

Examen 2015

Question 1 (15 points)

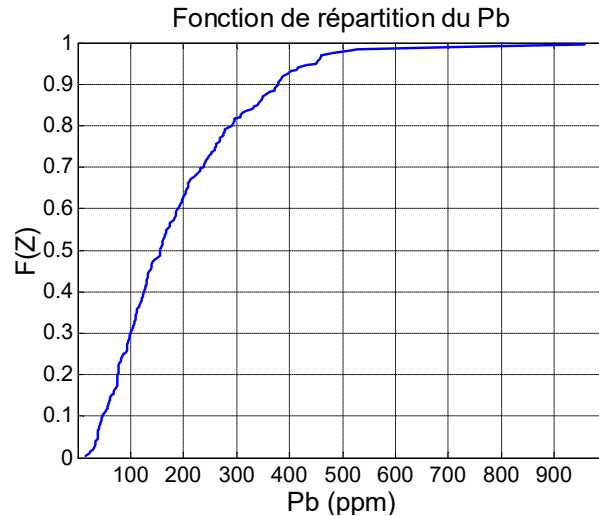
a) Expliquez pourquoi le post-conditionnement par krigeage des simulations non-conditionnelles peut être effectué beaucoup plus rapidement que le SGS sachant que l'on simule dans les deux cas le même nombre de points, et ce, en utilisant les mêmes données.

b) Par recuit simulé, on peut simuler des données dont le variogramme expérimental colle presque parfaitement au modèle souhaité. Expliquez en quoi ceci pourrait être en fait problématique?

c) La méthode SGS permet la simulation de variables continues gaussiennes. Proposez une modification à la méthode SGS qui permette la simulation avec de données non-gaussiennes (comme pour les données de contamination au Pb de Dallas).

Question 2 (12 points)

Un sol de la région de Montréal est contaminé au plomb dû à la présence d'une usine de recyclage de batteries automobiles. Deux cents analyses de la concentration en plomb sont disponibles et la fonction de répartition est illustrée à la figure suivante :



3 pts a) Quel devrait être approximativement le palier pour l'indicatrice correspondant au seuil 300 ppm?

Lors de l'estimation d'un point x_0 , le krigeage simple de l'indicatrice correspondant au seuil 200ppm retourne les poids suivants :

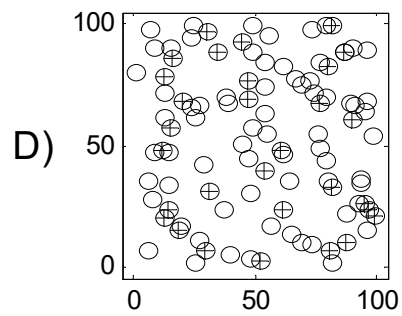
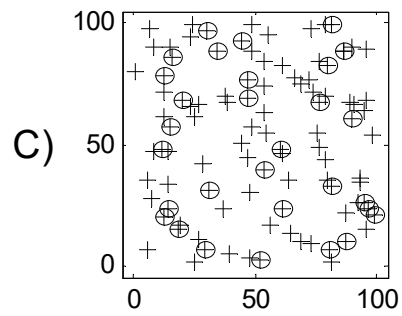
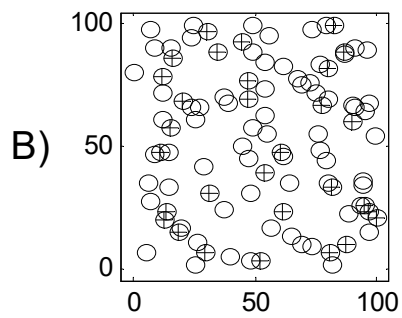
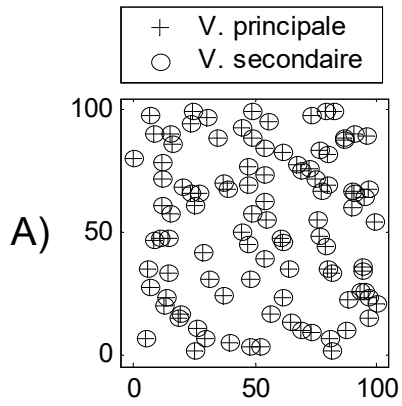
$$\lambda_1 = 0, \lambda_2 = -0.11, \lambda_3 = 0.22, \lambda_4 = 0.70$$

Les valeurs observées en ces points sont respectivement (en ppm) : $Z_1=127$, $Z_2=210$, $Z_3=185$, $Z_4=233$

9 pts b) Quelle est la probabilité (estimée par KI simple) que la concentration au point x_0 soit supérieure à 200 ppm ?

Question 3 (15 points)

Le tableau suivant présente, dans la colonne de gauche, le plan de localisation d'observations d'une variable principale (Z) et d'une variable secondaire (Y). Dans la colonne de droite, on donne le modèle linéaire de corégionalisation ajusté à ces données. Dans les matrices, Z apparaît en premier, Y en second. $\delta(h)$ représente la covariance effet de pépité.



$$C(h) = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 0.9 \end{bmatrix} \delta(h) + \begin{bmatrix} 8 & 8 \\ 8 & 9 \end{bmatrix} Sph(a=50, C=1)$$

$$C(h) = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 0.9 \end{bmatrix} \delta(h) + \begin{bmatrix} 8 & 8 \\ 8 & 9 \end{bmatrix} Sph(a=50, C=1)$$

$$C(h) = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 0.9 \end{bmatrix} \delta(h) + \begin{bmatrix} 8 & 8 \\ 8 & 9 \end{bmatrix} Sph(a=50, C=1)$$

$$C(h) = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 0.9 \end{bmatrix} \delta(h) + \begin{bmatrix} 8 & 0 \\ 0 & 9 \end{bmatrix} Sph(a=50, C=1)$$

Question 3 (suite)

12 pts *a) Indiquez pour chaque cas A) à D) si le cokrigage de $Z(x)$ (par Z et Y) est susceptible d'améliorer de façon importante la précision des estimations par rapport au seul krigeage de $Z(x)$. Justifiez vos réponses.*

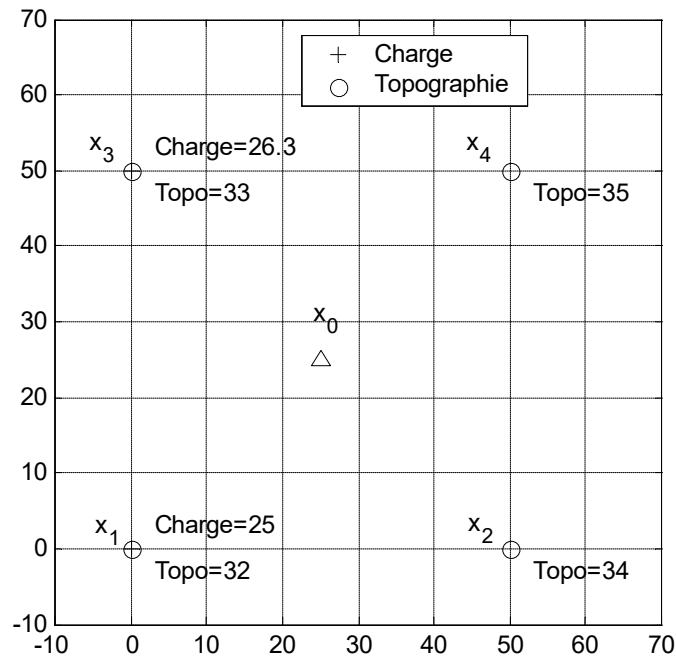
3 pts *b) Indiquez quelle méthode vous permettrait, avec les mêmes données, de confirmer ou d'infirmer vos réponses en a) ?*

Question 4 (18 points)

On désire estimer la charge hydraulique d'un aquifère à nappe libre. On observe une excellente corrélation entre la charge hydraulique et la topographie. Le modèle linéaire de corégionalisation pour ces deux variables (dans l'ordre : charge, topographie) est :

$$C(h) = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1.2 \end{bmatrix} \delta(h) + \begin{bmatrix} 7 & 9 \\ 9 & 12 \end{bmatrix} Sph(a = 100m, C = 1)$$

La figure suivante (coordonnées en m) montre les quatre points où les données (charge et/ou topographie, les deux en m) sont disponibles pour l'estimation de la charge au point x_0 .



3 pts a) Quelle est la corrélation (à $h=0$) entre charge et topographie selon le modèle de corégionalisation?

Question 4 (suite)

10 pts b) Construisez (sans le résoudre) le système de cokrigage ordinaire pour l'estimation de la charge hydraulique au point x_0 . Identifiez clairement les entrées de la matrice et des vecteurs. Le tableau suivant pourrait vous éviter des calculs :

Distance (m)	0	$25 \cdot 2^{0.5}$	50	$50 \cdot 2^{0.5}$
C(h), Sph(a=100m, C=1)	1	0.4918	0.3125	0.1161

Les poids de cokrigage et les multiplicateurs de Lagrange obtenus pour la question b) sont :

$$\lambda_{1, ch \arg e} = \lambda_{3, ch \arg e} = 1/2 \quad \lambda_{1, topo} = \lambda_{3, topo} = -0.19 \quad \lambda_{2, topo} = \lambda_{4, topo} = 0.19 \quad \mu_{ch \arg e} = -0.03 \quad \mu_{topo} = 0.31$$

5 pts c) Quelle est la valeur estimée et la variance de cokrigage pour la charge au point x_0 ?

Question 5 (12 points)

On effectue une simulation de moyenne 0 par méthode de Choleski. La matrice L obtenue est :

$$L = \begin{bmatrix} 7.75 & & & & \\ 3.82 & 6.74 & & & \\ 4.59 & 3.23 & 5.34 & & \\ 4.36 & 1.45 & 2.20 & 5.84 & \\ 4.77 & 1.79 & 2.42 & 2.86 & 4.48 \end{bmatrix}$$

3 pts a) *Quelle est la covariance entre les 2^e et 3^e variables aléatoires?*

3 pts b) *On a observé $Z_1=5$. Que vaut Y_1 ?*

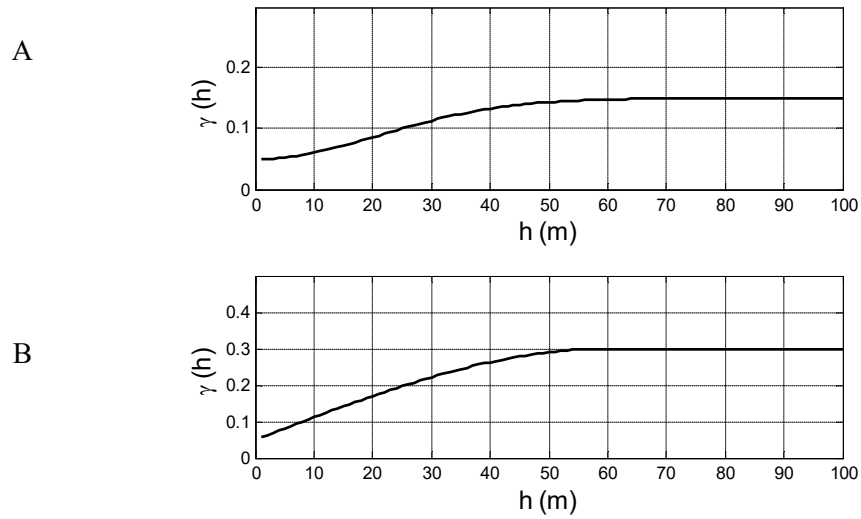
Question 5 (suite)

3 pts c) *Quelle est l'espérance conditionnelle de Z_2 sachant que $Z_1=5$?*

3 pts d) *En pratique, quelle est la principale limitation touchant la méthode de Choleski ?*

Question 6 (10 points)

La figure suivante montre des modèles de variogramme sensés être ajustés aux variogrammes des indicatrices correspondant à des seuils A et B pour la teneur en Pb (mesurée en ppm) du site de Dallas.

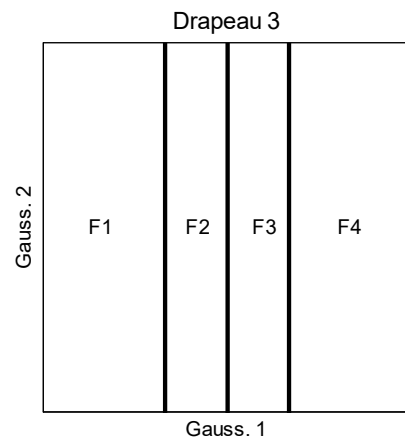
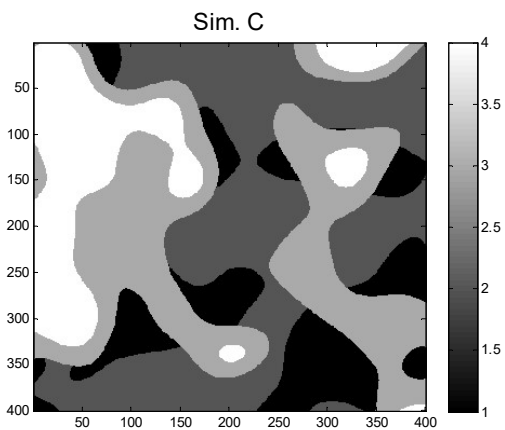
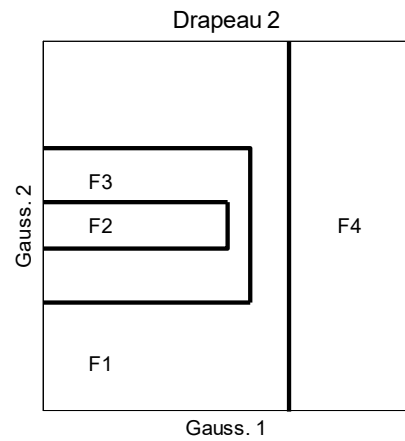
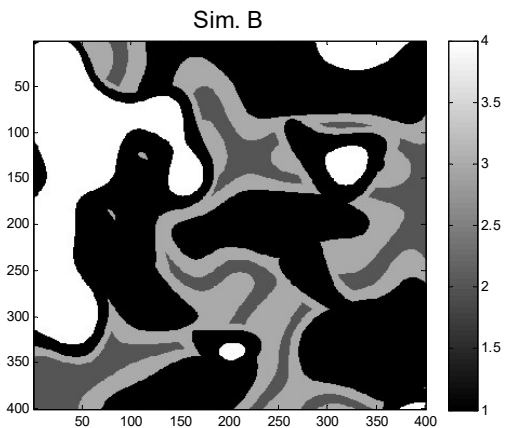
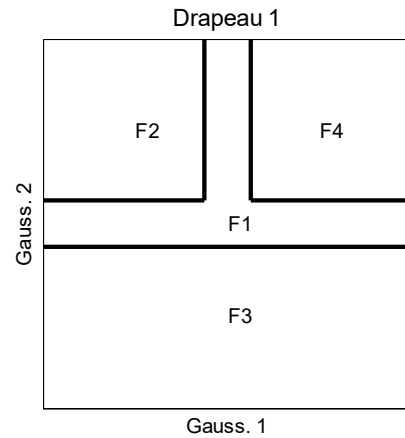
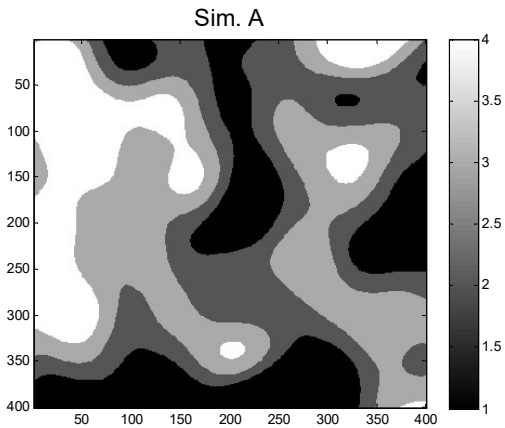


8 pts a) Quelles erreurs flagrantes ont été commises dans l'ajustement des variogrammes (justifier)?

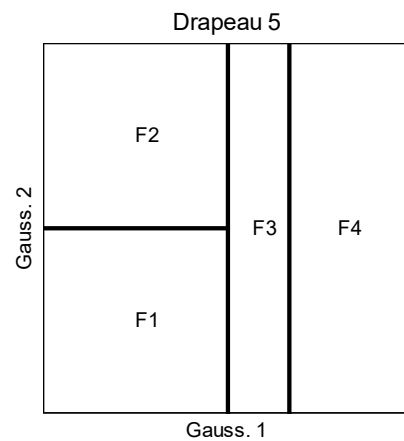
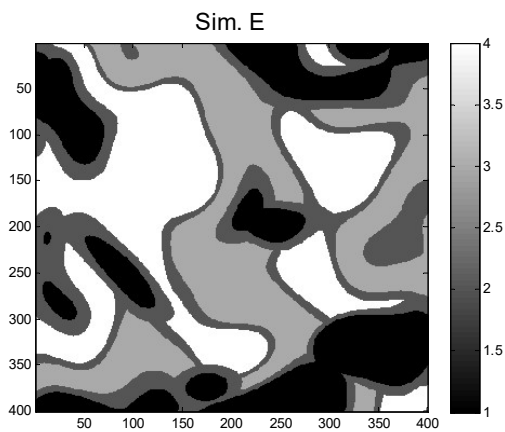
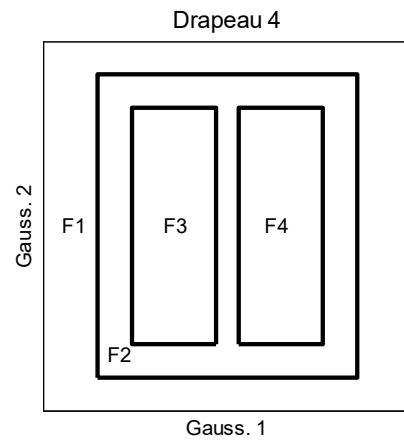
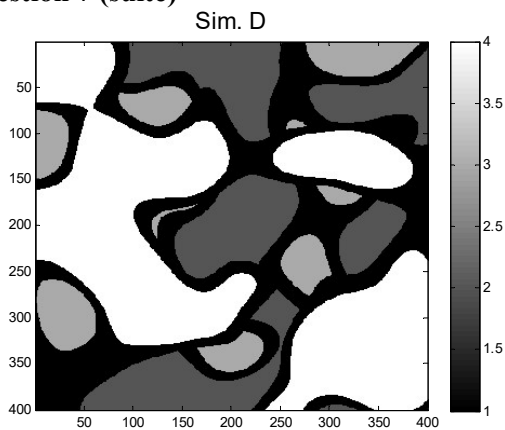
2 pts b) Quelles sont les unités des variogrammes illustrés?

Question 7 (18 points)

Les figures suivantes montrent 5 différentes simulations plurigaussiennes (A à E) et 5 différents drapeaux de codage (1 à 5) illustrés dans l'espace gaussien (entre -2 et 2). Dans chaque simulation, on retrouve 4 faciès représentés par des niveaux de gris différents (F1=1, F2=2, F3=3, F4=4). Les drapeaux ont été placés dans un ordre quelconque. On désire associer le bon drapeau à chaque image simulée.



Question 7 (suite)



10 pts a) Associez à chaque simulation le drapeau de codage (de 1 à 5) correspondant. Indiquez « aucun » si aucun drapeau de codage ne correspond selon vous à la simulation considérée.

Simulation	A	B	C	D	E
Drapeau					

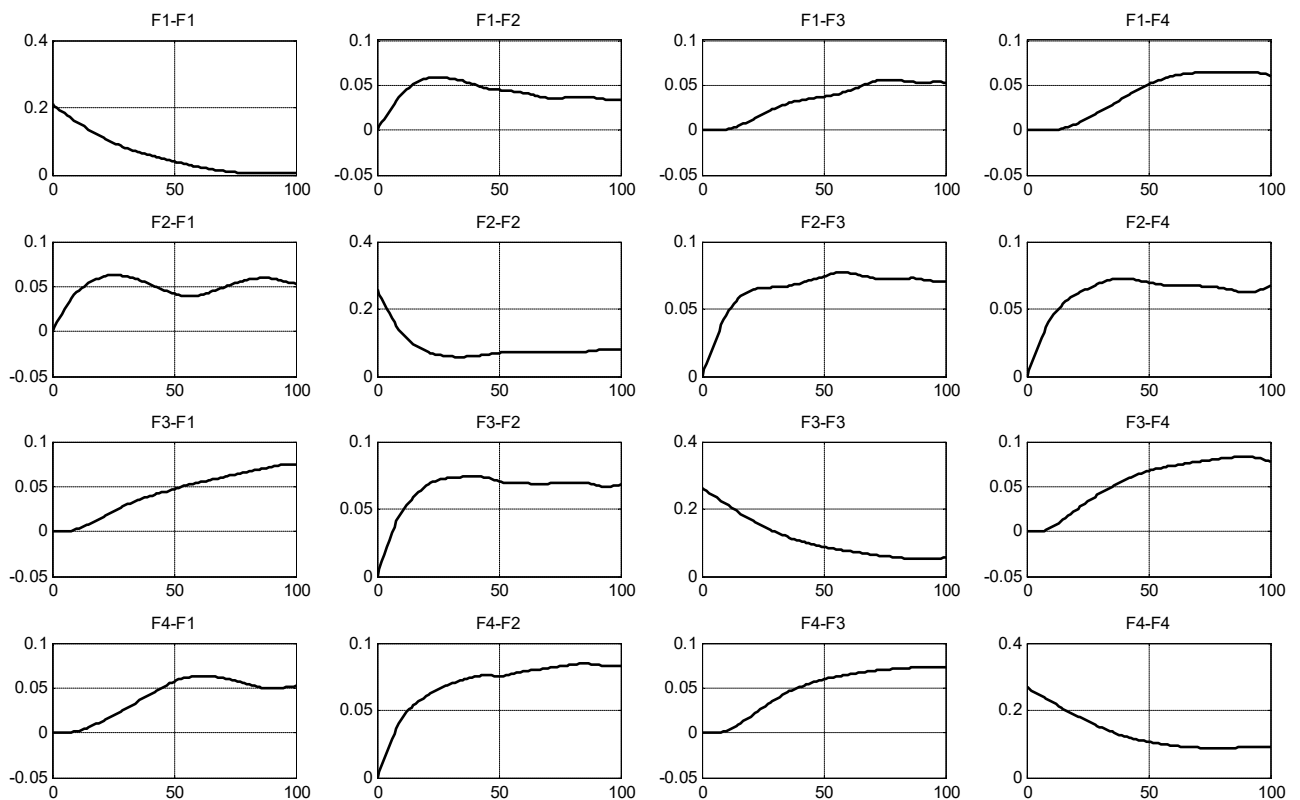
2 pts b) Une seule simulation de deux variables gaussiennes a été utilisée pour produire les simulations de faciès précédentes. Le modèle de variogramme utilisé pour les deux variables est le même. Identifiez le modèle utilisé parmi les modèles suivants (justifiez votre choix) :

- i. Gaussien isotrope avec portée de 100 pixels
- ii. Sphérique isotrope avec portée de 100 pixels
- iii. Gaussien anisotrope avec portées 100 horizontalement et 10 verticalement
- iv. Sphérique anisotrope avec portées 100 horizontalement et 10 verticalement

Question 7 (suite)

3 pts c) Identifiez la simulation (parmi A à E) qui aurait pu être obtenue par la méthode gaussienne tronquée plutôt que par la méthode plurigaussienne?

Les covariances non centrées expérimentales des indicatrices de faciès sont calculées pour l'une des simulations A à E. La figure suivante montre le résultat obtenu selon la direction verticale (positive vers le bas). On a considéré séparément $I_i(x)$ $I_j(x+h)$ et $I_j(x)$ $I_i(x+h)$.



3 pts d) Identifiez la simulation (parmi A à E) ayant fourni ces covariances non-centrées d'indicateurs. Justifiez votre choix.

Corrigé

Question 1

a) le post-conditionnement par krigeage est effectué avec toujours les mêmes données, si le nombre n'est pas trop grand, on peut faire le krigeage globalement et donc inverser la matrice de krigeage une seule fois. Avec SGS, le nombre de données inclut le nombre de points déjà simulés. Ce nombre normalement devient trop grand pour permettre l'inversion d'un seul coup.

b) Comme on l'a vu au cours, on doit s'attendre à voir des fluctuations du variogramme expérimental par rapport au modèle. Ne pas voir de fluctuations implique que statistiquement quelque chose n'est pas respecté.

c) au lieu de faire le KS, on effectue le krigeage d'indicatrices (seuils) et l'on tire une valeur de la distribution conditionnelle ainsi obtenue. On ajoute cette valeur aux données simulées et l'on poursuit.

Question 2

a) Au seuil 300 ppm la fonction de répartition vaut approximativement 0.82. Le palier devrait donc être assez près de $0.82 \times 0.18 = 0.15$.

b) On code les valeurs par rapport au seuil 200, on a $I_1=1, I_2=0, I_3=1, I_4=0$. La somme des λ donne 0.81. La valeur estimée vaut donc $0.22 + (1-0.81) \times 0.63 = 0.34$. $P(Z_0 < 200) = 0.34$, donc $P(Z_0 > 200) = 1 - 0.34 = 0.66$.

Question 3

4- A) Même si la corrélation entre les 2 variables est forte, le fait que la variable secondaire soit connue aux mêmes emplacements devrait faire que le gain de précision sera assez limité.

B) La variable secondaire est plus abondante et l'on a une forte corrélation selon le modèle (0.92). Le cokrigeage devrait améliorer beaucoup la précision des prédictions.

C) La variable principale est connue en plus d'endroits que la variable secondaire, le cokrigeage apportera très peu;

D) La disposition est comme en B) mais ici la corrélation est seulement de 0.07, ce qui indique que la variable secondaire n'est pas utile pour déterminer la variable principale.

b) En effectuant une validation croisée où l'on enlèverait les deux informations (var. principale et var. secondaire) au point d'estimation, on pourrait apprécier le gain procuré par le cokrigeage.

Question 4

6- a) corrélation : $9.8 / (8 \times 13.2)^{0.5} = 0.95$.

b) La matrice de cokrigeage est :

	Z ₁	Y ₁	Y ₂	Z ₃	Y ₃	Y ₄		
Z ₁	8	9.8	2.8125	2.1875	2.8125	1.045	1	0
Y ₁	9.8	13.2	3.75	2.8125	3.75	1.3934	0	1
Y ₂	2.8125	3.75	13.2	1.045	1.3934	3.75	0	1
Z ₃	2.1875	2.8125	1.045	8	9.8	2.8125	1	0
Y ₃	2.8125	3.75	1.3934	9.8	13.2	3.75	0	1
Y ₄	1.045	1.3934	3.75	2.8125	3.75	13.2	0	1
	1	0	0	1	0	0	0	0
	0	1	1	0	1	1	0	0

Le vecteur de droite est :

	Z_0
Z_1	3.4424
Y_1	4.4259
Y_2	4.4259
Z_3	3.4424
Y_3	4.4259
Y_4	4.4259
	1
	0

c) valeur estimée : $(25+26.3)/2 - 0.19*(33+32)+0.19*(34+35)= 26.41$ m
variance de krigage : $8-(3.44*.5-.19*.443+.19*4.43+.5*3.44-.19*4.43+.19*4.43-.03)= 4.59$

Question 5

a) on multiplie la 2^e ligne de L avec la 3^e colonne de L' soit : $3.82*4.59+6.74*3.23=39.30$

b) $Y_1=5/7.75=0.65$

c) $3.82*0.65=2.48$

d) On ne peut simuler de grands champs pour des raisons d'espace mémoire. La matrice de covariance est en effet de taille $n \times n$, où n est le nombre de points à simuler.

Question 6

a) erreurs flagrantes : A : pour une variable indicatrice, l'on doit ajuster un modèle avec comportement linéaire à l'origine, jamais un gaussien

B : le palier maximal ne devrait pas dépasser 0.25 en théorie.

b) sans unité puisque ce sont des variogrammes d'indicatrice.

Question 7

a) Associez à chaque simulation le drapeau de codage correspondant.

Simulation	A	B	C	D	E
Drapeau	3	2	5	1	4

b) Gaussien isotrope. On ne voit aucun étirement indicateur d'anisotropie et la texture est clairement celle d'un gaussien en raison de la régularité des contours des facies.

c) Celle correspondant au drapeau 3, donc A.

d) Pas de contact F3-F4, F1-F3 et F1-F4 => simulation E (drapeau 4)