ECOLE DES MINES
DE PARIS

SIMULATION CONDITIONNELLE DE MODELES ISOFACTORIELS

Xavier Emery
Février 2004

Introduction

La simulation en géostatistique

Importance de la simulation pour la prise de décision

Quelques exemples de modèles :

- modèle multigaussien
- modèles construits à partir d'objets
- ensembles aléatoires (booléen, gaussiennes seuillées…)
- variables catégorielles (plurigaussiennes…)

Qu'est-ce qu'un modèle isofactoriel ?

Une fonction aléatoire $\{Y(\mathbf{x}), \mathbf{x} \in \mathsf{R}^d\}$ est isofactorielle si une famille de fonctions (appelées ***facteurs***) forme une base orthonormée :

- toute fonction de $Y(\mathbf{x})$ s'écrit comme une combinaison linéaire des facteurs
- deux facteurs distincts n'ont pas de corrélation spatiale croisée

Cette propriété permet d'estimer des fonctions non linéaires de $Y(\mathbf{x})$ par la technique du ***krigeage disjonctif***.

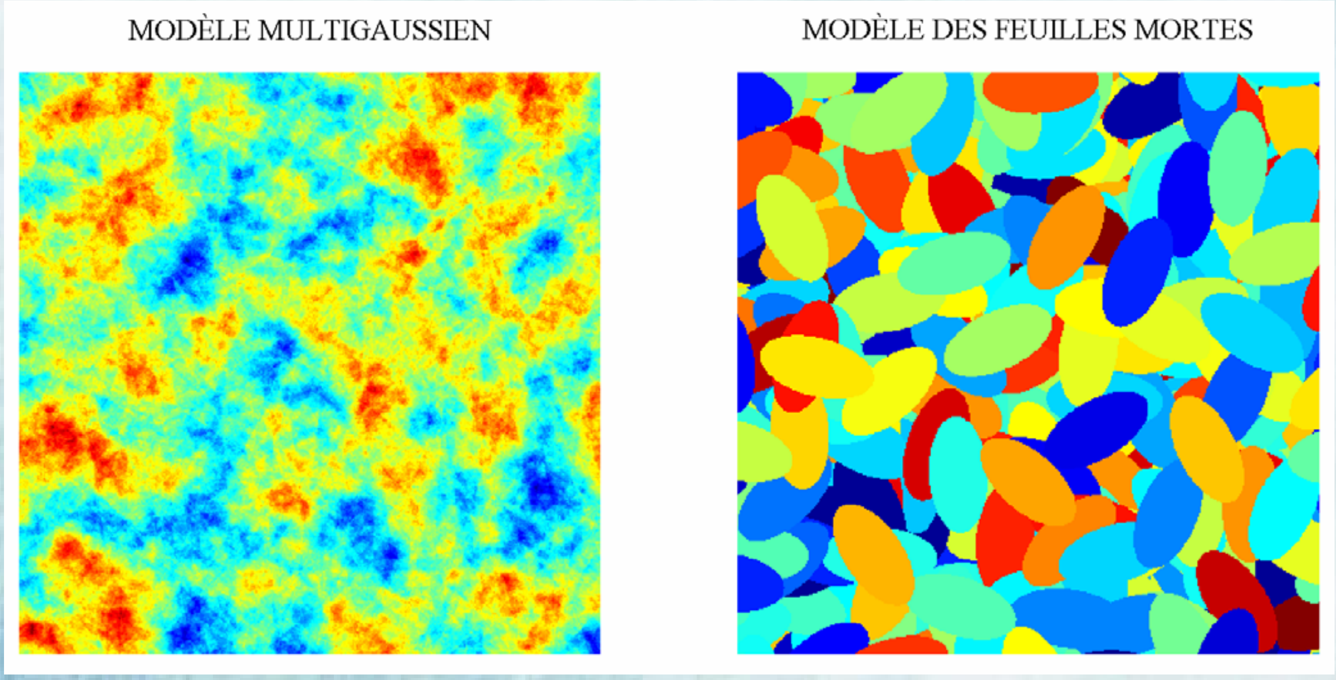
Exemples

modèle	**loi marginale**	**facteurs**
hermitien	gaussienne	polynômes d'Hermite
laguerrien	gamma	polynômes de Laguerre
beta	beta	polynômes de Jacobi
Meixner	binomiale négative	polynômes de Meixner
Charlier	Poisson	polynômes de Charlier
Krawtchouk	binomiale	polynômes de Krawtchouk
mosaïque	quelconque	

Classification des lois bivariables isofactorielles

On distingue deux cas extrêmes de modèles isofactoriels :

- ***modèles diffusifs*** : modèle bigaussien, bigamma…
- ***modèles mosaïques*** : Poisson, Voronoï, feuilles mortes…



Motivations de la thèse

La simulation de fonctions aléatoires se heurte à deux exigences :

- ***flexibilité*** : possibilité de décrire une vaste classe de phénomènes
- ***inférence*** des paramètres du modèle

Les modèles isofactoriels réalisent un compromis entre ces deux exigences.

L’algorithme séquentiel

Principe

Simulation séquentielle d’indicatrices

- simuler les différents sites de l’espace les uns après les autres, selon une séquence choisie au hasard ;
- en chaque site, simuler une valeur selon sa loi conditionnelle aux données, estimée par ***krigeage d’indicatrices***. Incorporer la valeur simulée au jeu de données.

Simulation d’un modèle isofactoriel

- remplacer le krigeage d’indicatrices par un ***krigeage disjonctif***

Questions soulevées (1)

1) Quel est le degré d'approximation de l'algorithme séquentiel lorsqu'il est appliqué à un modèle particulier ?

Pour plusieurs fonctions aléatoires et configurations de krigeage, on a comparé le krigeage d'indicatrices et le krigeage disjonctif à l'espérance conditionnelle.

Questions soulevées (2)

2) Existe-t-il des fonctions aléatoires que l’algorithme séquentiel simule parfaitement ?

Ceci ne se produit que dans l’espace à une dimension :

- modèle mosaïque à covariance complètement monotone
- modèles à covariances d’indicatrices complètement monotones
- modèle à résidus d’indicatrices (simulation hiérarchique)
- modèles isofactoriels markoviens

Questions soulevées (3)

3) Le modèle généré est-il un modèle spatial ?

Problèmes posés :

- les propriétés multivariables de l'image obtenue dépendent du nombre et de l'ordre des sites simulés
- conditionnement de la simulation à des données

Construction et simulation de modèles isofactoriels particuliers

Modèles étudiés

- modèles de substitution (non présentés)
- modèles construits à partir d'ensembles aléatoires (non présentés)
- modèles construits par addition de modèles élémentaires
  - ✓ bandes tournantes (non présentées)
  - ✓ dilution poissonienne (non présentée)
  - ✓ addition de mosaïques indépendantes
- modèle bigamma

Construction et simulation de modèles isofactoriels particuliers

Addition de mosaïques indépendantes

Principe (1)

Pour plusieurs lois isofactorielles, les couples de variables aléatoires $\{Y(\mathbf{x}+\mathbf{h}), Y(\mathbf{x})\}$ se décomposent sous la forme suivante :

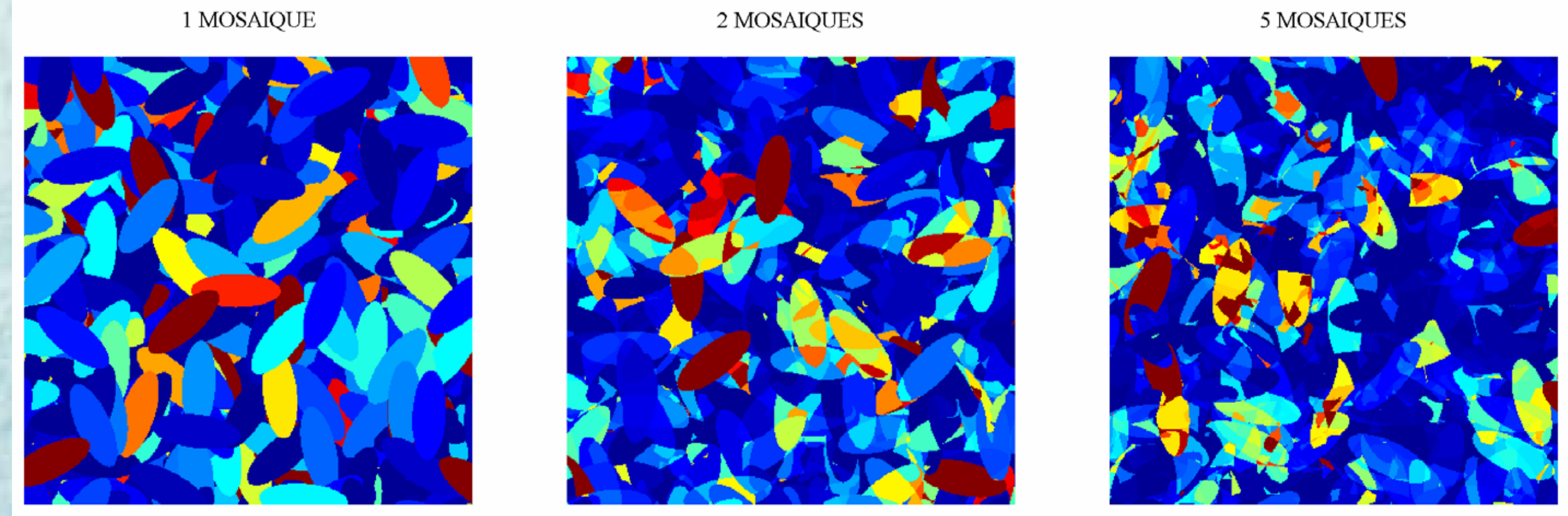
$$\begin{cases} Y(\mathbf{x}+\mathbf{h}) = U + V \\ Y(\mathbf{x}) = U + W \end{cases} \quad \text{avec U, V, W indépendantes}$$

Exemples

- loi bigaussienne
- loi isofactorielle de Laguerre
- loi isofactorielle de Poisson
- loi isofactorielle de Meixner

Principe (2)

L'identité précédente permet de définir des modèles isofactoriels par addition de N mosaïques indépendantes de même covariance.



Intérêts du modèle

On peut contrôler :

- la loi marginale
- la covariance
- le caractère mosaïque ou diffusif du processus

  si $N = 1$: processus mosaïque

  si N devient infini : processus diffusif

Conditionnement (1)

Addition de mosaïques des feuilles mortes

Une mosaïque des feuilles mortes est vue comme une population d'objets repérés par des germes poissoniens dans $D \times \mathbb{R}_+$.

On divise D en sous-domaines $\{D_i, i = 1\ldots I\}$ et $\mathbb{R}_+$ en intervalles $[j\tau, (j+1)\tau[, j \in \mathbb{N}\}$. Ceci définit des cellules partitionnant l'espace-temps.

Pour conditionner la simulation à un jeu de données, on utilise une procédure itérative de ***recuit simulé***, qui cherche à minimiser une fonction objectif.

Conditionnement (2)

1) partir d’une simulation non conditionnelle

2) choisir un indice n au hasard dans {1,… N}

3) choisir un indice i au hasard dans {1,… I}

4) choisir un indice j dans $\mathbb{N}$, e.g. selon une loi géométrique

5) dans la n-ième mosaïque, renouveler tous les objets germés dans $D_i \times [j\tau, (j+1)\tau[$. La transition est acceptée ou refusée selon la dynamique de Metropolis

6) retourner en 2)

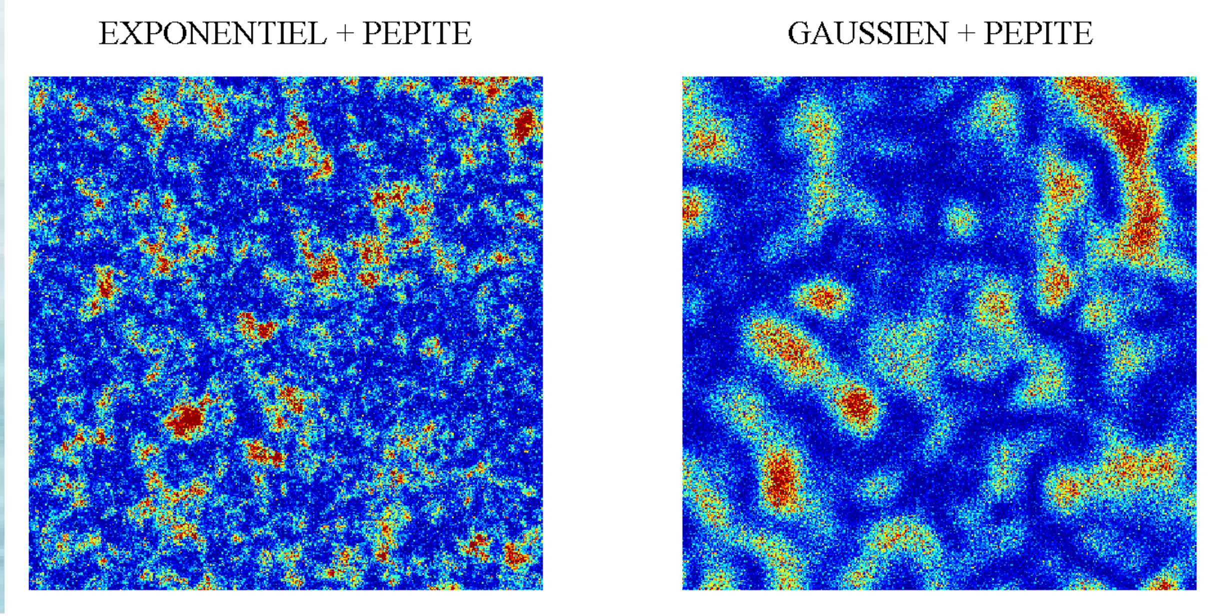
Construction et simulation de modèles isofactoriels particuliers

Modèle bigamma

Modèle du « chi carré »

Une fonction aléatoire à loi bigamma de paramètre demi-entier est obtenue en calculant la norme d'une fonction aléatoire vectorielle multigaussienne dont les composantes sont indépendantes et ont le même corrélogramme :

$$\forall \mathbf{x} \in \mathbb{R}^d, 2Y(\mathbf{x}) = \| \mathbf{U}(\mathbf{x}) \|^2$$



Conditionnement

Pour conditionner la simulation à un ensemble de données $\{Y(\mathbf{x}_\alpha), \alpha = 1\ldots n\}$, on utilise l'échantillonneur de Gibbs :

1) *Initialisation*: affectation de valeurs gaussiennes aux points de données $\{\mathbf{U}(\mathbf{x}_\alpha), \alpha = 1\ldots n\}$, telles que

$$\forall \alpha \in \{1...n\}, 2Y(\mathbf{x}_\alpha) = \| \mathbf{U}(\mathbf{x}_\alpha) \|^2$$

2) choisir un indice β au hasard parmi $\{1\ldots n\}$

3) remplacer $\mathbf{U}(\mathbf{x}_\beta)$ par un vecteur aléatoire simulé selon sa loi conditionnelle à $\{\mathbf{U}(\mathbf{x}_\alpha), \alpha \neq \beta\}$ et $Y(\mathbf{x}_\beta)$

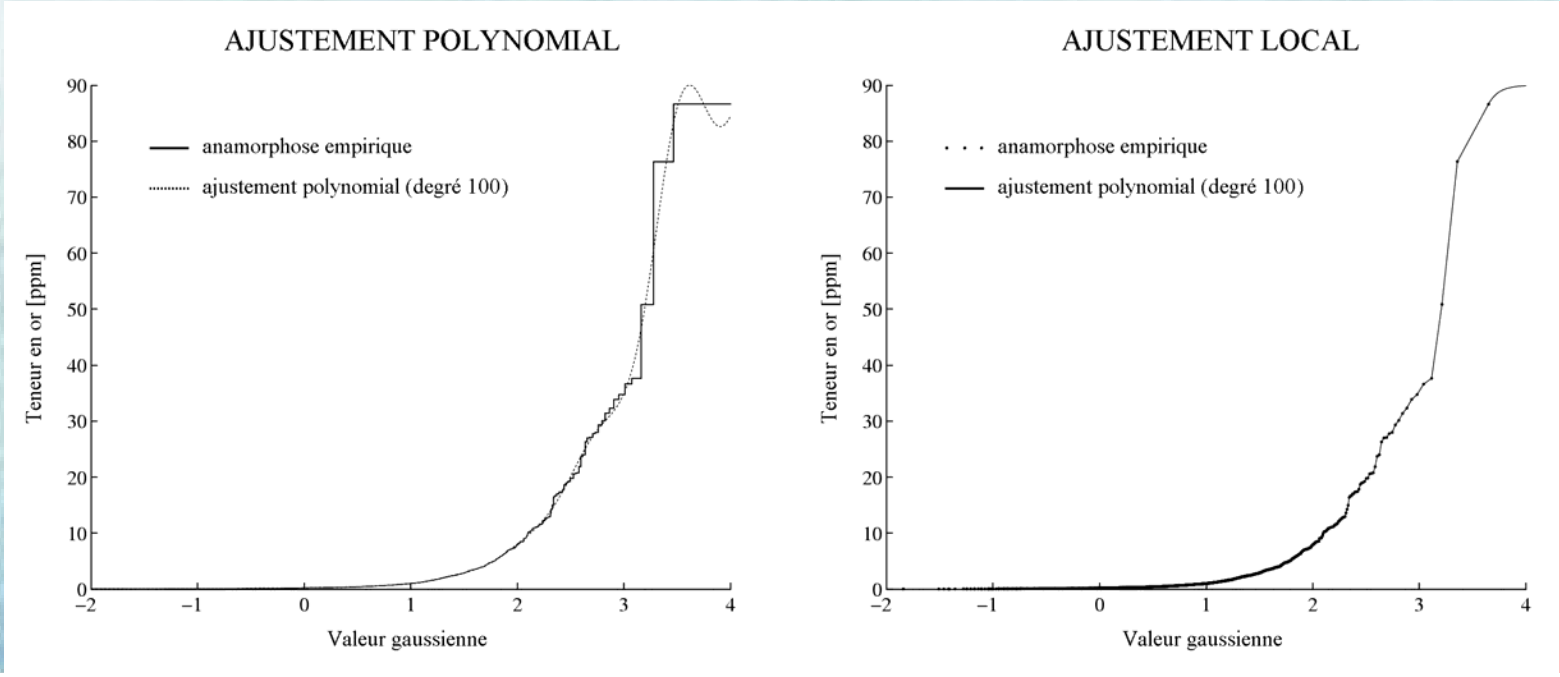
4) retourner en 2)

L’inférence statistique des modèles

Inférence de la loi marginale

Points critiques

- représentativité de l'histogramme des données
- modélisation des valeurs extrêmes
- conseil : ajustement local de l'anamorphose empirique



Inférence des lois bivariables (1)

Méthodes d'indicatrices

- analyse variographique de chaque indicatrice
- contraintes entre les diverses indicatrices

Modèles isofactoriels

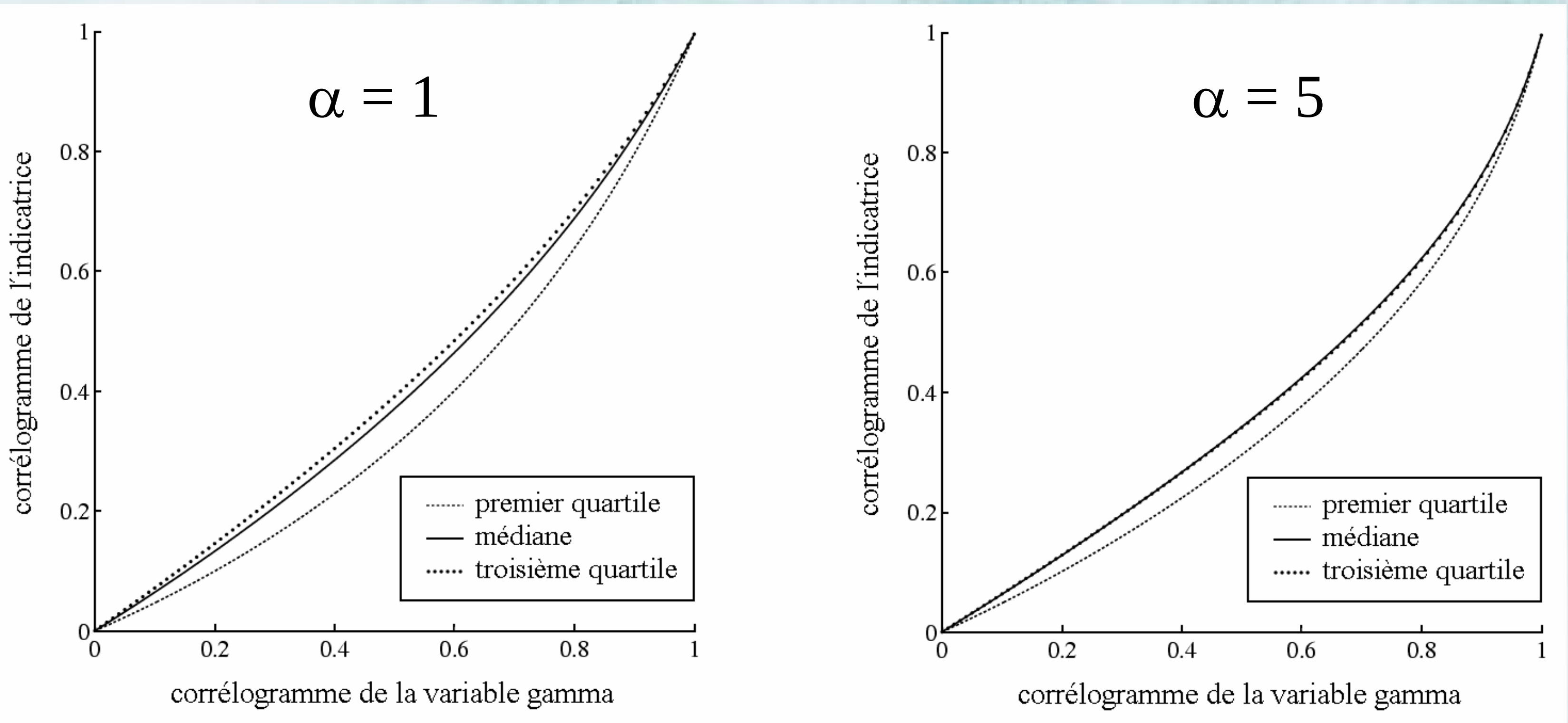
- les paramètres à définir sont réduits
- les critères généraux de cohérence interne restent inconnus

Inférence des lois bivariables (2)

Une propriété essentielle : la déstructuration des hautes teneurs

Ou comment se déstructurent les corrélogrammes d'indicatrices ?

Exemple : modèle de Laguerre betifié



Inférence des lois bivariables (3)

Conclusions

- comparer le variogramme et le madogramme ne suffit pas
- l'inférence et la validation du modèle doivent se baser sur des outils plus complets :
  - ✓ variogrammes de quelques indicatrices (quartiles)
  - ✓ variogrammes d'ordres inférieurs à deux :

  dans les modèles bigaussien et bigamma, on doit avoir

$$\forall \omega \in]0,2], \frac{\gamma_\omega(\mathbf{h})}{\gamma_\omega(\mathbf{h}_0)} = \left[\frac{\gamma(\mathbf{h})}{\gamma(\mathbf{h}_0)}\right]^{\omega/2}$$

Inférence des lois multivariables

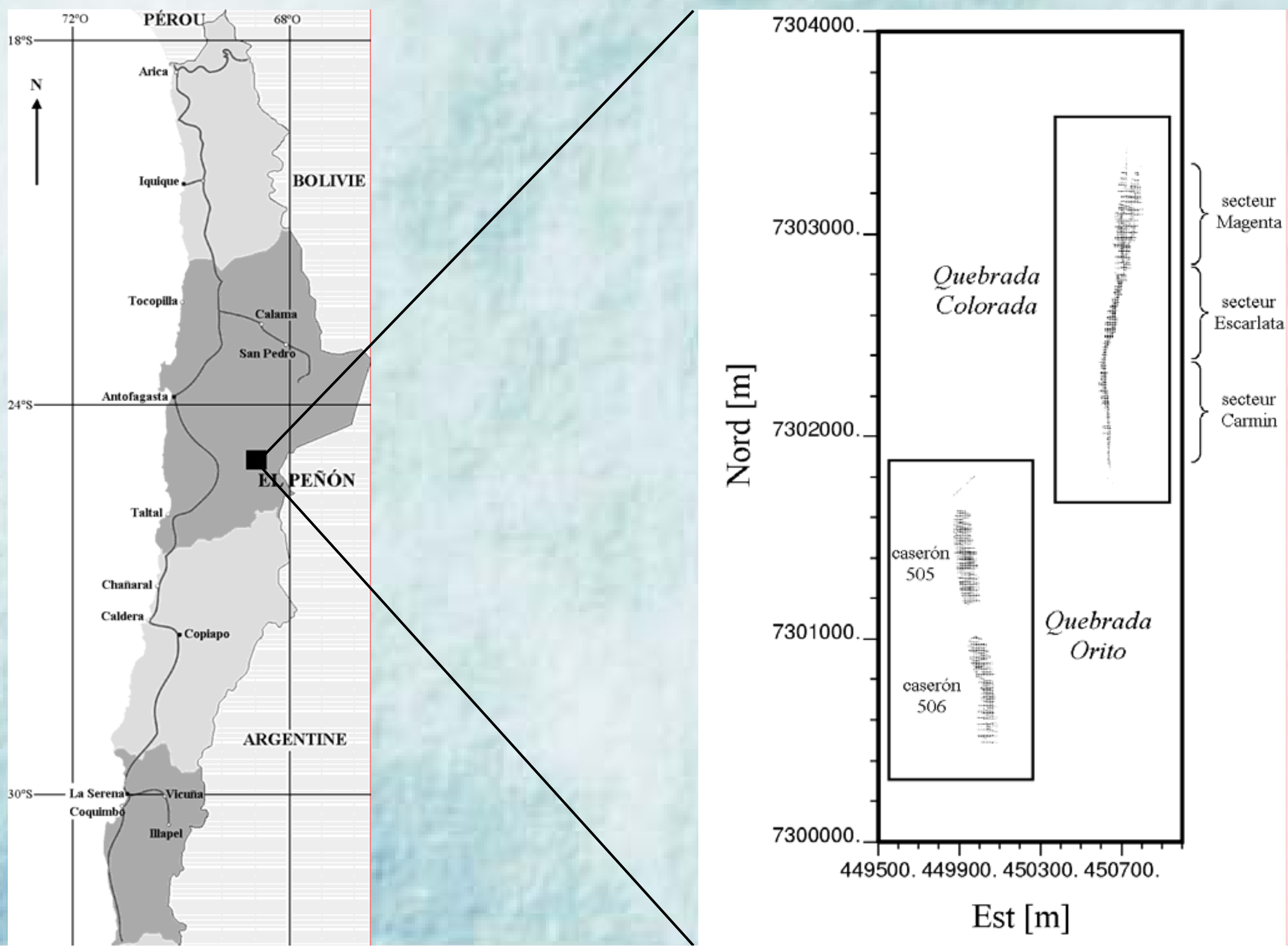
Ces lois sont laissées de côté dans l'inférence.

L'ajustement d'un modèle isofactoriel repose sur la loi marginale et les lois bivariables.

Application à des données minières

Présentation des données (1)

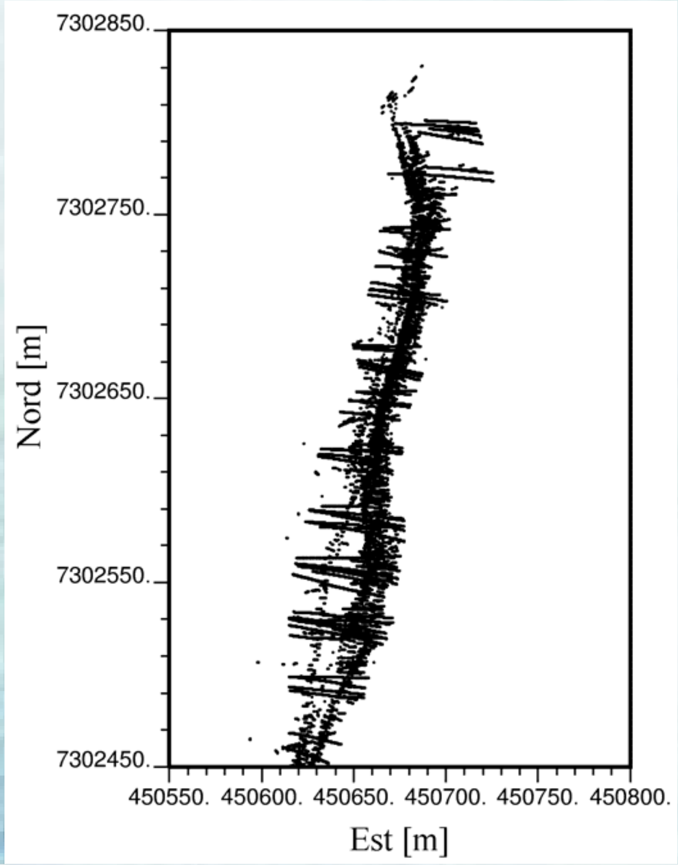
Gisement aurifère *El Peñón* (deuxième région du Chili)

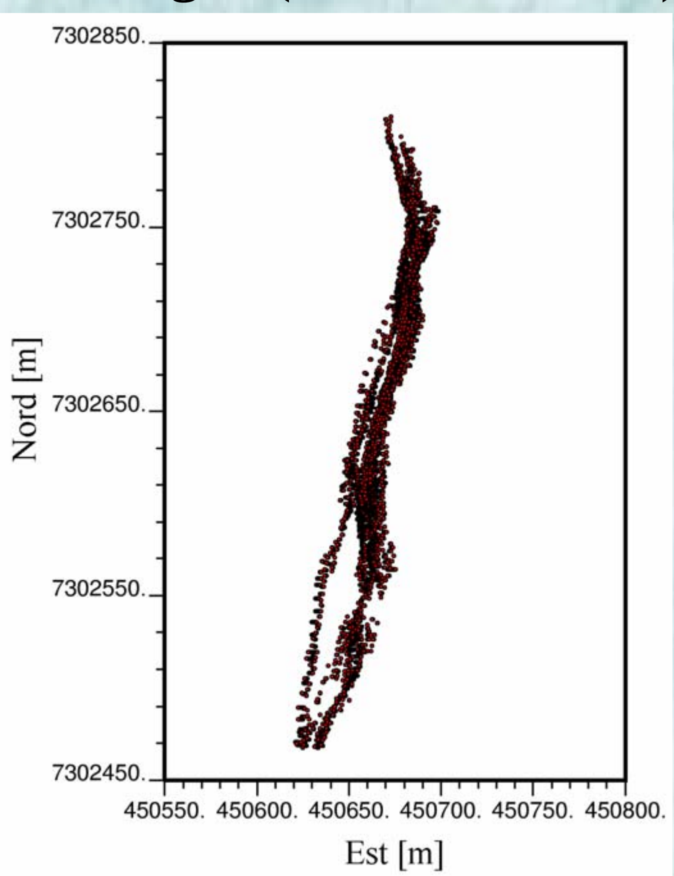


Présentation des données (2)

Etude du secteur *Escarlata* de la veine *Quebrada Colorada*

sondages (16 069 données) rainurages (5 094 données)

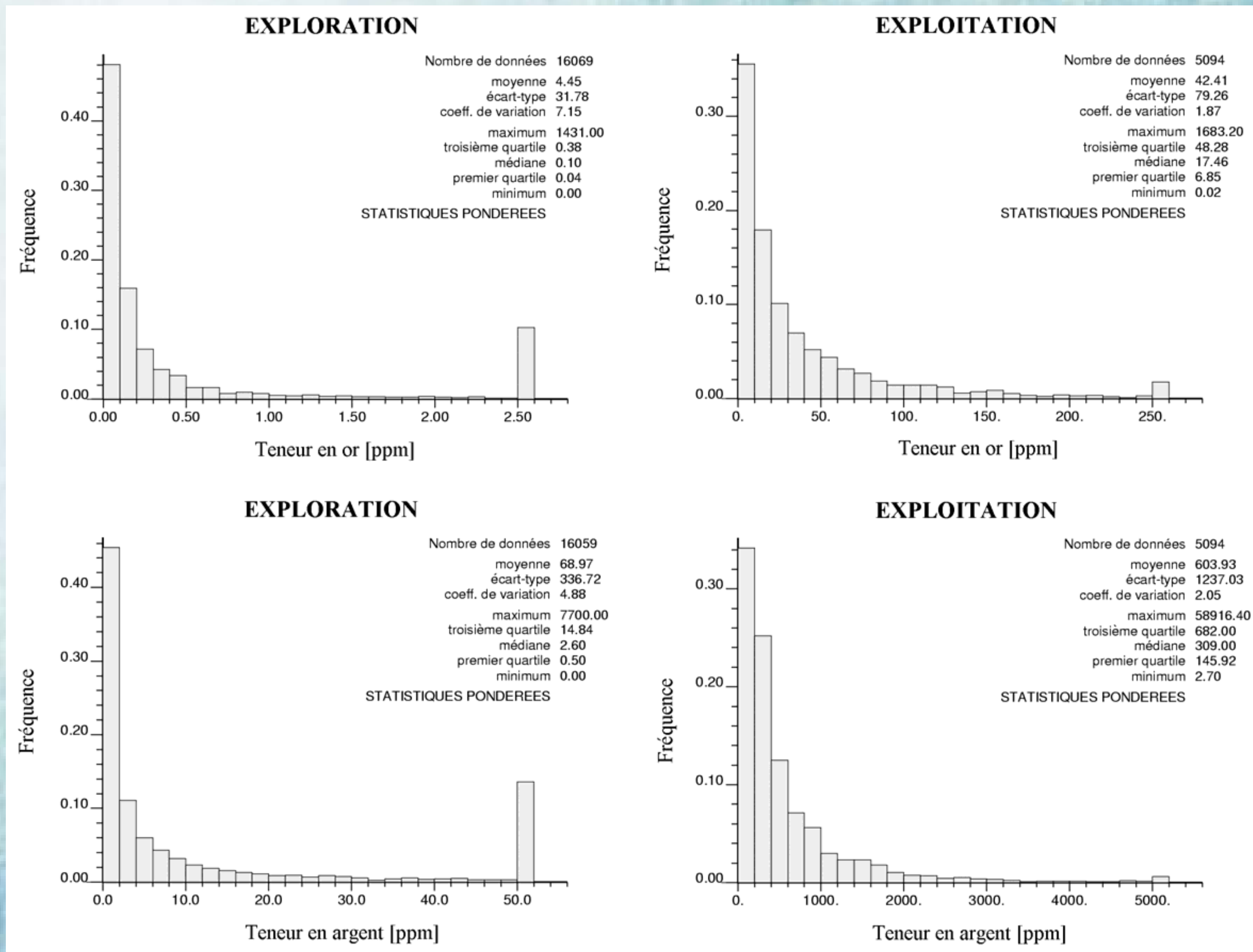




Etude exploratoire (1)

or
[ppm]

argent
[ppm]

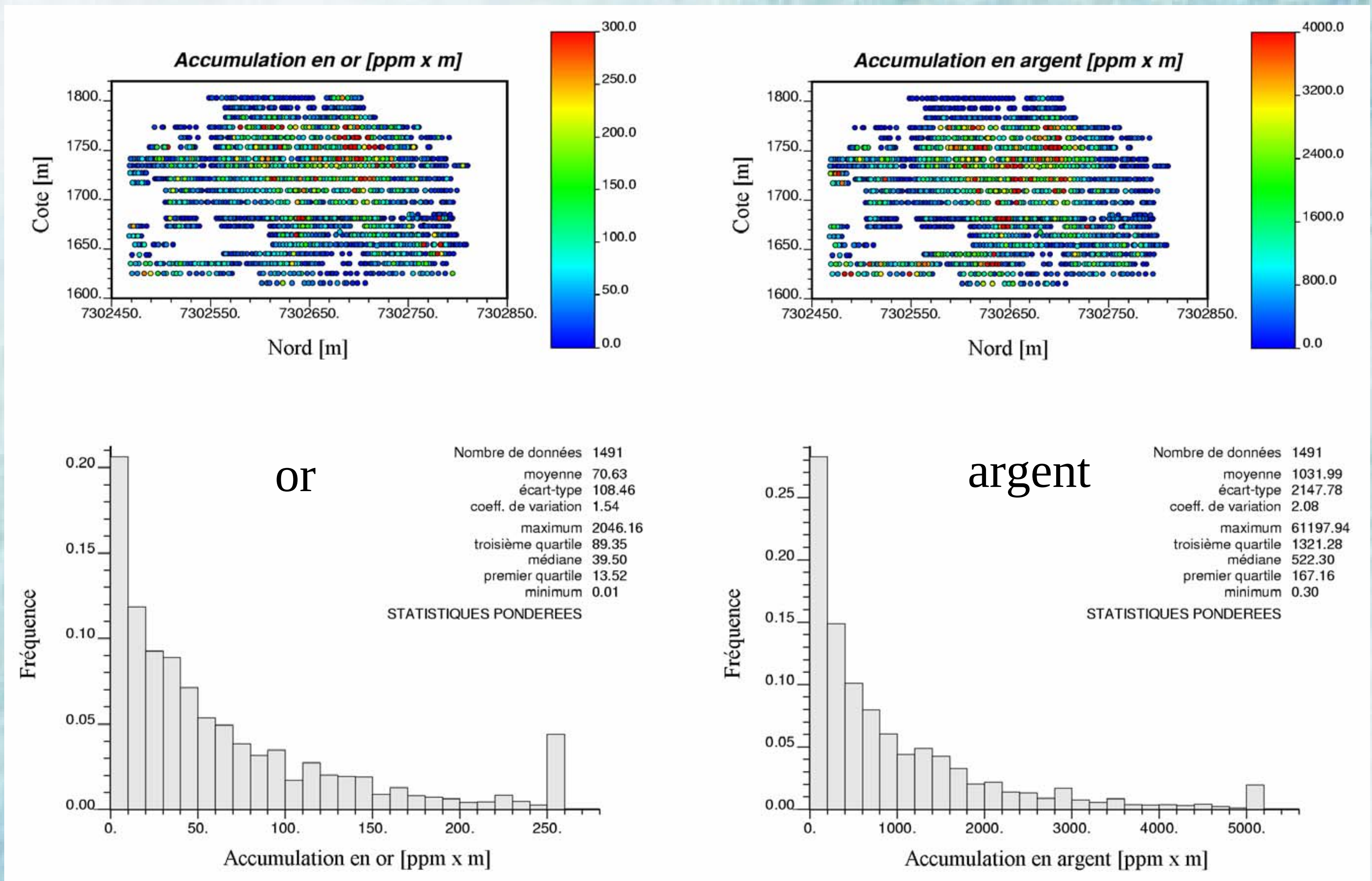


Etude exploratoire (2)

La décision est prise de passer en **accumulation** sur les données de rainurage :

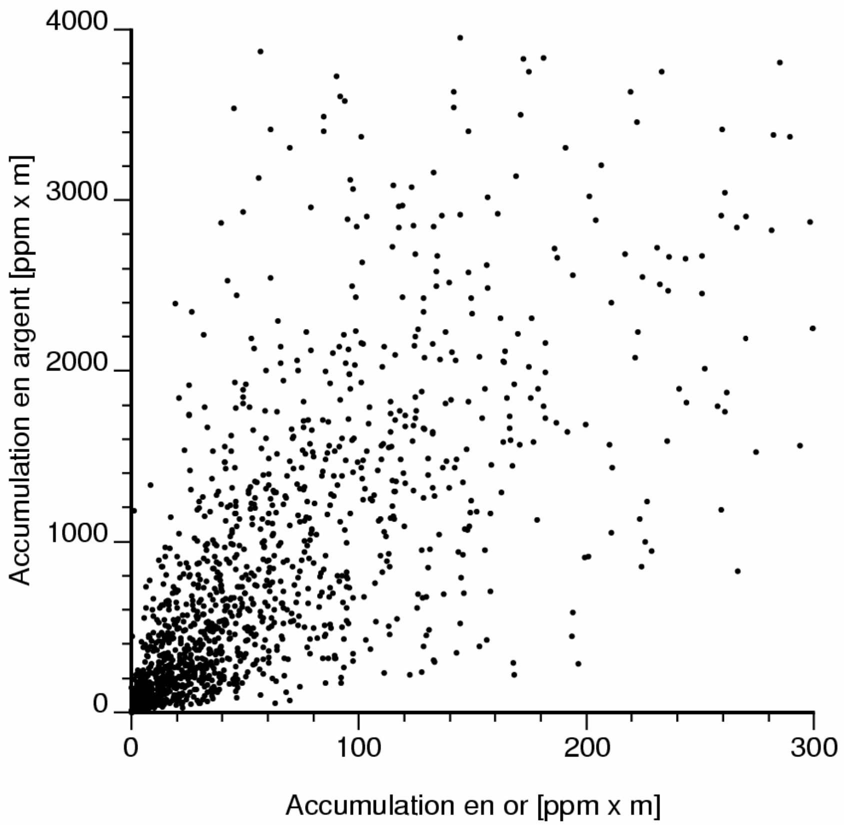
- on se ramène à une étude en deux dimensions
- on élimine un problème de stationnarité au sein de la veine
- on régularise les valeurs

Etude exploratoire (3)



Etude exploratoire (4)

Nuage de corrélation or / argent



Coefficients de corrélation

Valeurs brutes : 0.641

Transformées uniformes : 0.829

Indicatrices du premier quartile : 0.677

Indicatrices de la médiane : 0.651

Indicatrices du troisième quartile : 0.616

Choix du modèle (1)

On choisit d'utiliser un modèle de Laguerre, construit comme une somme de mosaïques indépendantes

- choix des paramètres d'anamorphose gamma : $\alpha_{Au} = \alpha_{Ag} = 0.5$
- extension du modèle au cadre bivariable

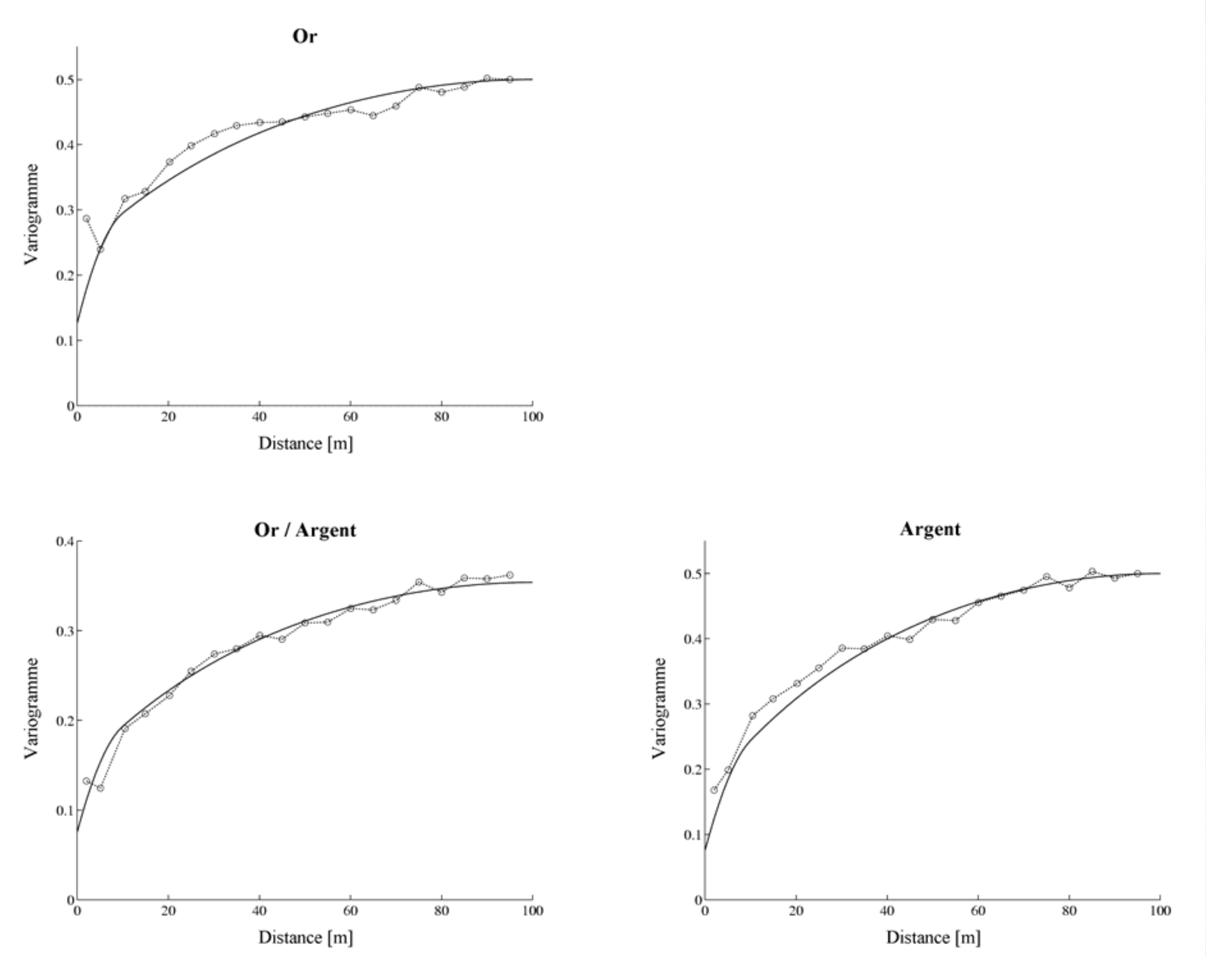
Choix du modèle (2)

Chaque variable est la somme de plusieurs mosaïques des feuilles mortes :

- chaque mosaïque contient plusieurs familles de grains primaires
- certains grains sont communs aux deux variables
- les valuations de ces grains se divisent en une partie commune aux deux mosaïques et une partie indépendante
- on ajoute un effet de pépite (une partie est commune aux deux mosaïques)

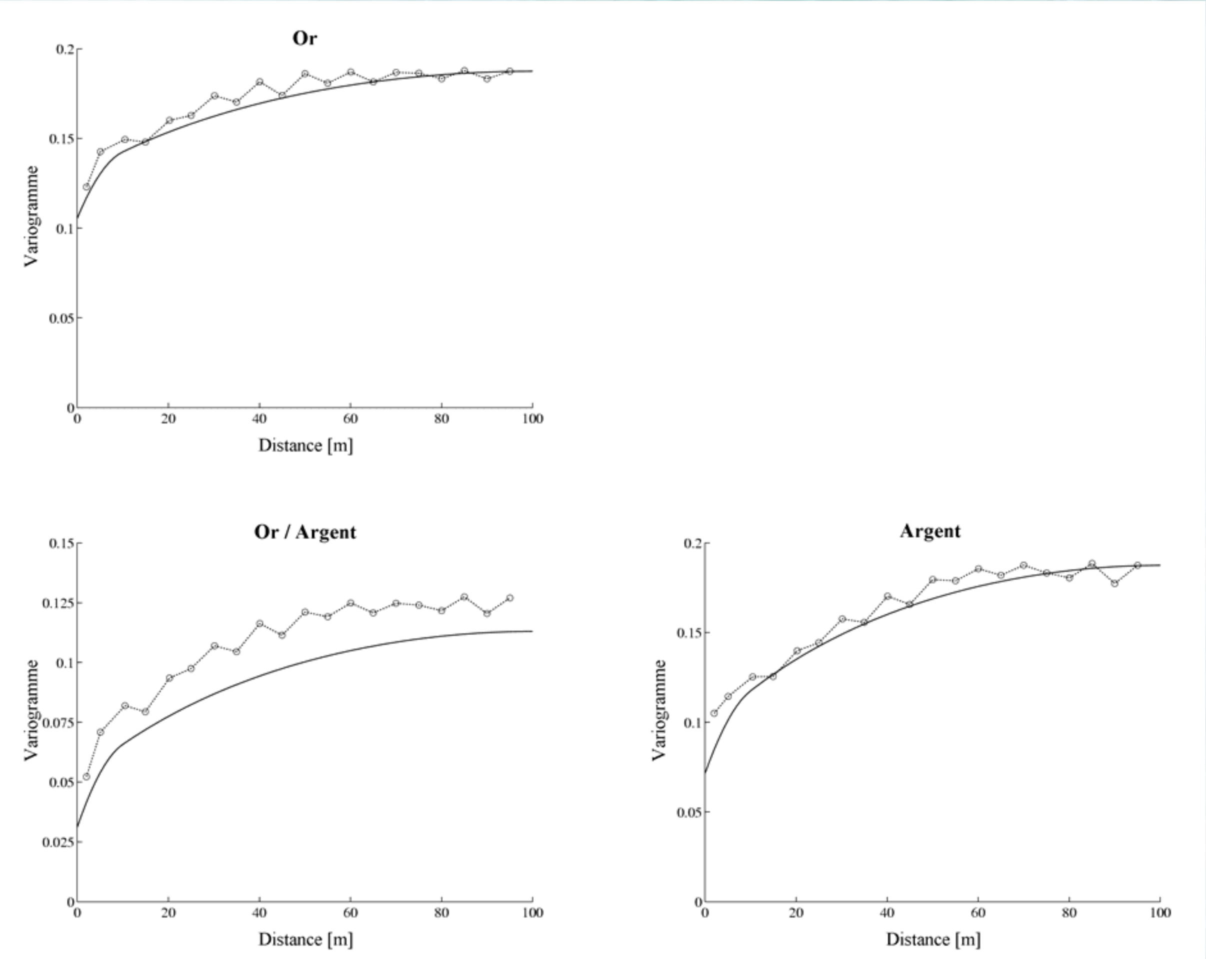
Choix du modèle (3)

Variogrammes simples et croisés des **variables anamorphosées**



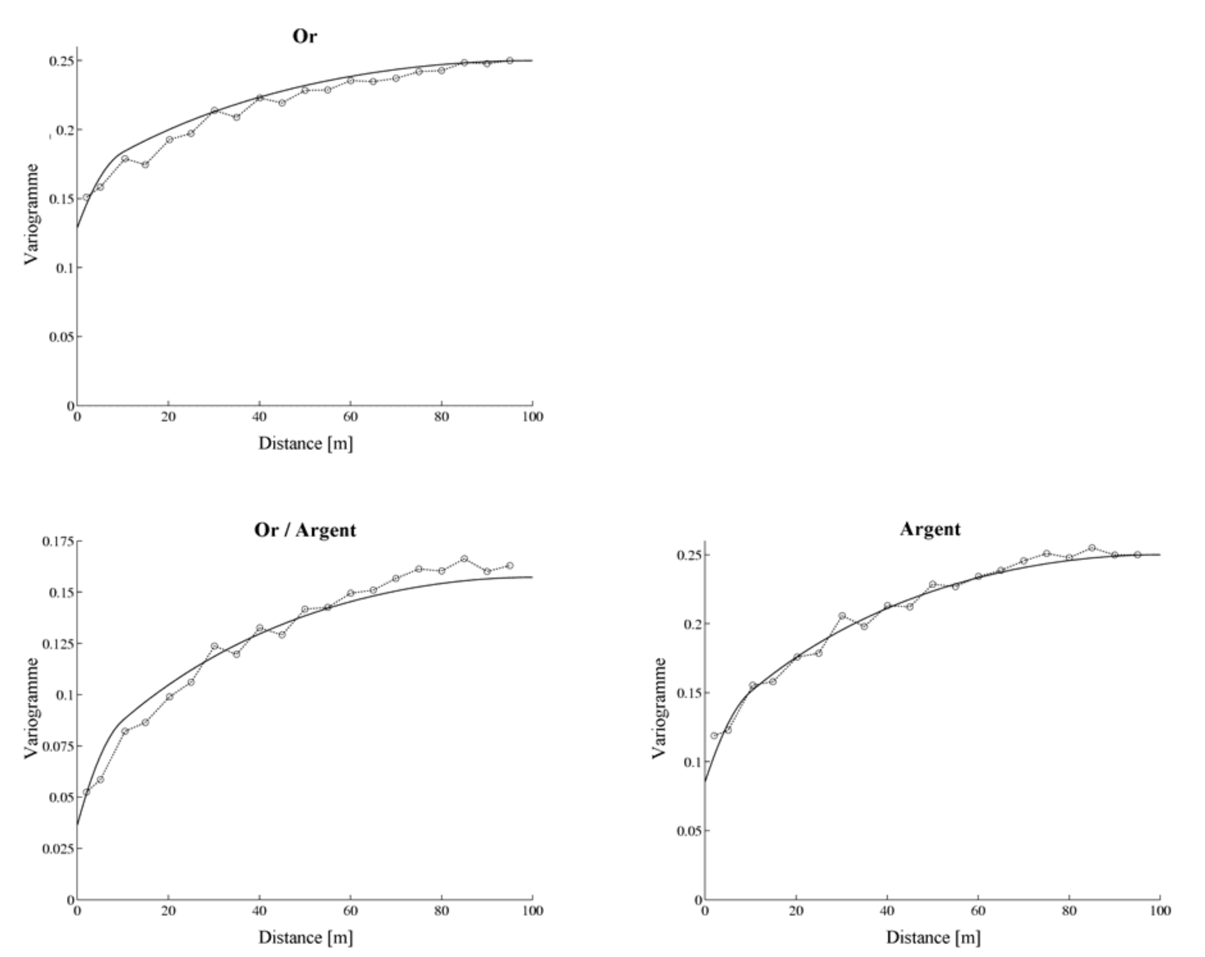
Choix du modèle (4)

Variogrammes simples et croisés d'**indicatrices** (premier quartile)



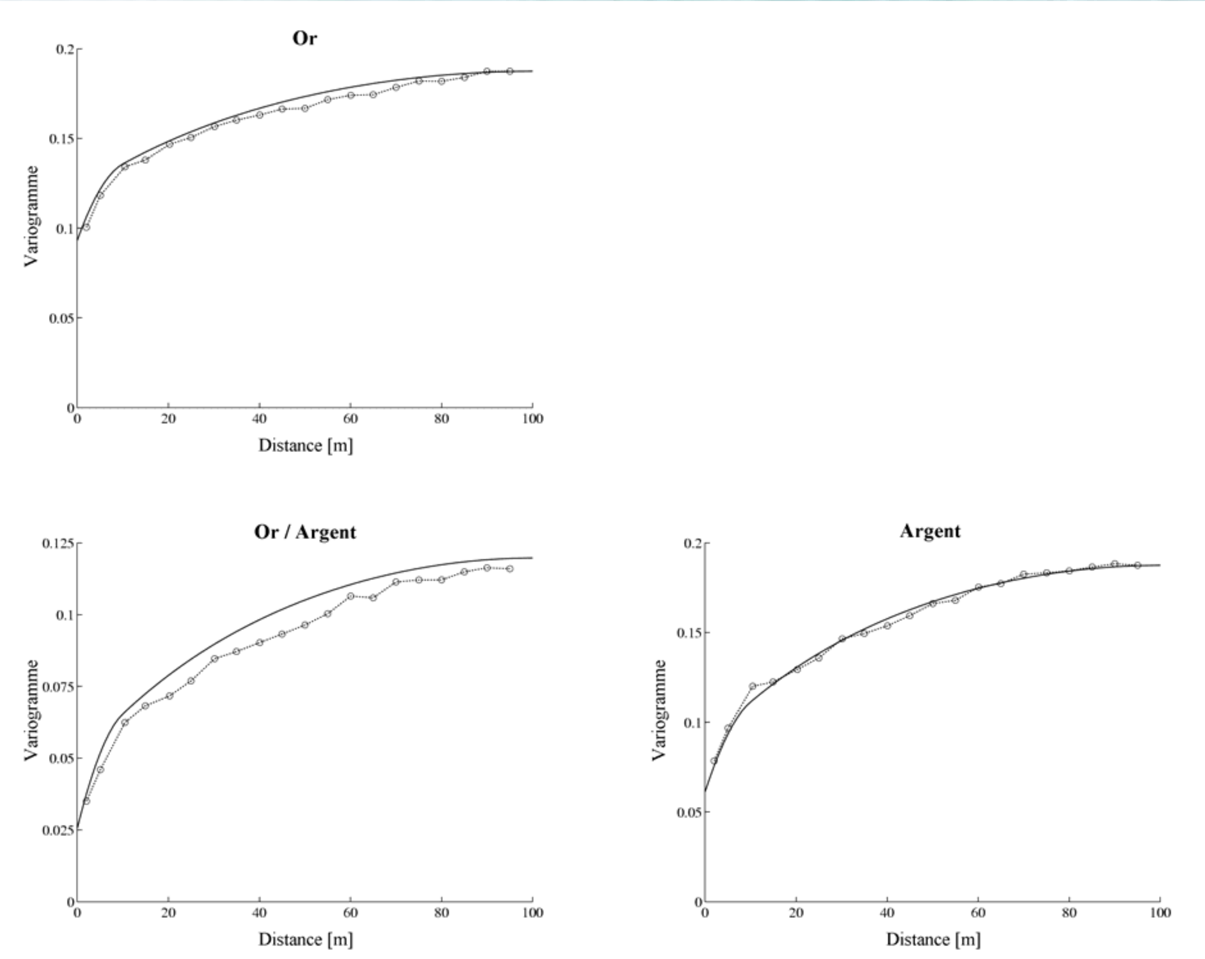
Choix du modèle (5)

Variogrammes simples et croisés d'**indicatrices** (médiane)



Choix du modèle (6)

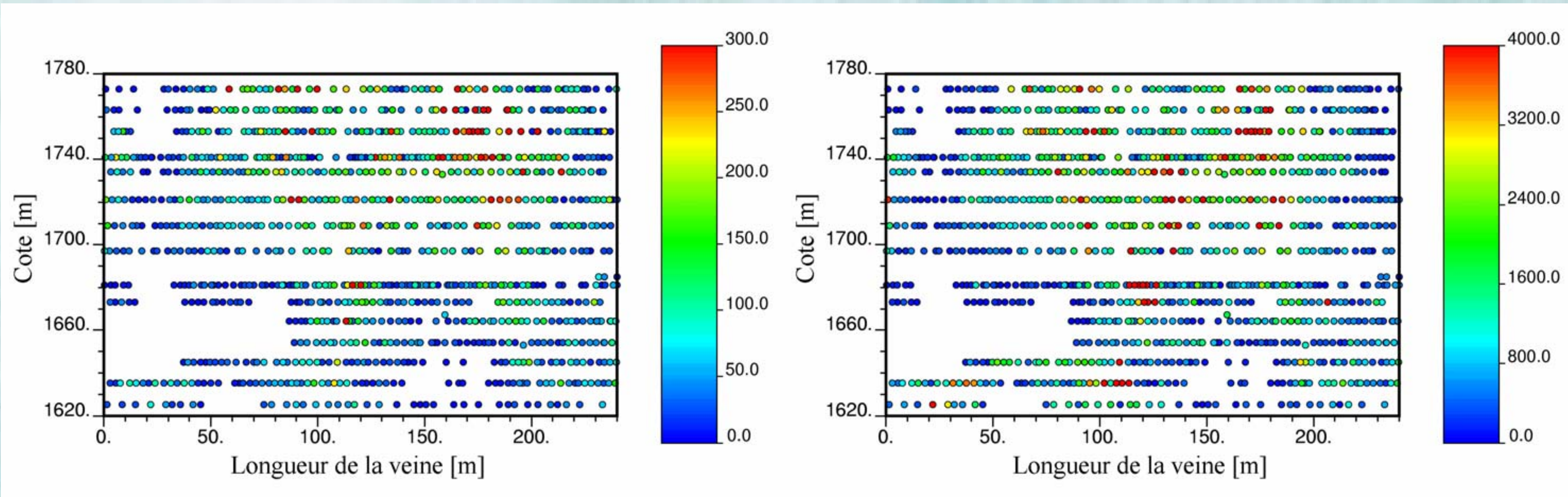
Variogrammes simples et croisés d'**indicatrices** (troisième quartile)



Simulation conditionnelle (recuit simulé)

Données conditionnantes

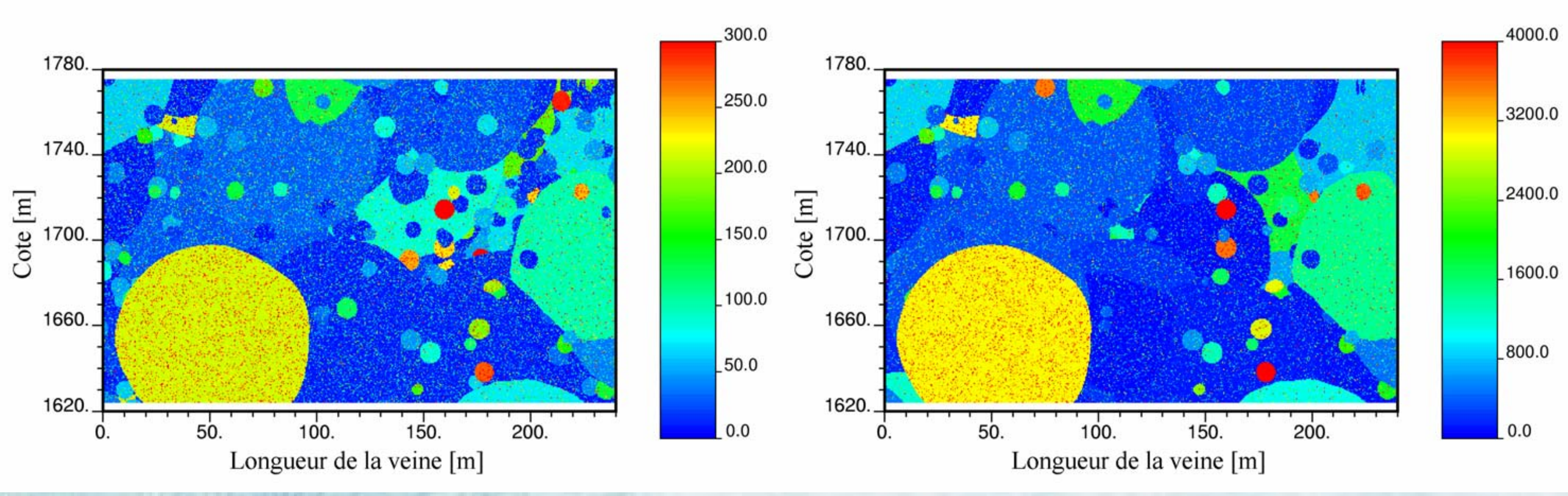
accumulation en or — accumulation en argent



Simulation conditionnelle (recuit simulé)

Etat initial : simulation non conditionnelle

accumulation en or — accumulation en argent

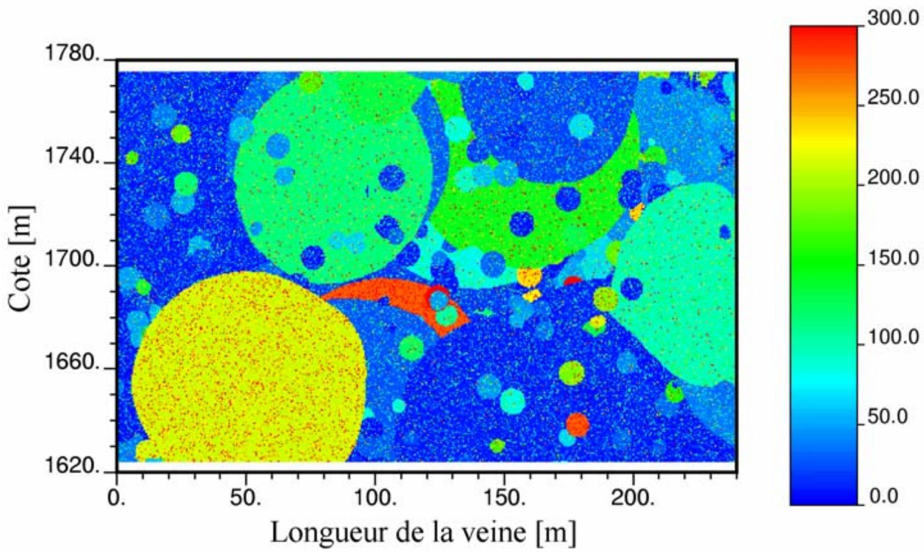


fonction objectif = 1898

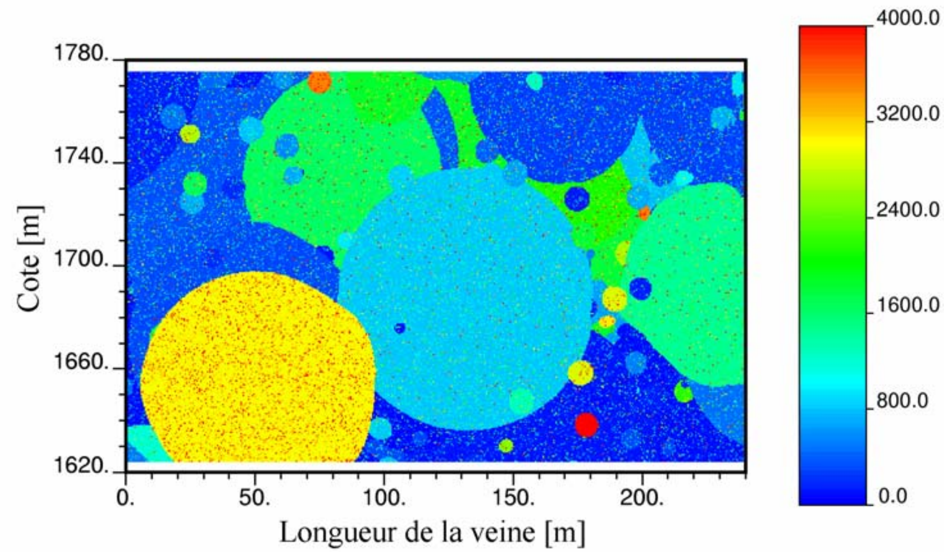
Simulation conditionnelle (recuit simulé)

10 000 itérations

accumulation en or



accumulation en argent

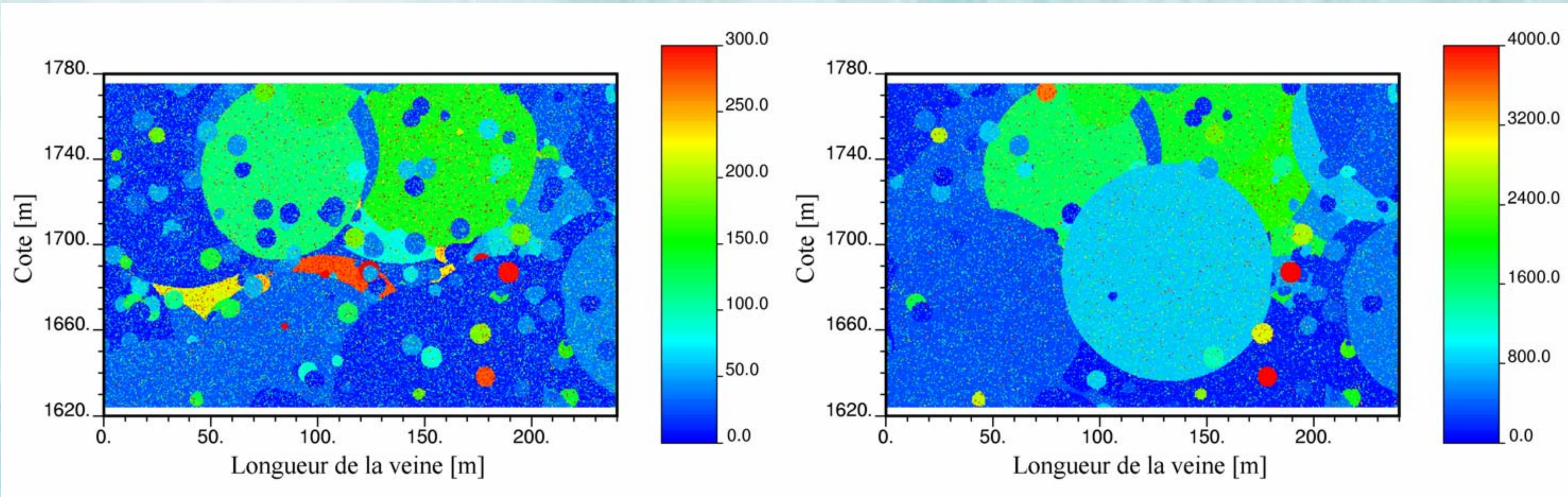


fonction objectif = 1656

Simulation conditionnelle (recuit simulé)

50 000 itérations

accumulation en or | accumulation en argent

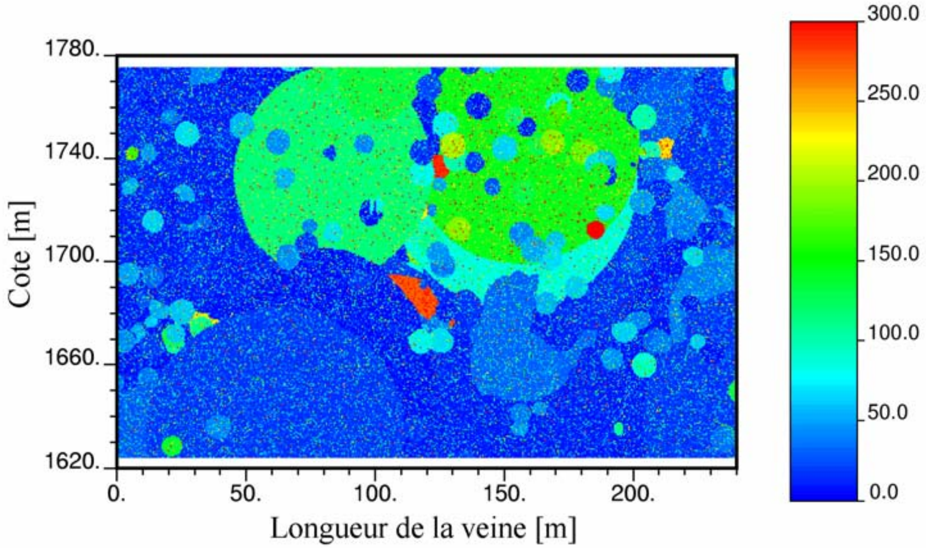


fonction objectif = 1235

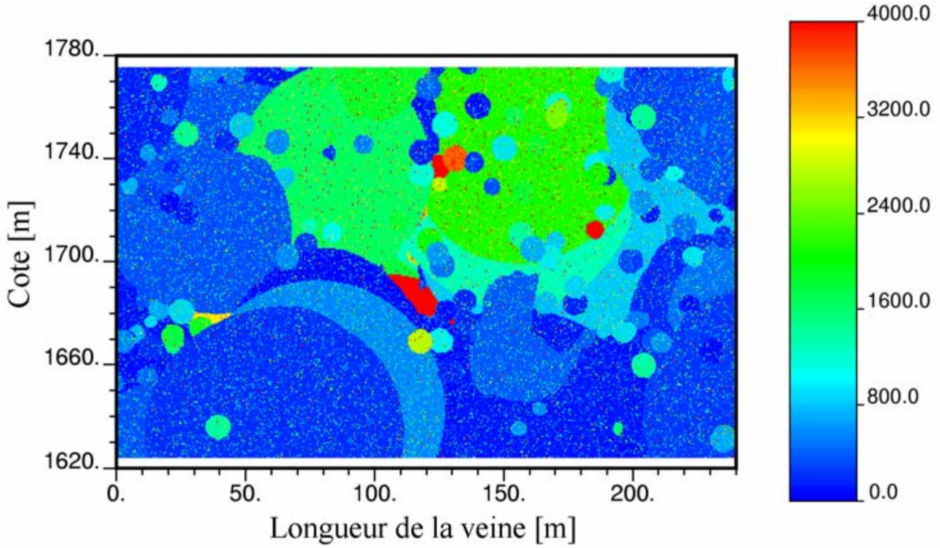
Simulation conditionnelle (recuit simulé)

100 000 itérations

accumulation en or



accumulation en argent

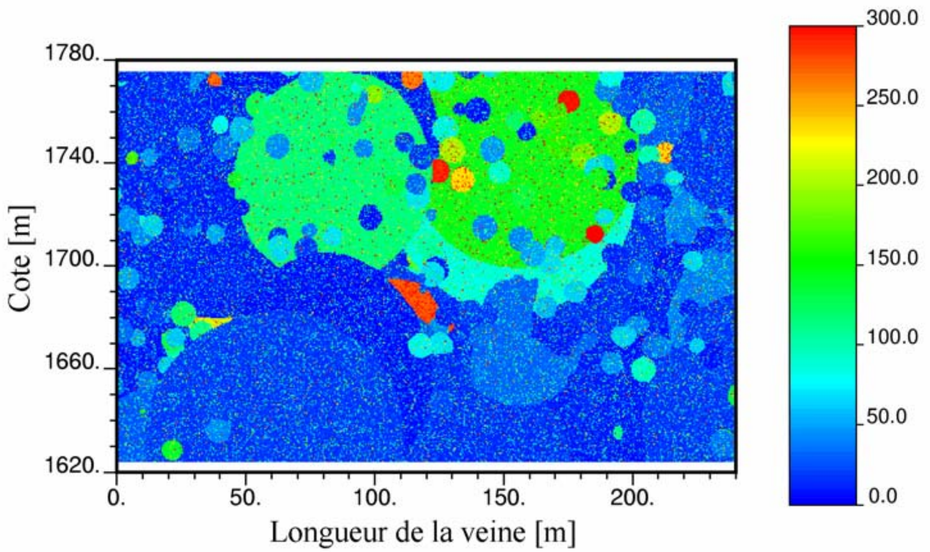


fonction objectif = 1036

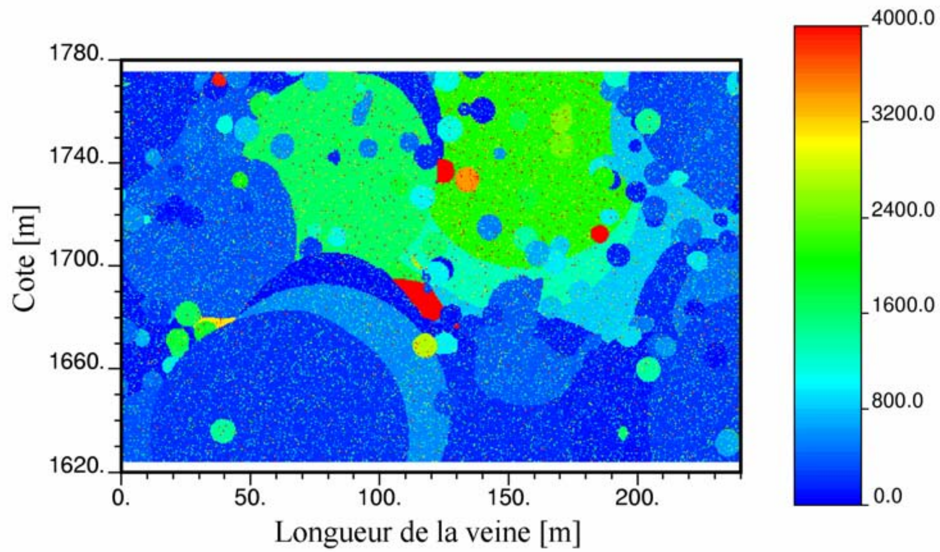
Simulation conditionnelle (recuit simulé)

150 000 itérations

accumulation en or



accumulation en argent



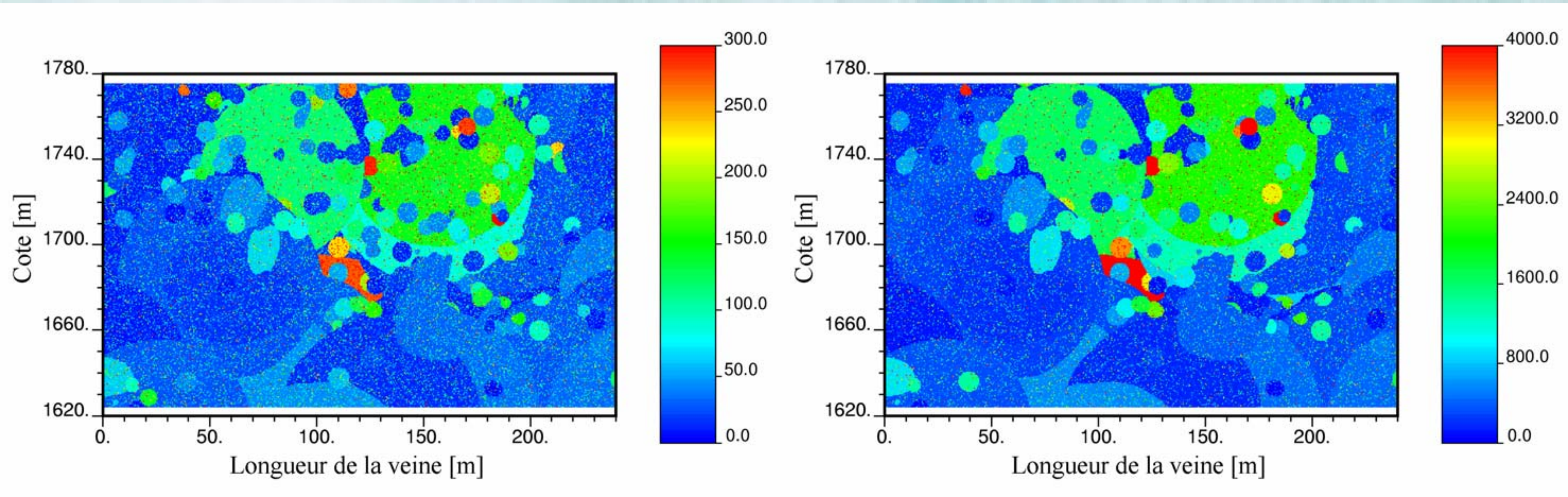
fonction objectif = 1006

Simulation conditionnelle (recuit simulé)

200 000 itérations

accumulation en or

accumulation en argent

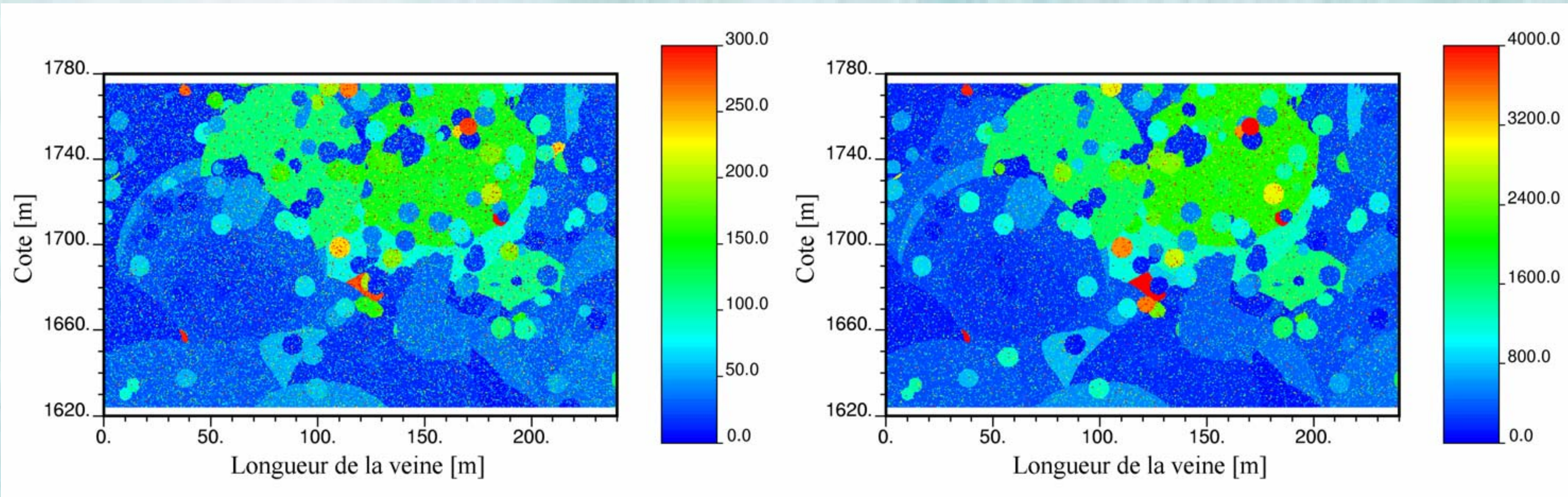


fonction objectif = 892

Simulation conditionnelle (recuit simulé)

500 000 itérations

accumulation en or accumulation en argent

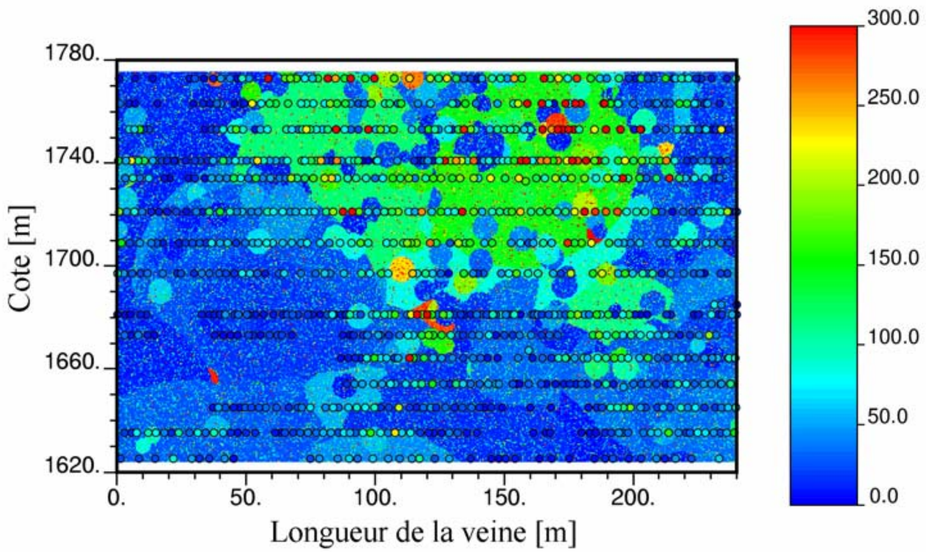


fonction objectif = 834

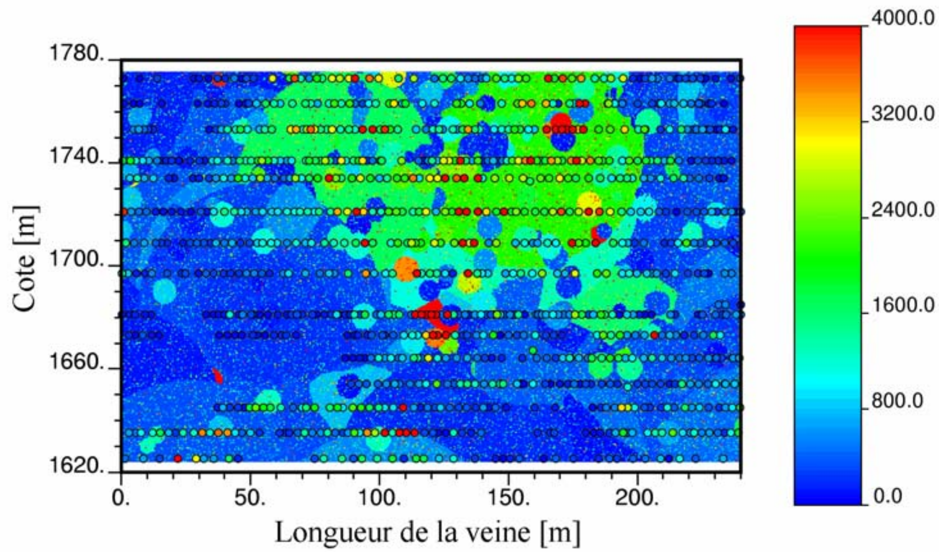
Simulation conditionnelle (recuit simulé)

500 000 itérations

accumulation en or



accumulation en argent



fonction objectif = 834

Conclusions

Conclusions (1)

Avantages et inconvénients de l'algorithme séquentiel :

- méthode « passe-partout »
- la simulation est directement conditionnée aux données
- les réalisations sont approximatives
- rien n'assure la cohérence interne du modèle
- les propriétés *a priori* des réalisations dépendent de facteurs "externes" (nombre total de sites simulés, nombre et position des données conditionnantes)

Conclusions (2)

Alternative : définition complète d’un modèle de fonction aléatoire

• outils pour l’inférence

• algorithmes pour le conditionnement à des données

Difficultés pratiques

• ajustement d’un modèle à une loi bivariable empirique

• temps de calcul (méthodes itératives)

• nécessité de rechercher d’autres modèles