse apparaître plusieurs classes de solutions. Dans ce cas, une simulation appartiendra toujours à la classe de l'image de départ. Il faut donc faire le choix de celles-ci de façon prudente, en sachant que les simulations conditionnelles reflètent fidèlement ce choix. Enfin, il reste que pour l'instant, cet algorithme est assez lent, nécessitant de nombreux calculs de composantes connexes. Dans sa version actuelle, une simulation 100×100 en 50 balayages nécessite environ 2 h CPU sur une station de travail usuelle.

Références

- [1] F. Alabert. *Stochastic imaging of spatial distributions using hard and soft data*. Msc. thesis, Stanford U, 1987.
- [2] K.S. Alexander et al. The Wulff construction and asymptotics of the finite cluster distribution for two dimensional Bernoulli percolation. *Comm. Math. Phys.*, 1989.
- [3] Cl. Deutsch. Conditioning reservoir models to well test information. In A. Soares, editor, *Geostatistics TRÓIA '92*, volume 1, pages 505–518, Dordrecht Holland, 1993. Kluwer Academic Press.
- [4] J. L. Doob. *Stochastic processes*. John Wiley & Sons, Inc., New York, 1953.
- [5] X. Freulon. Conditional simulation of a gaussian random vector with non-linear and/or noisy observations. In M. Armstrong, editor, *Workshop on geostatistical simulations (to appear)*, Dordrecht Holland, 1994. Kluwer Academic Press.
- [6] X. Freulon and Ch. de Fouquet. Conditioning a gaussian model with inequalities. In A. Soares, editor, *Geostatistics TRÓIA '92*, volume 1, pages 201–212, Dordrecht Holland, 1993. Kluwer Academic Press.
- [7] S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *I.E.E.E. transactions: Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.
- [8] J. Gómez-Hernández and A. Journel. Joint sequential simulation of multigaussian fields. In A. Soares, editor, *Geostatistics TRÓIA '92*, volume 1, pages 85–94, Dordrecht Holland, 1993. Kluwer Academic Press.
- [9] G. Grimmet. *Percolation*. Springer-Verlag, New York, 1992.
- [10] X. Guyon. *Champs aléatoires sur un réseau*. Masson, Paris, 1993.
- [11] H.H. Haldorsen and C.J. MacDonald. Stochastic modeling of underground reservoirs facies (smurf). *SPE, 62nd Annual Conference, Dallas, TX*, 1987.
- [12] A.G. Journel and F.G. Alabert. Focusing on spatial connectivity of extreme-valued attributes: Stochastic indicator models of reservoir heterogeneities. *SPE, 63rd Annual Conference, Huston, TX*, 1988.
- [13] G. Matheron et al. Conditional simulation of the geometry of fluviodeltaic reservoirs. *SPE, 62nd Annual Conference, Dallas, TX*, 1987.
- [14] G. Verly. Sequential gaussian co-simulation: A simulation method integrating several types of information. In A. Soares, editor, *Geostatistics TRÓIA '92*, volume 1, pages 543–554, Dordrecht Holland, 1993. Kluwer Academic Press.